The potential of automatic methods of classification to identify leaf diseases from multispectral images

Sabine D. Bauer · Filip Korč · Wolfgang Förstner

Abstract   Three methods of automatic classification of leaf diseases are described based on high-resolution multispectral stereo images. Leaf diseases are economically important as they can cause a loss of yield. Early and reliable detection of leaf diseases has important practical relevance, especially in the context of precision agriculture for localized treatment with fungicides. We took stereo images of single sugar beet leaves with two cameras (RGB and multispectral) in a laboratory under well controlled illumination conditions. The leaves were either healthy or infected with the leaf spot pathogen *Cercospora beticola* or the rust fungus *Uromyces betae*. To fuse information from the two sensors, we generated 3-D models of the leaves. We discuss the potential of two pixelwise methods of classification: *k*-nearest neighbour and an adaptive Bayes classification with minimum risk assuming a Gaussian mixture model. The medians of pixelwise classification rates achieved in our experiments are 91% for *Cercospora beticola* and 86% for *Uromyces betae*. In addition, we investigated the potential of contextual classification with the so called conditional random field method, which seemed to eliminate the typical errors of pixelwise classification.

Keywords   Pattern recognition · Gaussian mixture model (GMM) · Conditional random field (CRF) · *k*-nearest neighbour · Sugar beet · Sensor fusion

Introduction

This paper discusses the potential of three automatic methods of classification to detect leaf diseases of sugar beet plants. Information about the spatial distribution of infected areas in a field and possibly the distribution of the disease within a plant is a prerequisite for precision agriculture. The spatial distribution can be obtained through costly inspection by human experts that is usually performed on a small sample set. On the contrary, detection

S. D. Bauer (✉) · F. Korč · W. Förstner
Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn,
Nussallee 15, 53115 Bonn, Germany
e-mail: sabine.bauer@uni-bonn.de

by automatic methods of classification enables larger areas to be inspected at lower costs and also non-destructively.

The aim of precision agriculture is to identify and treat infected areas in one pass of the tractor over a field. To achieve this stereo cameras will be installed in front of the tractor. The images recorded must be analyzed with a suitable method to distinguish between healthy and infected leaves so that treatment can be directed to the infected plants. To realise this aim requires research in different areas of expertise. In this study we focus on methods to analyze the recorded images. In particular, we test the potential of three automatic methods of classification to detect leaf diseases correctly.

For fungicide treatment to be effective all diseased plants should be detected, but not at the expense of including too many healthy ones classified as diseased. A measure of the degree of misclassification of either healthy or diseased plants indicates the success of classification, i.e. the classification rate. This rate indicates the percentage of objects in a particular class that are detected correctly. For application in precision agriculture, the classification rate of the diseased plants (sensitivity) as well as the number of correctly classified healthy plants (specificity) must be high.

Our investigation focuses on leaf diseases of sugar beet plants because of the economic importance of this crop in Germany. However the methods can be transferred to other species. Leaves of sugar beet may be infected by several diseases, such as rust (*Uromyces betae*), powdery mildew (*Erysiphe betae*) and other leaf spot diseases (*Cercospora beticola* and *Ramularia beticola*). This investigation was restricted to single leaves, which were either healthy or infected with the leaf spot pathogen *Cercospora beticola* or the rust fungus *Uromyces betae*. We took images of leaves in a laboratory under well controlled illumination conditions that excluded reflections, shadows, occlusion and so on. Nevertheless, such problems must be solved in future research before the system can be applied in the field.

Previous publications dealing with the image-based automatic detection of leaf diseases are rare. To the best of the authors' knowledge only the work of Sanyal and Patel (2008), Huang (2007) and Pydipati et al. (2006) covers this area to date. All three studies used colour and texture features for the classification, but the methods of classification were different. Sanyal and Patel (2008) used a pattern recognition method to detect two diseases in rice plants using a multilayer perceptron. Huang (2007) applied an artificial neural network for detecting Phalaenopsis seedling diseases and Pydipati et al. (2006) used discriminant analysis to identify citrus diseases. In contrast to these studies, which were based on red, green, blue (RGB) images only, we also took the infrared channel into account. The infrared channel is of particular importance because it enables discrimination between areas with and without chlorophyll. Therefore, we photographed the leaves with two different cameras, a usual RGB and a multispectral camera. For the classification, we merged the images from these two cameras using 3-D models of the leaves. In addition, the image acquisition method of Pydipati et al. (2006) is destructive; they cut the leaves off the plants and photographed them in a laboratory. Our approach is non-destructive and in principle is applicable in the field. This means, for example, that we could investigate the evolution of a disease over time on the same leaf.

The analysis examines the potential of two independent pixelwise classification methods and one global pixel labelling approach:

- pixelwise *k*-nearest neighbour classification (*k*NN),
- pixelwise adaptive Bayes classification with minimum risk assuming a Gaussian mixture model (GMM),

- global pixel labelling based on conditional random fields (CRF).

We did not apply a region-based classifier on a gradient image because of the small size of the rust spots and lack of contrast between the healthy leaf area and *Cercospora* leaf spots (Bauer et al. 2009). Classification rates for three classes, 'healthy leaf area', '*Cercospora beticola*' and '*Uromyces betae*', were assessed. We hypothesised that classification rates, r, of above 90% can be attained for all three classes. In addition, we investigated how an enhancement of the feature vector with neighbourhood information and specific treatment of the local neighbourhood in a global CRF model can improve the classification result.

Materials and methods

Data

Plants and diseases

For our experiments we chose 30 sugar beet plants with four fully developed leaves. We inoculated 15 plants with the leaf spot pathogen *Cercospora beticola* and the other 15 plants with the rust fungus *Uromyces betae*. Both diseases are fungal infections, which are limited to the leaves. From each plant we selected two leaves and marked them to observe the development of the diseases on these leaves. The plant cultivation and inoculation strategy is described in detail in Mahlein et al. (2010).

