Classification of Traversable Surfaces for Navigation Combining Multiple Cues Toward a Visual-Based Learning System

Toby Low, Antoine Manzanera

Classification of Traversable Surfaces for Navigation

Combining Multiple Cues Toward a Visual-Based Learning System

Toby Low and Antoine Manzanera
UEI - Laboratoire d'Informatique et d'Electronique
École Nationale Supérieure de Techniques Avancées
Paris, France
tobeslow@gmail.com, antoine.manzanera@ensta.fr

***Abstract*—This paper describes a vision-based ground-plane classification system for autonomous indoor mobile-robot that takes advantage of the synergy in combining together multiple visual-cues. *A priori* knowledge of the environment is important in many biological systems, in parallel with their mutually beneficial reactive systems. As such, a learning model approach is taken here for the classification of the ground/object space, initialised through a new *Distributed-Fusion (D-Fusion)* method that captures colour and textural data using *Superpixels*. A Markov Random Field (MRF) network is then used to classify, regularise, employ *a priori* constraints, and merge additional ground/object information provided by other visual cues (such as motion) to improve classification images. The developed system can classify indoor test-set ground-plane surfaces with an average true-positive to false-positive rate of 90.92% to 7.78% respectively on test-set data. The system has been designed in mind to fuse a variety of different visual-cues. Consequently it can be customised to fit different situations and/or sensory architectures accordingly.**

***Index Terms*—image classification, image disparity, ground plane, obstacle avoidance, visual navigation, mobile robots.**

I. Introduction

Robot obstacle avoidance and navigation, although intuitive within nature, it is difficult in practice. The vast information provided by a single visual image, and moreover multiple visual images highlights a redundancy in having an additional active range-based sensory system. However, the sole use of vision for obstacle avoidance is also plagued with problems regarding robustness and environmental assumptions [1] [2] [3]. To tackle these issues, we approach the problem by developing a flexible framework that is able to fuse multiple visual cues for the task of image classification of traversable surfaces. As such, the completed system can be customised using various visual-cues and environments.

This paper describes the framework developed for the task of obstacle avoidance within an office environment that includes homogeneous surfaces, such as painted walls and carpet floors; as well as textured surfaces, such as patterned vinyl and stone walls. Furthermore, focus is placed on the use of monocular vision, although the framework can incorporate other sensory modalities (i.e stereo-vision). The most commonly used monocular approaches to obstacle avoidance involve the use of appearance-based features [3] [4] [5], or motion information [1] [6]. However, the sole use of either approach contains inherent assumptions and/or is highly sensitive to camera motion [2] [7]. Our approach takes advantage of both information sources, towards the goal of a more robust and adaptive system. In particular, the core of our system contains a learning algorithm (that maintains a population of Gaussian parameters) combined with a Markov Random Field (MRF) that is used to classify, regularise, integrate other visual-based cues (motion, etc), and provide feedback into the learning and modelling system. Similar learning algorithms have already been shown to provide a flexible system toward the task of obstacle avoidance [5] [8]. Moreover, MRF's have been shown to be a powerful yet flexible technique for the integration of multiple data sources [9] [10].

To aid the processing of multiple visual-cues, over-segmented image regions (superpixels) are chosen as the samples for image classification as they have several advantages [8] [11]. In our system, it is envisioned that a ground/obstacle calibration image set is provided for initial learning, from which it will be able to learn, adapt, and perform obstacle avoidance autonomously in a self-supervised learning (SSL) manner. The paper makes four main contributions:

- The development of an adaptive framework and algorithm for the task of mobile-robot obstacle avoidance.
- A novel image segmentation algorithm *Distributed-Fusion (D-Fusion)*, used to cluster superpixels to minimise the data within the learning model.
- A practical evaluation of the use of ground-plane modelling techniques for the task of classification, adaptation, and learning of new obstacles/ground plane surfaces.
- Experimental evaluation of a classification system that combines the use of appearance-based features, regularised over a gradient edge image.

II. Related Work

The abundant amount of information provided by images, coupled with the technological improvements in computing, has spawned the development of a large number of different visual-system approaches [7]. Of particular relevance to the work presented here, are the monocular mapless-based visual navigation systems in the sub-categories of optical-flow, ground-plane detection, and appearance-based approaches.

