Detection of weeds using image processing and clustering

Martin Weis, Roland Gerhards

Department of Weed Science, University of Hohenheim, Otto-Sander-Straße 5, 70599 Stuttgart, Germany
Corresponding author: Martin.Weis@uni-hohenheim.de

Abstract: *Knowledge about the distribution of weeds in the field is a prerequisite for site-specific treatment. Optical sensors make it possible to detect varying weed densities and species, which can be mapped using GPS data. The weeds are extracted from images using image processing and described by shape features. A classification based on the features reveals the type and number of weeds per image. For the classification only a maximum of 16 features out of the 81 computed ones are used. Features are used, which enable an optimal distinction of the weed classes. The selection can be done using data mining algorithms, which rate the discriminance of the features of prototypes. If no prototypes are available, clustering algorithms can be used to automatically generate clusters. In a next step weed classes can be assigned to the clusters. Such a procedure aids to select prototypes, which is done manually. Classes can be identified, that are distinct in the feature space or which are overlapping and therefore not well separable. Clustering can be used in some, less complex cases to establish an automatic procedure for the classification. Weed maps are generated using the system. These are compared to the results of a manual weed sampling.*

1 Introduction

The goal of site-specific weed control is the precise application of herbicides in highly infested areas of a field. Since the distribution of weeds is heterogeneous in most cases and stable across years (GERHARDS *et al.* 1997, MORTENSEN *et al.* 1998, GERHARDS & CHRISTENSEN 2003), site-specific weed control can reduce the amount of herbicides used. The spraying has to be controlled by the actual weed infestation. This way the selection and dosage of the herbicides can be optimized for each part of the field. In areas where the weed infestation is below the economic threshold no herbicides are used, in areas with a weed infestation above the threshold different herbicides in be used in varying dosages, adapted to the weed species. The first step therefore is to get information about the distribution of the different species. Manual weed sampling is time- and cost-intensive and therefore cannot be economic in a wider practice. SLAUGHTER *et al.* (2008) give an overview of the techniques for weed detection and find, that the robust weed detection remains the primary obstacle toward commercial develop-

ISSN 0947-7314
Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)

ment and industry acceptance of robotic weed control technology. Therefore a system was developed to measure the weed infestation.

2 Material and methods

To achieve an automatic classification of weeds from images a system was used, which is outlined in **Figure 1**. In the field two images were taken at the same time: one of the red (R), the other of the infrared (IR) spectrum of the light. Additionally the position was determined using an RTK-GPS. Both images are normalized and a difference image computed (IR-R). In this image the plants appear brighter than the background objects like soil, stones, mulch (SÖKEFELD *et al.* 2007). The background objects vanish in the difference image, because have a similar reflection in thr red and infrared spectrum.