Image acquisition

The images were taken in a laboratory under well controlled illumination conditions. We restricted ourselves in this investigation to single leaves. For each leaf we took four RGB images (FujiFilm FinePix S5600, 2592 × 1944) and one multispectral image (Tetracam ADC, 1280 × 1024) from a height of 30 cm and at different positions separated by about 10 cm. The MS camera has red, green and NIR channels (700–950 nm). On the RGB images one pixel refers to a resolution of ∼0.0967 mm squared at the leaf, and the resolution of the MS image was ∼0.2123 mm squared. After the inoculation, we took photographs of the sugar beet leaves every other day over a period of 3 weeks.

Fusion of the images from the RGB and the multispectral camera

We fused the images from the two cameras using a 3D-model of each single leaf. First, we calibrated the cameras using a 3-D test field to obtain the interior orientations, in particular to eliminate lens distortion, and adapted the resolution of the MS images to the higher resolution of the RGB images through bilinear interpolation. Next, we automatically determined the six parameters of the exterior orientation, i.e. the pose, namely position and rotation, of the cameras using the program AURELO (Läbe and Förstner 2006). Based on the interior and exterior orientation of the cameras we computed the surface model, i.e. the 3-D structure of the leaves, with INPHO-Software MATCH-T (Lemaire 2008). This fusion of the five images per leaf allows rectification of the leaf with respect to a reference plane as shown in Fig. 1 for three images. Each 3-D point, P_{xyz}, of the surface-model is mapped
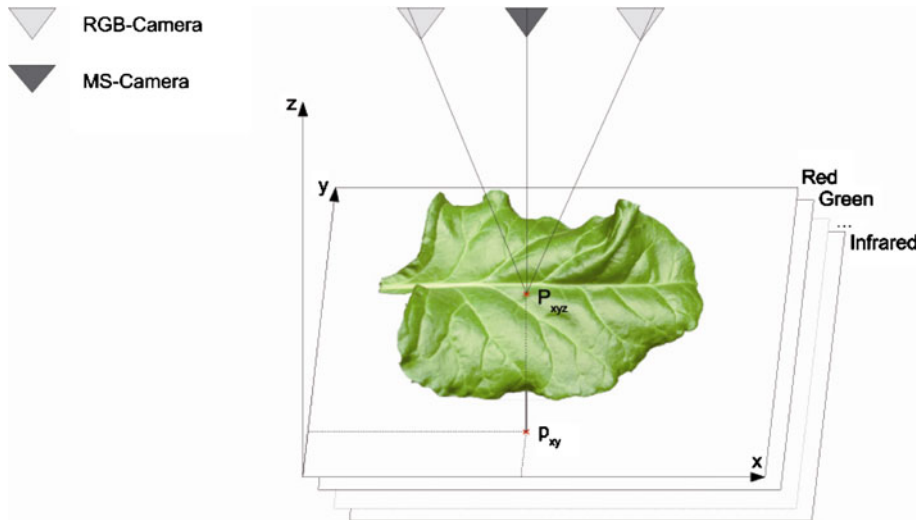
Fig. 1 Fusion of the images from the RGB and MS cameras to give a 3-D model of a leaf with points P_{xyz}. Colour information is known for all points P_{xyz} from each image. The 3-D point P_{xyz} is mapped to a 2-D point p_{xy} on the reference plane xy. This leads to a rectified 15-channel 2D image. For clarity only two RGB images are shown in the drawing, in reality there were four

to one point, p_{xy}, on the reference plane. For p_{xy} of this rectified image we obtained fifteen values:

- $4 \times$ blue from the four RGB images,
- $5 \times$ green from the four RGB images and the one multispectral image,
- $5 \times$ red same as with green,
- $1 \times$ infrared from the multispectral image.

From these 15 values we used only one blue, one green and one red value from one of the RGB images as well as the infrared channel from the multispectral image as classification features. We did not take into account duplicated information from the RGB camera because the acquisition conditions were standardized. We also omitted the red and green information from the MS camera because it showed no improvement in the results of classification.

Test-dataset

The test data were from the images of days 10–19 after inoculation. For this period we had 55 rectified images from the *Cercospora* dataset and 73 from the *Uromyces* dataset. We divided these images into patches of 64×64 pixels that showed only one of four different scenarios

- background, which is white,
- background with leaf area,
- healthy leaf area without any background and
- healthy as well as infected leaf areas.

Discrimination between the green leaf and white background was easy (Bauer et al. 2009). For training and testing we selected image patches that showed only healthy as well as infected leaf areas with no background. From the resulting image patches, we chose 350 randomly for both diseases so that our test-dataset finally contained 700 image patches.

In addition to the test-dataset there were ground truth or reference data that were generated by classifying the pixels in the images manually as healthy or infected by either *Cercospora* or *Uromyces*. We assumed that the ground truth data were correct.

Methods of classification

We applied three methods of classification; two pixel-based methods where pixel classes are determined pixel by pixel and assigned independently of each other, and a global approach, where pixel classes are determined and assigned simultaneously. For pixelwise classification we used the k-nearest neighbour (kNN) method because of its simplicity and independence of the data distribution, and an adaptive Bayes classification using Gaussian mixture models (GMM). The latter is a sophisticated version of the maximum posterior (MAP) classification that allows individual weighting of the different classes. In the global approach, we used a conditional random field (CRF) model.

The global approach, where pixel classes are assigned at all pixel locations jointly, is motivated by the following. Let us consider an example of ground truth pixel class labelling as in (Fig. 3), and in this example let us consider neighbouring pixel locations. First, most of the neighbouring pixel locations are assigned to the same class. Second, for pairs of neighbouring pixels belonging to two different ground truth classes some class combinations occur more often than others. Specifically, neighbouring pixels belonging to two different disease classes are far less common than the other class combinations. These two observations indicate that any two distinct pixel classes are statistically dependent events. Therefore, if a pixel belongs to a particular class, it is likely that a pixel at a neighboring location will belong to the same class. If a pixel at a neighboring location does belong to a different class, then some of these assignments are again more likely than others. Thus pixel class assignments at all locations depend on each other and this should be modelled accordingly in a global model. Our second observation further indicates that such a global model should allow different combinations of neighbouring pixel classes to be treated differently and should not simply impose smooth solutions.

