2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Classification of Natural Scene Multi Spectral Images using a New Enhanced CRF

Mohammad Najafi [1,2] Sarah Taghavi Namin[1,2] Lars Petersson[1]
[1]National ICT Australia (NICTA)*, Locked Bag 8001, Canberra ACT 2601, Australia
[2]The Australian National University (ANU), Canberra ACT 0200, Australia
{Mohammad.Najafi,Sarah.Namin,Lars.Petersson}@nicta.com.au

***Abstract*—In this paper, a new enhanced CRF for discriminating between different materials in natural scenes using terrestrial multi spectral imaging is established. Most of the existing formulations of the CRF often suffer from over smoothing and loss of small detail, thereby deteriorating the information from the underlying unary classifier in areas with a high spatial frequency. This work specifically addresses this issue by incorporating a new pairwise potential that is better at taking local context into account. Certain materials are very unlikely to appear next to each other in the scene and such configurations are penalised by employing the confusion matrix of the unary classifier. Similarly, horizontal as well as vertical configurations, which may be more or less likely for certain combinations of materials, are regarded in this formulation. Furthermore, the proposed pairwise potential also considers the length of boundaries between regions to account for the segmentation granularity issues and also uses class probabilities of the neighbouring regions to make up for the uncertainty of the unary classifier results. Seven band terrestrial multi spectral imaging were used due to its potential in distinguishing between different materials and objects. The proposed approach was evaluated using cross-validation, resulting in an average accuracy of 88.9% which is about 17% more than the accuracy of a standard CRF, which demonstrates the superiority of our approach in preserving local details.**

***Index Terms*—Multi Spectral, Classification, Fuzzy SVM, CRF, Pairwise Potential, Confusion Matrix**

I. Introduction

With the advent of multi/hyper spectral imaging in the past decades, a vast number of applications have benefited from the potential of this powerful imaging modality. A multi spectral image can reveal some of the properties of objects and materials, which can not otherwise be observed using conventional cameras, thanks to more frequency bands in the visible and invisible parts of the spectrum. Such an ability makes this type of data a great asset for object and material classification tasks.

In this paper, a new approach for preserving fine detail in the detection and classification of roadside materials using multi spectral imaging is devised. The resulting system has many applications in road and roadside objects assessment and robotics.

*NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

Multi spectral imaging has been widely used in land cover and environment classification using aerial surveying [1], [2], but there are also a limited number of works on terrain classification, in which, terrestrial multi spectral images have been employed. In [3], [4] NDVI (Normalised Difference Vegetation Index) feature was used to detect vegetation in the environment. Terrain classification was investigated in more detail by Taghavi et al. in [5], where the road side objects and materials were classified to 10 categories, using 7 band terrestrial multi spectral images. They utilised pixel-wise texture features such as GLCM and Fourier spectrum to make the system more robust to varying lighting conditions. Although they came up with some satisfactory results, the pixel-wise nature of their work has made their system vulnerable to noise and also somewhat slow and inefficient, especially for working with high resolution images. These issues can be addressed by combining similar pixels into regions and also benefiting from the neighbouring information.

One approach for exploiting this kind of information is CRF (Conditional Random Field), which is a probabilistic framework [6] and has been widely used in a number of multi spectral and RGB image classification systems [7]–[13]. CRF was applied to aerial spectral images in [7] and [8], but as it was stated by the authors, the undesired smoothing property of CRF is a challenging problem and it is even worse for the terrestrial images, which embody much more details. Other groups have tried to enrich the CRF framework with some discriminative terms to make it more "intelligent" in dealing with complex circumstances of neighbourhoods. Yang and Forstner in [9], worked on a region-wise building facade classification task for detection and recognition of different categories such as road, vegetation, sky, pavement, etc. They embedded the image location information into the their proposed system and then further improved it by proposing a hierarchical CRF framework [10]. However, there are a considerable number of misclassifications in their results which probably indicates that more information is required in order to handle the small details in the image and high similarities between the objects. Wojek and Schiele [11] incorporated the temporal information of the scenes into a dynamic CRF model to address the problem of over smoothing in classification of large scale categories such as road, grass, car, trees, etc, though this dynamic approach requires successively captured

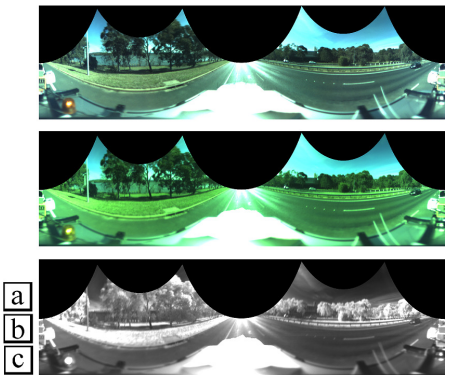

Fig. 1. A sample 7 band multi spectral image. a) Three band RGB image. b) Three band shifted RGB image in the visible part of the spectrum. c) NIR image in the invisible part of the spectrum.

images. Another issue of conventional CRF algorithms is that the pairwise function often depends on the absolute labels of the neighbours, which may lead to some incorrect context inference [12], [13].

In this paper, we propose an enhanced CRF framework which addresses existing problems in region-wise terrain classification using CRF, in particular that of preserving fine detail by preventing over smoothing.

Our main contributions are:

- Proposing a novel pairwise function that outperforms the standard CRF pairwise functions in terms of classification accuracy and preserving details.
- Introducing an efficient way to present the neighbourhood graph of the CRF in a region-based image classification problem.

