Automated Detection of *Mycosphaerella Melonis* Infected Cucumber Fruits

Danijela Vukadinovic, PhD*. Gerrit Polder, PhD. Gert –Jan Swinkels ***.**

**Wageningen University and Research Centre , Wageningen, the Netherlands*
(e-mail: danijela.vukadinovic@wur.nl).
*** Wageningen University and Research Centre , Wageningen, the Netherlands*
(e-mail: gerrit.polder@wur.nl).
**** Wageningen University and Research Centre , Wageningen, the Netherlands*
(e-mail: gert-jan.swinkels@wur.nl)

Abstract: In this paper we present a novel method for automated detection of *Mycosphaerella melonis* infected cucumber fruits. The two-step method consists of machine learning approach using: shape based features extracted from cucumber color images and light transmission spectra based features. The automated detection rate was compared to the manual detection rate of the human workers. Our automated method reached the 95% detection accuracy, which is comparable to the manual detection accuracy of 96%.

Keywords: Fruit disease detection, Machine learning, Computer vision, Sorting machines, Automation.

1. INTRODUCTION

Mycosphaerella melonis is one of the two main pathogens that cause two most destructive diseases to cucurbits crops, namely *Fusarium* wilt and gummy stem blight (Zhang, *et al.* 2005). Because few effective, economical, and environmentally safe management options are available for these diseases, they are a yield-limiting factor in cucurbit family (cucumber, squash, melon, pumpkin, vegetable marrows, bitter melon and zucchini) plant production worldwide. A major reason for this is the inability to accurately detect the presence and identity of the two fungal pathogens, especially in plant tissues and soil. Under favourable climatic conditions, *Mycosphaerella melonis* can infect all parts, except roots, of the cucumber plant at all stages of plant development (Fig. 1).

Fig. 1. Severe cases of *Mycosphaerella melonis* infection symptoms on the cucumber leaf (left), stem (middle) and fruits (right). Image source: http://www.shouragroup.com/v_cucumber_e.htm .

Infection at fruit development often leads to internal fruit rot that may go unnoticed at harvest. This raises concern among growers because the inevitable post-harvest fruit decay at the grocer creates a poor image for the distributor (packing house) and grower. Additionally, the infected cucumbers are not saleable and could infect the healthy fruits. The disease is especially not easy to manage in the greenhouses due to the humid climate conditions (*Olsen et al.*) that are favourable for the disease development and spread and because there is a constant occurrence of fresh wounds from pruning and harvesting that serve as new infection sites. During sorting, the contaminated cucumbers are difficult to recognize, even by experienced personnel who is trained to separate these as good as possible (Fig 2). Due to manual sorting on *Mycosphaerella melonis*, the processing speed must be reduced. Therefore, automation of this process would be highly desirable**.** For these reasons, we focus on the detection of the *Mycosphaerella melonis* in the infected fruits after harvesting.

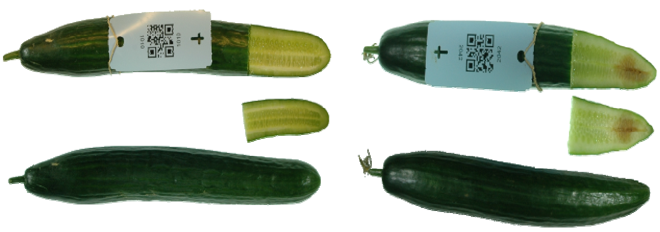

Fig. 2. Examples of the cucumber fruits used in our study: *Mycosphaerella melonis* healthy (left) and infected (right) cucumber fruits.