We now consider the above concepts more specifically. Let a set $y = \{y_s\}_{s \in \mathcal{S}}$ of elements y_s denote data from an observed image. An element y_s can be interpreted as a grey value or a colour vector. The index set $\mathcal{S} = \{0, \ldots, s, \ldots, S-1\}$ is a finite set of S pixel locations. For each scalar index, s, denoting a pixel location, we extracted a feature vector $h_s(y)$. Let a set $x = \{x_s\}_{s \in \mathcal{S}}$ denote the set of classes assigned to all pixel locations, $\mathcal{S}$, where the scalar $x_s \in \mathcal{L} = \{0, \ldots, l, \ldots, L-1\}$ denotes a single pixel class and where L is the number of classes in the set $\mathcal{L}$. We refer to the set x as the configuration of classes or, equivalently, the labelling. The configuration of classes x, as opposed to image data y, is unknown. Our task is to infer the unknown class configuration or, equivalently, to estimate the unknown configuration of classes from observed image data.

In the case of a pixelwise classifier, the inferred overall configuration of classes, x^*,

$$x^* = \left\{x_s^*\right\}_{s \in \mathcal{S}}$$

is the result of decisions x_s^* made *independently* at each pixel location s. Independence here is an implicit assumption. In the global approach, the overall configuration of classes, x^*, is a single global decision, x^*, made jointly for all pixel classes, $\overset{*}{x_s}$.

k-nearest neighbour classification

The k-nearest neighbour (kNN) classifier allocates a new sample $h_s(y)$, which denotes the feature vector on the pixel location s, to the class to which the majority of the k nearest neighbours belong. The proximity is measured by any convenient metric, such as the Euclidian (Fukunaga 1972). 'Learning' by the classifier involves keeping observations and ground truth classes in memory. The kNN classifier can estimate the bounds of the Bayes error, which is a general measure for the separation between classes. However, the number of training samples must be large enough so that a particular sample, $h_s(y)$, and its nearest neighbour are closely located in the feature space (Fukunaga 1972).

Adaptive Bayes classification using Gaussian mixture models

In contrast to the kNN classifier, the Gaussian mixture model (GMM) assumes that the data can be described by multiple Gaussian distributions. Our adaptive Bayes classifier is based on the Bayes classification with minimum risk from Fukunaga (1972) and allocates a new sample, $h_s(y)$, to the class x_s^*. This classifier minimizes the sum of the product from the weighting function, w, the a priori probability, $P(l)$, and the likelihood function, $L(h_s(y)|l)$, all of which must be determined in the training phase based on the training samples:

$$x_s^* = \arg \min_{x_s} \sum_{l=0}^{L-1} (w(l, x_s) - w(l, l))P(l)L(h_s(y)|l). \tag{1}$$

The main difference between our adaptive Bayes classifier and the Bayes classifier of Fukunaga (1972) is computation of the weighting function, w, for which we first created a basis weighting function w_{basis}. The basis weighting function enabled rare classes to be given more weight because previous investigations had shown that this increased the detection rate of the rare classes. We generated the basis weighting function based on the a priori probabilities of the different classes that were identified in the training process by

$$w_{\text{basis}}(l, x_s) = \begin{cases} \frac{1}{P(l)} & \text{if } l \neq x_s \\ 0 & \text{if } l = x_s \end{cases}. \tag{2}$$

To compute the actual weighting function, w, we initialized w as a zero matrix and classified the training data iteratively with our adaptive Bayes classifier. We saved the results in the confusion matrix o. In this confusion matrix the classification rates, r, of the different classes are on the diagonal from top left to bottom right where the estimated classes are denoted in relation to the ground truth classes. The misclassification rates, e, are on the secondary diagonal, i.e. on all other positions in the confusion matrix (see for example Table 1). As the diagonal in the basis weighting function, w_{basis}, is zero, we increase the weights or costs of misclassification with Eq. 3 depending on the rates of misclassification in the confusion matrix, o, and the weights, w_{basis}. Consequently the classification rates, r, of all classes converged towards 100%.

Table 1 Confusion matrix for the two *k*NN classifications with $k = 5$

%	Percentiles	$\hat{h}$	$\hat{i}_{cercospora}$	$\hat{i}_{uromyces}$
Results of the k-nearest neighbour classifier				
	Minimum	47.42	0.24	2.01
		68.45	0.15	0.52
	Lower quartile	76.84	2.07	8.68
		87.49	2.42	3.10
$\tilde{h}$	Median	83.65	4.09	12.54
		91.26	4.59	4.29
	Upper quartile	88.54	6.29	18.37
		93.83	7.24	6.22
6 786 902 px	Maximum	97.60	12.74	44.71
		98.86	19.07	13.63
	Minimum	0.41	43.50	1.98
		0.92	56.69	1.36
	Lower quartile	2.98	68.07	14.06
		4.52	73.79	7.64
$\tilde{i}_{cercospora}$	Median	6.39	74.05	18.23
		8.47	80.95	9.79
	Upper quartile	11.08	81.70	21.56
		14.35	86.82	12.71
361 955 px	Maximum	40.96	96.99	34.87
		31.54	96.13	22.80
	Minimum	3.41	1.64	30.53
		1.29	0.92	36.96
	Lower quartile	11.40	12.27	53.67
		5.90	5.83	60.67
$\tilde{i}_{uromyces}$	Median	16.31	20.47	59.25
		9.21	14.01	75.85
	Upper quartile	24.53	27.52	70.25
		18.53	23.11	84.55
19 143 px	Maximum	50.00	52.84	81.72
		46.65	43.65	94.51

