

Journal Classification System: **PA** (Precision Agriculture)

Computer Vision based Weed Identification under Field Conditions using Controlled Lighting

J. Hemming¹; T. Rath²

¹ Institute of Agricultural and Environmental Engineering (IMAG), Wageningen, The Netherlands;
e-mail: j.hemming@imag.wag-ur.nl

² Institute for Horticultural and Agricultural Engineering (ITG), University of Hanover, Germany;
e-mail: rath@itg.uni-hannover.de

Abstract

The methods of digital image analysis were used to develop an identification system for weeds in crops. Two vegetable crops (cabbage and carrots) and a number of naturally occurring weed species were used to develop the classification algorithms. Considering the rougher environment, special attention was given to the open field experiments. The images were obtained with a device that provided controlled lighting conditions. The analysis was carried out off line. Eight different morphological features and three colour features were calculated for each single object to build a joint feature space. On the basis of sample data sets of each class, statistics were carried out to determine the features, which are suitable for discrimination. A membership function based on a fuzzy logic approach was formed and used for the classification. The experiments showed that colour features can help to increase the classification accuracy. Moreover, colour was used successfully for the segmentation procedure of plants and soil. Depending on growth stage, weed density and method of calculation between 51 and 95% of the plants were classified correctly. Problems still exist by separating and allocating single plants in plant stands where the plants have grown together. Compared to other studies the plant identification system presented is an improvement, especially considering that the experiments were carried out under field conditions.

1. Introduction

Due to the rapidly increasing capacities of computers, image processing and online classification systems gain more and more importance in many areas. For production processes, quality assurance, security systems, in medicine and in surveying, the use increased within a few years exponentially. Such systems are by far superior to human eyes if reproducible detection and quantitative classification is necessary. Machine vision systems have been proposed for various agricultural applications by providing needed information about soil condition, residue cover, plant health, plant species identification, plant population density, and plant size. Weeds could be identified, counted, and mapped in fields for planning future herbicide applications (Woebbecke *et al.*, 1995). This would lead to the possibility of localised spraying of herbicides and thus reduce chemical waste, crop damage and environmental pollution. The main problem using image processing for such tasks is the fact that plants are very variable in size and colour, even within a species and within a single plant stand. Many of the methods used in industrial image processing cannot be used without adoption. Tillett (1991) gives a review of potential opportunities by image analysis for agricultural processes. Numerous research activities in the area of plant identification have been initiated in recent years. In the first investigations, different morphological features were

extracted from digitised images of plants. These features were compressed and used for identification models (Petry & Kühbauch, 1989). Others extended such models with knowledge-based rules (Guyer *et al.*, 1993). Sökefeld *et al.* (1996) managed to identify 70% of 22 different weed species common in sugar beet by computer vision using morphological features and Fourier descriptors. Rath (1997) presented a plant identification system for identifying native deciduous trees by their foliage leaves. A maximum of 96% computer-based identification of unknown leaves was achieved. However, the images were taken under laboratory conditions.

The use of such information for an automated weed control was investigated by Tillett *et al.* (1998). They developed a small autonomous vehicle for spot treatment with spray of transplanted cauliflower, which shows the potentials of saving in agro-chemicals. Tian *et al.* (1999) presented the development of a precision sprayer for site-specific weed management.

The objective of the work described here is to develop an identification and location system for weeds in crops with the methods of digital image analysis. Special attention is put on open field experiments. The paper concentrates on the methods and algorithm development and evaluation of an experimental system. Plant information gained with such a system can be used for a number of applications, including automated weed mapping and selective weed control.

2. Experimental details

2.1. Greenhouse and open field procedure

Two different crops were used in greenhouse and open field experiments. Cabbage (*Brassica oleracea* L. conv. *capitata* var. *capitata*) and carrots (*Daucus carota* L.). The foliage leaves of cabbage have a simple form. They are typically round to ellipse shaped with a smooth leafmargin. As a counterpart in shape, carrot plants with feathery leaves were used.

The greenhouse experiments took place on a ground bed with an extending to 3 m by 15 m. The crops were cultivated as in horticultural practice. Cabbage was grown with a row distance of 40 cm and a plant spacing of 30 cm. Carrots were sown with a row distance of 20 cm (greenhouse) and 30 cm (open field). For the greenhouse experiments, extra weed seed was added to create a weedy population. The main weed species emerged in the experiments were *Alopecurus myosuroides*, *Galinsoga ciliata*, *Lamium amplexicaule*, *Matricaria discoidea*, *Oxalis corniculata*, *Ranunculus arvensis*, *Taraxacum officinale* and *Urtica urens*. To simulate tractor operations, a mobile carrier rack was constructed. The rack carries three cameras, the video electronics and the illumination system. Three special plant lights with 400 W metal halogen lamps were used to illuminate the recording area. To block out the natural light, the rack was covered completely with lightproof polyethylene (PE) film.

For the open field experiments, a tractor add-on unit was constructed. This unit consists of a steel-tube rack covered with lightproof PE-film. Similar to the greenhouse test stand, lamps and cameras are mounted inside the unit (*Fig. 1*). As for the greenhouse experiments, the crops in the open field were cultivated as in horticultural practice. No extra weed was sown, the natural weed population was used.

2.2. Computing and video equipment

The digitisation and the analysis of the images was carried out with a standard personal computer (PC) with 350 MHz processor and 196 MB RAM. A Matrox Meteor red/green/blue (RGB) framegrabber card was used to capture the images. The image analysis was carried out with the software package HALCON 5.01 (MVTec, 1998) in combination with Microsoft Visual C++ Compiler 4.0 under the operating system Windows NT 4.0. 3-chip colour charge coupled device (CCD) cameras (Hitachi, HV-C20) served as image sensor. Three cameras next to each other were used to enlarge the width of the recording area in combination with a satisfactory image resolution. A multiplexer made it possible to switch between the video signals of the different cameras. With this system, it was possible to operate more than one camera on one individual framegrabber. The cameras were mounted in a way that a certain part of the image overlapped in horizontal orientation with the neighbouring images. The forward movement of the machine achieved the overlapping in vertical orientation. For the greenhouse experiments a time-adjustable pulse generator supplied the signals for the digitisation of new images. For the open field experiments, the distance covered by the tractor was determined by counting the impulses generated by a rotation angle encoder attached to the axis of the left front wheel.