Various studies have used different methods for the classification of plant diseases based on spectral data. For instance, Wu et al. (2008) used PCA-based back-propagation neural network (BPNN) model and PLS wavelength-based BPNN for detection of *Botrytis cinerea*-affected eggplant leaves under laboratory conditions. Prior to the visibility of symptoms the BPNN model predicted fungal infections with a maximum of 85% classification accuracy. Mahlein et. al. (2012) applied hyperspectral imaging to symptoms caused by different sugar beet diseases induced by different fungal pathogens in the laboratory. High detection accuracy was achieved when disease symptoms were externally visible. Silva et al. (2008) applied a neural network and principal component analysis to grayscale digital images to detect black sigatoka in banana plants (caused by the fungus *Mycosphaerella fijiensis*), however, the accuracy of the classification was not very high (51.67%). Pixia and Xiangdong (2013) used image processing and recognition technologies to study the disease of downy mildew, powdery mildew and anthracnose on cucumber leaves. The detection accuracy was very high: 96% for downy mildew and 100% for powdery mildew and anthracnose. However, the disease symptoms were visible externally and were clearly able to be identified by visual inspection. *Mycosphaerella melonis* shows no externally visible symptoms in infected cucumber fruits and therefore is a problem that is more difficult to solve by using conventional image processing techniques. To the best of our knowledge, an automated method on this topic has not been reported before.

The *Mycosphaerella melonis* infection always enters the fruit through the flower and therefore is always located at the tip of the cucumber fruit. The infection is externally invisible (Fig 2). Manual workers, trained to sort out the infected fruits, rely on the combination of the fruits tip hardness differences and barely visible difference in fruits tip shape between healthy and infected fruits (the infected fruits have a pointier tip). This manual method of detecting the infected fruits is highly unreliable and slow.

In this paper we present a novel method for automated detection of *Mycosphaerella melonis* infested cucumber fruits which aims at achieving both high speed and a high accuracy. The visual differences between the healthy and infected cucumbers are analysed using cucumber shape features. Similarly, the internal tissue changes in the infected cucumbers, i.e. less water content that causes firmness and internal browning, are detected using light transmission spectra analysis.

The method consists of three steps**.** First, the colour image based extraction of cucumber shape features and a cucumber classification that utilizes these features is performed. Subsequently, cucumber classification using transmission light spectra intensities as features is done. Finally, a combination of the two classifications resulted in the final infested cucumbers detection.

The remainder of the paper is organized as follows: in the section 2 the methodology used is described, followed by the experimental set up description in section 3 that includes data collection process and datasets characteristics. Results are presented in section 4, followed by a discussion in section 5 and conclusions in section 6.

2. METHODOLOGY

The first step of the method is a sample based linear classification that utilizes shape descriptors extracted from a colour image of a cucumber. The second step is a sample based linear classification that uses transmission light spectra intensities as features. In the last step, a combination of the results of the two classifications is used as the final resulting classification result.

2.1 Shape based features

The automated shape analysis method starts with individual cucumber segmentation, based on the excessive green transformation (Fig 3). The linear combination of the red, green and blue component of a RGB image:

$$I_{EG} = 2*I_G - I_R - I_B, \tag{1}$$

Results in the excessive green image I_{EG}. Here I_G, I_R and I_B denotes the green, red and blue component of the colour image respectively. The excessive green image is thresholded using an automatic threshold using the hysteresis (Canny, 1983) in order to segment a cucumber.

Fig. 3. Separate cucumber segmentation: RGB image (left), excessive green image (middle) and the final cucumber segmentation produced by the excessive green image thresholding (right).

Once the cucumber is segmented, we analyse only the tip of the cucumber where the infection is located. The length of the cucumber tip is experimentally chosen to be 20% of the cucumber length. This part of the cucumber is pointier for the *Mycosphaerella melonis* infected fruits. Therefore, the pointiness is described by the cucumber width distribution.