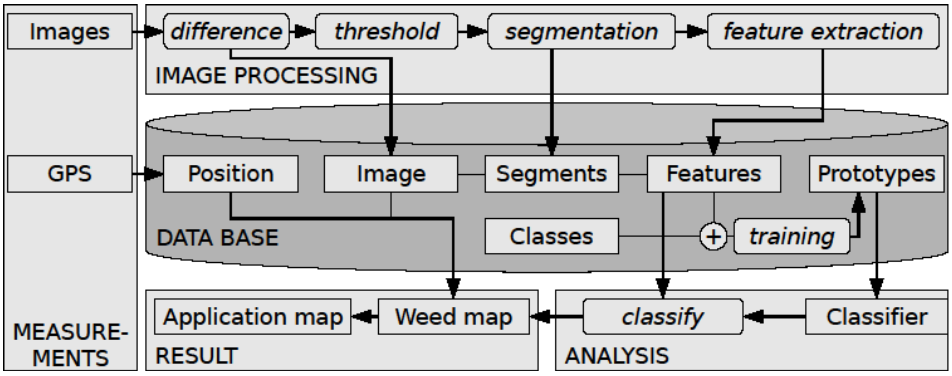

Figure 1: Schematic overview of the image processing

The difference image was analysed using digital image processing and the results are stored in a database (**Figure 1**). The first step is the binarisation with a grey level threshold that segments the foreground objects (white) from the background (black). A segmentation step identifies single foreground objects as objects, which are surrounded by background. Noise can be suppressed in this step, if small regions are filtered out with a size criterion. For the remaining objects features are computed, which characterise the shape of the plants. Geometric features are based on the object pixels, contour features, which are derived from the border pixels (fourier descriptors, curvature), and features based on the skeleton of the regions were computed. Size, compactness and Hu moments (HU 1962) are typical examples for the group of geometric features. The skeletonization was combined with a distance transformation of the regions and lead to a distance vector. Statistical measures from this vector describe then the thickness of the segments. With this step the image processing was complete and the features were used for the following analysis.

In the database all information of the measurement and image processing was stored. The positional information and other metadata (time, camera type, exposure time) are stored together with the file path in the database. During the segmentation step segments were created and their origin (image, image processing parameters), file path and the features were stored in the database. Each segment corresponds to a plant or parts thereof.

The following analysis is based on the values in the database. Prototypes have to be defined for the analysis and a classifier can be trained with this prototype information. Classes were defined within the database, which consist of three parts: the first is the EPPO code (EPPO 2007) denoting the species, the second the BBCH code (HESS *et al.* 1997) denoting the phenological growth stage and the third is an attribute which describes special cases that may occur due to the segmentation, e.g. single leaves or overlapping. In the training step defined classes are related to the segments, these are then prototypes with known class. This step has to be done manually by sighting the segments and assigning a class to selected segments. The training data then resides in the database and forms the training data set together with the features, which is used to train a classifier. The trained classifier then assigns all segments to classes. Together with the positional information the classification result can be assemled to weed distribution maps. The weed distribution maps are used to create application maps, which are used for the site-specific herbicide treatment. Different weed species, which are sensitive to the same herbicides, are grouped for the application map and economic thresholds are used to define the areas which have to be sprayed.

2.1 Feature selection

Not all features have the same relevance for the classification. Each feature adds a dimension to the feature space, which is high dimensional. In our system up to 81 features were computed, which are all numerical. Therefore the prototypes were located in a 81-dimensional feature space and it was necessary to reduce the number of dimensions. A feature selection or feature transformation can be used for this task. Selection algorithms weight features according to their discriminational abilities and select the ones that allow the best discrimination of the classes. The maximum number of features should not exceed 16 and the prototypes were used to select them. Selection algorithms can be grouped into two groups. One group uses the discriminative abilities of each feature or correlations of features to select the best ones. PCA weighting is one example, which uses the coefficients of a principal component analysis to weight the features. The other group of selection algorithms uses classification algorithms. Feature subsets are selected and the performance of a classifier, which uses only these subsets, is used.

Two different algorithms for a weighted selection were used, which are implemented in the data mining program RAPIDMINER (MIERSWA *et al.* 2006). The selection process consisted of a weighting using PCA weighting and info gain weighting with a following selection, that recursively did a (crossvalidated) nearest neighbor classification. The performance of the classification und therefore the feature subset was rated and feature subsets could be identified, which are optimized for the dicrimination of the classes.

2.2 Clustering

A supervised classifier that uses the training data of the prototypes is used to assign classes to the objects. Unsupervised classification algorithms, also known as clustering, can be used, if no training data exists. These kind of classifiers were used before class information was available and before the manual selection of prototypes has taken place. These algorithms aggregate similar objects to clusters according to the feature information. In this context clustering is interesting in two ways: they can be used to support the training and they show similarities or differences between classes, giving hints on the seperability. Clustering was used here to group plants with similar shapes. In a second step classes can be set for these automatically derived clusters and prototypes can be selected. The advantage of the approach is, that classes with similar features can be identified. The training for classes which are difficult to separate can be optimized this way. If class has a multimodal distribution in the feature space, leading to two or more clusters (see clusters 0, 1 and 2 in **Figure 4**), then an additional classes may be defined. The same applies, if there are still noise objects left: a noise class can be used for them. The resulting prototype definitions are used for the classification.