2.3. Image acquisition

Depending on the growth rate, images of the plants were taken with several days in between, 18 to 49 days after seeding. The digitised images were stored as 24 bit colour images with a resolution of 768 pixels by 576 pixels in the tagged image file format (TIF) on a computer hard disc during the drive. The recording area per image varied between 350 mm by 263 mm and 600 mm by 450 mm, depending on the crop and the used system. On average 60 single images per evaluation day were analysed, containing about 100 (experiments with cabbage) to 300 (experiments with carrots) crop plants. The analysis of the images was carried out as a second step in the office.

3. Algorithm development

The methods of image analysis are used to identify and localise crop and weed plants. For this the plant stand was digitised in single images touching each other. The recorded plants were separated by image processing operations. Several form and colour features were calculated for each object. These features were used to build a classification function to group the plants.

3.1. Colour space transformation

Preliminary tests showed that the separation of the green plants and the brown soil is insufficient by applying a threshold only on one channel of the RGB colour image. Much better results were obtained by first transforming the RGB image into a more perception orientated colour space and then applying a threshold using all colour channels. For this reason, each image was transformed from the RGB to the hue, saturation and intensity (HIS) colour space using the procedure described in Kopp, 1997.

3.2. Binarisation and preprocessing

Threshold values for the plants (*i.e.* the 'green' on the image) for the three channels H, S and I were determined by selecting the colours of interest. By applying a three-dimensional threshold to the image, the regions of interest (the plants) were separated and resulted in binary images. To eliminate digitisation errors caused by camera and framegrabber, holes in the

input regions with an area of less than 15 pixels were filled up and subsequently all regions smaller 15 pixels were removed.

3.3. Segmentation

After binarisation, plants grown together were represented as one connected region in the image. At this stage, it was impossible for the computer to allocate single plants or even single leaves. With the help of the morphological image processing operator erosion, it was possible – to a limited extend – to segment the single components. Thus, it was possible to split even somewhat overlapping leaves. To return to the original size and structure of the leaves, the eroded regions were iteratively processed with dilation and then an intersection with the original input image was performed. This procedure was successful in particular to segment plants with simple leaves, *e.g.* cabbage. *Figure 2* shows an example of singulating the leaves of a cabbage plantation with morphological operations. As a structuring element, a three pixel by three pixel cross mask was used for all morphological operations. A comprehensive description of the segmentation procedure is given in Hemming (2000).

3.4. Feature extraction

For each single region obtained in the segmentation procedure, eight different form features and three colour features were calculated. The calculation of most region features is directly provided by the HALCON-Software (MVTec, 1998); others are a simple combination of basic geometric units or are derived from own considerations. Table 1 shows the calculation of the parameters used. The calculation of the parameter spikes needs more explanation: it refers to the leaf margin structure, *i.e.* smooth or toothed leaf margin. The method is to count the number of indentations of the leaf margin. From empirical preliminary investigations, the area of an item representing a spike was defined as having an area of between 1/1000 and 1/100 of the area of the leaf itself. To make the parameter more independent in size, the number of counted spikes n on a leaf was divided by the contour length of the convex hull of the leaf region C_h . *Figure 3* shows examples for different leaf margins.

3.5. Obtaining the training data sets

Prior to the automatic classification the system was trained with the objects to identify. Four different object classes were established. The first class is the crop, and for different weeds up to three additional classes are possible. The analytical program has a special mode that permits the selection of objects manually identified with a mouse-pointer on the screen. All shape and colour features presented in the previous section were calculated for these objects and stored in a file. In this way, a training data file for each object class was established. The data for the training sets were derived from a separate part of the recording area on the same day the evaluation took place. The number of samples per class can be chosen freely; in the investigations presented here, between 5 and 15 samples per class were made available.

3.6. Parameter selection

3.6.1. Statistics

Not all of the parameters were equally taken into account for the classification process. A statistical test was carried out to find the most suitable parameters, showing significant differences between the crop class and the weed classes. At the beginning, the system loads the parameters from the prebuilt training data files. A test for significantly different means between the crop class and each of the weed classes was carried out. The t-test for unknown variances which are possibly unequal was used for this purpose (Behrens-Fisher problem, see

Sachs , 1997). With the Student's distribution probability function (detailed calculation described in Press *et al.*, 1992) the error probability was determined.

For the classification process, the analytical program used this information either for parameter selection (non-weighted parameters) or for weighting the parameters for the subsequent establishment of the membership function. With these steps, it was taken into account that use of parameters without significant differences will reduce the discrimination of the classes.

3.6.2. *Non-weighted parameters*

Only parameters with significantly different means in the training data sets of crop and weed classes and with less than a certain error probability are used for the subsequently calculation of the membership. Throughout the investigations presented here, a threshold of 5% ($\alpha < 0.05$) was used. The selection was carried out dynamically with the use of the training data sets each time the program was run. Alternatively the analytical program can be configured to use a fixed number of parameters. The selection depends on the error probability, beginning with the parameter that shows the smallest error probability.

3.6.3 *Weighted parameters*

As an option, all parameters are incorporated in the classification and weighted depending on their error probability. Instead of an exclusion of a parameter, the weighting method takes into account the statistically calculated suitability. The more clearly the differences between the classes can be revealed by a parameter, the more importance it gets in the calculation of the membership. Two different weighting functions were applied: weak and strong weighted parameters.

For the weakly weighted function, all parameters with more than 50% error probability have no influence on the membership calculation. For the strongly weighted function, parameters with more than 5% error probability are already of no consequence. All others are weighted in a linear way between 0 and 1 for the membership according to their error probability.