First, the cucumber length is normalized to 1000 pixels, then the first 200 cucumber widths starting from the top are extracted and finally, each of the widths represented one feature for the machine learning method (Fig 4). Widths are defined as distances between two cucumber borders along the lines orthogonal to the cucumber centreline. The centreline is determined by fitting an ellipse to the binary object (Haralick, *et. al.*)

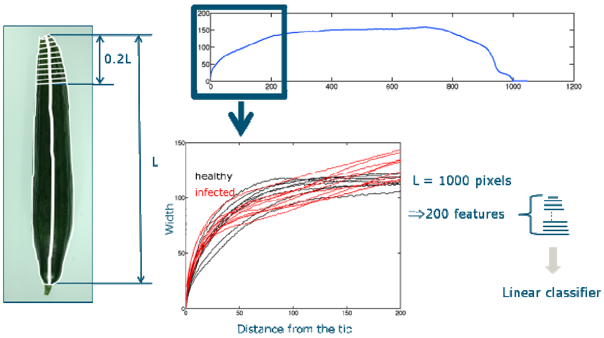

Fig. 4 Shape based classification method: cucumber widths extracted along the centreline (left) and the plot of a cucumber widths vs the distance from the cucumber tip (right). Each of the 200 widths represents a feature that is used by a linear classifier for the shape based cucumber classification.

2.2 Transmission spectra intensities features

There are many infected cucumber samples that do not have clearly and externally distinguishable visual difference compared to the healthy samples (Fig 5).

Fig. 5 Infected cucumbers that resemble healthy fruits.

In order to detect fruits' internal changes, we measure the light transmission spectra through the fruits using near infrared (NIR) spectrophotometer. The spectra are measured in three points near the cucumber tip (Fig 6).

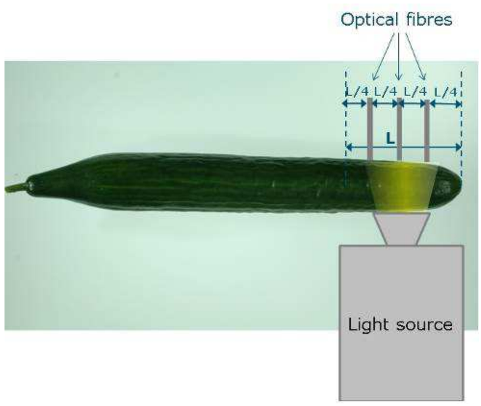

Fig. 6 Light transmission measurement near the cucumber tip: L ~ 0.2*Cucumber length.

The raw spectra are normalized with respect to dark and white reference and with respect to intensity between 0 and 1. Signal sampling frequency was 1nm and the spectrometer integration time was 10ms. The signals are resampled to 51 wavelengths. The sampling frequency is determined using a small tuning set that contains 50 samples. In figure 7 averaged spectra of all healthy and infected cucumbers in our dataset are presented in green and blue respectively.

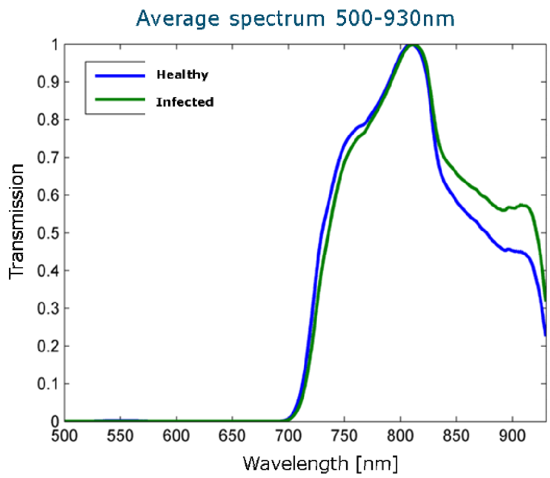

Fig. 7 Averaged light transmission spectra of all healthy and infected cucumbers in the dataset used in green and blue respectively. Wavelengths are presented on the x-axis and the responses of the spectrometer on the y-axis.

There is an obvious difference in shape that indicates possible separation between the healthy and the infected cucumbers in the feature space of spectra intensities. The biggest difference is in the visible spectrum range from around 400nm – 700nm.

The sampled spectral intensities are used as features for a machine learning algorithm. The performance of the classifier is tested using leave-one-out cross validation (Fig. 8).

