Computers and Electronics in Agriculture 80 (2012) 89–96

Contents lists available at SciVerse ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Sequential support vector machine classification for small-grain weed species discrimination with special regard to *Cirsium arvense* and *Galium aparine*

Till Rumpf[a,*], Christoph Römer[a], Martin Weis[b], Markus Sökefeld[b], Roland Gerhards[b], Lutz Plümer[a]

[a] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Germany
[b] Institute of Phytomedicine (360), Department of Weed Science, University of Hohenheim, Germany

A R T I C L E I N F O

Article history:
Received 7 June 2011
Received in revised form 24 October 2011
Accepted 28 October 2011

Keywords:
Early weed detection
Cirsium arvense
Galium aparine
Feature selection
Sequential classification
Support vector machines

A B S T R A C T

Site-specific weed management can reduce the amount of herbicides used in comparison to classical broadcast applications. The ability to apply herbicides on weed patches within the field requires automation. This study focuses on the automatic detection of different species with imaging sensors. Image processing algorithms determine shape features for the plants in the images. With these shape descriptions classification algorithms can be trained to identify the weed and crop species. Since weeds differ in their economic loss due to their yield effect and are controlled by different herbicides, it is necessary to correctly distinguish between the species. Image series of different measurements with plant samples at different growth stages were analysed. For the classification a sequential classification approach was chosen, involving three different support vector machine (SVM) models. In a first step groups of similar plant species were successfully identified (monocotyledons, dicotyledons and barley). Distinctions within the class of dicotyledons proved to be particularly difficult. For that purpose species in this group were subject to a second and third classification step. For each of these steps different features were found to be most important. Feature weighting was done with the RELIEF-F algorithm and SVM-Weighting. The focus was on the early identification of the two most harmful species *Cirsium arvense* and *Galium aparine*, with optimal accuracy than using a non-sequential classification approach. An overall classification accuracy of 97.7% was achieved in the first step. For the two subsequent classifiers accuracy rates of 80% and more were obtained for *C. arvense* and *G. aparine*.

1. Introduction

Weed populations have been found to be distributed heterogeneously within agricultural fields (Marshall, 1998; Johnson et al., 1996; Christensen and Heisel, 1998; Gerhards and Christensen, 2003; Christensen et al., 2009). Due to the lack of automatic weed detection techniques and site-specific herbicide application, the majority of farmers spray herbicide uniformly across the field. An exact herbicide application to the weeds within a weed patch requires not only detailed information on the weed density but also on weed species distribution. Based on this information and the application of economic weed thresholds selective herbicides can be sprayed site-specifically. Gerhards and Oebel (2006) realised herbicide savings in cereals, maize and sugar beet field from 6% to 81% with site-specific herbicide application based on weed species distribution maps. A site-specific application of a mixture of the individual herbicides on the same field achieved only savings of 19% compared to a uniform treatment of the whole field (Gerhards and Sökefeld, 2003).

According to the demand of Christensen et al. (2009) the identification of single weed species a species discrimination of dicotyledons is necessary. A discrimination of species yields information on weed distribution and weed species composition, laying the foundations for the site-specific application of selective herbicides. A suitable application technology, which allows a simultaneous application of several herbicidal agents on-the-go, based on sensor signals or weed distribution maps, is a further requirement for the adoption of site-specific and selective herbicide application. Possible technical solutions for patch spraying with several herbicides were outlined by Schulze-Lammers and Vondricka (2010) and kefeld (2010).

A major step towards a practical solution for site-specific weed management is the development of precise and powerful data acquisition techniques to automatically and continuously determine in-field variation of weed populations. The most promising techniques to identify weed species in arable crops are based on image processing (Weis and Sökefeld, 2010). Infrared, multispectral and RGB (red, green, blue channel) cameras were used to take pictures of crops and weed species from a low distance above the ground. Plant properties were then extracted from the images by image processing algorithms. Those properties were computed as features, which were used to separate species from each other.

* Corresponding author. Tel.: +49 228 736335.
E-mail address: rumpf@igg.uni-bonn.de (T. Rumpf).

0168-1699/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2011.10.018

Åstrand and Baerveldt (2004) combined colour and shape features, which were derived after an initial plant segmentation step. The colour features were computed as standard deviation and mean of the RGB values. Shape features were computed as area of segment, form factors (distance variance to centre of gravity, compactness and moments). Segments were merged if the distance between them was small. In addition to the colour and shape features a row distance measure was introduced to locate the position 'in-row' or 'between-row'. Classification was finally done with a Bayesian approach. Burks et al. (2000) computed 33 unique colour texture features (co-occurence) from a hue, saturation and intensity (HSI) representation to distinguish between five weed species and the soil. Different neural network classifiers were tested for their performance based on the data given in Burks et al. (2005), resulting in a backpropagation training algorithm with high classification success. Blasco et al. (2002) implemented a vision system for the detection of weeds in a lettuce crop. Segmentation into plant/soil was based on a Bayes-classifier for RGB colours and size features were used to separate weeds from lettuce plants. Cho et al. (2002) used a discriminant function for shape feature selection and neural networks to identify weeds in a radish crop. Tellaeche et al. (2011) successfully applied Support Vector Machines to identify weeds between crop rows based on weed and crop cover measures for image parts. Zhu and Zhu (2009) present a Support Vector Machines approach for weed detection based on shape and texture features for single leaves.