Stanley [5], winner of the DARPA Grand Challenge 2005, employs a visual-model learning speed control system combining the techniques founded by Ulrich [12] and Thorpe

[13]. Stanley modelled the road terrain using a mixture of Gaussians in the RGB space with training also conducted using a trapezoidal region, although it was verified using a laser sensor. Two recent appearance-based works by Kim [8] and Alencastre-Miranda [14] are of particular inspiration to the system developed here. In Kim's approach, patch-based image regions were compared with superpixel representations for outdoor traversability classification. Kim found that superpixels produced more accurate classifications, able to recognise small thin tree trunks that were not seen in the patch-based results. Alencastre-Miranda created an outdoor colour classification system using MRF's, which allowed for the integration of contextual information of the scene. As such, the system was able to correct mis-classifications and produce correctly labelled images under different lighting conditions during the day.

Another important visual-cue is the image motion sampled using optical-flow. This method is invariant under many assumptions inherently associated with appearance-based methods. Although the qualitative use of optical-flow has had good success for navigation through techniques such as balancing optical-flow fields [15], we focus on those techniques which are beneficial toward the task of image classification. The idea of ground-plane detection using optical-flow is not new [16] [17]. The main idea of these systems revolves around modelling the ground-plane optical-flow field from the perspective of the robot's camera image. Consequently, when the expected flow field is compared to the computed optical-flow field, non-ground regions will produce a high disparity. Work by Chen [18] has used the same ground-plane motion principles. However, as opposed to the calculation of optical-flow on the conventional image format, a reciprocal-polar (R-P) format was used. Thus, coplanar motions in the R-P image space were found to lie on sinusoid, and then fitted to the R-P motion data for ground-plane detection. In a more computational efficient manner, Braillon [19] investigated ground-plane modelling for the detection of dynamic obstacles. Braillon achieved this by using simple similarity measures between the original image and the consecutive image patches shifted according to the expected optical-flow field. The inherent pitfalls that exist with using optical flow and ground-plane modelling include the computational accuracy and time to compute both the real-time optical-flow and expected flow; the frequency of frames required to meet the motion requirements; and the restriction to environments with flat surfaces.

Our approach takes advantages of both the appearance-based and the motion-based principles, and combines them using a flexible and adaptable MRF model. The system makes use of a semi-supervised model for learning and adaptation by requiring a initial *a priori* of the environment. Although the focus in this paper is placed on combining appearance-based features with ground-plane motion information, the system has been developed with a view toward combining many different visual features, which are envisioned to work in synergy to achieve a robust system. The approach here differs from many learning-based approaches because it encompasses the idea of using multiple visual-cues to produce an adaptable self-supervised system.

III. System Outline

Figure 1 shows the system block diagram for our approach. The system can be divided into three main phases, (i) Initialisation Phase, (ii) Operation Phase, and (iii) Update Phase.

– The initialisation phase is a one-off phase that creates a *world model* used for the image classification. The world model consists of a population of hyper-dimensional Gaussian distributions categorised into ground-plane or object space. To construct the initial world model, a *Distributed-Fusion (D-Fusion)* process is employed with the aid of hand-segmented labels of an initial training set. This phase is equivalent to the training phase in many learning algorithms, although it envisioned here to allow the robot to begin in a 'safe' state, from which it can then explore.

– In the operation phase, new images are classified in the MRF system using the world model, regularised using gradient images, and improved using additional visual-cues (ground-plane motion).

– The update phase is highlighted by the dotted-line boxes, and works in parallel with the operation phase to update and feedback classification confidences to the learning model. This phase is employed to update the world model when confident visual-cues exist, providing feedback for system adaptation. This phase is currently in development.

Important overlapping processes in the system's three phases include the description of visual features, and the image segmentation. These processes are heavily dependent on each other, as the image regions must contain significant statistical data of the chosen visual features for the D-Fusion algorithm to be effective.