The class 'healthy leaf area' is *h* and the diseased classes are *i*. The particular disease can be identified from the subscript. An estimated value is represented as $\hat{h}$ and a ground truth value as $\tilde{h}$. The number of pixels per class is px. The first row of the classification result for each parameter is without neighbourhood information and the second row gives the result with neighbourhood information

$$w(l, x_s) = w(l, x_s) + w_{basis}(l, x_s) \, o(l, x_s). \tag{3}$$

The iteration was stopped if the change factor, *a*, was smaller than 0.01, where

$$a = \sum_{lx_s} |o(l, x_s) - o_{old}(l, x_s)|. \tag{4}$$

To compute the likelihood function $L(h_s(y)|l)$ we used the expectation–maximization (EM) algorithm (Bilmes 1998) to obtain the parameters of the GMM. For each class, *l*, it is

possible to set the number of Gaussian distributions individually. In preliminary experiments, the best classification results were achieved with the following number of Gaussian distributions per class:

- healthy leaf area: 2 distributions,
- *Cercospora beticola*: 3 distributions and
- *Uromyces betae*: 1 distribution.

The a priori probability, $P(l)$, is usually determined in the training phase of the classifier, but the probability of occurrence of a class varies considerably for different leaves. Those determined in the training phase do not generally provide a good approximation for an individual test-image. Therefore, we estimated the a priori probability, $P(l)$, individually for each image in the classification process by preclassifying the image with the likelihood function, $L(h_s(y)|l)$, given above.

Conditional random field

We used the conditional random field (CRF) that enables local classifiers that use arbitrary overlapping features to be combined with local data-dependent class interaction models. The CRFs were proposed by Lafferty et al. (2001) in the context of segmentation and labelling of 1-D text sequences. This concept has been extended and made applicable to images by Kumar and Hebert (2006). The CRF models the posterior probability of a class configuration, x, of all pixel classes given the observed data, y, as

$$P(x|y, \theta) = \frac{1}{Z(\theta)} \exp(-E(x|y, \theta)), \tag{5}$$

where we denote explicitly the dependence on unknown model parameters θ. The function $E(x|y, \theta)$, Eq. 7, is the so called energy function. The parameter dependent term $Z(\theta)$ is given by

$$Z(\theta) = \sum_{x \in \mathcal{L}^S} \exp(-E(x|y, \theta)), \tag{6}$$

where S is the number of pixels, the so called normalization constant (also known as the partition function). The energy function, $E(x|y, \theta)$, is expressed as

$$E(x|y, \theta) = \sum_{s \in \mathcal{S}} E_1(x_s|y, \theta_1) + \sum_{\{s,s'\} \in \mathcal{S}_2} E_2(x_s, x_{s'}|y, \theta_2), \tag{7}$$

where $\theta = (\theta_1, \theta_2)$. In Eq. 7 we extend the two-dimensional rectangular grid of pixel locations to the 4-neighbourhood system. The set $\mathcal{S}_2$ then denotes the set of *unordered* pairs of neighbouring pixel locations in the latter. The term $E_1(x_s|y, \theta_1)$ can be thought of as a cost of assigning a single pixel location s to a particular class, x_s. Because its value depends on a single pixel class, it is called an unary term. Specifically, this term can be a local classifier using arbitrary overlapping features. The term $E_2(x_s, x_{s'}|y, \theta_2)$ can be thought of as a cost of assigning two pixel locations s, s' to respective classes x_s and $x_{s'}$. Because its value depends on a pair of classes, it is called a pairwise term. This term enabled us to model specific local class interaction.

There are two major challenges when applying CRF models. The first is to estimate the unknown model parameters, θ, and the second is compute a class configuration, x^*, that maximizes the posterior probability. Both tasks are difficult to solve exactly, but good

approximate solutions can be found efficiently (see the Appendix for more information on this).

Experiments

The kNN experiments

To evaluate the *k*NN method we did two experiments with different feature vectors. First, we used only the colour information red, green, blue and NIR from a particular pixel y_s (Fig. 2a). Second, we used additional information from the left, right, upper und bottom neighbouring pixels so that the vector $h_s(y)$ contained 20 features in total, i.e. red, green, blue and NIR for each pixel (Fig. 2b). This enabled us to investigate how an enhancement of the feature vector with the 4-neighbourhood information affects the classification result.

To train the *k*NN classifier we had to define the number of training samples per class, i.e. in our case the number of training pixels. For a good estimate of the Bayes error this number should approach infinity, but the more training samples the longer is the time taken for computation because of the larger search space. Previous *k*NN results showed that the best results were obtained if each class had the same number of training samples. Therefore, we chose 2000 pixels randomly for each class. We selected the pixels from half of the 700 image patches in our test-dataset as each patch had an average of only 12 *Uromyces* pixels. For the test we classified the other half of the image patches. We repeated each *k*NN experiment 5 times only because of the long computation time.
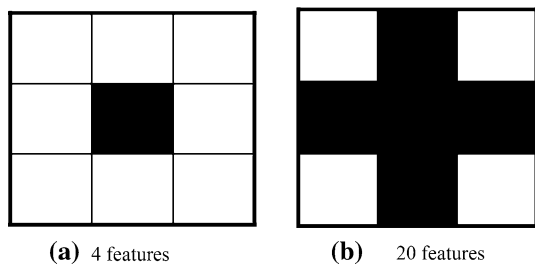
The GMM experiment

The procedure for the GMM analyses was almost the same as for the *k*NN method. Each GMM analysis was done first on the same image patches as for *k*NN, and then on a smaller training set with a larger testing set because this model requires fewer training image patches than the *k*NN classifier. Thus, we could test more image patches which increases the statistical accuracy of the result. In this case we randomly selected 1 in 10 of the 700 image patches and used their complete information for training. We tested the model on the residual 630 image patches and repeated this GMM analysis 10 times.