The approach is tested on terrestrial multi spectral images of road and road side scenes, captured in seven frequency bands including six visible bands and one NIR band, which is the same setup as in [5]. The imaging system is composed of a *FluxData*^{TM} camera along with a panoramic mirror (*GoPano+*) which provides a full 360 degree view. The resulting panoramic images are then dewarped using a post-processing software which is included in the *GoPano+* package. The outputs of this step are multi spectral images of 1241×4176 pixels in size (Fig. 1). The whole dataset consists of 1497 multi spectral images.

In this work, instead of working on image pixels, we partition each image into superpixels to reduce the computation time of the algorithm and also get more meaningful context information. Then we extract the features of each superpixel and classify them into different material categories using a fuzzy SVM classifier [14]. Finally we use the SVM results as input to our proposed CRF which takes the local context into account, thereby further improving the classification system.

II. Conditional Random Field

A Conditional Random Field is used to express a probabilistic model which attempts to predict the label of a region, given information about that region as well as its neighbours. A complete description of CRF can be found in [6]. A CRF with *unary* and *pairwise* terms is expressed with the following standard equation [13]:

$$\mathbf{P}(\mathbf{Y}|\mathbf{X},\phi,\psi) = \frac{\mathbf{exp}\Big(\sum_{i=1}^{M}\big[\mathbf{\Phi}(\mathbf{x_i},\mathbf{y_i},\phi)+\sum_{j\in N_i}\mathbf{\Psi}(\mathbf{x_i},\mathbf{x_j},\mathbf{y_i},\mathbf{y_j},\psi)\big]\Big)}{\mathbf{Z}(\phi,\psi,\mathbf{X})} \tag{1}$$

in which $\mathbf{Y}$ is the set of data labels to be predicted, $\mathbf{X}$ is the set of features extracted from the data and $\mathbf{Z}$ is the partition function. In this equation, M is the number of data items, N_i is the neighbourhood space of data and $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are the unary and pairwise potentials with ϕ and ψ as their parameters, respectively.

The unary term which is sometimes called *local potential*, associates the data features with the labeling set. In other words, it indicates the chance of selecting a label for a data item, solely based on the features of that item. In contrast, the pairwise potential determines how the neighbouring labels can influence each other.

The ultimate goal is to find the best compromise between these two terms to maximise the classification results. In order to achieve this, the CRF should be trained to maximise the probability in Eq. (1) for the true labels. This problem turns into an energy minimisation problem by taking the negative logarithm of this probability:

$$\mathbf{E}(\mathbf{Y}|\mathbf{X},\phi,\psi) = \log\big(\mathbf{Z}(\phi,\psi,X)\big) - \sum_{i=1}^{M}\big[\mathbf{\Phi}(\mathbf{x_i},\mathbf{y_i},\phi)+\sum_{j\in N_i}\mathbf{\Psi}(\mathbf{y_i},\mathbf{y_j},\mathbf{x_i},\mathbf{x_j},\psi)\big] \tag{2}$$

The above energy or cost will be minimised for the optimal labeling of the data.

III. Approach

Our approach is comprised of the following main steps. Initially, the images undergo a region segmentation process. Then, the appropriate features are extracted from each region. In the next step, the regions are classified into some predefined labels using a probabilistic SVM classifier. Subsequently, a new CRF formulation is devised and applied to the system to gain the final classification results. At the end, *saturated* and *vague* regions in the image are classified into the most relevant categories. The steps of our approach are depicted in Fig. 2.

A. Segmentation

In the first step, the images are segmented into regions covering uniform areas, so called superpixels. The reason for this is two fold; Firstly, contextual information is more prominent considering larger regions of similar appearance rather than just looking at individual pixels and their neighbours only. Secondly, if the number of regions for which a label needs to be estimated can be reduced significantly, it also means that the overall computational need is similarly shrunk.

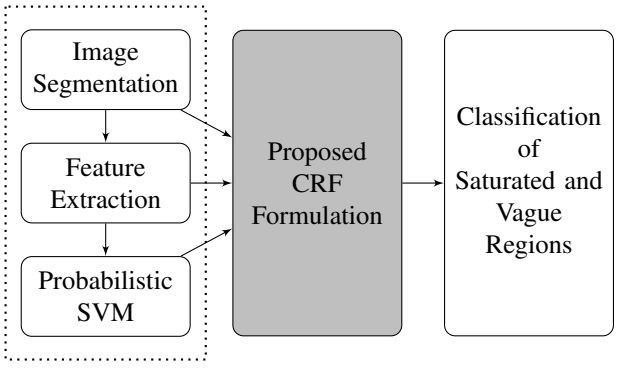

Fig. 2. The overall view of the proposed approach. After the segmentation step, some discriminative features are extracted from the image regions. Then, the regions are classified using a probabilistic SVM classifier. Next the proposed CRF is applied to the classification results and finally, appropriate class labels are assigned to the Saturated and Vague regions.

We firstly segment the whole image into two major parts of vegetation and non-vegetation using NDVI feature [3] in order to achieve more consistent regions. Afterwards, we sub-segment the RGB images of each part into superpixels using the Mean Shift algorithm [15]. In total, the algorithm is tuned to produce around 2000 to 3000 regions for each multi spectral image. A sample of a segmented image can be seen in Fig. 7-a.

B. Feature Extraction

Three types of features that are incorporated into our system, are explained in this section.