For the calculation of the weakly weighted factors:

$$\phi_i = \begin{cases} 2(0.5 - \alpha_i) & \text{if } 2(0.5 - \alpha_i) > 0 \\ 0 & \text{else} \end{cases} \quad (1)$$

For the calculation of the strongly weighted factors:

$$\phi_i = \begin{cases} 20(0.05 - \alpha_i) & \text{if } 20(0.05 - \alpha_i) > 0 \\ 0 & \text{else} \end{cases} \quad (2)$$

Where ϕ_i is the weighting factor for the i^{th} parameter; and α_i is the error probability for the i^{th} parameter for significantly different means between the crop class and the weed class

3.7. Classification

Based on the fuzzy logic orientated model introduced by Rath , 1997), a simplified membership function for each parameter and each class was formed. This function is based on the standardised Gaussian curve and the statistical values ‘mean’ and ‘standard deviation’ of the training data set.

$$M_i = \begin{cases} 1 & \text{if } \sigma_i = 0 \\ -0.5 \left(\frac{\Omega_i - \bar{\Omega}_i}{\sigma_i} \right)^2 & \text{else} \\ e & \end{cases} \quad (3)$$

where i is the index of calculated parameter of a classification class; M_i is the membership for i^{th} parameter to the training data set; Ω_i is the value of the i^{th} parameter of a classification class; $\bar{\Omega}_i$ is the arithmetic mean of the i^{th} parameter in the training data set; and σ_i is the standard deviation of the training data set of a classification class.

For all segmented objects in the image, the membership values of the class of the crop and of the different weed classes were calculated for each parameter used. The assessment of the total membership depended on the weighting of the parameters. Eventually, the object was classified as an object of the class of which it had the highest membership.

3.7.1. Non-weighted parameters

As described earlier, either an ‘error probability threshold dependent selection’ or a ‘fixed number of parameters selection’ was carried out. The parameters selected by these methods were then used equally for the calculation of the total membership. The calculation of the total membership was done as follows:

$$M = \frac{\sum_{i=1}^n M_i}{n} \quad (4)$$

where: M is the total membership to a classification class; and n is the number of selected parameters.

3.7.2. Weighted parameters

With the weighting functions, all available parameters together with the determined weighting factor were used for the calculation of the total membership. The calculation of the total membership with weighted factors was carried out as follows:

$$M = \frac{\sum_{i=1}^n \phi_i M_i}{n} \quad (5)$$

3.8. Calculation of identification success

In order to recognise not only single leaves but also total plants, a cluster analysis was executed with the objects identified as leaves of the crop. Two objects were put into the same

cluster if their minimal distance was smaller than a calculated threshold value (for details see Hemming, 2000). All remaining and still unclassified plant areas on the image separated by the binarisation procedure were finally also classified as weeds. For each analysis, several methods of calculation were carried out. The method of parameter selection/weighting was altered. To probe the influence of the colour-based parameters the colour-based parameters could be excluded. To calculate the identification success, the result of the automatic classification was compared with the result of a visual classification (by a human).

For each analysis day, the mean values of the following parameters were calculated:

- (a) total area of the regions after binarization and pre-processing (the area covered by plant material)
- (b) area of the manual identified crops
- (c) area of automatically identified crops
- (d) type 1 error (percentage of not identified crop plant area)
- (e) type 2 error (percentage of weed plant area classified as crop plant area)

4. Results

4.1. Suitability of classification parameters

The percentage of use of the classification parameters in the non-weighted calculations is summarised in *Fig. 4* for the experiments with cabbage and in *Fig. 5* for the experiments with carrots. On the one hand, some parameters for the experiments with cabbage turned out to be very suitable for the classification. The feature 'area/circumference' was always used in the classification and 'area' and 'hue' still in more than 90% of the calculations. On the other hand, the characteristics 'length/width' and 'spikes' were not very appropriate to make a distinction between cabbage plants and weeds: in less than 20% of the calculations, these parameters were used. All other parameters were used in 50 to 75% of the cases.

For the experiments with carrots, the most frequently used parameter was 'maximum diameter' in 78% of the calculations. Most of the other parameters were equivalent and used in 56 to 67% of the cases. Insignificant parameters for the discrimination of carrots and weeds were 'spikes' (22%), 'length/width' and 'intensity' (33% each).

4.2. Plant classification

4.2.1. Experiments with cabbage

For greater clarity, a closer look is taken at the method of calculation with non-weighted parameters, an error probability threshold dependent parameter selection and a training data set with 15 samples per class. A mean classification rate of 88.15% correctly classified crop plants was achieved in conjunction with a mean of 11.85% for type 2 error. *Figure 6* shows an overview of the classification results for this method of calculation. For the greenhouse experiments, the best crop identification was achieved in the mean growth stage, 36 days after sowing. In the early stage, not all plant features were adequately developed and the resolution of the image concerning the object size was comparatively low. After a certain age of the crop, more and more misclassification occurred and there was an especially sharp rise in type 2 errors. This result was attributed to the fact that the single plants were growing closer together and that leaves touched or overlapped each other. If the leaves overlapped too much,

the use of morphological operations to separate the components failed. Moreover, the weed density increased which made it also more complicated to isolate and analyse single leaves.

4.2.2. Experiments with carrots

The classification of carrot plants turned out to be more difficult than the classification of cabbage plants. A mean classification accuracy of 72.46% correctly classified crop plants was achieved with a mean of 35.61% for type 2 error. *Figure 7* shows an overview of the classification results for carrots. Due to limited plant material, only 5 samples were used to build the training data set 34 days after sowing (greenhouse experiments) and only 10 samples after 58 days (open field experiments), respectively. As for cabbage, the best crop plant identification was achieved in the mean growth stage where the first foliage leaf of the carrot is already developed but there is still some distance between the single plants.

4.3. Effect of number of samples in the training data set

The expected improvement in classification by a higher number of samples in the training data set cannot be confirmed definitely. However, there is a correlation between the two error types: the higher the type 1 error the lower the type 2 error and vice versa. For the greenhouse experiments the number of samples played only a minor role for the classification success. For the open field trials only at the earliest growth stage the triplication of the number of samples enhanced the identification rate of the crop of more than 10% by accepting a doubling of type 2 error.