The CRF experiment

One of our central aims was to investigate whether results of pixelwise classification can be improved with a global model. For this purpose we performed three experiments. In the first we integrated the GMM into a global CRF model that prefers class configurations, where neighbouring pixel locations are assigned to the same class. We refer to such class

Fig. 2 **a** The central pixel (*black*) represents four colour values (red, green, blue and NIR) and **b** there are four additional neighbouring pixels with four colour values each to give a total of 20 features, i.e. the enhanced feature vector with 4-neighbourhood information

(a) 4 features   **(b)** 20 features

configuration as being smooth. A class configuration, where all pixels are assigned to the same class is an example of a smooth class configuration. We used the so called Potts model (CRF1) that has a long tradition in many fields that involve spatial statistical modelling (for more detail see the Appendix). We computed a class configuration using the GMM and CRF1 and compared the results qualitatively by observation.

In the second experiment we used an extension of CRF1, referred to as CRF2, in which the pairwise term now assigns each distinct class pair with a different cost. Specifically, we extended the CRF1 model by changing the pairwise term in a way that assigns pairs that we considered previously to be uncommon with twice as high a cost compared to other pairs that are more common. Details of the CRF2 model are given in the Appendix. We computed a class configuration using the GMM, CRF1 and CRF2, and compared the results qualitatively by observation.

In the third experiment we treated all class pair costs in the CRF2 as unknown model parameters. We used two training images and automatically learned all unknown model parameters from the training data. We refer to the model with automatically learned parameters as CRF3. We computed a class configuration using the GMM and CRF3 method and compared the results qualitatively by observation.

Results and discussion

k-nearest neighbour classification

The classification results for *k*NN with and without 4-neighbourhood information are given in Table 1. On the left side are the ground truth classes; h is the 'healthy leaf area' class and the diseased classes are $i_{cercospora}$ and $i_{uromyces}$. An estimated value is given as $\hat{h}$ (see top row of table) and a ground truth value as $\tilde{h}$. Our aim is to maximize the classification rates, r, from the particular classes that are denoted on the diagonal (top left to bottom right) in the confusion matrix, Table 1. The misclassification rates, e, on the secondary diagonal, i.e. on all other matrix positions, should be minimized. The classification and misclassification rates were based on the result from a single pixel, therefore, we compared the classification result of a pixel with its ground truth class.

The results in Table 1 are based on the classification rates obtained from the five experiments, but because there were few *Uromyces* pixels in each image patch we combined the results of ten image patches to compute the classification rates.

Without taking account of the 4-neighbourhood information (first row in Table 1 for each parameter), the median classification rate is 84% for the healthy leaf area, 74% for the leaf spot disease, *Cercospora beticola* and only 59% for the rust fungus, *Uromyces betae*. In the classification that took account of the 4-neighbourhood information, the classification rates of all classes are much better (second row in Table 1 for each parameter). For the healthy leaf area, the median is 91%, for *Cercospora beticola* 81% and for *Uromyces betae* 76%. The large difference between the *k*NN results with and without neighbourhood information for *Uromyces betae* of 59 to 76%, respectively, was probably caused by the small spots of this disease which result in a great difference in the neighbourhood colour information. In contrast, the neighbourhood colour information in a homogeneous region is almost the same.

Although the *k*NN classification with the neighbourhood information is promising, it did not achieve the threshold of our central hypothesis of being above 90% for all classes.

Gaussian mixture model

The results of the GMM experiment for the same image patches as the *k*NN classifier and those from the larger set are almost the same. The results from the larger set of image patches have greater statistical accuracy and we focus only on these below.

Table 2 gives the results from the GMM-based adaptive Bayes classifier on 6300 classified image patches; this value relates to the number of classified image patches and iterations. As for *k*NN, we combined the results of ten image patches to compute the classification rates. The median for the healthy leaf area is 94%, for the leaf spot disease *Cercospora beticola* it is 91% and for the rust fungus *Uromyces betae* it is 86%.

The GMM results are considerably better than expected based on those of *k*NN. It seems that the feature space for the *k*NN classifier with 2000 pixels per class and up to 20 features was too sparsely populated. Nevertheless, the results do not attain the hypothesised 90% threshold for the class *Uromyces betae*, although they do for *Cercospora beticola* and the healthy leaf area class. However, the classification result with the GMM is better than the enhanced *k*NN result in Table 1 in terms of the median values and also the lower and upper quartiles.

Examples of a suboptimal classification result are shown in Fig. 3 for both diseases. This figure shows the comparison between ground truth data and the results of GMM classification, and indicates the misclassification by the GMM. To eliminate these errors, we explored the suitability of the CRF.

Table 2 Confusion matrix for the GMM-classification with 4-neighbourhood information

%	Percentiles	$\hat{h}$	$\hat{i}_{cercospora}$	$\hat{i}_{uromyces}$
Result of the Gaussian mixture model				
	Minimum	60.96	0.00	0.03
	Lower quartile	90.44	0.27	1.41
$\tilde{h}$	Median	94.44	2.00	2.76
	Upper quartile	97.33	5.25	4.38
24 369 846 px	Maximum	99.82	32.98	21.61
	Minimum	0.00	54.22	0.00
	Lower quartile	1.03	86.76	1.76
$\tilde{i}_{cercospora}$	Median	3.68	91.26	3.75
	Upper quartile	7.66	96.47	6.52
1 370 349 px	Maximum	39.62	99.86	21.84
	Minimum	0.00	0.00	27.36
	Lower quartile	1.72	0.46	69.86
$\tilde{i}_{uromyces}$	Median	3.94	3.41	85.97
	Upper quartile	12.70	12.75	95.42
64 605 px	Maximum	72.64	66.36	100.00

