A Machine Learning Approach for Recognising Woody Plants on Railway Trackbeds

Roger G. Nyberg*

**Dalarna University, Dept. of Informatics, Borlange, Sweden, rny@du.se*

Keywords: Machine learning, computer vision, maintenance planning, railway, weeds.

Abstract

The purpose of this work in progress study was to test the concept of recognising plants using images acquired by image sensors in a controlled noise-free environment. The presence of vegetation on railway trackbeds and embankments presents potential problems. Woody plants (e.g. Scots pine, Norway spruce and birch) often establish themselves on railway trackbeds. This may cause problems because legal herbicides are not effective in controlling them; this is particularly the case for conifers. Thus, if maintenance administrators knew the spatial position of plants along the railway system, it may be feasible to mechanically harvest them. Primary data were collected outdoors comprising around 700 leaves and conifer seedlings from 11 species. These were then photographed in a laboratory environment. In order to classify the species in the acquired image set, a machine learning approach known as Bag-of-Features (BoF) was chosen. Irrespective of the chosen type of feature extraction and classifier, the ability to classify a previously unseen plant correctly was greater than 85%. The maintenance planning of vegetation control could be improved if plants were recognised and localised. It may be feasible to mechanically harvest them (in particular, woody plants). In addition, listed endangered species growing on the trackbeds can be avoided. Both cases are likely to reduce the amount of herbicides, which often is in the interest of public opinion. Bearing in mind that natural objects like plants are often more heterogeneous within their own class rather than outside it, the results do indeed present a stable classification performance, which is a sound prerequisite in order to later take the next step to include a natural background. Where relevant, species can also be listed under the Endangered Species Act.

1 Introduction

Vegetation that grows on railway trackbeds presents potential problems. The presence of vegetation threatens the safety of personnel inspecting the railway infrastructure [1]. In addition, vegetation growth clogs the ballast and results in inadequate track drainage; in turn, this could lead to the collapse of the railway embankment [2, 3, 4].

Woody plants like pine, spruce and birch often establish themselves on railway trackbeds (see figure 1).

Figure 1: Established conifer clusters on a trackbed

Woody plants are problematic because legal herbicides are not effective in controlling them; this is especially the case for conifers. If maintenance administrators would know where the woody plants are located, i.e. their spatial position along the railway system, it may be feasible to mechanically harvest them. Often, however, such information is not available. This depends on whether or not administrators and tenders are trained to recognise plants and map them.

Generally, the amount of vegetation growing on railway trackbeds and embankments is controlled by applying herbicides like Glyphosate (commercial name: Roundup Bio) [5]. However, Glyphosate does not affect conifer trees that much. Whilst it will kill new shoots, the plant itself will continue to grow [6].

Even though some types of herbicides are claimed to be less harmful than others, the existence of numerous environmental organisations and greater environmental awareness among the public make it hard to argue in favour of herbicides. For example, in a survey conducted by the International Union of Railways (UIC), Jan Skoog, Environmental Coordinator in the Strategic Department at the Swedish National Rail Administration stated that: “Concerns for groundwater protection and public opinion have led to the decision not to use chemicals where other possible alternatives are available” [7]. Together with the desire to cut the costs of vegetation control, this has motivated several railway companies to engage in various activities to reduce the amount of herbicides used [7].

The purpose of this study was to test the concept of recognising plants using images acquired by image sensors in a controlled noise-free environment. Noise could refer to: noise produced by the image sensor because of poor camera configuration, or poor pre-processing of the image, or (on a higher abstraction level) the noise could refer to the fact that the scene contains many objects that are not of interest (i.e. the background). At the time of acquiring the image, the objects of interest may not have been known. Hence, there is no defined background and no defined foreground.

Sometimes it is very hard or even impossible to extract natural objects, e.g. leaves, out of natural scenes. This is because most of the natural objects are very heterogeneous in shape, colour, and texture.

2 Methodology

Primary samples were collected outdoors, comprising 592 leaves from nine different deciduous tree species. In addition, 120 conifer seedlings from two species were acquired from a plant nursery (see table 1). The selection of plant data was based on standing volume statistics in Sweden [8] and advice from academics in forestry and comprised species likely to be found on railway trackbeds.

After the physical collection, each leaf and each conifer plant were then documented by acquiring images from a nadir perspective in a laboratory environment.