4.4. Effect of parameter weighting

The influence of the different parameter weighting functions on the classification errors is presented in Table 2. All results are compared to the method of calculation with non-weighted parameters and an error probability threshold dependent parameter selection. It can be seen that the application of the weakly weighted parameter function does not lead to a better classification result. By applying the strongly weighted parameter function instead, both type 1 and type 2 errors are successfully reduced for both crops.

4.5. Effect of colour analysis

As already shown, the colour-based parameters were selected rather often. For cabbage, mostly 'intensity' was used (in 75% of the cases), 'hue' in 67% and 'saturation' in 58% of the cases. For carrots, 'hue' and 'saturation' were selected in 67% of all evaluations, 'intensity' was only chosen in 33% of the cases. To probe the influence of the colour-based parameters ('hue', 'saturation' and 'intensity'), a method of calculation was carried out where the use of these parameters was blocked. For the greenhouse experiments with cabbage the classification error was increased by more than 8% to 11.30% for type 1 error (not identified crop) in the early growing stage. The error was lowered by more than 2% even at a later date. For the open field experiments, the use of colour was more indifferent but the positive effects always predominated. In the carrot crop, the use of the colour-based parameters did not show clear advantages. Without these parameters, the type 1 error was even lowered by nearly 3% on the average.

5. Discussion

In contrast to most other research in the investigations presented, the images were taken from plants under field conditions. Accordingly, weeds grew at random positions and leaves overlapped each other quite often or were partially occluded by adjacent plants. Therefore, an important task before feature extraction was the partition and separation of the single plants. Recognition of occluded objects is considered one of the toughest problems in image processing and pattern recognition and also other authors came across this problem (Whittaker *et al.*, 1987; Franz *et al.*, 1991; Tian & Slaughter, 1993; Lee *et al.*, 1997). The algorithms available are still cumbersome and can only be used with special limitations. As shown, the use of simple morphological image processing operators (erosion and dilation) achieved a rather good segmentation of single leaves for plants with simple leaves, such as cabbage. Using morphological operators as well, Lee *et al.* (1997) and Shatadal *et al.* (1995) achieved comparably good results. The separation of single plants in the carrot crop developed more complex and has not yet been solved for all growth stages. Other methods, such as the Fourier-Mellin based curvature method developed by Franz *et al.* (1991) do not work if the leaves have a variable leaf serration and cannot be used because of this limitation.

The classification model is based on the investigations of Rath (1997) where an in-depth analysis of different data interpretation models for identification and classification of foliage leaves of deciduous trees was carried out. The study compared knowledge based systems, fuzzy logic orientated approaches, artificial neuronal networks and statistically based systems. It is concluded that data interpretation models for plant identification systems should use statistically based algorithms, such as discriminant analysis or a fuzzy logic based classifier. It is mentioned that the use of linear discriminant analysis may turn out very labour-intensive and recommended fuzzy methods instead. Fuzzy logic based methods were also found to be promising for plant identification by Simonton & Graham (1996). The used statistical t-test for parameter selection and/or parameter weighting is based on the values 'mean' and 'variance' as well as the membership function. Using the weighting functions presented, features with low variances in the sample data (and therefore suitable for a discrimination) are attributed a high influence on the membership calculation. The approach of automatically weighting the features turned out to be very suitable. This differs from other publications dealing with plant identification where the weights had to be selected manually after a statistical analysis (Tian & Slaughter, 1993; Sökefeld *et al.*, 1996).

It is difficult to compare the classification results with the ones published somewhere else because of the extremely different boundary conditions in each study. Others who worked in the area of individual plant identification achieved an average identification accuracy of 68% (Sökefeld *et al.*, 1996) or 30 to 77% (Meyer *et al.*, 1998) for the discrimination of different weeds (laboratory conditions); 61% for identifying tomato seedlings in the open field (Tian & Slaughter, 1993) and 88% accuracy for the classification of deciduous trees from their leaves (Rath, 1997). The experiments presented in this work were carried out under practically orientated conditions. Therefore, the achieved average identification accuracy of 88% for cabbage and 72% for carrots is an improvement.

Due to the fact that there are no fixed classification classes and that own training data sets had to be created from a small sample of the crop, a maximum of variability was achieved. The system can be trained to identify one species among all other plants in the stand and it does not make any difference if this is a weed plant or a crop plant. Once recorded, the images can be processed several times to identify different species. Out of the available classification features, only the descriptors, which are suitable for a discrimination of the classification

classes, were selected automatically (or used with a higher weighting). Problems may occur if there is a very high variability within one of the classes and if the features are not normally distributed because the statistics used are based on the assumption that the distribution of the parameters within each group is multivariately normal. One major limitation for the transferability to different crops is the segmentation procedure for plants with fine structured leaves. As described, it will not be possible to separate single plants or single leaves of such plants in dense plant stands, resulting in a decrease of the classification accuracy.

6. Conclusions

A system with high classification accuracy for plants grown under field conditions was established using morphological and colour features in a joint feature space. The experiments showed that colour features can help to increase the classification accuracy. Moreover, colour was used successfully for the segmentation procedure of plants and soil. As the methods presented are not specialised in a single species, they should be transferable easily to other crops and weeds. According to the results of the statistics, the features suitable for discrimination were ranked or weighted dynamically. Data derived from such a system can be used for a number of applications. These include automated weed mapping, individual spot treatment, autonomous vision-guided agricultural vehicles and after adaptation also for more specialised tasks, such as selective harvesting.

Acknowledgements

This work was supported by grant from the German research association (Deutsche Forschungsgemeinschaft).

