Ralf Tautenhahn, Alexander Ihlow and Udo Seiffert

Adaptive feature selection for classification of microscope images

Original published in:
Fuzzy logic and applications : 6th international workshop, WILF 2005, Crema, Italy, September 15 - 17, 2005 ; revised selected papers. - Berlin [u.a.] : Springer, 2006. – S. 215-222.
Print ISBN 978-3-540-32529-1 Online ISBN 978-3-540-32530-7
(Lecture notes in computer science ; 3849 : Lecture notes in artificial intelligence)

DOI: 10.1007/11676935_26

URL: http://link.springer.com/chapter/10.1007/11676935_26

[Visited: 2014-03-20]

Adaptive Feature Selection for Classification of Microscope Images

Ralf Tautenhahn[1], Alexander Ihlow[2], and Udo Seiffert[2]

[1] Leibniz Institute of Plant Biochemistry,
Weinberg 3, D-06120 Halle (Saale), Germany
ralf.tautenhahn@ipb-halle.de
[2] Leibniz Institute of Plant Genetics and
Crop Plant Research (IPK) Gatersleben,
Corrensstr. 3, D-06466 Gatersleben, Germany
{ihlow, seiffert}@ipk-gatersleben.de
http://mue.bic-gh.de

Abstract. For high-throughput screening of genetically modified plant cells, a system for the automatic analysis of huge collections of microscope images is needed to decide whether the cells are infected with fungi or not. To study the potential of feature based classification for this application, we compare different classifiers (kNN, SVM, MLP, LVQ) combined with several feature reduction techniques (PCA, LDA, Mutual Information, Fisher Discriminant Ratio, Recursive Feature Elimination). We achieve a significantly higher classification accuracy using a reduced feature vector instead of the full length feature vector.

1 Introduction

Recent biomolecular methods produce large amounts of raw data exceeding all limitations of currently used manual or semiautomatic analysis. To study resistance mechanisms of crop plants against fungi a high-throughput screening of genetically modified cells is performed and the desired automated process should be able to analyse an immense number of microscope images without human interaction. An overview of computerized cell image analysis can be found in [1]. Automated classification of cell images – from a medical point of view – has been documented in e.g. [20,15,18] and the recognition of plankton images from an underwater video microscope system has been described in [16].

This paper focuses on a feature based classification of biological objects which have been previously segmented in high-resolution microscope images. The biological relevant object [14] to be automatically detected is a so called *haustorium* – a complex object consisting of a "waist" with "fingers" (see Fig. 1 for some typical samples). In the underlying processing pipeline, regions of interest containing relevant biological cells (more precisely, genetically transformed cells characterized by a greenish blue dye) are extracted from the acquired images [9]. Next, these individual transformed cells are checked for potential *haustoria*, using advanced image segmentation methods [11,10]. This step leads to a rather

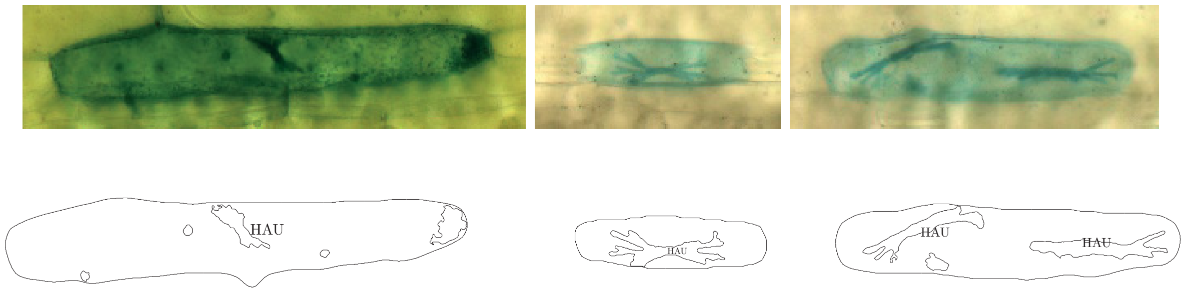

Fig. 1. Three different regions of interest containing biologically relevant cells extracted from original microscope images. The segmented objects inside those cells have to be classified into *haustoria* (marked by "HAU" in the sketches) and other objects. As can clearly be seen, the contrast may be rather poor and the objects differ very much in colour, shape, size and orientation.

large number of objects which might be either *haustoria* or similar image structures being any other objects. Because this segmentation does not provide a sufficiently correct recognition of *haustoria*, classification has to be done to distinguish between real *haustoria* and *similar* objects.