2.3 Data set

The data for the clustering was derived from images taken in December 2007. The field had a size of 3.5 ha with a winter wheat (*Triticum aestivum* L.) crop. The crop was not emerged at that time and the weeds were grass weeds (mainly *Alopecurus myosuroides* Huds.) and dicotyledon weeds (*Veronica persica* L., *Matricaria chamomilla* L.). The image series contains of 3367 images and their DGPS-coordinates. 160 images were selected from the series, which were near manual sampling points, and used for the training. The number of weeds, separated for each species, were counted manually from the images and compared to the results of the image classification. Additionally a manual field sampling was done using a frame to count the weed densities for each species. The position of the images and manual sampling points are not exactly the same, but differ up to two meters.

A map was created from the results of the supervised and unsupervised classification and the manual sampling. It can be seen that the weed patches can be found with each of the methods (**Figure 2**).

ISSN 0947-7314
Leibniz-Institut für Agrartechnik Potsdam-Bornim e.V. (ATB)

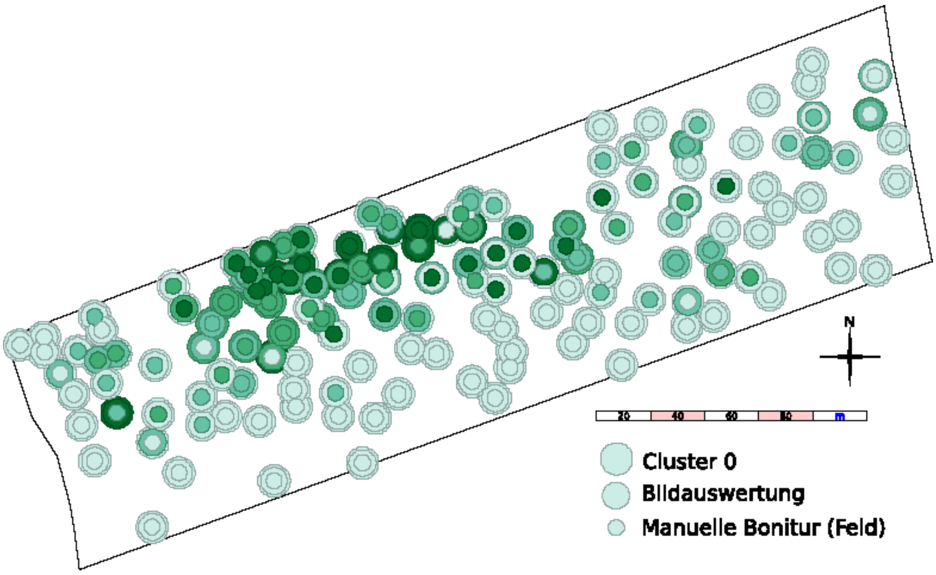

Figure 2: Map of the weed densities using different measurements: outer rings contain results of unsupervised classification (Cluster 0), middle ring shows the supervised classification using prototypes, inner circle manual sampling results

3 Results and discussion

The images were analysed following the schema from **Figure 1**, additionally there was the data of manual measurements from the images. The weeds were counted from the subset of the images, they were also used to select the prototypes. The result of the clustering can be compared to the training data. The clusters and object classes can be visualised in the feature space using three features. **Figure 3** shows the trainingdata of the prototypes on the left and the result of the clustering on the right (clusters numbered from zero to four).