The goal was to let a machine learning algorithm classify the 11 species without any interaction from humans. Since we had prior knowledge about which species is in which image, respectively, a supervised learning approach was chosen. This approach is about learning the relationship between two data sets with help of a "supervisor": the observed data X and an external variable Y (the output) that are to be predicted. The observed data X in this study consists of features extracted from the images. The observed data X is denoted as the training dataset, and the output Y is usually called the targets. In this study, the targets are the 11 plant species mentioned above. Typically, the classifier was trained with the training dataset and a set of desired target outputs.

After the training session, the classifier was then validated by testing with previously unseen input data, called the test dataset. The performance was then measured based upon the classification error rate.

In this study, a machine learning approach called Bag-of-Features (BoF) was used. This approach involves the detection and extraction of visual features from the acquired image data set. All extracted features (the observed data set) are then clustered and represented in order to create a visual feature "dictionary". Thus, each image can be represented by a histogram of the visual features in the dictionary. Based on this information, previously unseen images can be classified as belonging to a plant species category using supervised machine learning methods.

2.1 Machine learning using the Bag-of-Features approach

In this work, the so called BoF model was investigated and applied in order to classify images by treating image features as *visual words*. In computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.

The difficulties are that objects in our three dimensional world look different from different angles and under different lighting conditions when mapping them into two dimensions, such as in images (or frames in a video sequence).

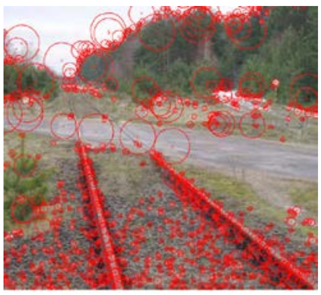
The same type of object often appears to be different, e.g. Scots pines (Pinus sylvestris) most often appear to be different from each other. This is known as intraclass variation. Interclass variations are variations between different types of objects, e.g. Scots pines and Norway spruces are different types of objects, or species in this context. Natural objects (as opposed to man-made objects) often have an asymmetric shape and colour, and therefore experience a high degree of intraclass as well as interclass variation.

When using a BoF model to learn the features of different deciduous trees (here represented by their leaves) and conifer species in images, a set of 'negative images' should also be presented. This negative category should not contain any leaves or conifers. See table 1, for the distribution of images per category (species), including negative images.

The BoF approach deals with image classification in a sequence of three phases, as follows:

Phase 1: Feature extraction, including feature detection (feature sampling) and description

First, the features have to be detected. These features are corners, detected using Harris-Laplacian corner detection [9]. This feature detection is performed as a sampling process from the training or test image. Local sub images around the detected features are then extracted. Each part becomes a feature of the original image and is then represented using: Scale Invariant Feature Transform (SIFT) [10], Dense SIFT (DSIFT), or Multi-scale Dense SIFT (MD-SIFT) [11]. The results of this first phase are denoted as feature vectors (or feature descriptors). Examples of detected image features are visualised in figure 2a-d, where 2a is an Alder leaf, 2b is a negative image, 2c is a Silver Birch leaf and 2d is a Norway Spruce.

a)  b) 

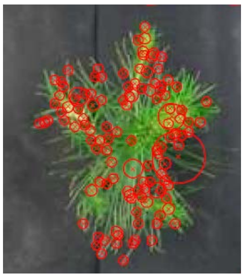
c)  d) 

Figure 2: Detected SIFT features

Phase 2: *Codebook formation and image representation*
a: The features are clustered into *k* number of clusters using k-means clustering. Optimally, similar features are clustered together with other similar features. The similarity is dependent on the quality of the representation, as in phase 1 above. The clustered features in this step build up the visual vocabulary, which is the content of the so called codebook. In this study, the number of clusters *k* was experimentally set to: $k = 500$.
b: Next, the clustered features are assigned to the most representative visual word in the visual vocabulary. This process is known as vector quantisation (VQ). An alternative to VQ is, for example, the Locality-constrained Linear Coding (LLC) proposed by [12].

Phase 3: Learning and recognition
A support vector machine (SVM) classifier, originally by [13], with linear or non-linear kernels, is used to learn the category models. Here, the following four classifiers were used: A. Nearest neighbour classification (1-NN) using L2 distance (denoted NN L2); B. Support vector machine with a linear kernel (denoted SVM Linear); C. Support vector machine with a linear kernel using a coding scheme called Locality-constrained Linear Coding (LLC) (denoted SVM Linear LLC max-pooling), and finally D. A support vector machine with a Chi^2 kernel (denoted SVM Chi^2).