Since the objects stand out only slightly against the background, the object recognition has to be rather sophisticated. Furthermore, the objects differ in colour, shape, size, and orientation (see Fig. 1). Thus, for example, template based approaches or any solution requiring model assumptions or a-priori knowledge will not be suitable. A common and very flexible approach is to extract a number of features from labeled examples for all different object classes (here: *haustorium* or not) from the image and to perform training and classification subsequently. Since the impact of particular features often depends strongly on the subsequent classification method – a fact that is often highly underestimated, both the feature selection and the classification have to be considered together. By means of the above mentioned quite challenging real-world application of *haustoria* recognition, this paper investigates a number of common statistical and neural network based classification methods in conjunction with several common feature selection algorithms and comes up with some expected results but also some maybe unanticipated ones.

2 Feature Generation

A total number of 38 features is generated, characterizing shape as well as colour and texture. An overview of the features is given in table 1. During the segmentation procedure described in [11], a contrast enhancement is done using the morphological top-hat operations. Features can be extracted from the original or the enhanced images: the average colour values of every object were measured in RGB and HSV from both image versions and texture features were also calculated for original and enhanced image.

From the objects curvature [12] the normalized multiscale bending energy NMBE can be calculated. This measure is 1 for a circle and larger for every other,

Table 1. Overview of features from different categories which were generated for classification purposes. The texture and colour features were calculated from both the original and the morphological contrast enhanced image.

Category	Feature	Number	Comment
Simple geometric	Area	1	
	Roundness metric	1	$R = \frac{4\pi F}{U^2}$
Shape	Hu-Moments	7	[8]
	Granlund-Descriptors	7	[6]
	NMBE	1	[3]
Texture	Contrast	2	[5]
	Correlation	2	
	Energy	2	
	Homogeneity	2	
Colour	RGB	3	
	RGB (enhanced)	3	
	HSV	3	
	HSV (enhanced)	3	
Other	CSAT	1	see text
	Total	38	

more 'twisted' object, independent of its size. Before calculating the curvature, the contour is smoothed using a Gaussian function with $\sigma = 1.5$. We constructed another feature, CSAT, which is calculated from the enhanced image by counting each object's pixels with saturation value = 1.

3 Feature Selection

Dimensionality Reduction with PCA and LDA

Principal component analysis (PCA) and linear discriminant analysis (LDA) are two common techniques for feature reduction. While the PCA provides axes with maximal variance, the aim of the LDA is to find vectors which maximize the separability of predefined classes. More precisely, a vector d is obtained such that the ratio of the between-class variance to the within-class variance is maximized. This criterion C can be expressed as

$$C = \frac{d^T\, B\, d}{d^T\, W\, d} \ ,$$

with B being the between-class covariance matrix and W the within-class covariance matrix. The best discriminant vector d_1 is provided by

$$W^{-1}\, B\, d_1 = \lambda\, d_1 \ ,$$

where d_1 is the eigenvector of $W^{-1}B$ associated with the largest eigenvalue. It is well known as the *Fisher linear discriminant*. However, if K classes were defined, at most $K-1$ eigenvectors exist. To obtain an orthogonal set of more than $K-1$ vectors, a method proposed in [4] was applied.

Three different techniques for feature selection were used: the recursive feature elimination as a method of measuring the influence of features on the weight vector of the classifier, the mutual information to quantify correlation between several features and classes as well as the Fisher's discriminant ratio to rate individual features.

Recursive Feature Elimination
The RFE [7] is a different version of the *Sequential Backward Selection* [17]. It can be performed with classifiers which rely on minimizing a cost function of a weight vector $\mathbf{w}$, e.g. $\gamma(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$ for a support vector machine. The idea is to quantify the influence of the feature i by measuring the absolute value of the weight w_i. The process consists of the following steps:

– Train the classifier (optimize the weight vector $\mathbf{w}$ with respect to $\gamma(\mathbf{w})$).
– Compute the ranking criteria $c_i = (w_i)^2$ for all i.
– Remove the feature j with smallest ranking criterion c_j.

The result of this algorithm is a feature ranking, but the top ranked (most recently eliminated) features are not necessarily the ones that are individually most relevant, only their combination in terms of a feature vector allows an assessment of their relevance [7].