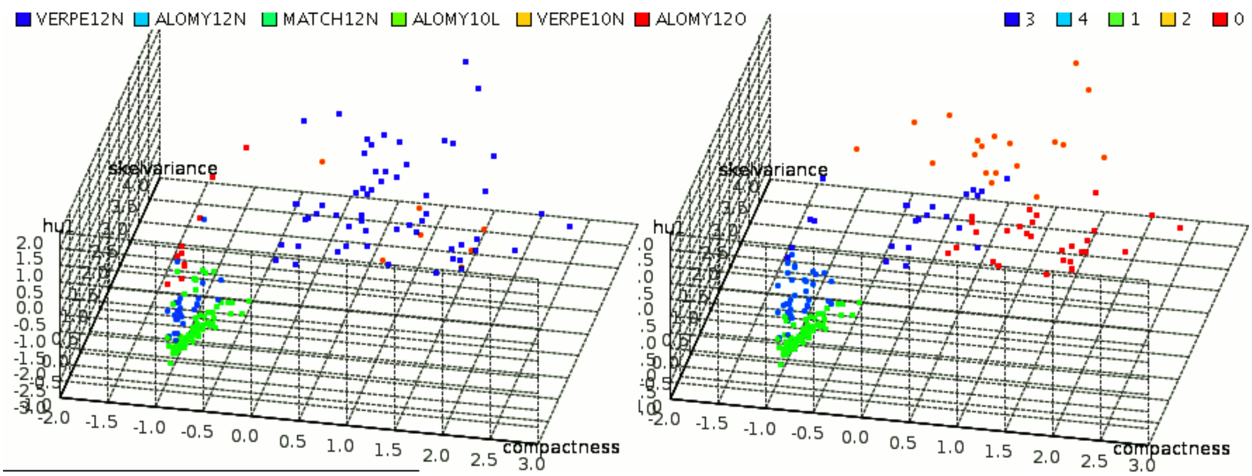

Figure 3: Classes (manually selected prototypes) and clusters (unsupervised classification) in the feature space. The features are compactness, variance of the skeleton distance vector and the first Hu moment

The segments were labeled with colors according to the clusters and combined to label images. The combination of these with the (inverted) difference images is shown in **Figure 4**. Dicotyledon weeds can be found in cluster four, the monocotyledons are in cluster zero to three. The unsupervised classification can distinguish between these important weed classes. By assigning classes to the clusters the training step can be simplified, prototypes can be marked as belonging to that class. This way an efficient training is possible. It can also be seen, that the monocotyledon weeds have a multimodal distribution (single, elongated leaves, overlapped leaves). This can be taken into account in the class definitions for the prototypes, this class can be separated into two subclasses.

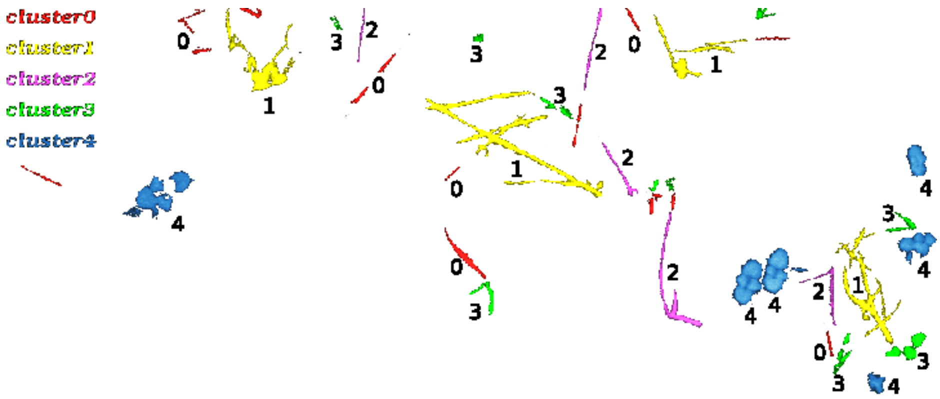

Figure 4: Cluster assignments in an example image, cluster 4 contains dicotyledonous, cluster 0-3 monocotyledonous weeds

Figure 5 relates the manual class assignments for *Alopecurus myosuroides* Huds. to cluster zero. There were some images without objects in the cluster (points on the ordinate), but in most of the images the numbers show the same tendency as the manually determined ones.