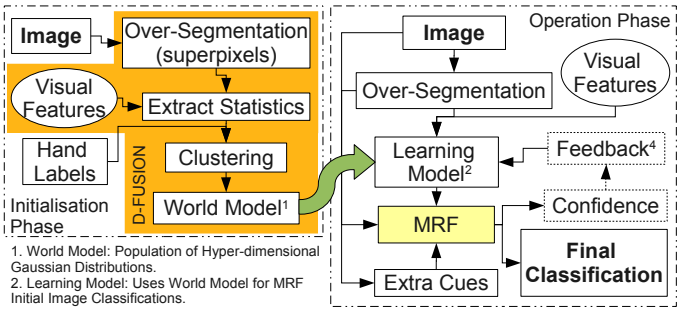

Fig. 1. Overall System Diagram. Left - Initialisation Phase, Right - Operation Phase and Update Phase (dotted-line boxes).

IV. The World Model using D-Fusion

For the learning and construction of the world model, we created a novel clustering/segmentation technique called *D-Fusion*. The motivation for this was four-fold:

1) Biological-inspiration from complex cells learning to distinguish natural scenes using statistical patterns [20].

2) Ability to better encompass spatial-textural properties in a scene.
3) Simplification of computational processing and structure during later stages in classification and regularisation in the MRF.
4) Ability to provide better spatial coherence for the association of additional visual-cues.

Distribution-Fusion is a fusion of the statistical distributions formed from an over-segmentation technique, in this case Gaussians with mean and covariance. D-Fusion consists of three steps, (i) an initial over-segmentation of the image, (ii) a choice of visual features, and (iii) a clustering method. The image regions are the representents of the world model. To support the distributions, the D-fusion is used to limit the redundancy in the world model to create more reliable 'individuals' in the population of Gaussian distributions.

A. Over-segmentation

Over-segmentation allows for the grouping of more coherent pixels over a feature-space, to produce what is also commonly known as a superpixel. From this, statistical representations (mean and covariance) over a number of chosen visual feature-spaces can be extracted from each superpixel. Consequently, these superpixels encompass important spatial statistical features, such as textures. Here, the Efficient Graph-Based Image Segmentation (EGBIS) technique [11] is used to construct the superpixels. EGBIS meets key requirements, such as being fast to compute, and allows control of superpixel size and connectivity, while maintaining the important boundary and object lines within the image. Furthermore, EGBIS also provides an Region Adjacency Graph (RAG) that provides the structure for the MRF model.

B. Appearance Features

Given an over-segmented image, a choice of the visual feature-space is required to aid the fusion of superpixels. Thus, a feature-space must be chosen to describe the environment appropriately, in our case by discriminating the ground from the rest of the scene. By using a Gaussian distribution model, each region is characterised by a hyper-ellipsoid with a mean vector μ_n and a $n \times n$ covariance matrix. As the robot operates in an indoor office environment, the RGB colour-space and texture features computed using 20 Gabor filters (5 freq.$\times$4 orient.) were chosen. Gabor filter frequencies sample a range of high-end frequencies, whilst orientations cover four directions (45° separation). These visual-features were chosen based on the visual-examination of the filtered images and the D-Fusion segmentations over several different feature-space sets.

C. Merging Super-Pixels

With the new feature statistics generated, superpixels must be merged before being included into the world model. As such, superpixels are clustered by means of the hyper-ellipsoid clustering method developed by Kelly [21]. Kelly uses an geometric-based metric r_{ij} to provide an effective distance measure between two hyper-ellipsoid clusters i and j. Thus a threshold R is chosen to determine the final number of clusters. In the initialisation phase, the D-Fusion algorithm iterates through every superpixel over the training images, merging those Gaussians whose $min(r_{ij}) \leq R$. After merging, the new segments mean and covariance are updated accordingly. Although the process of creating an initial world model is computationally costly, once initialised, an incremental approach is taken to update the models. Superpixels are only merged together if they have common labels (as determined from hand-labelled segmentations of the ground/obstacle space). Furthermore, only those Gaussian distributions that are derived from the largest image regions are included in the final world-model. An example D-Fusion segmentation image is shown in Figure 2.