Mutual Information
Mutual information $MI(X, Y)$ is a measure of relative entropy between the joint probability $p(x, y)$ of two random variables X, Y and the product of their marginal probabilities $p(x)p(y)$ [2]:

$$MI(X, Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}.$$

In the context of classification the mutual information for features υ_i and classes ω_j is given as:

$$MI(\upsilon_i, \omega_j) = p(\upsilon_i, \omega_j) \log \frac{p(\upsilon_i, \omega_j)}{p(\upsilon_i)p(\omega_j)}$$

To evaluate the feature υ_i, the MI-values for all classes ω_j weighted with their priors $p(\omega_j)$ are summarized:

$$MI(\upsilon_i) = \sum_{\omega_j \in \Omega} p(\omega_j)\ MI(\upsilon_i, \omega_j)\ .$$

Fisher Discriminant Ratio
The FDR can be used to quantify the separability capabilities of individual features [17]. For the two class case, the FDR of feature υ is given as

$$FDR(\upsilon) = \frac{(\mu_{\upsilon 1} - \mu_{\upsilon 2})^2}{\sigma_{\upsilon 1}^2 + \sigma_{\upsilon 2}^2},$$

where $\mu_{\upsilon 1}$ is the mean and $\sigma_{\upsilon 1}$ the variance of class 1 and $\mu_{\upsilon 2}$ the mean and $\sigma_{\upsilon 2}$ the variance of class 2 corresponding to the feature υ.

4 Results

After applying the mentioned selection techniques, feature rankings can be calculated. The rankings reflect the diversity of the selection methods. In our experiments the RFE rates the colour features very high, whereas the Mutual Information tends to place form attributes on top of the list.

Table 2. Classification accuracies (with standard deviations) measured using the full sized feature vector. Most classifiers show similar, moderate performance. LVQ fails classification if the feature vector is used with full length.

Classifier	KNN3	kNN5	kNN7	SVM RBF2	SVM POLY3	MLP	LVQ
Classification accuracy	0.90 ± 0.05	0.90 ± 0.05	0.90 ± 0.05	0.89 ± 0.05	0.86 ± 0.06	0.88 ± 0.05	0.56 ± 0.08

Table 3. Classification accuracies achieved with reduced dimensionality. The combinations of feature reduction techniques and classifiers with best results are shown. All classifiers show an improved accuracy compared to classification using feature vectors with full length. The values of SVM-RBF2, SVM-POLY3 and LVQ are significantly increased.

Classifier	kNN3	kNN5	kNN7	SVM RBF2	SVM POLY3	MLP	LVQ
Reduction method	FDR	RFE	RFE	RFE	LDA	LDA	LDA
Dimensionality	18	22	21	19	5	36	7
Classification accuracy	0.92 ± 0.05	0.93 ± 0.04	0.93 ± 0.04	**0.95** ± 0.03	**0.92** ± 0.04	0.90 ± 0.05	**0.91** ± 0.05

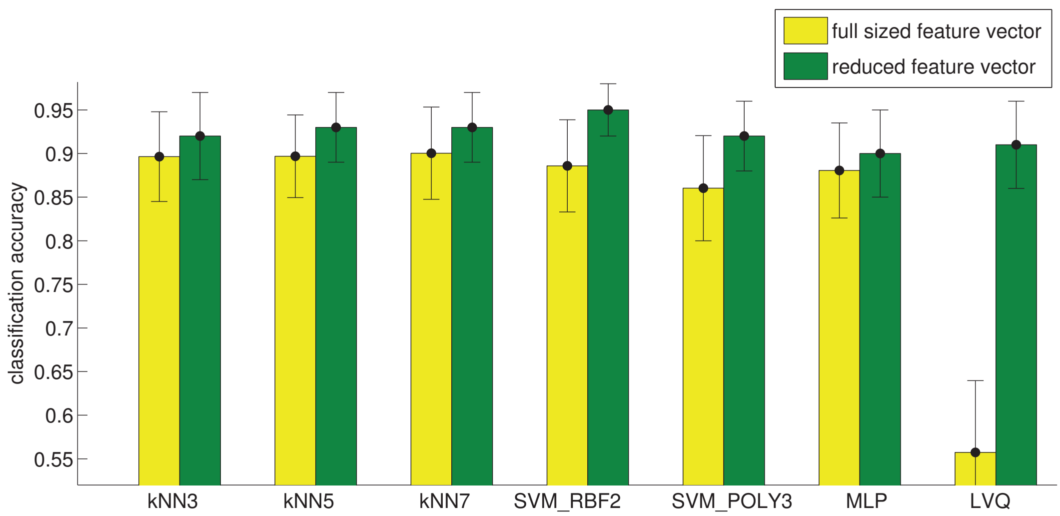

Fig. 2. Comparison of classification accuracies achieved with reduced feature vectors and feature vectors with full size. All classifiers benefit from feature reduction.