In the training set there were both positive training images containing an object class (i.e. leaves or conifer seedling), and negative images, which did not contain any objects of interest (see table 1). Training and testing procedures depended on the object class that has the least number of images. Thus, the ratio of training and test images was computed by first checking how many images were in each object class and then by choosing the class with the fewest. After this, 65% percent (rounded downward to its nearest integer) were randomly chosen to belong to the training image set and the rest (35%) were allocated to the test image set. For example, the lowest number of collected images belonged to the Picea abies category with 59 images. Out of these 59 images, 38 were assigned as training images and 21 were assigned as test images. The same number of training and testing images were then randomly chosen (without replacement) for every species class.

Latin name	**English name**	**Number of images**
Betula pubescens	Downy Birch	70
Betula pendula	Silver Birch	60
Alnus incana	Grey Alder	60
Alnus glutinosa	Alder	60
Populus tremula	Aspen	67
Quercus robur	Pedunculate Oak	70
Salix caprea	Goat Willow	65
Prunus padus	Bird Cherry	80
Acer platanoides	Norway Maple	60
Negative images	Negative images	168
Picea abies	Norway Spruce	59
Pinus sylvestris	Scots Pine	61
Total		*880*

Table 1. Number of images in the data set

3 Results

The feature extraction of the training and test image set were performed by SIFT, DSIFT, and MDSIFT, respectively, as shown in table 2, where the classifier kernels are denoted as: A - NN L2, B - SVM Linear, C - SVM Linear LLC max-pooling, D - SVM Chi^2 kernel.

Each combination, as seen row-wise in table 2, was executed 10 times each. The average overall classification accuracy (in %) and the average execution time per run (in seconds) were computed. The accuracy is here defined as in equation. 1:

$$\textit{Accuracy rate} = 1 - \textit{Error rate}. \qquad (1)$$

The results in table 2 show that if both the classifiers performance and execution speed are to be taken into account, then the SVM Chi^{2}kernel with DSIFT feature extraction was the best. The confusion matrix for this alternative is presented in table 3. The columns from left to right, denoted A to L, as well as the rows A to L are A: Alnus incana, B: Acer platanoides, C: Alnus glutinosa, D: Betula pendula, E: Betula pubescens, F: Negative image, G: Picea abies, H: Pinus sylvestris, I: Populus tremula, J: Prunus Padus, K: Quercus robur, and L: Salix caprea. In table 3, the columns denote the predicted species by the classifier. The rows denote the true species, i.e. the true class, as seen and assigned by a human.

Classifier	Accuracy (%)	Feature extraction	Avg. execution time (sec)
A	74.5	SIFT	18
B	85	SIFT	21
C	85.5	SIFT	70
D	88	SIFT	24
A	88.5	DSIFT	80
B	88.4	DSIFT	91
C	94.5	DSIFT	285
D	95	DSIFT	87
A	91.5	MSDSIFT	208
B	84.3	MSDSIFT	224
C	93.7	MSDSIFT	950
D	95.5	MSDSIFT	203

Table 2: Classifier performance dependent on the feature extraction method.

	A	B	C	D	E	F	G	H	I	J	K	L
A	100	0	0	0	0	0	0	0	0	0	0	0
B	0	100	0	0	0	0	0	0	0	0	0	0
C	0	0	100	0	0	0	0	0	0	0	0	0
D	0	0	0	81	19	0	0	0	0	0	0	0
E	0	0	0	5	95	0	0	0	0	0	0	0
F	0	0	0	0	0	100	0	0	0	0	0	0
G	0	0	0	0	0	0	91	9	0	0	0	0
H	0	0	0	0	0	0	9	91	0	0	0	0
I	0	0	0	5	9	0	0	0	86	0	0	0
J	0	0	0	0	0	0	0	0	0	95	0	5
K	0	0	0	0	0	0	0	0	0	0	100	0
L	0	0	0	0	0	0	0	0	0	5	0	95

Table 3.Confusion matrix (%) for the SVM Chi^{2}kernel with DSIFT feature extraction.

4 Conclusion

The maintenance planning of vegetation control could be improved if plants were recognised and localised. If plants can be identified and spatially mapped, it may be feasible to mechanically harvest them (in particular, woody plants).

In this work in progress study, one of the goals was to recognise those woody plants that are most likely to grow on railway embankments. Bearing in mind that natural objects are often more heterogeneous within their own class rather than outside it, the results present a stable classification performance.

All investigated classifiers with different feature extractions methods recorded a mean in accuracy of 88.7±3.8% (at 95% confidence level), see table 2.