1) Mean and Standard Deviation: The first 14 feature types are the Mean and Standard Deviation of the intensity values of each region, computed for each band. It should be noted that all the features undergo a normalisation process to have a mean of zero and a standard deviation of one.

2) GLCM: There are different techniques for extracting the texture of an image, among which, GLCM (Gray Level Co-Occurrence Matrix) [16] has been one of the most popular methods in the past decades. This algorithm is very powerful and easy to implement. The advantage of this matrix is that it is independent of gray-level scalings, which makes it very useful in recognising similar textures under different lighting conditions [5]. Each component of GLCM stands for the number of occurrences of a specific adjacency for a pair of gray levels. Here we consider vertical and horizontal pixel adjacencies within each region and extract three properties of Contrast, Energy and Homogeneity from each of the two computed GLCMs. In total, 6 features from each spectral band are obtained for each region, which results in 42 GLCM features.

3) Histogram of Hough Orientations: Among the roadside material categories that we consider, there are some classes like *Light Poles and Road Guards* and also *White Lines* which typically show up as parallel lines in the images. This property can be exploited as a clue for detection and recognition of these types of categories. For this purpose, the Hough Transform is applied to the image region and after setting a relative threshold (50%) on the intensity in Hough space, the main orientations of the local edges inside the region are identified. Then a 7-bin histogram of these orientations ranging from 0 to 180 degrees is calculated for the region. A uniform histogram indicates that there is no major set of parallel edges in the region. Conversely, a sparse histogram implies that the region contains an object or part of an object with a significant set of parallel boundaries.

C. SVM Classification

An SVM can be used to categorise non-linearly separable data points by using appropriate kernels. We use fuzzy SVM with an RBF (Radial Basis Function) kernel [5]. The fuzzy SVM constructs a probabilistic model using the training data, which is later used to predict the class label of unknown data. We employ the method presented in [14] to compute the SVM probabilistic outputs. The *LIBSVM* toolbox [17] in *MATLAB*^{TM} is used for the training and evaluation of SVM.

D. Proposed CRF framework

In this section, we first describe our CRF neighbourhood graph. Then we define each potential function and at the end, explain the CRF training and inference processes.

1) Neighbourhood Graph: The CRF operates on a graph describing the connectivity between neighbouring regions. This connectivity graph is built using an approach based on GLCM to find the neighbours. First, an identical and unique intensity is given to all the pixels within each region. The result will be an image with N gray levels, where N is the number of regions. The GLCM of this image indicates the number of occurrences of the adjacencies between each pair of gray-level intensities. Since each region is represented by a unique intensity, the neighbourhood relationships of the regions can also be determined using the computed GLCM. By performing this process for both horizontal and vertical adjacencies using GLCM and comparing the adjacency values for a pair of neighbouring regions in the two resulting GLCMs, one can determine if the two regions are largely horizontally or vertically adjacent. These two different modes of adjacency are treated differently which provides direction dependent context information. Fig. 3 demonstrates the process of finding the neighbourhood graph for an example image with 9 regions. A sample GLCM for horizontal neighbourhood is illustrated in this figure.

2) Unary Potential: The unary term computes the cost of selecting a label for each region based on its features. This cost should be higher for labels that have a lower class probability. Here we take the negative logarithm of the probabilistic output of the fuzzy SVM classification to adapt it as the unary cost function:

$$\mathbf{\Phi}(\mathbf{y_i}, \mathbf{x_i}) = -\log(P(\mathbf{y_i}|\mathbf{x_i}) + \epsilon) \tag{3}$$

The probability score is generated using the approach presented in [14]. An ϵ is added to the equation to ensure a non-zero value as the input of the logarithm.

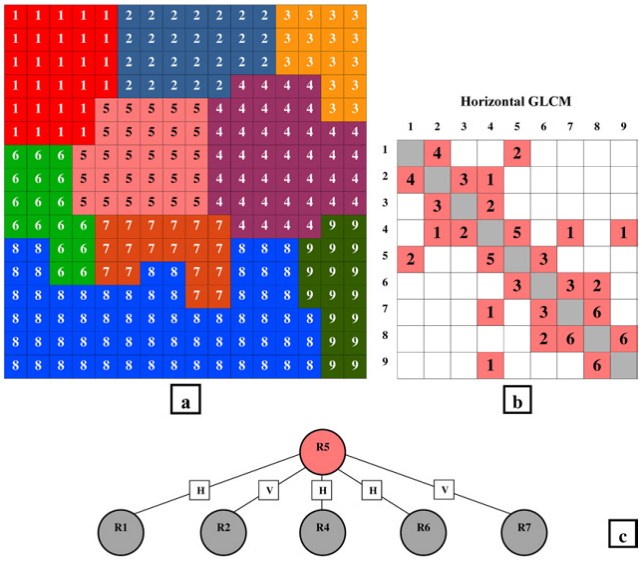

Fig. 3. The procedure of finding the horizontal neighbourhood graph for the segmented regions. a) A sample image segmented into 9 regions. A unique gray-level intensity is assigned to the pixels of each region. b) The horizontal GLCM of the image reveals the number of occurrences of horizontal adjacencies between each pair of gray-level intensities, which in turn indicates the neighbourhood of the regions (Vertical GLCM is not shown). c) The CRF graph for region 5. The adjacency direction between each pair of regions (V or H) is computed by comparing the number of adjacent pixels in each direction, or comparing their corresponding horizontal GLCM and vertical GLCM value.