The class 'healthy leaf area' is *h* and the diseased classes are *i*. The particular disease can be identified from the subscript. An estimated value is represented as $\hat{h}$ and a ground truth value as $\tilde{h}$. The number of pixels per class is px
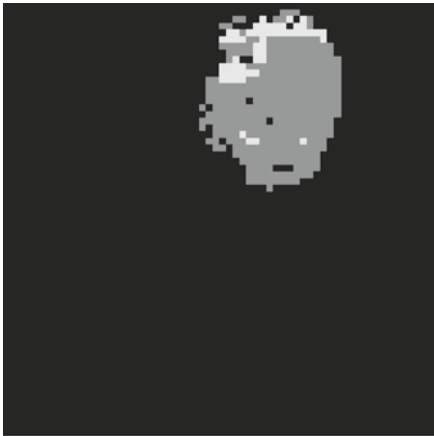
Cercospora beticola

(a) GMM Result   **(b)** Ground Truth

Uromyces betae

(c) GMM Result   **(d)** Ground Truth

Fig. 3 Results of the GMM-classification with 4-neighbourhood information for both diseases. **a**, **c** The classification result with the pixel errors and the misclassification between the two diseases, and **b**, **d** the corresponding ground truth images

Conditional random fields (CRF)

The results of classification by GMM and CRF1 and CRF2 are shown in Fig. 4 for comparison; those for CRF1 are more similar to those for GMM than CRF2. Most of the pixels in Fig. 4b have been identified correctly by the CRF classification when compared with the ground truth in Fig. 4d. We also note that compared to the result for GMM, the CRF1 classification is more smooth or more homogeneous as we would expect. Ground
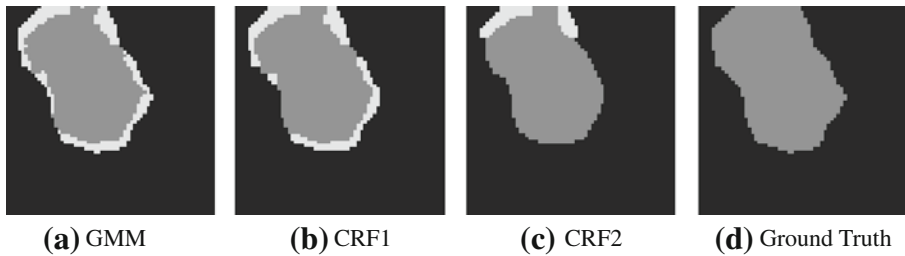
(a) GMM   **(b)** CRF1   **(c)** CRF2   **(d)** Ground Truth

Fig. 4 Classification results by: **a** Gaussian mixture model (GMM), **b** CRF Potts model (CRF1), **c** CRF interaction model (CRF2) and **d** ground truthclasses. The classes are: healthy leaf area (*black*), *Uromyces betae* (*light grey*) and *Cercospora beticola* (*dark grey*)

truth class configuration in Fig. 4d is even smoother; it contains only two smooth, homogeneous regions. Furthermore, the boundary of the disease region indicates that there is a minor improvement in terms of false positives, i.e. there are fewer light grey pixels that should not be there. This illustrates that it is possible to improve on the result of an independent pixelwise classifier with a global smoothing model. On the other hand, if the pixelwise classification provides relatively smooth results the global smoothing model will probably result in only a minor improvement. The improvement may still be worthwhile, however, as the CRF1 approach has been well studied, it requires only simple (single parameter) parameter learning, it can be done manually and inference algorithms to compute the solution are readily available.

The result of the CRF2 model (Fig. 4c) is closer to the ground truth (Fig. 4d). The pairwise term introduced in the CRF2 model improves on the result of CRF1(Fig. 4b). This approach appears promising and in principle it is feasible, however, manual estimation of an increased number of model parameters is impractical.

The results of the third experiment involving global models, CFR3, are given in the middle column of Fig. 5. The classification results with CRF3 are close to the ground truth and are an improvement on those from the GMM. Here the pairwise term of the CRF model is learned fully automatically from training data by adopting an approximate learning approach described in the Appendix. Our results indicate that automatic approximate parameter learning is feasible and that it can be done in a way that notably improves on the results of the GMM in Fig. 5a. However, this result is based on a controlled experiment of limited size and further study of this last approach is needed before an application in field conditions can be considered.

Conclusions

The results of this research show that a differentiation between healthy leaf areas and those infected with *Uromyces betae* or *Cercospora beticola* on single leaves of sugar beet plants is possible. We achieved classification rates of 86% for *Uromyces betae*, 91% for *Cercospora beticola* and 94% for the healthy leaf area. Therefore, we attained the 90% threshold for all classes except for *Uromyces betae*. Enhancement of the feature vector with neighbourhood information had a beneficial effect on these classification rates.

Our results show the typical failures of a pixelwise classifier, i.e. isolated pixels are misclassified and neighbouring pixels have been allocated to different classes, although they belong to the same class.
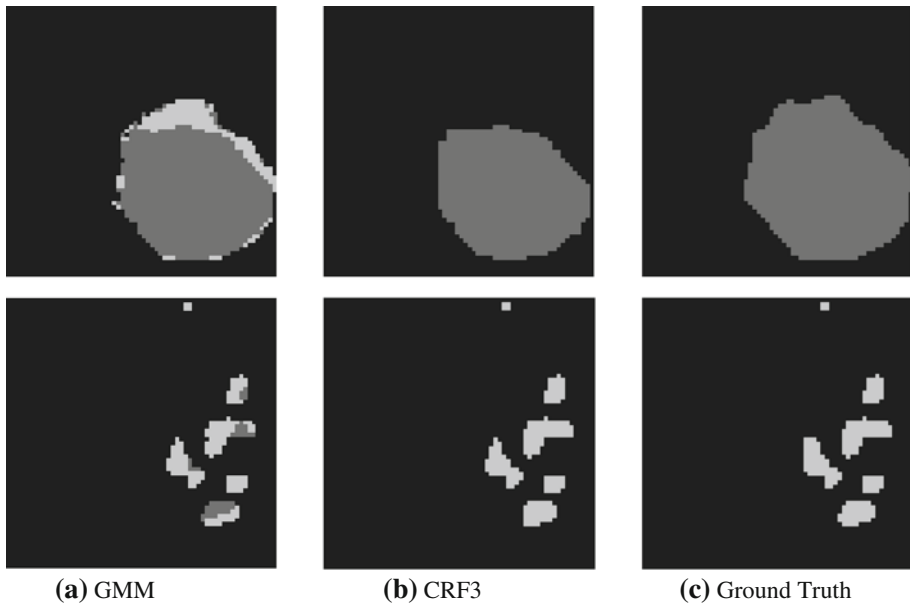
(a) GMM   **(b)** CRF3   **(c)** Ground Truth