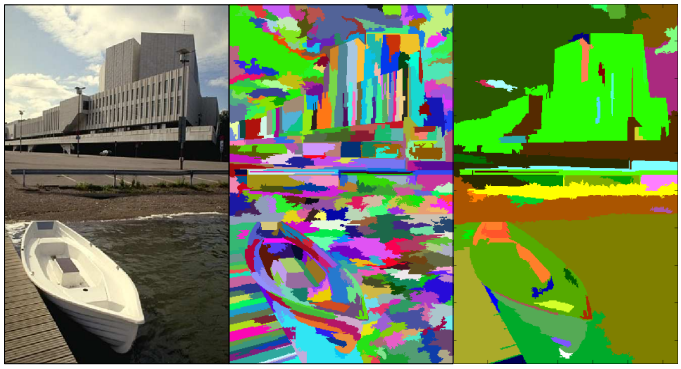

Fig. 2. D-Fusion Segmentation Example. Right: Original, Middle: EGBIS, Left: D-Fusion

V. MRF Classification and Regularisation

For the task of producing the final classification images, a non-regular planar graph MRF model is associated with the superpixels of the image. A MRF system has the benefits of classifying superpixels while considering the their neighbourhoods as well as any additional visual-cues. Figure 3 shows a diagram of the MRF model. Three main information sources are used within the MRF model: the world model Γ, edge strength image x, and a visual-cue confidence image z.

The principle of MRF based classification is to model the final labelling (here ground/non-ground) as the most probable value taken by a random field (here the set of superpixels) under the assumption of local dependence (corresponding here to the adjacency relation between superpixels). This can be modelled by a local potential function defined on every superpixel and its neighbourhood, such that maximizing the probability of a label field is equivalent to minimizing the sum of the corresponding potential function over all superpixels. This optimization is usually performed through an iterative process, here a Gibbs sampler, which operates by modifying iteratively superpixel labels, starting from an initial label field.

A. Initial labelling

As described before, our world model is given by a set of N Gaussian distributions represented by a collection of triplets:

$$\Gamma = \{G_n\}_{n<N} = \{\mu(n), C(n), \lambda(n)\}_{n<N} \qquad (1)$$

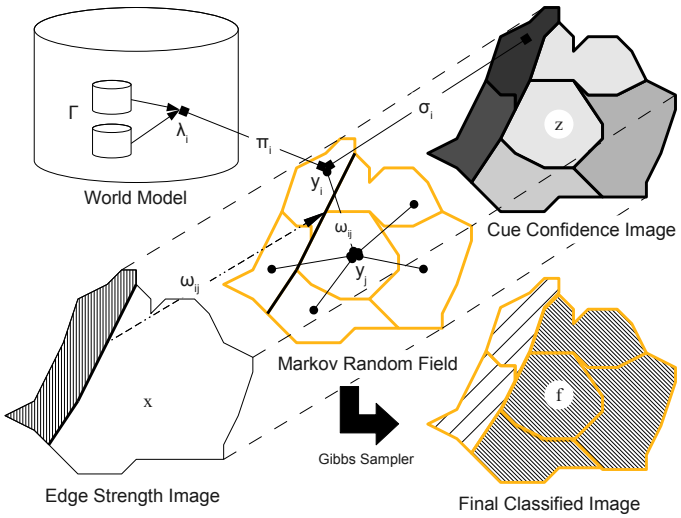

Fig. 3. MRF Model.

where $\mu(n)$ and $C(n)$ are the mean vector and covariance matrix in the feature space, and $\lambda(n)$ is the ground label obtained from the learning phase.

As a trade-off between a fully global Bayesian classification where every label would be represented by one single distribution, and a fully local nearest neighbour method, where every pixel is modelled according to its features, our approach effectively combines the two strategies by classifying a superpixel according to the nearest superpixel in terms of Gaussian distributions. Let R be an unclassified superpixel, with mean vector μ_R in the feature space. The Mahalanobis distance [22] between R and the Gaussian distribution n is defined as:

$$D_M(R, G_n) = \sqrt{(\mu_R - \mu(n))^T C(n)^{-1} (\mu_R - \mu(n))} \quad (2)$$

Given a set of superpixels provided by image segmentation $\{R_i\}_{i \in L}$, the initial (unregularised) classification of every superpixel R_i is then found by:

