Fast and Spatially-smooth Terrain Classification using Monocular Camera

Chetan J., Madhava Krishna, C. V. Jawahar
International Institute of Information Technology, Hyderabad, India-500032.

Abstract

In this paper, we present a monocular camera based terrain classification scheme. The uniqueness of the proposed scheme is that it inherently incorporates spatial smoothness while segmenting a image, without requirement of post-processing smoothing methods. The algorithm is extremely fast because it is build on top of a Random Forest classifier. The baseline algorithm uses color, texture and their combination with classifiers such as SVM and Random Forests. We present comparison across features and classifiers. We further enhance the algorithm through a label transfer method. The efficacy of the proposed solution can be seen as we reach a low error rates on both our dataset and other publicly available datasets.

1 Introduction

The goal of terrain classification [3, 10] is to recognize various terrains that occur in urban and rural environments in an automated fashion. An automated solution to terrain classification is very crucial in various domains such as (i) advanced driver assistance systems [13], (ii) autonomous navigation, (iii) remote sensing, (iv) urban and rural planning. For instance, a mobile robot navigating outdoors, comes across various terrains such as soft, slippery, hard, smooth, rocky or undulating ones. The navigation strategy for the robot differs mainly based on the kind of terrain it traverses and the limits on its velocities vary according to these surfaces. An algorithm capable of prior judgment of the terrain provides the much needed time for the robot to adapt its velocity planner and thus becomes a vital cog in outdoor navigation systems.

Various methods have been proposed in literature for the problem of terrain classification. In particular Vibration-based methods [7, 12] (which use accelerometers, IMU etc.,) have been very successful. Yet, the main drawback of those methods is that they classify terrain only while the sensor attached to robot is traversing the terrain and not beforehand.

Camera-based methods follow a canonical form of using images from camera as training data along with lasers or stereo-rig for obtaining ground truth. Among the literature we surveyed, the work reported in [1], is closest to our problem. However it partially relies on time consuming texture features. Bradley *et al.* [2] uses multi-spectral camera to detect chlorophyll content for recognizing grass and trees. Recent work includes, Blas *et al.* [9] employing pre-segmentation algorithm based on clustering using Local binary patterns. Vernaza *et al.* [10] uses Markov random fields framework. Procopio *et al.* [8] adds memory to the machine learning model by using ensemble of classifiers. They report an accuracy of around 90% on their datasets, but almost all of them detect only navigable region and does not characterize the terrain.

While the problem can be approached by using combination of various sensing modalities such as 2D and 3D lasers, multiple cameras etc., this paper explores the extent of scene interpretation ability vested in a single camera and is thus different. This investigation is especially crucial since cameras are often less expensive, compact, and are not power hungry like laser range finders. Unlike many previous approaches, which deals with the problem of detection of navigable regions, where the terrain characterization is neglected, our goal is to detect and characterize the terrain ahead into commonly observed terrains.

In this paper, we propose a new partition algorithm and a temporal label transfer method that enhances the performance of baseline classifiers. The novel partition algorithm partitions the image into various regions. A patch in an image thus belongs to various partitions based on the partitioning scheme. The patch is classified for each such partitioning scheme and the eventual classified label for that patch is based on a majority rule across such partitioning schemes. We show that such a partitioning is indeed generic as it enhances the classifier accuracy of various classifiers such as Random forests, SVMs and K-Nearest neighbours. The efficacy of the proposed algorithm can be vindicated as we report highly efficient terrain classification on our dataset and other two datasets by Procopio *et al.* [8]. We show that our partition algorithm inherently segments the image smoothly. The temporal transfer method efficiently uses temporal information from already predicted labels of the previous frames. We also show that by using temporal label transfer, we save considerable amount of computation time per image.

2 Problem Parameters

Terrain classification was modeled as a classification problem of pixels and smaller windows in the past [4, 11], where the important parameters were features, classifiers and datasets. In this section, we analyze the relative importance of these parameters and demonstrate that the problem can be solved upto greater extent using state of the art features and classifiers.

Features. For any learning based method, selecting meaningful features is very important. We use popular RGB histogram [8, 11] and LBP histogram [9] as our features considering the computational cost and performance. We use the optimal weighted combination of these features that best suits the classifier.