3) Pairwise Potential: In this work, we propose a comprehensive pairwise potential,

$$\begin{aligned}\mathbf{\Psi}(\mathbf{y_i},\mathbf{y_j},&\mathbf{x_i},\mathbf{x_j}) = \\ &\big(1-h_i(\mathbf{y_i})\big)\Big(\frac{\psi_1 D_{ji}}{1+\|(\mathbf{x_i}-\mathbf{x_j})\|}\Big)\delta(\mathbf{y_i}\neq\mathbf{y_j}) \\ &+\psi_2\Big((1-P_j(\mathbf{y_j}))\Big)\end{aligned} \quad (4)$$

which takes into account several factors as described below. This pairwise function is superior over a standard CRF in terms of classification accuracy and also preserving image details against the over smoothing property of the CRF.

i) Feature similarity: The term $\frac{1}{1+\|\mathbf{x_i}-\mathbf{x_j}\|}$ considers the feature difference between two regions. As a result, a higher cost will be assigned if the regions with similar features attain different labels.

ii) Smoothing term: $\delta(\mathbf{y_i}\neq\mathbf{y_j})$ acts as a smoothing term which prefers identical labeling for neighbouring regions.

$$\delta(\mathbf{y_i}\neq\mathbf{y_j}) = \begin{cases} 1 & \mathbf{y_i}\neq\mathbf{y_j} \\ 0 & o.w \end{cases} \quad (5)$$

iii) Uncertainty of the neighbouring labels (P_j): The major problem which is raised by the delta function is that the algorithm favors a label similar to the adjacent regions, regardless of how certain we are about the neighbours' labels. In other words, it assumes that the neighbours are correctly labeled, which may lead to over smoothing in some regions. To tackle this problem, we insert $(1-P_j(\mathbf{y_j}))$ as a function of unary probability of the neighbour in the pairwise term to make it more knowledgeable about the surrounding regions. Therefore, the system gives more cost to identical labeling with the adjacent regions that have a rather low class probability.

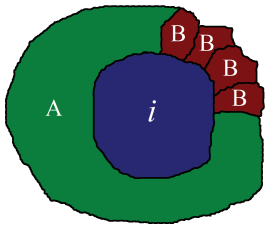

Fig. 4. An example that illustrates the effect of neighbourhood length. Region i has more neighbouring regions from class B, but it has a longer adjacency with class A, so class A should have a stronger influence on it.

iv) Neighbourhood length (h_i): A problem in the standard region-based pairwise potentials is that each of the surrounding regions are treated equally, irrespective of their amount of neighbourhood. As indicated in Fig. 4, region i has four neighbours of class B and one neighbour of class A. In consequence, the effect of class B on region i is about four times greater than the influence of class A on region i. However, it does not seem like a fair decision, since the length of neighbourhood between i and A is much larger than the neighbourhood length of i and B.

We embed $(1-h_i(\mathbf{y_i}))$ in the pairwise term to compensate for this problem. The parameter $h_i(\mathbf{y_i})$ stands for the proportion of the boundary pixels from a neighbouring region with the class label of $\mathbf{y_i}$, to all of the boundary pixels of the region. For example, in Fig. 4, $h_i(A)$ is greater than $h_i(B)$, where A and B are the class labels of the neighbouring regions. Adding this term to the pairwise equation will decrease the pairwise cost of selecting class A and increase its effect on region i.

v) Local context matrix (D): The image context provides a rich source of information which can be utilised in the pairwise potential to improve the classification accuracy. For this purpose, a contextual cost matrix can be devised in order to take the relationships between the neighbouring regions with different class labels, into account. Such a matrix can be designed by setting variable parameters and finding their optimal values via a minimisation process. However, it might lead to over-fitting due to the large number of parameters and also the high level of complexity in our image dataset. These parameters can also be assigned manually, but it requires a deep knowledge of the application and also much trial and error.

Here, we simplify this problem by using the confusion matrix of the SVM classifier to build a contextual cost matrix. The confusion matrix indicates the number of misclassifications for each pair of labels. A large value for a non-diagonal component $(\mathbf{y_i},\mathbf{y_j})$ shows that these two labels have a significant conflict with each other. In order to diminish this misclassification error, a large cost, proportional to the conflict rate, is needed for the interaction of these labels. To this aim, we utilise the normalised confusion matrix to build such a cost matrix as the local context matrix D.

For instance, in our experiments, there might be some regions on grass that are incorrectly classified into leaves and should be turned into grass using CRF. A possible side effect of this process is disappearing of, eg, a thin sign pole beside the grass due to over smoothing. This problem can be resolved by incorporating the matrix D into the CRF. If the misclassification rate of grass and leaves is higher than that of grass and the sign pole (which is in our experiments), the smoothing effect of the CRF would be more significant on leaves rather than on the sign pole and the pole is more likely to be preserved.

vi) Horizontal and vertical pairwise potentials: In this paper, we consider two types of horizontal and vertical neighbourhoods for the regions. Hence, we have two pairwise potentials that consider the interactions of the region with the horizontal and vertical neighbours, separately:

$$\mathbf{E} = \log(\mathbf{Z}) - \sum_{i=1}^{M} \left[\mathbf{\Phi}(i) + \sum_{j \in {N_i}^H} \mathbf{\Psi_H}(i,j) + \sum_{j \in {N_i}^V} \mathbf{\Psi_V}(i,j) \right] \tag{6}$$

ψ_{1H}, ψ_{1V}, ψ_{2H} and ψ_{2V}: These parameters determine the degree of contribution of each term in the pairwise potential, and will be trained in the next step.