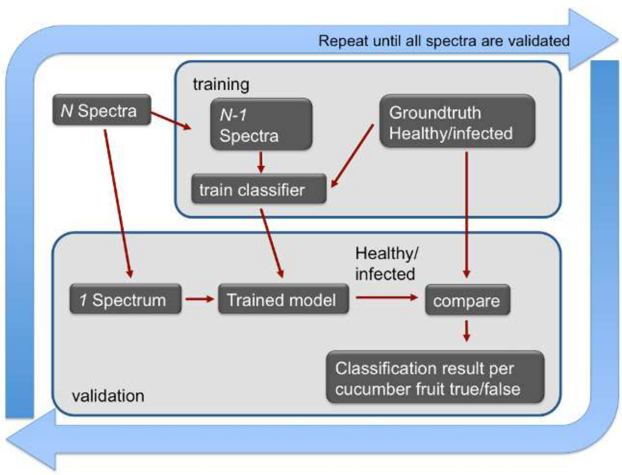

Fig. 8 Block diagram of the light transmission spectra based classification part using leave-one-out cross validation.

2.3 Classification

The classification between healthy and infected samples is done using the features described in sections 2.1 and 2.2. The classification is performed in a two-step way. Firstly, the shape based classification is done using fisher linear classifier. Consequently, the spectral based classification is performed using the spectral features with the fisher linear classifier. Finally, the cucumbers that are classified as infected by both shape and spectral based classification are labelled as infected.

Fisher linear classifier is used since it performed the best out of several classifiers evaluated on a small tuning set containing 50 samples. The classifiers compared were: Parzen windows, k-means, support vector machines and Fisher classifier.

3. EXPERIMENTAL SET UP

The total of 200 cucumber samples used in our work were collected in two time points: in April 2014 140 cucumber and in May 2014 60 cucumbers. Among those, there were 87 and 15 infected cucumbers collected in April and May respectively. The colour imaging and the spectral measurements were done 2 days after the harvesting. Colour camera used was 3008x2000 pixels Nikon D70. The spectrometer used for the spectral measurements was an Ocean Optics S2000 fibre optic spectrometer with a spectral range of 340-1030 nm. Each cucumber was QR labelled and after the measurements were finished, each was cut open and the ground truth with respect to infection presence was determined.

Due to few errors that occurred during the signals measurement phase a few samples had to be excluded from the method development and evaluation. For the shape based classification method correctly imaged colour images of 191 cucumbers where used (Table 1) and for the spectra based classification correctly measured transmission signals of 193 cucumbers were used (Table 2). The final classification algorithm that utilizes the combination of the shape and spectral measurements was evaluated on 184 samples (Table 3).

4. RESULTS

Leave-one-out analysis was performed on the total amount of data collected and the results are presented in the Table 1, Table 2 and Table 3. The confusion matrices with true positive and true negative rate values in brackets are shown for the shape-based classification method (Table 1), spectra-based classification method (Table 2) and for the combined approach (Table 3).

Table 1. Confusion matrix with true positive and true negative rate for shape based classification: ground truth presented in the columns and the automated classification results in the rows.

	Healthy_am	Infected_am	Total
Healthy_gt	**85 (94%)**	5	90
Infected_gt	7	**94 (93%)**	101
Total	92	99	191

Table 2. Confusion matrix with true positive and true negative rate for spectra based classification: ground truth presented in the columns and the automated classification results in the rows.

	Healthy_am	Infected_am	Total
Healthy_gt	**87 (96%)**	4	91
Infected_gt	14	**88 (86%)**	102
Total	101	92	193

Table 3. Confusion matrix with true positive and true negative rate for the final classification method: ground truth presented in the columns and the automated classification results in the rows.

	Healthy_am	Infected_am	Total
Healthy_gt	**81 (92%)**	7	88
Infected_gt	2	**94 (98%)**	96
Total	83	101	184

The final method confusion matrix (Table 3) translates into the following success rates: healthy cucumbers are detected with the 92% success, infected cucumbers are detected with the 98% success, while the overall accuracy rate was 95%.

The manual overall detection rate was 96%, where all the infected cucumbers were correctly classified and the 4% error was made when the healthy cucumbers were sorted out as infected and hence the false positive rate was 0%.