Using the SIFT feature extraction was faster than the other two approaches, in particular the MDSIFT. Based on the results, the MSDSIFT feature extraction together with the SVM Chi^2-kernel classifier exhibited the best performance. One drawback was the long execution time.

If time constraints apply, then it may be better to use the DSIFT feature extraction together with the SVM Chi^2-kernel classifier, which only used 39% of the execution time used by the MSDSIFT feature extraction with the SVM Chi^2-kernel classifier.

The suggested type of feature extraction and classifier, the SVM Chi^2 kernel with DSIFT feature extraction, was able to classify a previously unseen plant correctly with an accuracy of 95±4% (at 95% confidence level) of the cases, and at minimum 81% (for plant D, i.e. Betula pendula), see table 3. There was some confusion in classification between Norway spruce (labelled G) and Scots Pine (labelled H) where a mix-up aroused in 9% of the cases. Also, Aspen (labelled I) leafs were mixed up with Silver Birch (labelled D) and Downy Birch (labelled E) in 5% and 9% (respectively) of the cases. Also, the two Birch species (labelled D and E) were mixed up to some degree by the classifier, see table 3 again.

Future work: This preliminary study provides arguments for continued work in the future. Future work includes images (or video clip frames) that contain natural scenes showing the same species as the objects of interest with real world backgrounds (i.e. various kinds of ballast, gravel, sleepers, rails, bolts, soil, etc). Invariance issues like scale, rotation, translation and occlusion must be further investigated.

Given the growing interest and demands for biodiversity (e.g. the European Union Biodiversity Strategy, which aims to halt the loss of biodiversity), it would be desirable to add sets of images that contain species listed under the Endangered Species Act.

Acknowledgements

This work was sponsored by the Swedish Transport Administration (STA), Sweden. Special thanks go to Jan-Erik Lundh at the STA for his continuous support during the various stages of this work. This article builds on the work carried out in [14] by the same author. Mathworks MATLAB® in conjunction with LIBSVM[15] were used to realise the BoF implementation.

References

[1] Trafikverket. Sakerhetsbesiktning av fasta järnvägsanläggningar. TDOK 2014:0240. v 4.0. 2015

[2] Briggs K. "Charing embankment: climate change impacts on embankment hydrology". *Ground Engineering* 43: pp. 28–31.(2010)
[3] Scott J, Loveridge F, O'Brien A. "Influence of climate and vegetation on railway embankments" *In 14th European Conference on SoilMechanics and Geotechnical Engineering*, Madrid, ES. (2007)
[4] Smethurst J, Clarke D, Powrie W. "Factors controlling the seasonal variation in soil water content and pore water pressures within a lightly vegetated clay slope". *Geotechnique* 62:429–446. (2012)
[5] Torstensson L. "Use of herbicides on railway tracks in sweden". *Pesticide Outlook.* 12. pp. 16-21. 2001
[6] Guiseppe, K.; Drummond, F.; Stubbs, C. & Woods., S. "The use of glyphosate herbicides in managed forest ecosystems and their effects on non-target organisms with partial reference to ants as bioindicators": *Technical Bulletin 192. Maine Agricultural and Forest Experiment Station,* 2006.
[7] Below, M. "UIC Vegetation Control Project Final Report" *Technical report. International Union of Railways (UIC)*, Paris. 2003
[8] Swedish University of Agricultural Sciences. "Statistics from the Swedish National Forest Inventory". Internet http://www.slu.se/nfi (Retrieved 10 Dec 2015).
[9] Mikolajczyk, K. and Schmid, C. "Indexing based on scale invariant interest points". *Inproceedings 8th International Conference on Computer Vision*, Vancouver, Canada. pp. I:525-531. 2001
[10] Lowe DG. "Distinctive image features from scale-invariant keypoints". *International. Journal of Computer Vision*. 60: pp. 91 – 110. 2004
[11] Bosch A, Zisserman A, Munoz X. "Image Classification using Random Forests and Ferns" *IEEE 11th International Conference on Computer Vision*, pp. 1–8. IEEE. 2007.
[12] Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y. "Locality-constrained linear coding for image classification." *In Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on 2010 Jun 13 (pp. 3360-3367).
[13] Boser, B. E.; Guyon, I. M. and Vapnik, V. N. "A training algorithm for optimal margin classifiers" *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 144-152. 1992
[14] Nyberg, R.G.: "Automating Condition Monitoring of Vegetation on Railway Trackbeds and Embankments". Doctoral thesis, Edinburgh Napier University, 2015.
[15] C.C. Chang and C.J. Lin. "LIBSVM: A library for support vector machines". *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1 - 27:27, 2011.