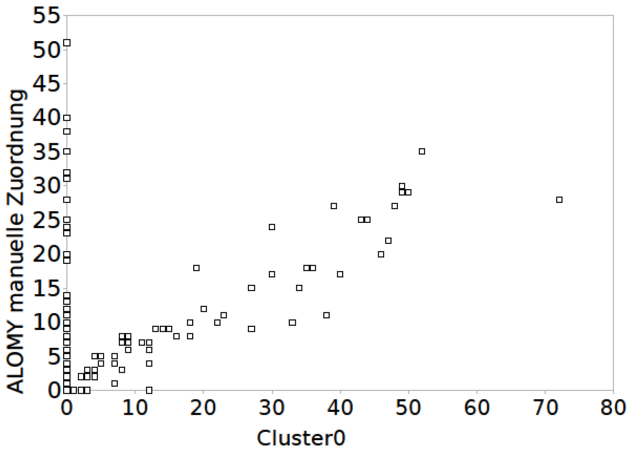

Figure 5: Correlation of the manual class assignments (ALOMY: *Alopecurus myosuroides* Huds.) and cluster zero of the unsupervised classification (clustering)

ISSN 0947-7314

4 Conclusions

Weed sampling from camera images, as described in this approach, can be used to generate weed maps with high spatial density, which are necessary for a site-specific weed management. The approach uses bi-spectral images, which allow a good separation between plants and background. The analysis is based on the shape of single plants, which are parametrised using shape features. Supervised classifiers need training data, which are selected prototypes of weed and crop plants. The selection of the prototypes can be supported by unsupervised classification (clustering) by assigning classes to automatically derived clusters. This way the separability of the classes according to the shape features can be visually assessed already in the training step and the classes and training data can be adjusted to the situation.

Unsupervised classification could be used to establish a fully automatic approach, if prior information about the weed species are introduced as starting values for the clustering.

References

EPPO (2007): Harmonized classification and coding of the uses of plant protection products. *EPPO Bulletin*, 37(1):25–28, doi: 10.1111/j.1365-2338.2007.01069.x. European and Mediterranean Plant Protection Organization.

GERHARDS R., CHRISTENSEN S. (2003): Real-time weed detection, decision making and patch spraying in maize (zea mays l.), sugarbeet (beta vulgaris l.), winter wheat (triticum aestivum l.) and winter barley (hordeum vulgare l.). *Weed Research*, 43:1–8

GERHARDS R., WYSE-PESTER D.Y., MORTENSEN D.A., JOHNSON G.A. (1997): Characterizing spatial stability of weed populations using interpolated maps. *Weed Science*, 45:108–119

HESS M., BARRALIS G., BLEIHOLDER H., BUHR L., EGGERS TH., HACK H., STAUSS R. (1997): Use of the extended bbch scale - general for the descriptions of the growth stages of mono- and dicotyledonous weed species. *Weed Research*, 37:433–441

HU M.K. (1962): Visual pattern recognition by moment invariants. *IRE Transactions Information Theory*, 8(2):179–187, ISSN 0018-9448. URL http://ieeexplore.ieee.org/xpl/abs_free.jsp%3 FarNumber=1057692

MIERSWA I., WURST M., KLINKENBERG R., SCHOLZ M., EULER T. (2006): Yale: Rapid prototyping for complex data mining tasks. In Lyle Ungar, Mark Craven, Dimitrios Gunopulos, and Tina Eliassi-Rad, editors, *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940, New York, NY, USA, ACM. ISBN 1-59593-339-5. doi: http://doi.acm.org/10.1145/1150402.1150531. URL http://rapid-i.com/ component/option,com_docman/task,doc_download/gid,25/Itemid,62/

MORTENSEN D.A., DIELEMAN J.A., JOHNSON G.A. (1998): Weed spatial variation and weed management. In J. L. Hatfield, D. D. Buhler, and B. A. Steward, editors, *Integrated weed and soil management*, pages 293–309, Chelsea, MI, USA, Ann Arbour Press

SLAUGHTER D.C., GILES D.K., DOWNEY D. (2008): Autonomous robotic weed control systems: A review. *Comput. Electron. Agric.*, 61(1):63–78, ISSN 0168-1699. doi: http: //dx.doi.org/10.1016/ j.compag.2007.05.008

SÖKEFELD M., GERHARDS R., OEBEL H., THERBURG R.-D. (2007): Image acquisition for weed detection and identification by digital image analysis. In J.V. Stafford, editor, *Precision agriculture '07*, volume 6, pages 523–529, The Netherlands, 6th European Conference on Precision Agriculture (ECPA), Wageningen Academic Publishers. ISBN 978-90-8686-024-1