Classifiers. Performance of selected features are evaluated on a set of popular and promising classifiers. The classifiers which we consider in our experiments are Naïve Bayes(NB), K-Nearest Neighbor(K-NN), Artificial Neural Networks(ANN), Support vector machines(SVMs) and Random Forests(RF) [6]. Random forest is a classification algorithm that uses an ensemble of unpruned decision trees, each of which is built on a bootstrap sample of the training data using a randomly selected subset of feature space dimensions. Experiments were conducted by changing important parameters like number of epochs and number of nodes in the hidden layers in ANNs, number of trees and size of node in RF. In case of SVMs, we conduct experiments with linear SVM using 1 vs 1 multiclass classifier (SVM-L) and non-linear SVM (SVM-K).

Data sets. We experiment with three datasets in this study: our own dataset and two other datasets by Procopio *et al.* [8]. For collecting our data, monocular camera was mounted on the top of the vehicle, and videos were recorded at 7.5 fps. We collected the data in and around a radius of 10km navigating at various speeds ranging from 0.2m/s to 4m/s. We observe that the data is challenging, as it contained wide variations in illumination. We also observed that the data varied from unpaved or damaged rural roads to paved urban roads. We collected 25 videos, each of 1 min. Figure 1 shows some of the sample frames from the videos and their corresponding ground truth images. Five distinctly different terrains were identified in the data collection[1].

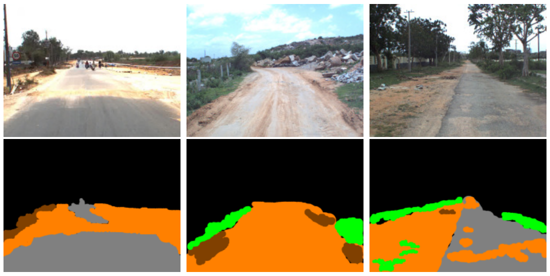

Figure 1: Sample frames and their Ground truth

D	NB	ANN	K-NN	SL	SK	RF
O	43.6	35.6	28.3	29.0	28.7	25.5
A	18.9	32.3	33.8	31.2	38.4	18.2
B	13.7	26.2	17.8	27.9	39.8	18.9

Table 1: Base line error-rates on Our dataset(O) and two datasets(A and B) of Procopio et al. [8]. Where **D:Dataset**, O:Our Dataset, A:DS3A, B:DS3B, SL:SVM-L and SK:SVM-K.

Empirical evaluation. For the empirical studies, we consider a part of our data set (200 images). We use 50% of the data for training and the rest for testing. We extract multiple, non-overlapping, patches of size 16×16 from these images. Thus we have around $2 * 185000$ patches[2] for training and testing.

From Table 1, we observe that RF's outperformed all other classifiers because of its capability to handle large number of input variables and data samples [6]. Additionally RF classifiers are computationally efficient for training and testing, compared to SVMs. We also observed that with only training on our datasets, the performance on other datasets were appreciable, which clearly shows the superiority of our dataset. Since the data sets and details of the earlier reports are not completely available, a direct comparison of results may not be applicable. However, it may be noted that the quantitative results, which we report in Table 1, are comparable to the results reported in literature [8, 10], which use non-visual sensors and stereos along with appearance clues. This advantage comes out of the fact that monocular cameras that are currently in use provide much richer sampling in space and dynamic range, and are hence useful for such tasks.

3 Proposed Methodology and Results

In the last section, we have shown that the monocular camera based terrain characterization is a feasi-

[1] Data contains regions of road, muddy-road, rough-terrain, grass (Note that the class grass contains only traversible grass or very small plants, big plants and trees are considered obstacles.) and obstacles (which contains static objects like trees, rocks etc., and dynamic objects like moving vehicles), which are labelled with colors black-grey, orange, brown, green and black respectively.

[2] The number of patches in all the five classes are equal.

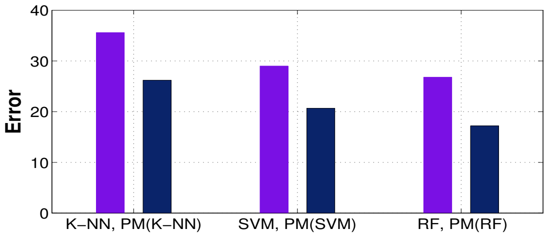

Figure 2: Comparison of base-line classifiers with Partition-based algorithm operated over them. Where PM(K-NN),PM(SVM) and PM(RF) represents the error rate of partition-based algorithm operated on classifiers K-NN, SVM and RF respectively.

ble and promising paradigm for out door navigation. We have observed that RF classifier is performing best among several classifiers. In this section, we describe two enhancements for terrain classification. Initially we describe our partition based algorithm and several experiments which indicate that, the algorithm is robust and spatially smooth. Secondly we describe our label transfer method along with experiments showing that, it saves considerable amount of computation time.