The aggregate pairwise potential is very flexible and can compromise very well between keeping correctly classified details and smoothing out the wrongly classified regions.

4) CRF Training: In the training step, the aim is to optimise the CRF parameters in a way that the true labels become the most probable labels in the training data (Eq. (1)). In other words, the energy in Eq. (2) should be minimised for the training data.

One of the most notable approaches for this type of optimisation is *Maximum Log-Likelihood*. However, the problem with this method is that the computation of global partition function $\mathbf{Z}(\phi, \psi_1, \psi_2, \mathbf{X})$ is intractable in our case. In order to resolve this issue, various approximations have been proposed [18], among which, *Maximum Pseudo-Likelihood* has been reported to be one of the most efficient approaches with satisfactory results [18]. In this method, Eq. (7) which represents the pseudo-likelihood objective function, is maximised during training and in the limit, it results in the same values of parameters as maximum likelihood:

$$\mathbf{P}(\mathbf{Y}|\mathbf{X}) \simeq \prod_{i=1}^{M} P_i = \prod_{i=1}^{M} \frac{\mathbf{exp}\left[\mathbf{\Phi}(\mathbf{x_i}, \mathbf{y_i}) + \sum_{j \in N_i} \mathbf{\Psi}(\mathbf{y_i}, \mathbf{y_j}, \mathbf{x_i}, \mathbf{x_j})\right]}{\sum_L \mathbf{exp}\left[\mathbf{\Phi}(\mathbf{x_i}, \mathbf{y_i}) + \sum_{j \in N_i} \mathbf{\Psi}(\mathbf{y_i}, \mathbf{y_j}, \mathbf{x_i}, \mathbf{x_j})\right]} \tag{7}$$

The optimal parameters can be computed by taking the negative logarithm of this probability function and then performing an energy minimisation process on the training data. We use the *trust-region-reflective* algorithm [19] in *MATLAB*^{TM} to find the optimal values for ψ_{1H}, ψ_{1V}, ψ_{2H} and ψ_{2V}.

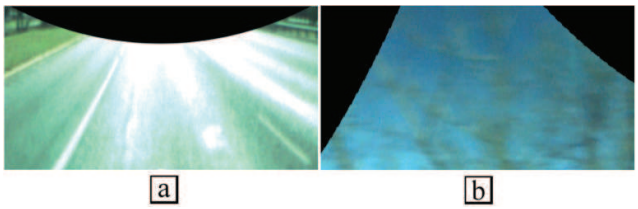

Fig. 5. a) An example of saturation in the image, where no information about the material can be extracted from these regions (extracted from the centre of Fig 7-a). b) Sample vague regions around the tree branches.

5) CRF Inference: This step uses the previously learnt CRF model to predict the labels of unseen regions. Due to the non-submodularity of the pairwise function, we use the ICM (Iterated Conditional Modes) approach for the inference. The following procedure (inspired by [20]) is chosen due to its straightforward inference concept and also its high convergence speed.

1. Set the probabilistic outputs of the SVM as the initial probabilities of regions. Also set the SVM labeling outputs as the initial labeling of the regions.
2. Update the probabilities P_i from Eq. (7) for all regions in the image and also update the region labels by finding their maximum class probability.
3. Repeat step 2 until no change in the labels of the regions is observed or the maximum number of iterations is reached.

E. Saturated and Vague Regions

We ultimately aim to classify the images into the first 10 prominent categories in Table I. However, as it is apparent in Fig. 5-a, there are some saturated regions in the image (such as the mid parts of the road) which does not convey any useful information about the scene. This makes it very hard for the classifier to recognise the real materials and objects in those regions. In order to handle this issue, we define a new class for these saturated regions to discriminate them from other parts of the image.

In addition, there are some vague regions between different adjacent materials in the images. This is due to the partial averaging effect of the pixels in these areas, predominantly found around leaves and tree branches. Fig. 5-b shows an example of such obscure regions where it is uncertain if the pixels belong to sky or tree branches. Labeling these regions as belonging to one or the other class is often impossible, even manually, so we assign a new label to be in charge of these uncertain regions.

The above class labels do not represent any real world object or material, so we may utilise the information from their neighbourhood to identify them. Since the system is optimised for a 12 class problem, CRF is unable to dissolve all the regions from these two categories into the 10 real classes. A comparison of the accuracies of these two categories in Tables II and III together with Fig. 7-c, shows that CRF merges some of these saturated and vague regions into one of the 10 main categories, but not completely.

To handle this problem, once the classification of all 12 classes is finished, the actual materials in the saturated and