$$\lambda_i = \lambda(\arg \min_{n<N} D_M(R_i, G_n)) \quad (3)$$

B. MRF Model Description

In the MRF model, a random field Y on a segmented image defines in each superpixel a random variable Y_i, whose value $y_i \in \{-1, +1\}$ represents ground or object space. Every Y_i is associated to a graph node of the RAG provided by the EGBIS segmentation. Every node index i is then associated to a set of neighbours indexes $\mathcal{N}(i)$ such that $\{Y_j\}_{j \in \mathcal{N}(i)}$ are the random variable adjacent (and then dependent) to Y_i.

The world model is first used to obtain classification confidence weights and labels in the MRF system. The MRF also provides smoothing and enforcement of the spatial consistency through using image edge strengths. Additionally, motion ground-cue confidence information developed (but not yet integrated) in Section VI is expected to improve classifications and update the world model. The current MRF is defined through the following equations:

1. The potential component modelling attachment to the data is defined on 1-order cliques, and is of the form:

$$\Psi = -\sum_{i \in L} \pi_i (y_i . \lambda_i) \quad (4)$$

where L is the set of superpixel indexes over the image, y_i is the MRF node (unobserved) mirroring the superpixel i, λ_i is the initial classification label provided by criterion 3, and π_i is the confidence weight associated with that classification label for node i.

1a. The confidence weight π_i for attributing the label λ_i to superpixel R_i is determined through:

$$\pi_i = \frac{|\min_{n<N}\{D_M(R_i, G_n); \lambda(n) = -1\} - \min_{n<N}\{D_M(R_i, G_n); \lambda(n) = +1\}|}{(\min_{n<N} D_M(R_i, G_n))^2} \quad (5)$$

2. The potential component modelling classification smoothness prior is defined on 2-order cliques, and is of the form:

$$\Phi = -\sum_{i \in L} \sum_{j \in \mathcal{N}(i)} \omega_{ij} (y_i y_j) \quad (6)$$

where y_j is the MRF node (unobserved) mirroring the superpixel j, and ω_{ij} is the weighting factor between the two nodes y_i and y_j.

2a. The weighting factor ω_{ij} provides a smoothing link to the image layer through an edge strength image. Weights are used to decrease or increase the smoothing across edges seen in the image. The weights are calculated as follows:

$$\omega_{ij} = \exp(-c * u_{ij}) \quad (7)$$

$$u_{ij} = \frac{1}{\text{Card}(P(i,j))} \sum_{k \in P(i,j)} E(k) \quad (8)$$

where $P(i, j)$ is a set of superpixels boundary coordinates that are common to the superpixels i and j. $Card(X)$ is the cardinality of set X, $E(k)$ is the edge strength at that pixel coordinate k found from the gradient magnitude from Sobel filters, and c is a constant used to control the smoothing.

3. Using these potentials, the conditional distribution over y is:

$$p(Y = y | O = o) = \frac{1}{Z_T} \exp(-\frac{1}{T}(\Phi + \Psi)) \quad (9)$$

where Y is the hidden field, O the observed (image) field, Z_T is the partition function, with T the temperature parameter, which is used in the Gibbs process to control the level of determinism.

Figure 4 shows two test images classified by the MRF system in a seen environment. Table I presents the mean true-positive and false-positive rates for the classification over a set of training and test images, each set containing 25 images.

TABLE I
CLASSIFICATION RESULTS FOR IMAGE SETS.