3.1 Partition based algorithm

The proposed algorithm partitions the training images and trains different classifiers on different parts of the image independently. This is repeated for partitions of different sizes. Training different classifier from different part of the image handles the problem of perspectivity of the imaging process, i.e., it learns the fact that near and far image patches show different textural characteristics. Also learning from fixed partition over several training images has two main advantages. The first advantage is that it helps the classifier to learn new facts about associativity of classes, such as occurrences of grass along with mud is more probable than that of grass along with tar road. The second advantage is that it helps the algorithm to be dependent upon the position of the partition of the image and thus learns the spatial context. By training a classifier from larger sized partitions, global properties of the class are learnt and as the size of the partition decreases, more local properties are learnt. Our algorithm is a generic framework that can be operated on any classifier.

In training phase, as summarized in Algorithm 1 we build N classifier-sets using all the training images, let us call them $S = \{C_1, C_2, C_3, ...C_\mathrm{N}\}$. Note that a classifier-set C_i contains i^2 classifiers. To characterize the terrain of the given image, for each patch of the image, we get N labels from each of the N classifier-sets in S. From these N labels, most occurring label is declared as the final label of the patch.

Algorithm 1 Partition based algorithm

– Training

```
Goal: To build $N$ classifier-sets
Input: $M$ Training images, $S \leftarrow \emptyset$
for $k = 1$ to $N$ do
    Partition training images into $k^2$ parts, $C_k \leftarrow \emptyset$
    for $p = 1$ to $k^2$ do
        Train a Classifier on $p^{th}$ partition over all training images, call it $KF_p$
        $C_k \leftarrow C_k \cup \{KF_p\}$
    end for{ Now $C_k = \{KF_1, KF_2, ...KF_{k^2}\}$ }
    $S \leftarrow S \cup \{C_k\}$
end for{ Now $S$ contains $\{C_1, C_2, ...C_N\}$ }
```

Experiment 1: Comparison with baseline classifiers. Figure 2 shows the error-rates of our partition-based algorithm operating on baseline classifiers K-NN, SVM and Random Forests. We observe that our algorithm always decreases the error-rates by approximately 10%, this is an appreciable decrease in the error rate. This also shows that our algorithm is generic, i.e., it improves the performance of classifier independent of the classifier choosen. To show the superiority of our algorithm across other databases, we conduct an experiment in which our partition-based algorithm operating over RF is tested on (i) Our dataset (ii) DS3A and (iii) DS3B datasets of Procopio et al. [8]. We report the error rates in first and second column of Table 2, from the table, we observe that our algorithm compared to baseline RF classifier, decreases the error rate by approximately 10% on all three datasets. We also observe that even without training on any of the images of DS3A or DS3B datasets, we get error rates as low as 6.8%, the superiority of our algorithm is thus clearly evident.

Experiment 2: Effect on number of Classifier-sets (N). In this experiment, we check the effect of varying number of classifier-sets(N) on the algorithm. N is a parameter which controls both efficacy and speed. We

D	RF	PM	RF	PM	AVG	Err
O	26.8	**17.2**	08.7	**01.0**	35.5	05.6
A	18.2	**07.9**	06.9	**00.6**	42.3	04.3
B	18.9	**06.8**	05.2	**00.4**	45.1	04.3

Table 2: 1^{st} and 2^{nd} column represents error-rates of RandomForest(RF) and our partition based algorithm(PM). 3^{rd} and 4^{th} column represents smoothness-error rates, which corresponds to experiment-3. 5^{th} and 6^{th} column represents the percentage of images, that were labelled just by using Temporal-label-transfer method in Section 3.2, where AVG: Average of percentages of portion of labels that are transferred over sequence of 100 images and Err: Error in label transfer.

observe that as N increases, the speed of the algorithm decreases, the error-rate decreases and then slowly increases. From our experiments we found that, the optimal choice for N is 5, which has high efficacy without compromising speed.