5. DISCUSSION

The accuracy of the method based only on the shape colour images features was improved by incorporating spectral features. This was expected given that the most prominent symptom of the *Mycosphaerella melonis* infection is the internal browning of the fruits. From the growers perspective, the priority is to have less false positives since cucumbers with an internal rot present can cause high financial losses for the grower if appear on the market. The combined shape-spectral approach reduces the false positive rate from 7% (Table 1) and 13% (Table 2) to only 2% (Table 3). This is still worse than the manual score where the number of false positives is zero. However, the automated approach false positive rate is significantly reduced. To the best of our knowledge, this study is the first to evaluate the automated detection of *Mycosphaerella melonis* infected fruits.

The width distribution was based on the centreline of the fitting ellipse. This is not a suitable method for the small number of curved cucumbers in the dataset we used. In order to overcome this issue, we calculated the width distribution using distance transform (Danielsson, 1980). However, the final classification accuracy decreased. This was possibly caused by the discrete nature of the distance transform approach causing the noisy and less smooth results. Usage of the improved shape descriptor would further increase the classification accuracy.

Additionally, humans rely on the cucumber tip hardness in order to sort out infected fruits. Therefore, a logical step towards the presented algorithm improvement would be to incorporate automated measurements of the cucumber tip firmness. Features measured by, for example acoustic durometer, can be used in the machine learning framework presented in this paper.

6. CONCLUSIONS

In this paper we presented automated method for detection of *Mycosphaerella melonis* infected cucumber fruits. The method is evaluated on a second day after cucumbers harvesting. The overall classification accuracy of the automated method is equal to 95% which is a high score and comparable to the manual classification accuracy of 96%. This is a very promising result with respect to the automated *Mycosphaerella melonis* sorting machine installation in the cucumber packaging production line.

REFERENCES

Canny, J. (1983). Finding Edges and Lines in Images. *Report, AI-TR-720, M.I.T. Artificial Intelligence Lab.*, Cambridge, MA.

Danielsson, P.E. (1980). Euclidean Distance Mapping, *Computer Graphics and Image Processing*, Vol. 14, pp 227-248.

Haralick, R., Shapiro, L. (1992). *Computer and Robot Vision*, Addison-Wesley, pp. 73-75

Laurance D. Hall, Stephen D. Evans, Kevin P. Nott, Measurement of textural changes of food by MRI relaxometry, *Magnetic Resonance Imaging*, Volume 16, Issues 5–6, June–July 1998, Pages 485-492

Mahlein, A. K., Steiner, U., Hillnhütter, C., Dehne, H. W., & Oerke, E. C. (2012). Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods*, 8(1), 3.

Olsen, M. W. and Stanghellini M. E. *Mycosphaerella melonis* on Greenhouse cucumber, *Plant Disease*, 65: 157 – 159.

Pixia, D. and Xiangdong, W. (2013). Recognition of Greenhouse Cucumber Disease Based on Image Processing Technology, *Open Journal of Applied Sciences*, Vol. 3 No. 1B, pp. 27-31.

Silva SHMG da, Moraes W da., S, Bendini H do N, Padovani CR, Jorge LA de C, Sevegnani KB. (2008). Digital image processing for identification of black sigatoka. Central theme, technology for all: sharing the knowledge for development. *Proceedings of the International Conference of Agricultural Engineering, XXXVII Brazilian Congress of Agricultural Engineering, International Livestock Environment Symposium – ILES VIII*, Iguassu Falls City, Brazil.

Wu, D., Feng, L., Zhang, C., He, Y. (2008). Early detection of Botrytis cinerea on egg- plant leaves based on visible and near-infrared spectroscopy. *Transactions of the American Society of Agricultural and Biological Engineers*, 51 (3), pp 1133–1139.

Zhang, Z, Zhang, J., Wang, Y., Zheng, X. (2005). Molecular detection of *Fusarium oxysporum f. sp. niveum* and *Mycosphaerella melonis* in infected plant tissues and soil *FEMS Microbiol Lett*. Oct 15;251(2):357