Image Set	Ground TP-R	Ground FP-R
Training	95.44%	6.22%
Test	90.92%	7.78%

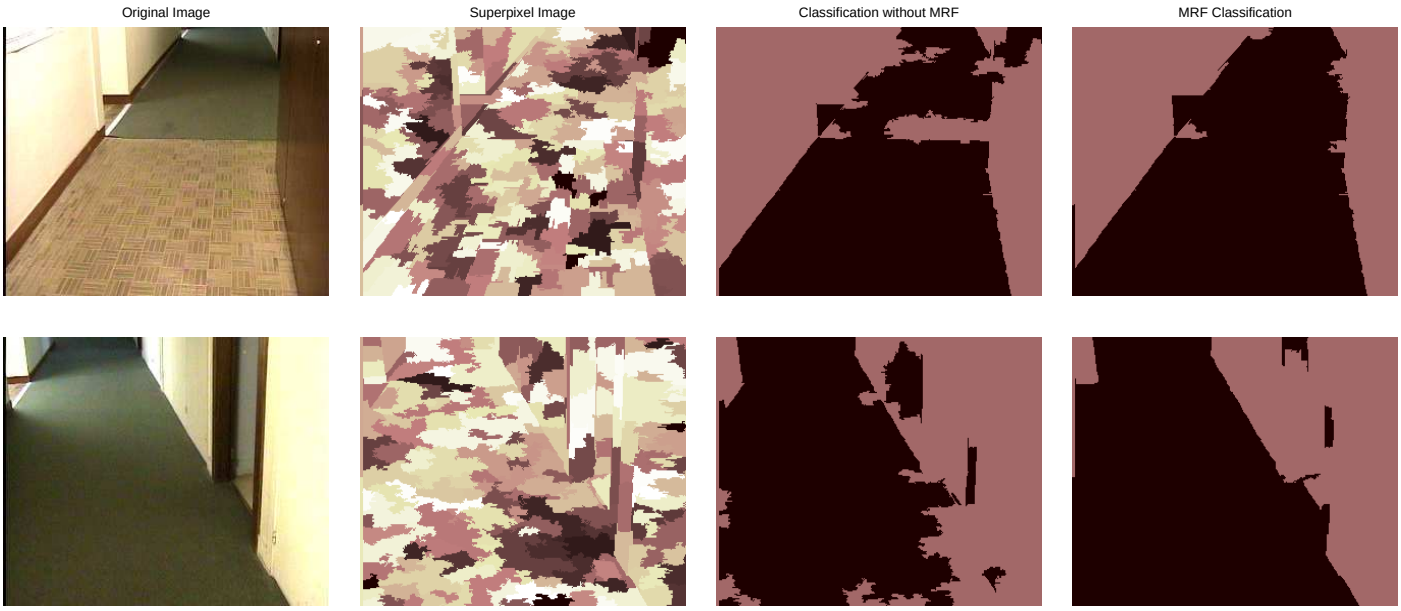

Fig. 4. Example Classification Images.

VI. Learning/Adapting Using Ground-Plane Cues

The fundamental flaw of appearance-based modelling is the introduction of an unseen surface in the image, as seen in Figure 5. Here, a red trolley is unidentifiable using the world model and thus incorrectly classified. In order for the system to learn new features, ground-plane modelling was investigated with a view toward its ability to provide sparse but important obstacle information. The idea behind ground-plane modelling is to fit an expected ground-plane flow field (using odometric data and camera parameters) to the optical-flow field computed from the robot images. From this, areas of high disparity should theoretically correspond to non-ground points in the environment, and vice versa. However, current optical-flow algorithms are not ideal, with trade-offs occurring between noise, accuracy, and computational time [23]. We avoid the explicit computation of the optical-flow field by utilising a technique proposed by [19], that matches patches shifted according to the expected ground-plane motion in the first image, to the corresponding patches in the second image. From these comparisons, areas of low similarity correspond to an obstacle and vice versa.

A. Finding the Expected Motion Field

Before obtaining the expected motion field, the problem of detecting pure translational motion must first be tackled. Even if the robot is commanded to perform a translation motion, in practice the robot and camera motion instabilities corrupt the image motion. Thus, in order to detect a pure translation camera motion, a similar technique to [18] is used to estimate the Focus of Expansion (FOE) of image motion. By estimating the FOE and comparing it to the expected FOE from pure translation camera motion, bad frames can be discarded. To detect the FOE, a Harris corner detector combined with a pyramidal Lucas and Kanade optical-flow technique is first

Fig. 5. Classification with Unseen Obstacle.

employed. Flow vectors are then extrapolated to an intersection point. Provided the mean and variance of the intersection points are within the given thresholds, the similarity image is computed. The expected motion field of the ground-plane is then calculated through the homography described in [19].