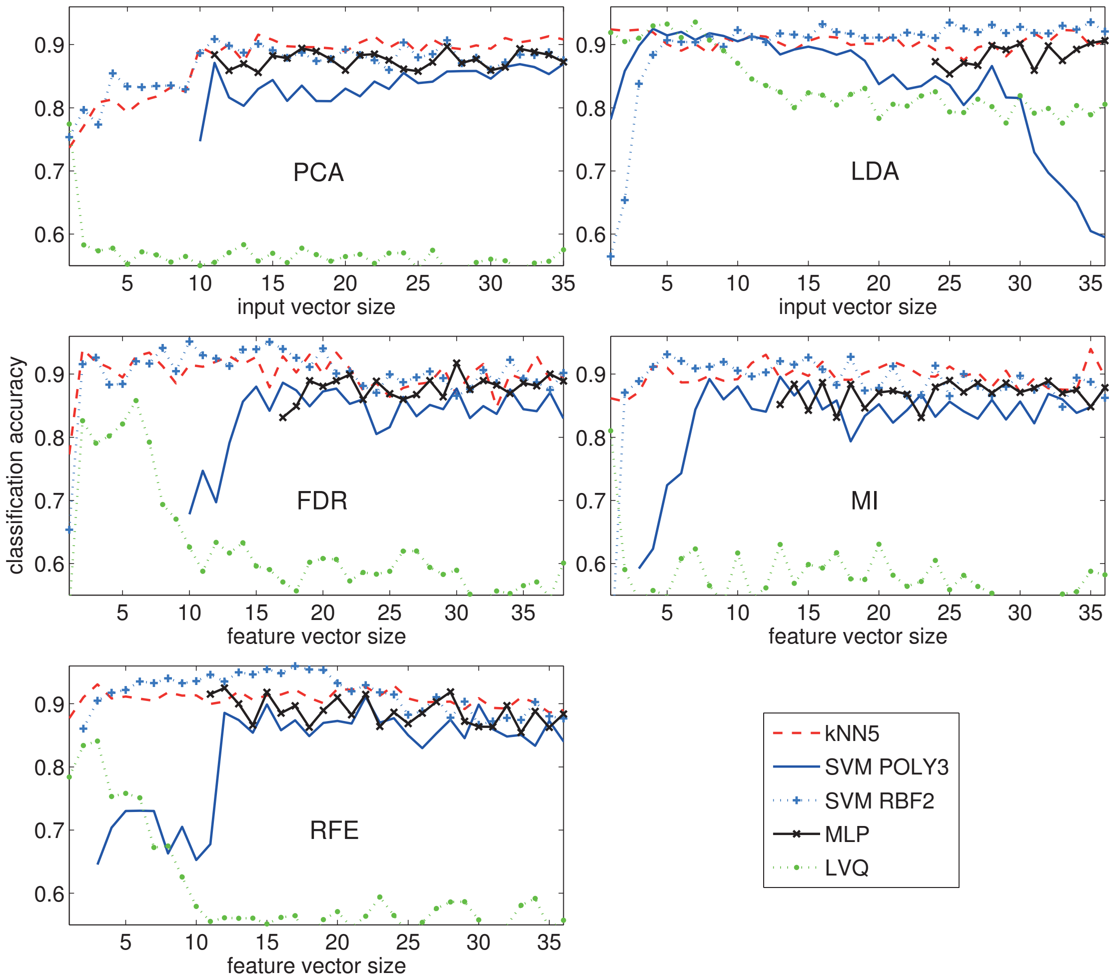

Fig. 3. Comparison of different feature reduction techniques combined with several classifiers. SVM-POLY3 and LVQ achieve their greatest values with a low dimensional feature vector, calculated with LDA, while the kNN5 classifier and the SVM-RBF reach their maxima with a medium sized feature vector, containing features obtained by recursive feature elimination. Some combinations could not be calculated due to bad convergence. The kNN3 and kNN7 classifier behave similar like kNN5 and are not drawn for clearness reasons.

To get an impression of the performance on our dataset we use different classifiers: a k-nearest-neighbor classifier ($k = \{3, 5, 7\}$), a multilayer perceptron with two hidden layers (12 neurons in the first and 3 in the second hidden layer), learning vector quantization (16 neurons in the hidden layer) and support vector machines with polynomial ($n = 3$) kernel and also with a radial base function ($\sigma = 2$) [19]. The specified parameters are the result of preselection and optimization.

Our sample set consists of 364 annotated micrographs of single plant cells. It was split into training- and test sets using 10-fold cross validation. To compare the results of several classification results on one sample set, the *corrected re-sampled t-test* [13], which takes into account the variability due to the choice of the training sets, is used.