Fig. 5 Classification results by: **a** Gaussian mixture model (GMM), **b** Conditional random field interaction model (CRF3), where all parameters were automatically learned from training data and **c** ground truth class configuration. The classes are: healthy leaf area (*black*), *Uromyces betae* (*light grey*) and *Cercospora beticola* (*dark grey*)

Our first limited experiments with CRFs suggest that modelling class neighbourhood in a global probabilistic model is a feasible approach to eliminate the artefacts of pixelwise classification. We have shown that simple CRF models can be used readily to smooth the result of independent pixelwise classification. More investigation is needed to develop readily transferable approximate learning methods that would open up ways to apply more complex, and hence even more useful, CRF models.

In the future we aim to investigate the conditions under which we can apply the CRF model to a larger dataset, which would be more realistic. There is also a need to automate the decision as to whether a whole leaf is healthy or not, and if not which pathogen has caused the infection. Such an approach could replace the human expert, which would be advantageous for site-specific management in precision agriculture. The method of classification presented could broaden the scope for disease diagnosis of a crop by non-experts in plant diseases. Furthermore it would be possible to test considerably more leaves. The eventual aim is to apply the automatic detection of leaf diseases in the field with a tractor equipped with appropriate cameras and facilities for the analysis of the photographed images. This would be of benefit to the farmer and the environment as it would reduce the amount of pesticides applied.

Appendix

Learning unknown model parameters

In principle we adopted the standard maximum likelihood approach and, using training images and the corresponding ground truth class configuration, we estimated the unknown CRF model parameters, θ, automatically from the training data.

As a result of the large size of the set $\mathcal{L}^S$, the partition function $Z(\theta)$ in Eq. 6 is in general intractable to compute and it is not feasible to compute the likelihood of the CRF model parameters. Therefore, we computed an approximation based on the work of Kumar et al. (2005) and Korč and Förstner (2008). Korč and Förstner (2008) showed that parameter learning based on a pseudolikelihood approximate scheme in combination with the maximum a posteriori inference gives results comparable to the state of the art approaches in Kumar et al. (2005).

Inference of unknown class configuration

For a given formulation of the posterior probability model, $P(x|y, \theta)$, for learned model parameters, θ, and for a given unseen image y we want to find a class configuration x that maximizes the posterior probability. This is equivalent to the problem of finding a vector with components that are elements of the label set $\{0, \ldots, L-1\}$ that minimizes the energy function in Eq. 7. The set $\mathcal{L}^S$ of possible class configurations is finite (it contains L^S points) and, in principle, the problem can be solved by simply enumerating the set and evaluating the energy function at each point. However, $\mathcal{L}^S$ grows exponentially with the number of pixel locations, S, and such an approach is prohibitive even for small problems. In general, the above problem is very difficult to solve.

Recently, several approaches have proved to be efficient in finding approximate solutions to the above problem, Szeliski et al. (2008). In our experiments we adopted an approach based on convex relaxation, in which the original problem is approximated by a convex one that can be solved efficiently. Convex relaxation has proved a powerful alternative to other existing approaches. We implemented the linear programming (LP) relaxation proposed by Schlesinger (1976) for a special case and independently by Chekuri et al. (2001), Koster et al. (1998), Wainwright et al. (2005) for the general case.

Details of the adopted model (CRF1, CRF2, CRF3)

We adopted the following form of the unary term $E_1(x_s|y, \theta_1)$,

$$E_1(x_s|y, \theta_1) = 1 - \delta\big(x_s - x_{s,CLS}(y, \theta_1)\big),$$

where the function $\delta(\cdot)$ is the Kronecker delta and the scalar $x_{s,CLS}$ is the class label from a pixelwise classifier (CLS) at location s as a function of the data y and of the pixelwise classifier parameters, θ_1. Here, $x_{s,CLS}$ is produced by either the kNN or GMM model at s. The above unary term assigns zero if $x_s = x_{s,CLS}$ and one otherwise. The overall unary energy function, i.e. the sum of all unary terms, has a unique minimum.

In the following we describe two specific forms of the pairwise energy term. The first form is a well studied model in physics and will serve mainly as a means to introduce the second model that we adopt to evaluate the potential of global models for the task under consideration.

Potts model (CRF1)

We used the Potts model to specify the pairwise terms $E_2(x_s, x_{s'}|y, \theta_2)$ in which there is no dependence on data y and the model parameters, θ_2, are reduced to a single scalar denoted by β. The Potts model form of the pairwise term is then given by

$$E_2(x_s, x_{s'}|\beta) = \beta(1 - \delta(x_s - x_{s'})). \tag{8}$$

The above term assigns zero if $x_s = x_{s'}$ and β otherwise. There are L label configurations that give the same overall pairwise energy value. For $\beta > 0$, any other configuration of labels would give an energy function value that is larger and so these L label configurations are global minimizers of such overall pairwise energy. For $\beta < 0$, the overall pairwise energy would be minimized by a chequerboard type of pattern. For instance, for $L = 2$ there would be exactly two such equivalent chequerboard solutions. For $\beta = 0$, any label configuration would minimize the overall pairwise energy.