B. Similarity Image and Confidence Computation

To compare the similarities between the original image patch and the shifted image patch, the mean-centered correlation measure was employed. To diminish the effect of less informative surfaces that contain low or no gradient changes, the measure was multiplied by the standard deviation of each patch. The construction of the confidence cue image is defined as follows:

$$\zeta_i = \frac{Card(\{s \in S_i; s > T\})}{Card(S_i)} \tag{10}$$

where ζ_i is the confidence of node/superpixel i, $Card(X)$ is the cardinality of set X, s is the similarity measure, T is a similarity threshold, S_i indexes all pixels in superpixel i. The final confidence cue image relates to the degree of which an

obstacle is present. Figure 6 shows the creation of the cue confidence image for an image that met the imposed motion requirements, and demonstrates a frame that can be used to teach and update the world model to identify the new obstacle seen in Figure 5.

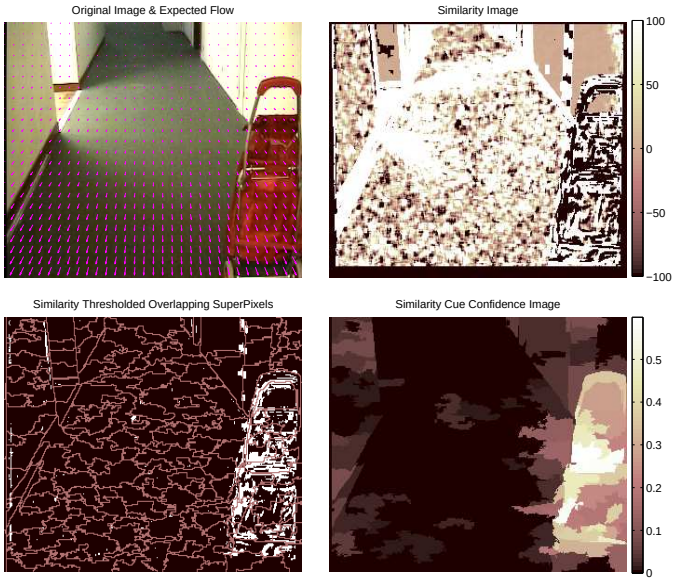

Fig. 6. Cue Confidence Image Creation - Corridor.

VII. Discussion and Future Work

Our results from the current MRF implementation demonstrates some promising abilities for traversability classification, with test-set images averaging a true-positive to false-positive rate of 90.92% to 7.78% respectively. Such results can be attributed to the use of the D-fusion for creation of the world model, that is able to identify several types of different obstacle surfaces alongside the two distinct ground-plane surfaces. Furthermore, in difficult areas with distinct lighting changes or similar ground/obstacle surfaces, the MRF is able to perform regularisation using gradient images of the scene, to form an overall better classification image more usable for obstacle avoidance.

Additionally, for the task of adapting and learning new surfaces in a scene, similarity/disparity images found through using ground-plane motion modelling have also shown great potential in improving MRF classifications and training of the world model. The ground-plane cue confidence images here are able to highlight areas of high-obstacle probability. These probability calculations can be used within in the MRF and feedback loop to adapt the world model, with the help of image classification confidences. Ground-plane cues are only envisioned as one of the many visual-cues used in the system. The flexibility of the developed system structure allows it to take advantage of a number of weak visual-features to improve its overall reactivity and robustness.

Future work includes: (i) Integration of the ground-plane cue confidence images into the MRF system. (ii) Creation of a learning algorithm using ground-cues to update the world-model. (iii) Investigation of other visual-cues that can be integrated into the system, in particular, vertical line cues that may help in identifying homogeneous ground-plane regions.

References

[1] T. Camus, D. Coombs, M. Herman, and T.-H. Hong, "Real-time single-workstation obstacle avoidance using only wide-field flow divergence," in *Journal of Computer Vision Research*, 1996, pp. 323–330.

[2] T. Low and G. Wyeth, "Obstacle detection using optical flow," in *Proceedings of the 2005 Australasian Conf. on Robotics & Automation*, Syd, Australia, December 2005.