Table 2 shows the results using the feature vector of full length (38 features). The classification accuracies of kNN, SVM-RBF2 and the MLP are similar in the range between 0.88 and 0.90. LVQ shows very poor performance.

The situation changes considerably when the feature reduction algorithms are applied. The achieved classification accuracies are shown in figure 3 as a function of the feature vector size. The classifiers respond differently to the reduction techniques: LVQ and the SVM with POLY3 kernel show great improvements with LDA-transformed input data. The accuracy of the kNN classifier and the SVM with RBF kernel can be enhanced using the RFE-selected features[1].

5 Conclusion

For the automatic classification of microscope images of plant cells we studied the influence of feature selection and -reduction techniques on several classifiers. Using reduced feature vectors the classification accuracy of learning vector quantization, a support vector machine with a radial base function and also with a polynomial kernel could be significantly improved compared to the classification accuracy achieved with a feature vector of full length. In our tests, the highest accuracy (95%) was obtained by a support vector machine with RBF-kernel in conjunction with recursive feature elimination.

Acknowledgements. We thank Patrick Schweizer and his staff for their expertise in interpretation of LM-micrographs. This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant 0312706A.

References

1. Ewert Bengtsson. Computerized cell image analysis: Past, present, and future. In *SCIA*, pages 395–407, 2003.
2. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
3. Luciano da Fontoura Costa and Roberto Marcondes Cesar Jr. *Shape Analysis and Classification*. CRC Press, 2001.
4. J. Duchene and S. Leclercq. An optimal transformation for discriminant and principal component analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 10(6):978–983, 1988.
5. Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts, 1993.
6. Gösta H. Granlund. Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers*, C-21(3):195–201, March 1972.
7. Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1-3):389–422, 2002.

[1] The RFE-feature ranking is computed with the SVM-RBF2, but also evaluated with other classifiers.

8. Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187, February 1962.
9. Alexander Ihlow and Udo Seiffert. Microscope color image segmentation for resistance analysis of barley cells against powdery mildew. In *9. Workshop "Farbbildverarbeitung"*, ZBS Zentrum für Bild- und Signalverarbeitung e.V. Ilmenau, Report Nr. 3/2003, pages 59–66, Ostfildern-Nellingen, Germany, October 2003.
10. Alexander Ihlow and Udo Seiffert. Automating microscope colour image analysis using the Expectation Maximisation algorithm. In *Pattern Recognition: 26th DAGM Symposium*, pages 536–543, Tübingen, Germany, September 2004. Springer.
11. Alexander Ihlow and Udo Seiffert. Haustoria segmentation in microscope images of barley cells. In *10. Workshop 'Farbbildverarbeitung'*, pages 119–126, Koblenz, October 2004. Der andere Verlag.
12. Farzin Mokhtarian and Alan Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(1):34–43, January 1986.
13. Claude Nadeau and Yoshua Bengio. Inference for the generalization error, 2000. Advances in Neural Information Processing Systems 12, MIT Press.
14. Patrick Schweizer, Jana Pokorny, Olaf Abderhalden, and Robert Dudler. A transient assay system for the functional assessment of defense-related genes in wheat. *Molecular Plant-Microbe Interactions*, 12(8):647–654, 1999.
15. W. Nick Street, William H. Wolberg, and Olvi L. Mangasarian. Nuclear feature extraction for breast tumor diagnosis. *Biomedical Image Processing and Biomedical Visualization*, 1905:861–870, 1993.
16. Xiaoou Tang, W. Kenneth Stewart, Luc Vincent, He Huang, Marty Marra, Scott M. Gallager, and Cabell S. Davis. Automatic plankton image recognition. *Artificial Intelligence Review*, 12(1-3):177–199, February 1998.
17. Sergios Theodordis and Konstantinos Koutroumbas. *Pattern Recognition*. Elsevier Academic Press, San Diego, 2003.
18. Jean-Philippe Thiran and Benoît M. Macq. Morphological feature extraction for the classification of digital images of cancerous tissues. *IEEE Transactions on Biomedical Engineering*, 43(10):1011–1020, October 1996.
19. Jason Weston, Andre Elisseeff, Gökhan BakIr, and Fabian Sinz. The SPIDER: object-orientated machine learning library. http://www.kyb.tuebingen.mpg.de/bs/people/spider/index.html.
20. William H. Wolberg, W. Nick Street, and Olvi L. Mangasarian. Breast cytology diagnosis via digital image analysis. *Analytical and Quantitative Cytology and Histology*, 15(6):396–404, 1993.