Experiment 3: Spatial smoothness test. In Table 2, third and fourth column show the smoothness-errors of RF and PM operated on RF, on three datasets. Smoothness-error is the difference between error rates before and after applying smoothing algorithm modeled by MRF [14] on the classifier predictions. We observe that our algorithm has a negligible smoothness-error compared to RF classifier, which clearly shows that PM itself is capable of characterizing the image smoothly in spatial context.

3.2 From Image to Video

Temporal label transfer. Most of the previous methods in literature deal with single image. They do not use the fact that they are dealing with a sequence of continuous video stream. When robot navigates through terrain, the camera captures sequence of frames, any two consecutive frames have lot of common image regions i.e., they look visually much similar. Inorder to characterize the terrain of the image using traditional machine learning based algorithm some kind of feature is extracted from each patch. The feature vector is fed to a classifier, which returns the label of the patch. Note that in this process, feature extraction is computationally expensive. In our case, when a new frame is captured by the camera, fast coarse optical flow [5] between the previously captured frame and current frame is calculated, then for each patch, if there is flow present in the current-frame-patch, we transfer the corresponding patch-label from the previous frame to the current frame. If there is no flow available for that patch, feature is extracted from the patch and fed to our partition-based algorithm described in section 3.1. In this way without even extracting features from the current frame, we can label considerable portion of the frame.

We conduct an experiment to see, how much portion of the image can be labelled by just using temporal label transfer. The average percentage of image that is labelled using temporal label transfer over testing images and their corresponding error are reported in fifth and sixth column of Table 2 respectively. We observed that by just using temporal label transfer, we can label approximately 40% of the image on three datasets with very less error. This automatic transfer of label resulted in considerable decrease in computation time. The decrease was to the tune of 40% on an average computed over several experiments, which is crucial in real time systems like mobile robots.

4 Conclusions and future work

This paper presented a novel partition-based algorithm for classification of outdoor terrains using monocular camera. The proposed algorithm is generic and enhanced the error-rates of base-line classifiers by approximately 10%. The algorithm was extensively tested on our and on other publicly available datasets. The computational time of the whole system is reduced by applying the partition-based algorithm to only those regions of the image, where the temporal label transfer is not applicable. The future scope of the work includes much better processing of the video data using complex temporal clues along with fusing geometric and appearance clues in an optimization framework.

References

[1] D. H. A. Angelova, L. Matthies and P. Perona. Fast terrain classification using variable-length representation for autonomous navigation. In *CVPR*, 2007.

[2] R. U. Bradley and J. Bagnell. Vegetation detection for driving in complex environments. In *ICRA*, 2007.

[3] H. T. Christian Weiss and A. Zell. Combination of vision- and vibration-based terrain classification. In *IROS*, 2008.

[4] N. V. Cristian Dima and M. Hebert. Classifier fusion for outdoor obstacle detection. In *ICRA*, 2004.

[5] R. C. H Liu and A. Rosenfeld. Fast two-frame multiscale dense optical flow estimation using discrete wavelet filters. In *Journal of the Optical Society of America*, 2003.

[6] B. L. Random forests. In *Machine Learning*, 2001.

[7] P. Leang and B. Bhanu. Learning integrated perception-based speed control. In *ICPR*, 2004.

[8] J. M. Michael J. Procopio and G. Grudic. Learning in dynamic environments with ensemble selection for autonomous outdoor robot navigation. In *IROS*, 2008.

[9] A. S. Morten Rufus Blas, Motilal Agrawal and K. Konolige. Fast color/texture segmentation for outdoor robots. In *IROS*, 2008.

[10] B. T. P. Vernaza and D. D. Lee. Online self-supervised terrain classification via discriminatively trained submodular markov random fields. In *ICRA*, 2008.

[11] A. T. R. Manduchi, A. Castano and L. Matthies. Obstacle detection and terrain tlassification for autonomous off-road navigation. In *IROS*, 2005.

[12] N. F. M. S. Weiss C. and A.Zell. Comparison of different approachs to vibration-based terrain classification. In *European Conf. on Mobile Robotics*, 2007.

[13] N. T. Yamaguchi K., Watanabe A. and N. Y. Road region estimation using a sequence of monocular images. In *ICPR*, 2008.

[14] T. C. P. Zoltan Kato and J. C. M. Lee. Color image segmentation and parameter estimation in a markovian framework. In *Pattern Recognition Letters*, 2001.