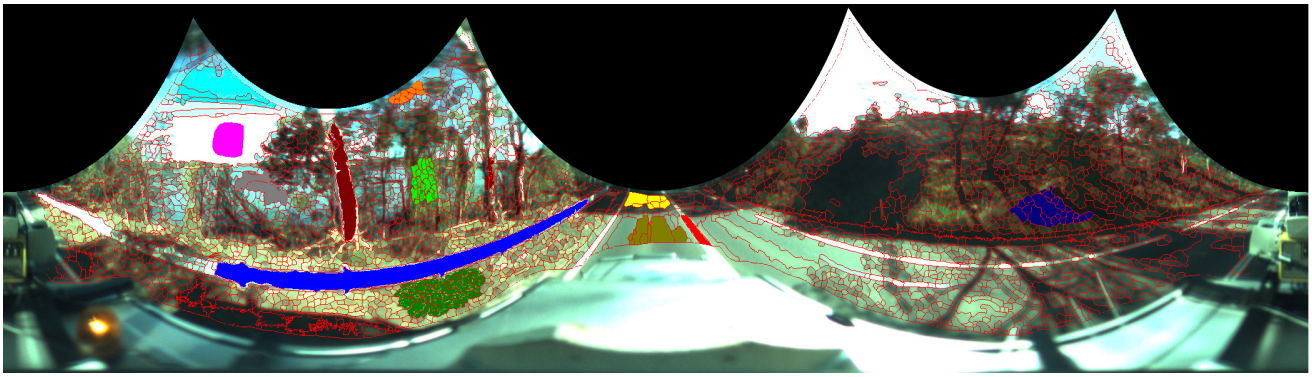

Fig. 6. A manually labelled image with colour codes for 12 classes as used in the classification system (Table I).

TABLE I
THE LIST OF THE TARGETS IN THE CLASSIFICATION SYSTEM

1 - Tree Trunks: Dark Brown	7 - Shadow on Road: Yellow
2 - Light Poles,Road Guards: Blue	8 - Leaves: Green
3 - Shadow on Grass: Dark Blue	9 - Sky: Light Blue
4 - Grass: Dark Green	10 - Lake: Gray
5 - Road: Brown	11 - Saturated Regions: Purple
6 - White Lines on Road: Red	12 - Vague Regions: Orange

vague regions are inferred using the information of their surrounding regions that have real world class labels. This is done by applying a majority voting rule to the labels of adjacent pixels from the neighbouring regions.

IV. EXPERIMENTAL RESULTS

We classified the materials in the environment into 10 primary classes and also dedicated two extra classes to the saturated and vague regions (Table I). Fig. 6 displays a manual labeling for a sample image in which colours of the classes are chosen according to Table I. Although texture features are to some extent tolerant against different lighting conditions, we introduced two classes for *Shadow on Grass* and *Shadow on Road* to facilitate the classification [5].

For this experiment, 90 multi spectral images were randomly selected and the performance of the proposed approach was investigated using a three-fold cross-validation. The images were divided into 3 random partitions of 30 images and three validation runs were performed.

Initially, the images underwent a segmentation process using the NDVI feature and Mean Shift algorithm, where each image was segmented to 2500 regions on average. After the segmentation process, manual labeling was performed for some selected regions in all the images to constitute the datasets. Then, 1041 regions from each class were randomly selected for the classification process. Thereafter, a feature extraction process was applied to these data items to obtain 70 features from each region as described in Section III-B.

In the next step, three-fold cross-validation using the fuzzy SVM classifier ($C = 3$, $gamma = 0.0189$) was performed on the above data, which resulted in an average accuracy[1] of 84.2% for the ten primary classes. It should be noted that in each fold, the same number of samples from each class were taken into account.Table II presents the confusion matrix which was computed by the SVM classification.

[1]Accuracy: The number of correctly classified data divided by the total number of data

TABLE II
THE CONFUSION MATRIX COMPUTED USING SVM APPLIED TO THE VALIDATION DATA (RESULTS ARE IN PERCENT AND ROUNDED)

	1	2	3	4	5	6	7	8	9	10	11	12
1	79	2	2	2	0	1	2	7	0	0	1	4
2	8	74	1	0	1	5	4	0	0	3	1	3
3	2	2	75	0	0	0	6	11	0	0	0	4
4	0	1	1	91	0	1	0	3	0	0	0	3
5	0	3	0	3	83	1	2	0	1	5	1	1
6	1	9	0	1	0	87	1	0	0	0	0	1
7	2	2	5	0	2	0	88	0	0	0	0	1
8	1	0	2	3	0	0	0	89	0	0	0	5
9	0	0	0	0	0	0	0	0	93	2	2	3
10	1	4	0	0	4	1	0	1	4	83	0	2
11	1	1	0	0	0	0	0	0	2	0	94	2
12	1	1	0	0	0	0	0	3	2	3	0	90

The approach was followed by applying the proposed CRF to the SVM results. The confusion matrix that was calculated during the SVM training was used to compute matrix D in the pairwise potential. Then 10 new images were randomly picked and were segmented and then the resulting regions were manually labeled for the CRF training. The achieved CRF model was applied to the results of SVM classification through an inference process with a maximum of 20 iterations. The maximum number of iterations was already determined in a validation process on the training data. The average accuracy of the CRF output was 88.9% for the ten primary classes and the computed confusion matrix can be seen in Table III. The rest of saturated and vague regions were then investigated and classified into one of the 10 primary classes using the rules presented in III-E.

Furthermore, the system was reevaluated using a general formulation of CRF by disregarding the 3 introduced terms in the pairwise potential; neighbourhood length, neighbour certainty and local context matrix. The average accuracy of the system that featured a traditionally formulated CRF was 71.5% which is dramatically lower than the accuracy of our system. It is noticeable that this result is worse than the accuracy of a pure SVM classifier. The main reason behind this outcome is the presence of some detailed objects such

TABLE III
THE CONFUSION MATRIX COMPUTED USING CRF APPLIED TO THE VALIDATION DATA (RESULTS ARE IN PERCENT AND ROUNDED)