Still, there was no approach for reliable and robust classification between weed species yet. This, however, is a necessary prerequisite for site-specific weed management.

The objective of this paper was the automatic classification of crop and the four weed classes *Galium aparine*, *Cirsium arvense*, other dicotyledonous weeds and monocotyledonous weeds using image processing and classification. Diverse economic thresholds for these weed classes and the availability of selective herbicides against these weeds are the reason for this separation. Gerowitt and Heitefuss (1990) determined weed thresholds for *G. aparine* between 0.1 and 0.5 plants m^{-2}, 40–50 plants m^{-2} for dicotyledonous weeds in total and 20–30 plants m^{-2} for grass weeds. For the very competitive weed species *C. arvense* Börner (1995) indicated a threshold of 2 plants m^{-2}.

The approach in this study is based on an imaging system capturing two images of different wavelengths to differentiate plants from background. The plants are then extracted with image processing algorithms and classified according to their shape. The shape description is expressed as shape features. While the differences in shape features between crop and weed were rather large, features of weeds were highly similar and therefore specific weeds are hard to identify. In addition different growth stages amplify the problem, especially during two-leaf stage. For instance, the dicotyledons *G. aparine* and *Veronica persica* appear to be nearly identical, but due to the high economic loss caused by *G. aparine* accurate classification is critical. This means that very specific features and classifiers are needed to solve this problem. This, however, is not possible in a single multi-classification approach with all weed species and crops. Hence, a sequential classification approach was developed. The main idea is to separate between similar subgroups of the dataset, like crop, monocotyledons and dicotyledons, which are well separable. In next steps dicotyledons were identified by features specialised for the current subgroup. Relevant features were determined by SVM-Weighting (Guyon et al., 2002) or RELIEF-F (Kononenko et al., 1994). Depending on the current task for classification linear and non-linear Support Vector Machines (SVMs) were used. This way it is possible to differentiate weed species of similar appearance.

The image analysis system in combination with automatic algorithms for plant species discrimination can be included into real-time and map-based approaches for site-specific weed control.

2. Materials and methods

2.1. Data acquisition

Images were taken from greenhouse series and in the field: weed and crop species were grown in pots in the years 2006 and 2008, field data were acquired in maize (2008), winter wheat (2007) and sugar beet crops (2007). Images of the red (R, ca. 580 nm) and infrared (IR, >720 nm) spectrum of the light were taken simultaneously and subtracted from each other (IR-R) to generate difference images (Sökefeld et al., 2007). Most of the red light is absorbed by plants for the photosynthesis, whereas the infrared light is reflected. All other materials in field (soil, mulch, stones) have a similar reflection at both wavelengths. Plant material therefore appears bright in the difference images due to the typical 'red edge' in the reflectance spectrum.

The data set consisted of samples for 10 species, which were to be differentiated: two monocotyledonous, seven dicotyledonous weed species and summer barley were chosen because of their relevance for weed management. All species were in early growth stages, ranging from germination to two-leaf stage, only *C. arvense* appeared with up to five leaves. The different growth stages of each weed and crop (*Hordeum vulgare*, summer barley) were merged, because they have no economic relevance in the management practice, since the management thresholds are set according to the number of plants per m^2.

2.2. Segmentation and feature extraction

The acquired difference images are converted using image processing techniques: a grey value threshold is used to separate plants and background, resulting in binary images with two values, one for the foreground (plant) and the other for the background. In the binary images objects are identified by segmentation of connected foreground components. These segments correspond to plants or parts thereof, if single leaves are not connected after the thresholding step. Fig. 1 shows the resulting objects after binarisation and segmentation for some of the training samples of this study.

Overlapping plants in this step lead to complex objects, containing parts of different plants. These objects would be difficult to separate into their components and even then a proper shape description is likely to fail for the following analysis. Therefore they are put into classes for overlapped plants and as such handled parallel to the other species classes in the system. Overlaps can also be identified according to their shape description, as they lead to large objects and their complexity expresses itself in some of the shape features. Since the monocotyledonous species with their long leaves naturally tend to overlap, these objects are usually assigned to a monocotyledonous class, which especially for *H. vulgare* crops leads to a valid decision. In Fig. 1 some overlapping of *H. vulgare* plants (HORVS) can be seen. The general assumption for the application of this approach is, that the plants are measured in early development stages, where overlapping does not affect the overall sampling accuracy. The shape of single plants cannot be extracted from cluttered scenes, limiting the analysis to early growth stages. The most important herbicide applications, which can benefit from this technology, take place shortly after germination, the problems due to overlapping are limited during this period.

To identify different weeds and crop species, shape parameters were computed for the objects in the image. Some of the shape parameters were derived from the set of pixels belonging to an object, like areasize, inertia values according to main axes of the different object, central moments and moment invariants. Other features were derived from the border representation, like border