References

- Franz E; Gebhardt M R; Unklesbay K B** (1991). Shape description if completely visible and partially occluded leaves for identifying plants in digital images. *Transactions of the ASAE*, **34**(2), 673-681
- Guyer D E; Miles G E, Gaultney L D; Schreiber M M** (1993). Application of machine vision to shape analysis in leaf and plant identification. *Transactions of the ASAE*, **36**(1), 163-171
- Hemming J** (2000). Computer vision for identifying weeds in crops. PhD Thesis, Gartenbautechnische Informationen, Institut für Technik im Gartenbau und Landwirtschaft. **Heft 50**. University of Hanover, Germany
- Katayama T; Okamoto T; Imou K; Torii T; Mukai T** (1998). Identification of plants for wild flower garden. In: *Proceedings from the 3rd IFAC/CIGR Workshop on Artificial Intelligence in Agriculture*. (Kozai T; Murase H; Hoshi T eds), 164-169. Elsevier.
- Kopp H** (1997). *Bildverarbeitung interaktiv [Interactive image processing]*. B. G. Teubner Verlag. Stuttgart
- Lee W S; Slaughter D C; Giles D K** (1997). Robotic weed control system for tomatoes using machine vision system and precision chemical application. *ASAE Paper No 97-3093*
- Meyer G E; Mehta T; Kocher M F; Mortensen D A; Samal A** (1998). Textural imaging and discriminant analysis for distinguishing weeds for spot spraying. *Transactions of the ASAE* **41**(4), 1189-1197
- MVTec** (1998). *HALCON Reference Manual C++ Version 5.01*. MVTec Software GmbH. München

- Petry W; Kühbauch W** (1989). Automatisierte Unterscheidung von Unkrautarten nach Formparametern mit Hilfe der quantitativen Bildanalyse [Automated discrimination of weed species with quantitative image analysis]. *Journal of Agronomy & Crop Science* **163**, 345-351
- Press W H; Teukolsky S A; Vetterling W T; Flannery B P** (1992). *Numerical Recipes in C*. 2nd ed. Cambridge Press, USA
- Rath T** (1997). Methoden zur computerbildanalytischen Pflanzenidentifikation am Beispiel dendrologischer Bestimmungen [Methods for plant identification by image processing demonstrated in the area of dendrological determinations]. *Gartenbautechnische Informationen*, Institut für Technik im Gartenbau und Landwirtschaft. **Heft 42**. University of Hanover, Germany
- Sachs L** (1997). *Angewandte Statistik [Applied statistics]*. 8. ed. Springer Verlag, Berlin
- Shatadal P; Jayas D S; Bulley N R** (1995). Digital image analysis for software separation and classification of touching grains: I. disconnect algorithm. *Transactions of the ASAE*, **38**(2), 635-643
- Simonton W; Graham D** (1996). Bayesian and fuzzy logic classification for plant structure analysis. *Applied Engineering in Agriculture*, **12**(1), 89-97
- Sökefeld M; Gerhards R; Kühbauch W** (1996). Automatische Erkennung von Unkräutern im Keimblattstadium mit digitaler Bildverarbeitung [Automated identification of weeds in the stage of cotyledon using image analysis]. KTBL-Arbeitspapier. In: *Innovative Verfahren zur Unkrauterkenennung* **236**, 47-58. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. (KTBL). Darmstadt
- Tian L; Slaughter D C** (1993). Computer vision identification of tomato seedlings in natural outdoor scenes. ASAE Paper No 93-3608
- Tian L; Reid J F; Hummel J W** (1999). Development of a precision sprayer for site-specific weed management. *Transactions of the ASAE*, **42**(4), 893-900
- Tillett R D** (1991). Image analysis for agricultural processes: a review of potential opportunities. *Journal of Agricultural Engineering Research*, **12**, 247-258
- Tillett N D; Hague T; Marchant J A** (1998). A robotic system for plant-scale husbandry. *Journal of Agricultural Engineering Research*, **69**, 169-178
- Timmermanns A J M; Hulzebosch A A** (1996). Computer vision system for on-line sorting of pot plants using an artificial neural network classifier. *Computers and Electronics in Agriculture*, **15**, 41-55
- Whittaker A D; Gaultney L D; Miles G E; Mitchel O R** (1987). Fruit location in a partially occluded image. *Transactions of the ASAE*, **30**(3), 591-596
- Woebbecke D M; Meyer G E; Mortensen D A; Von Bargaen K** (1995). Shape features for identifying young weeds using image analysis. *Transactions of the ASAE*, **38**(1), 271-281