	1	2	3	4	5	6	7	8	9	10	11	12
1	80	3	3	2	0	0	1	8	0	0	0	3
2	6	82	1	0	2	2	4	1	0	1	0	1
3	2	1	89	1	0	0	3	3	0	0	0	1
4	0	1	1	94	0	0	1	2	0	0	0	1
5	0	2	0	2	92	1	1	0	0	1	0	1
6	1	2	0	1	6	85	5	0	0	0	0	0
7	0	2	3	0	1	0	93	0	0	0	0	1
8	1	0	1	2	0	0	0	93	0	0	0	3
9	0	0	0	0	0	0	0	0	92	1	4	3
10	2	1	0	0	2	1	0	2	2	89	0	1
11	1	1	0	0	2	0	0	0	5	0	88	3
12	2	1	2	1	1	0	0	8	4	1	1	79

as *power poles* and *road guards* in the images which have been smoothed out by the ordinary CRF. Since the number of samples from each category in the evaluation process were equal, a low detection rate in such classes led to a poor classification performance. This result illustrates the potential of our proposed pairwise function, specially in the classification of objects with fine details.

The proposed algorithm was applied to the entire set of image regions to get a fully labeled classification result. Fig. 7 demonstrates the results of subsequent steps of the algorithm for a sample image. It can be seen that a significant improvement has been made by applying our CRF formulation to the SVM results. As evident in Fig. 7, the degree of smoothness is controlled very well in most regions and many fine details are still present in the final result.

V. CONCLUSION

We proposed a novel addition to a classic CRF for material classification in the context of road and roadside objects using multi spectral imaging. The addition addressed, in particular, the issue of over smoothing and loss of fine detail. A new CRF pairwise function was introduced which uses different factors to reach a more purposeful and context dependent smoothing. This function is shown to be very adaptive to changes in the contextual relationships, the region features, the amount of neighbourhood with the adjacent regions, and also the certainty of the labels of the regions.

We utilised the confusion matrix of the unary classifier to calculate matrix D and embedded it into our classification system to represent a certain kind of contextual information. This is an efficient method to take into account the relationships between all the classes, as it removes the need for training a large number of context parameters. Since no knowledge of the dataset is needed to design this matrix, it can be applied to a variety of applications.

In addition, the presence of several misclassified neighbours might result in an erroneous decision in the CRF. Addressing this issue, we equipped the pairwise function with the class probabilities of the neighbouring regions. According to Eq. (4), we give more cost to the cases where the neighbouring labels have a lower degree of certainty.

We also used a *neighbourhood length* parameter to make up for the difference in the amount of neighbourhood with the adjacent regions. As described in III-D-3 and Fig. 4, this parameter acts as an equaliser between the number of neighbouring regions and the number of neighbouring pixels.

Moreover, we specified two more classes for the saturated parts and vague boundaries of the image. Due to the lack of useful information in these regions, putting them in one of the primary classes will degrade the classification performance. The results demonstrate that these regions were successfully identified using the probabilistic SVM (Fig. 7-b) and then converted to the relevant classes (Fig. 7-d).

The primary advantage of region-wise processing is the significant increase in the computation speed compared to a pixel-wise algorithm with context awareness. The number of regions for each image is around 2500 on average, which is considerably less than the number of image pixels (more than 2 millions). This huge difference makes a pixel-wise classification much more demanding than our implementation. Apart from the computation time, the regions present more locally consistent information about the materials and objects in the image, so they can provide more context information, compared to individual pixels.

The proposed approach was evaluated using a large scale dataset of road and roadside objects and led to an average classification accuracy of 88.9% which was about 5% more than the accuracy of SVM classifier. This experiment was also carried out for a CRF framework with an ordinary pairwise function (lacking the terms introduced in III-D-3). The resulting classification accuracy of 71.5% demonstrates the superiority of our proposed pairwise potential.

A major limitation of our work was in the segmentation step. Although we attempted to improve the superpixels using the information in the NIR band, there were still some regions that expanded over two or more objects and materials. Since the regions are the basic blocks of input to our work, we intend to improve this step in the future and also test our approach on some publicly available datasets.

REFERENCES

[1] J. Lee and O. Ersoy, "Consensual and hierarchical classification of remotely sensed multispectral images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 9, pp. 2953–2963, Sept. 2007.

[2] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 11, pp. 3804–3814, 2008.

[3] S. Tarrant, G. Piercey, D. Hart, and P. McGuire, "Automatic road extraction from multispectral high resolution satellite images," *Defence R&D Canada, Suffield, Ralston ALTA (CAN), C-Core*, 2009.

[4] D. Bradley, R. Unnikrishnan, and J. Bagnell, "Vegetation detection for driving in complex environments," in *Robotics and Automation, IEEE International Conference on*, Apr. 2007, pp. 503–508.

[5] S. Namin and L. Petersson, "Classification of materials in natural scenes using multi-spectral images," in *Intelligent Robots and Systems, IEEE/RSJ International Conference on*, Oct. 2012, pp. 1393–1398.

[6] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML 2001, 2001, pp. 282–289.