[3] J. Michels, A. Saxena, and A. Y. Ng, "High speed obstacle avoidance using monocular vision and reinforcement learning," in *Proc. of the 22nd Int. Conf. on Machine Learning*. NY, USA: ACM, 2005, pp. 593–600.

[4] Y. Guo, V. Gerasimov, and G. Poulton, "Vision-based drivable surface detection in autonomous ground vehicles," in *Int. Conference on Intelligent Robots and Systems*, October 2006, pp. 3273–3278.

[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*. The MIT Press, 2006.

[6] Y. G. Kim and H. Kim, "Layered ground floor detection for vision-based mobile robot navigation," in *Proc. of the ICRA*. IEEE, 2004, pp. 13–18.

[7] F. Bonin-Font, A. Ortiz, and G. Oliver, "Visual navigation for mobile robots: A survey," *J. Intell. Robotics Syst.*, vol. 53, no. 3, pp. 263–296, 2008.

[8] D. Kim, S. M. Oh, and J. M. Rehg, "Traversability classification for ugv navigation: a comparison of patch and superpixel representations." in *IROS*. IEEE, 2007, pp. 3166–3173.

[9] J. Diebel and S. Thrun, "An application of markov random fields to range sensing," in *Proceedings of Conference on Neural Information Processing Systems (NIPS)*. Cambridge, MA: MIT Press, 2005.

[10] A. Saxena, M. Sun, and A. Y. Ng, "Learning 3-d scene structure from a single still image," in *IEEE 11th Int. Conference on Computer Vision*, 2007, pp. 1–8.

[11] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vision*, vol. 59, no. 2, pp. 167–181, 2004.

[12] I. Ulrich and I. R. Nourbakhsh, "Appearance-based obstacle detection with monocular color vision," in *Proceedings of the 17th National Conf. on A.I. and 12th Conf. on Innovative App. of A.I.*, 2000, pp. 866–871.

[13] C. Thorpe, M. Hebert, T. Kanade, and S. Shafer, "Vision and navigation for the carnegie-mellon navlab," *IEEE Trans. on Pattern Analysis & Machine Intelligence*, vol. 10, no. 1, pp. 362–373, May 1988.

[14] M. Alencastre-Miranda, L. Munoz-Gomez, R. Swain-Oropeza, and C. Nieto-Granda, "Color-image classification using mrfs for an outdoor mobile robot," *J. Systemics, Cybernetics & Informatics*, vol. 3, p. 52, 2004.

[15] D. Coombs, M. Herman, T. Hong, and M. Nashman, "Real-time obstacle avoidance using central flow divergence and peripheral flow," in *IEEE Transactions on Robotics and Automation*, 1995, pp. 276–283.

[16] W. Enkelmann, "Obstacle detection by evaluation of optical flow fields," in *ECCV 90: Proceedings of the first european conference on Computer vision*. NY, USA: Springer-Verlag NY, Inc., 1990, pp. 134–138.

[17] M. Zucchelli, J. Santos-Victor, and H. I. Christensen, "Multiple plane segmentation using optical flow," in *British Machine Vision Conf*, 2002, pp. 313–322.

[18] Z. Chen, N. Pears, and B. Liang, "Monocular obstacle detection using reciprocal-polar rectification," *Image Vision Comput.*, vol. 24, no. 12, pp. 1301–1312, 2006.

[19] C. Braillon, C. Pradalier, J. Crowley, and C. Laugier, "Real-time moving obstacle detection using optical flow models," in *Proc. of the IEEE Intelligent Vehicle Symp.*, June 2006, pp. 466–471.

[20] Y. Karklin and M. Lewicki, "Emergence of complex cell properties by learning to generalize in natural scenes," *J. Nature*, November 2008.

[21] P. M. Kelly, "An algorithm for merging hyperellipsoidal clusters," Los Alamos National Laboratory, Tech. Rep., 1994.

[22] R. De Maesschalck, "The mahalanobis distance," *J. of Chemometrics & Intell. Lab. Systems*, vol. 50, no. 1, pp. 1–18, January 2000.

[23] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, "Performance of optical flow techniques," *Int. J. of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.