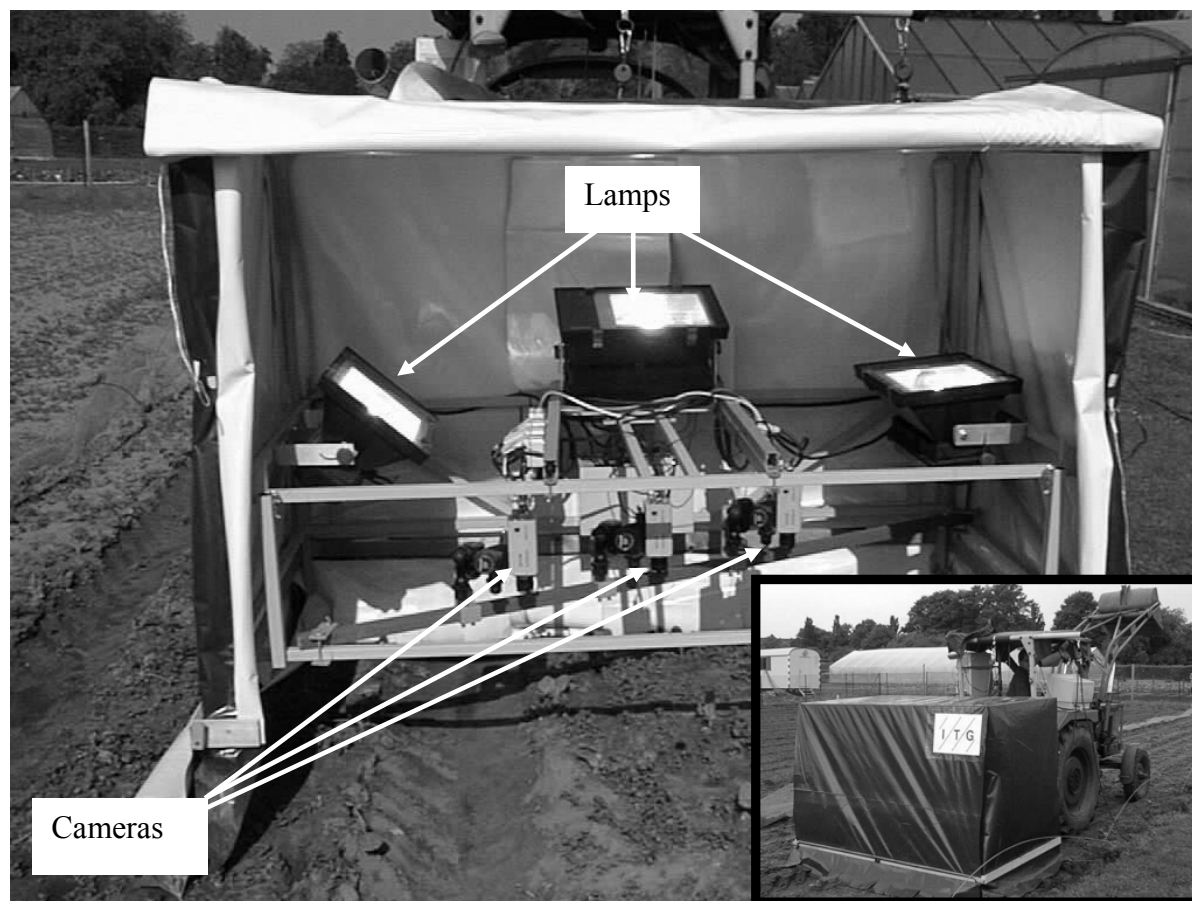


Fig. 1. The open field image processing system

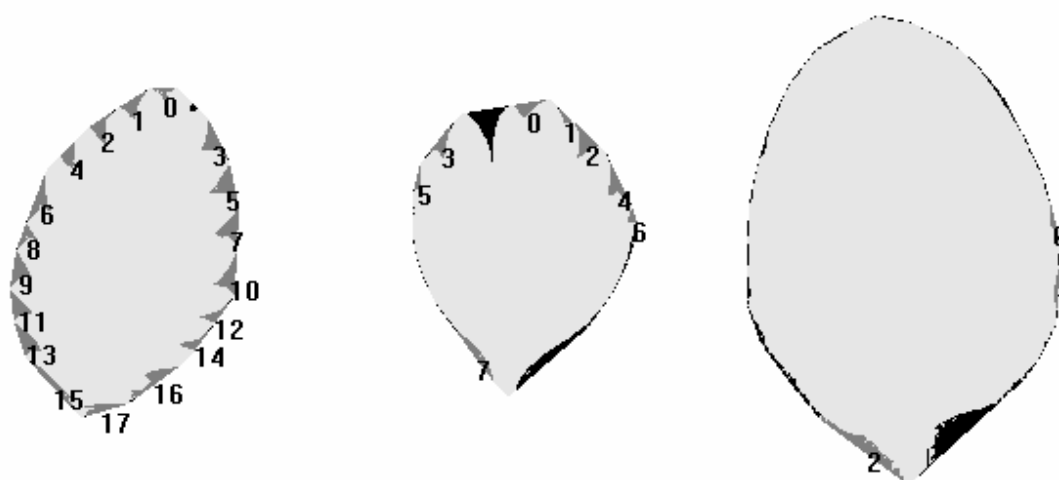


Fig. 3. Examples of procedure 'spike' for different leaf margins; input regions in light grey; difference of convex hull and input region in dark grey and black; regions classified as spike in dark grey (with numbers); regions not classified as spike (too small or too big) in black