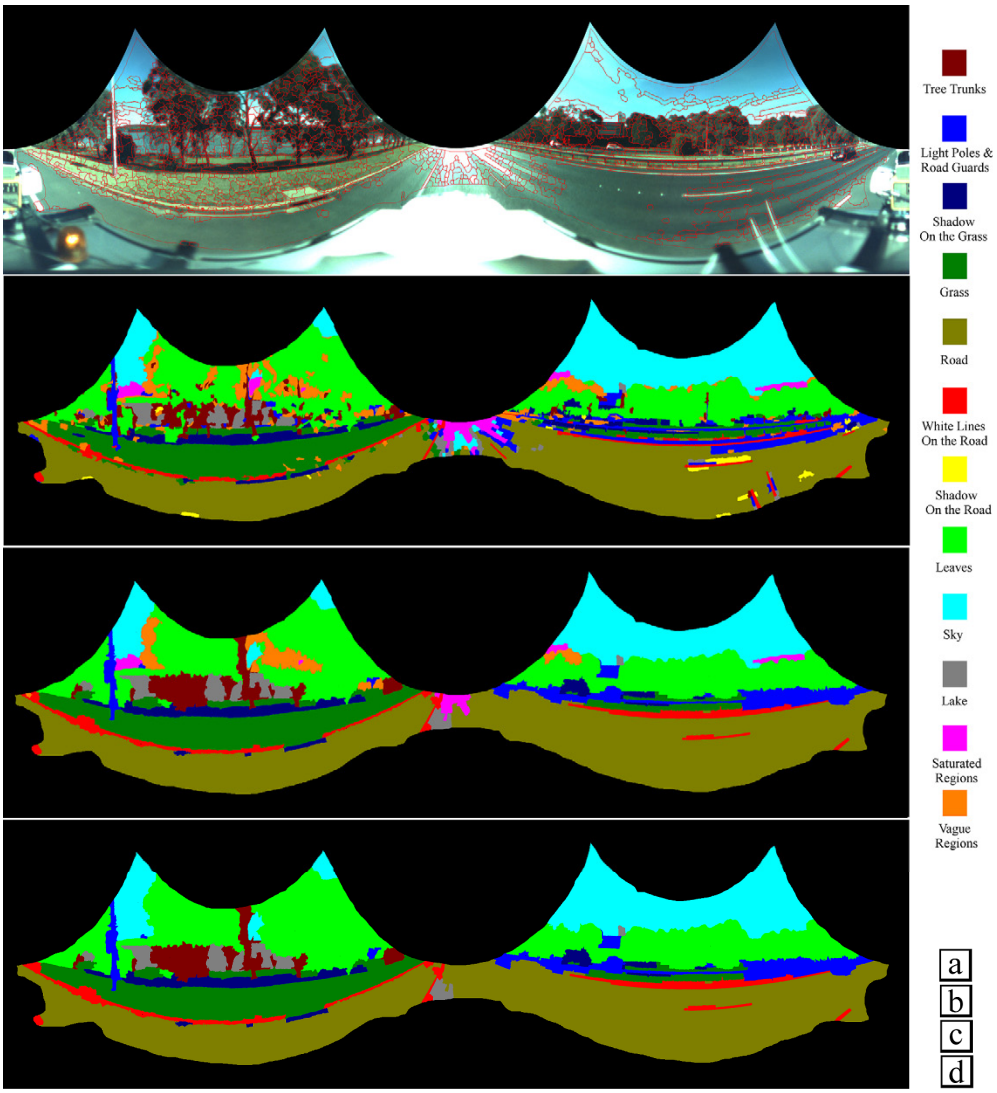

Fig. 7. The results of four main steps of our proposed approach: a) The output of segmentation using the Mean Shift method. b) The result of SVM classification (12 classes). c) The result after applying CRF. d) The final classification result after reducing the number of classes to 10. The SVM result in Fig. (b), contains many misclassifications which have been resolved using CRF in Fig. (c). For example, it is apparent that some parts of the *Road* are misclassified with *Road Guards*, *Lake*, *Sky* and *Grass* using SVM, while CRF has been able to classify these regions correctly. As it can be seen in Fig. (d), the results of classification for the saturated and vague regions are satisfactory.

[7] P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *Image Processing, IEEE Transactions on*, vol. 19, no. 7, pp. 1890–1907, July 2010.

[8] G. Zhang and X. Jia, "Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 9, no. 5, pp. 856–860, Sept. 2012.

[9] M. Y. Yang and W. Forstner, "Regionwise classification of building facade images," in *Proceedings of the ISPRS conference on Photogrammetric image analysis*, ser. PIA 2011. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 209–220.

[10] M. Yang and W. Forstner, "A hierarchical conditional random field model for labeling and classifying images of man-made scenes," in *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on*, Nov. 2011, pp. 196 –203.

[11] C. Wojek and B. Schiele, "A dynamic conditional random field model for joint labeling of object and scene classes." in *ECCV (4)*, ser. Lecture Notes in Computer Science, D. A. Forsyth, P. H. S. Torr, and A. Zisserman, Eds., vol. 5305. Springer, 2008, pp. 733–747.

[12] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *International Journal of Computer Vision*, vol. 80, no. 3, pp. 300–316, Dec. 2008.

[13] X. He, R. S. Zemel, and D. Ray, "Learning and incorporating top-down cues in image segmentation," in *Proceedings of the 9th European conference on Computer Vision - Volume Part I*, ser. ECCV 2006. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 338–351.

[14] T. F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, Dec. 2004.

[15] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, May 2002.

[16] R. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786 – 804, May 1979.

[17] C. C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *Intelligent Systems and Technology, ACM Transactions on*, vol. 2, pp. 27:1–27:27, 2011.

[18] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. Springer Publishing Company, Incorporated, 2009.

[19] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Mathematical Programming*, vol. 89, pp. 149–185, 2000.

[20] S. Shetty, H. Srinivasan, M. Beal, and S. Srihari, "Segmentation and labeling of documents using conditional random fields," *Proceedings of SPIE*, pp. 65 000U–65 000U–9, 2007.