 We now discuss the role of the parameter β in the overall energy function involving the sum of both the unary and the pairwise terms. This will provide us with an intuitive idea of how to set β when applying the model. For $\beta = 0$, this equals the overall unary energy. Otherwise, for finite nonzero β the solution is a compromise between the cases described above. The degree of compromise is controlled by the value of β.

Model with label specific interaction (CRF2, CRF3)

We now adopt the following form of the pairwise energy term

$$E_2(x_s, x_{s'}|\theta_2) = \theta_{2;x_s x_{s'}},$$

where $\theta_{2;x_s x_{s'}}$ is a smoothness factor that depends on values of pair of labels, and where $\theta_{2;x_s x_{s'}} = \theta_{2;x_{s'} x_s}$, i.e. the order of two neighbouring pixels, does not play a role. We note that by setting $\theta_{2;x_s x_{s'}} = \beta$ for $x_s \neq x_{s'}$ and $\theta_{2;x_s x_{s'}} = 0$ for $x_s = x_{s'}$ we recover the previously described Potts model. We refer to the model as the model with label pair specific interaction as for each combination of class labels $x_s, x_{s'}$ there is a specific factor $\theta_{2;x_s x_{s'}}$. Such generalization allows us to deal with some of the specific aspects of our application.

Details of the CRF experiments

In all three experiments involving global models, given image data y, we first computed the pixel class $x_{s,GMM}(y, \theta_1)$ for each pixel location s by the GMM. By setting $x_{s,GMM} = x_{s,CLS}$ we used the pixel class $x_{s,GMM}$ in the unary energy term of the CRF models described above.

 In the first experiment, we set $\beta = 1.2$ and computed the class configuration by LP relaxation.

 In the second experiment we used an extension of the model from the first experiment. The Potts pairwise energy term can be written equivalently as $E_2(x_s, x_{s'}|\beta) = \beta \cdot I_{x_k, x_{k'}}$, where $I_{x_k, x_{k'}}$ are elements of the matrix

$$\left(I_{x_k, x_{k'}}\right) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

For the experiment, we retained the value of β and then set

$$\left(I_{x_k, x_{k'}} \right) = \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

In this way we assigned the 0-1 and 1-0 class combination with a cost of 2. This is twice as large a cost as the 0-2, 2-0, 1-2 and 2-1 class combination that we assigned a cost of 1. Previously we argued that the class combination we assigned with the higher cost is very uncommon. Class combinations assigned with lower costs were the more common events.

References

Bauer, S. D., Korč, F., & Förstner, W. (2009). Investigation into the classification of diseases of sugar beet leaves using multispectral images. In E. J. van Henten, D. Goense, & C. Lokhorst (Eds.), *Precision agriculture '09* (pp. 229–238). Wageningen: Wageningen Academic Press.

Bilmes, J. A. (1998). *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and Hidden Markov models.* Technical report. Berkeley: University of California at Berkeley, International Computer Science Institute, Department of Electrical Engineering and Computer Science.

Chekuri, C., Khanna, S., Naor, J., & Zosin, L. (2001). Approximation algorithms for the metric labelling problem via a new linear programming formulation. In *Proceedings of the twelfth annual ACM-SIAM symposium on discrete algorithms* (pp. 109–118). Washington, DC: ACM/SIAM.

Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* New York: Academic Press.

Huang, K.-Y. (2007). Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture, 57,* 3–11.

Korč, F., & Förstner, W. (2008). Approximate parameter learning in conditional random fields: An empirical investigation. In G. Rigoll (Ed.), *Pattern recognition.* LNCS 5096 (pp. 11–20). Berlin: Springer.

Koster, A. M. C. A., van Hoesel, S. P. M., & Kolen, A. W. J. (1998). The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters, 23,* 89–97.

Kumar, S., August, J., & Hebert, M. (2005). Exploiting inference for approximate parameter learning in discriminative fields: An empirical study. In A. Rangarajan, B. Vemuri, & A. L. Yuille (Eds.), *Energy minimization methods in computer vision and pattern recognition.* LNCS 3757 (pp. 153–168). Berlin: Springer.

Kumar, S., & Hebert, M. (2006). Discriminative random fields. *International Journal of Computer Vision, 68,* 179–201.

Läbe, T., & Förstner, W. (2006). Automatic relative orientation of images. In L. Gründig & M. O. Altan (Eds.), *Proceedings of the 5th Turkish-German joint geodetic days,* Berlin.

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In C. E. Brodley & A. P. Danyluk (Eds.) *Proceedings of the 18th international conference on machine learning* (pp. 282–289). Massachusetts: Morgan Kaufmann.

Lemaire, C. (2008). Aspects of the DSM production with high resolution images. In J. Chen, J. Jiang, & S. Nayak (Eds.) *Proceedings of the XXIst ISPRS congress,* Beijing, China.

Mahlein, A.-K., Steiner, U., Dehne, H.-W., & Oerke, E.-C. (2010). Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture, 11,* 413–431.

Pydipati, R., Burks, T. F., & Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture, 52,* 49–59.

Sanyal, P., & Patel, S. C. (2008). Pattern recognition method to detect two diseases in rice plants. *The Imaging Science Journal, 56,* 319–325.

Schlesinger, M. I. (1976). Sintaksicheskiy analiz dvumernykh zritelnikh singnalov v usloviyakh pomekh (Syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika, 4,* 113–130.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., et al. (2008). Comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30,* 1068–1080.

Wainwright, M. J., Jaakkola, T. S., & Willsky, A. S. (2005). MAP estimation via agreement on trees: Message-passing and linear programming. *IEEE Transactions on Information Theory, 51,* 3697–3717.