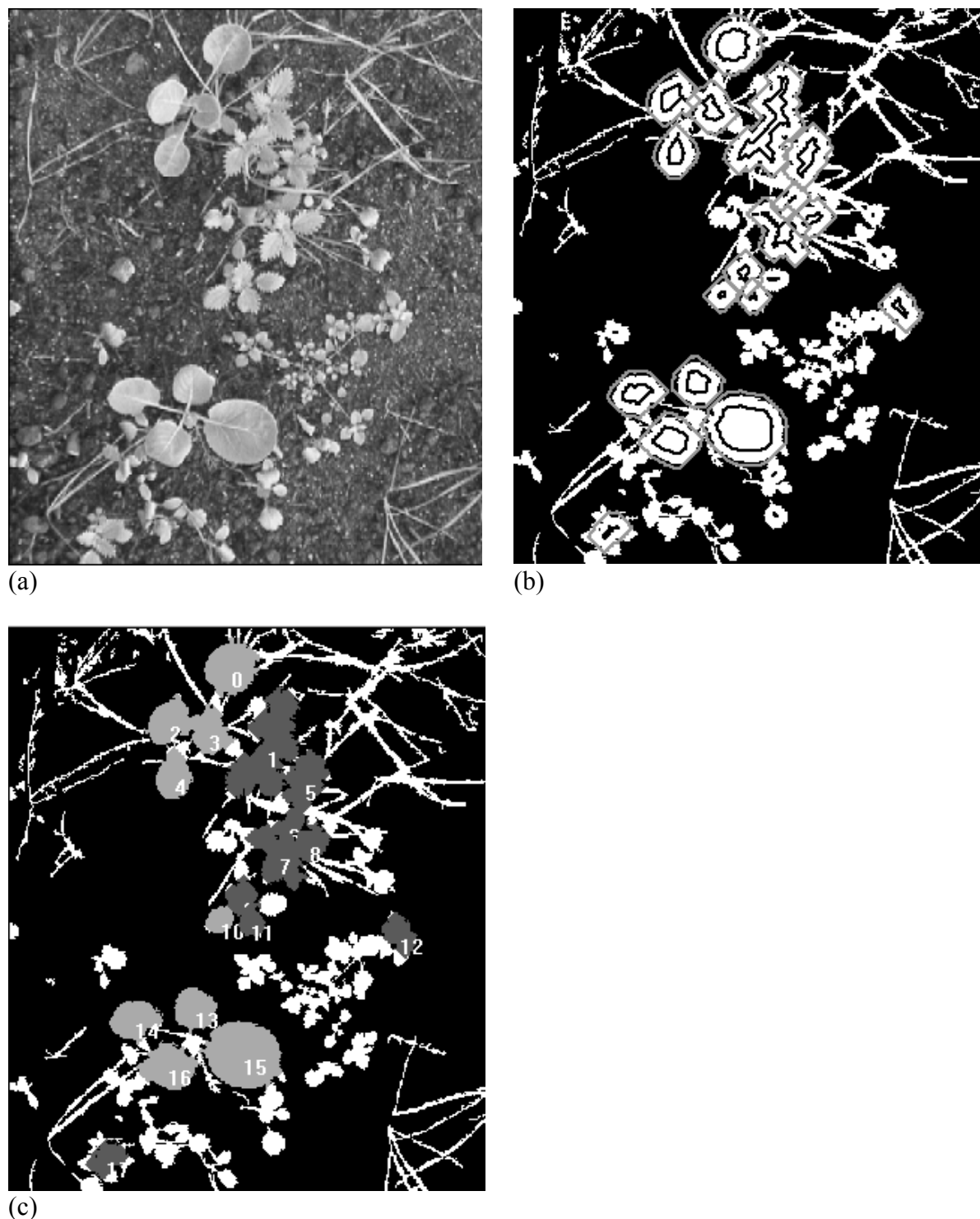


Fig. 2. Example of singulating the leaves with morphological operations: (a) detail of original colour image (cabbage and different weeds, 36 days after sowing); (b) image after binarisation and preprocessing showing original regions in white, regions after applying erosion 8 times with black borders, and eroded regions after applying dilation 12 times with grey borders; and (c) result after segmentation and classification procedure showing objects classified as crop in light grey, and objects classified as weed in dark grey

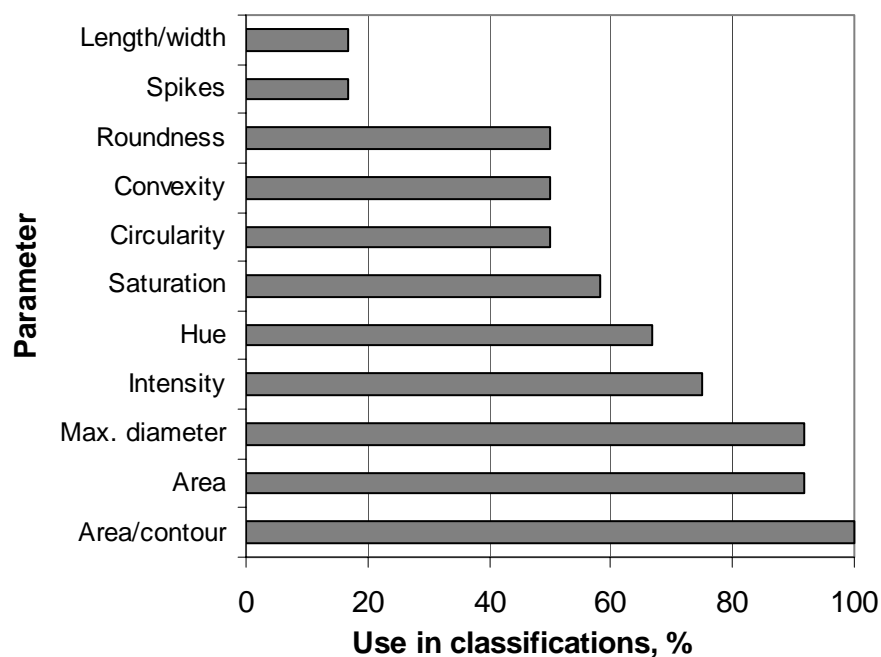


Fig. 4. Use of classification parameters for the identification of cabbage (result of 12 different classifications, greenhouse and open field experiments)

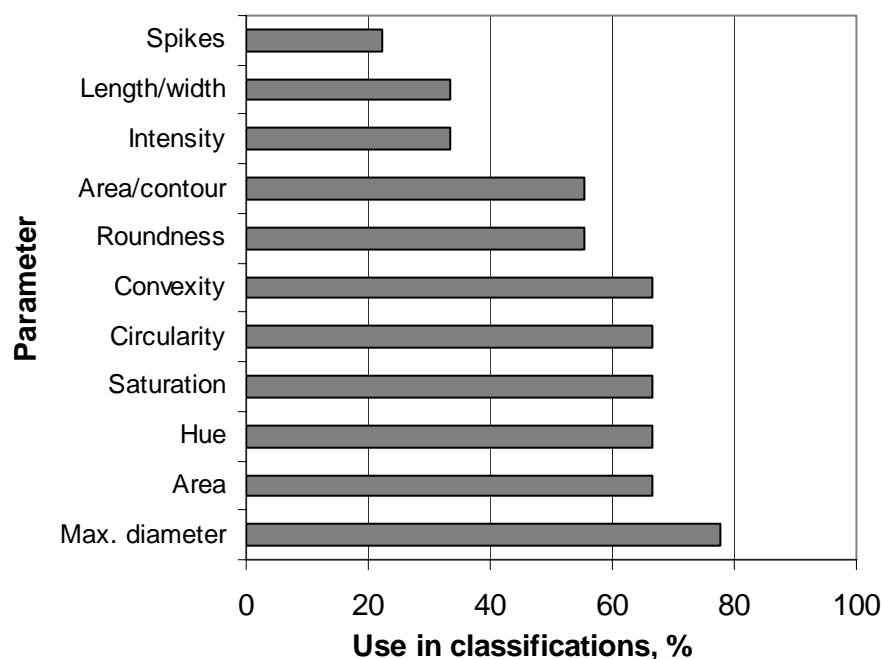


Fig. 5. Use of classification parameters for the identification of carrots (result of 9 different classifications, greenhouse and open field experiments)

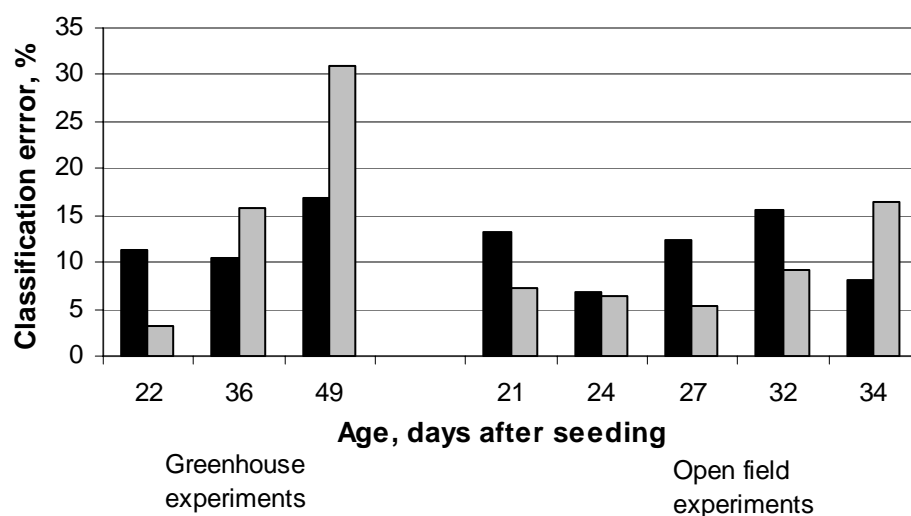


Fig. 6. Summary of classification results for cabbage; ■, type 1 error (not identified crop); ■, type 2 error (weed classified as crop)

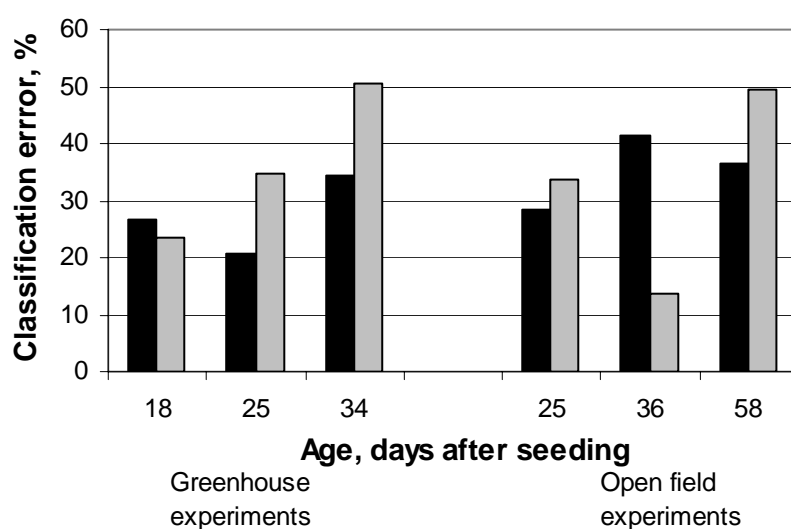


Fig. 7. Summary of classification results for carrots; ■, type 1 error (not identified crop); ■, type 2 error (weed classified as crop)

Table 1
Classification parameters

7. Parameter	Formula
Area	$\Omega_0 = \sum_r f(\mathbf{p}_r)$
Area/contour length	$\Omega_1 = \frac{\Omega_0}{C}$
Length/width	$\Omega_2 = \frac{L}{W}$
Circularity	$\Omega_3 = \frac{\Omega_0}{\max(\ \mathbf{p}_c - \mathbf{p}_i\ ^2) \pi}$
Convexity	$\Omega_4 = \frac{\Omega_o}{A_h}$
Maximum diameter	$\Omega_5 = \max(\ \mathbf{p}_i - \mathbf{p}_j\)$
Roundness	$\Omega_6 = 1 - \frac{\sigma}{\bar{D}}$ $\bar{D} = \frac{\sum_i \ \mathbf{p}_c - \mathbf{p}_i\ }{A_c}$ $\sigma = \sqrt{\frac{\sum_i (\ \mathbf{p}_c - \mathbf{p}_i\ - \bar{D})^2}{A_c}}$
Spikes	$\Omega_7 = 100 \frac{n}{C_h}$ (see also 3.4)
Colour-based parameters 'hue', 'saturation' and 'intensity'	$\Omega_{8/9/10} = \frac{\sum_r g(\mathbf{p}_r)}{\Omega_0}$

where: σ is the deviation from the mean distance; A_c is the area of the contour; A_h is the area of the convex hull of the region; C is the contour length of the region; C_h is the contour length of convex hull of input region; \bar{D} is the mean distance; $f(\mathbf{p}_r)$ is the binary value of pixel \mathbf{p}_r of the region; $g(\mathbf{p}_r)$ is the grey value of the pixel \mathbf{p}_r of the region (for the corresponding colour channel); L is the greater diameter of the smallest surrounding rectangle of the region (length); n is the number of components selected as spikes in the procedure 'spike' (see 3.4); \mathbf{p}_c is the center of the region (calculated as the mean of the line and column coordinates of all pixels of the region); $\mathbf{p}_i, \mathbf{p}_j$ are pixels of the contour of the region; $\|\mathbf{p}_i - \mathbf{p}_j\|$ is the distance between vector \mathbf{p}_i and \mathbf{p}_j ; \mathbf{p}_r is the pixel of the region; W is the smaller diameter of the smallest surrounding rectangle of the region (width)

Table 2
Effect of parameter weighting on classification errors (mean of all analysis days)

<i>Parameter</i> weighting	Crop	<i>Change in classification error, %</i>	
		Type 1 error	Type 2 error
Weak	Cabbage	-0.15	0.35
	Carrots	-3.98	4.08
Strong	Cabbage	-0.11	-0.08
	Carrots	-1.41	-1.61