From Experiments to Realizations: Hyperspectral Systems

Matthias Michelsburg[1], Tan-Toan Le[1], Kai-Uwe Vieth[2], Thomas Längle[2], Günter Struck[2], Fernando Puente León[1]

[1]Institut für Industrielle Informationstechnik (IIIT), Karlsruher Institut für Technologie (KIT), Hertzstraße 16, 76187 Karlsruhe

[2]Fraunhofer Institut für Optronik, Systemtechnik und Bildauswertung IOSB, Fraunhoferstraße 1, 76131 Karlsruhe

Corresponding author: kai-uwe.vieth@iosb.fraunhofer.de

1 Introduction

Expectations on sorting quality have increased tremendously over the last two decades, regardless of what has to be sorted: minerals, food, or waste. Although the drivers might depend on the type of material, the goal is always the same: companies want to or have to increase the sorting quality while keeping or increasing the throughput. Thus, modern fast optical inspection systems based on visual, e.g., RGB line cameras, are integrated into many production lines.

For about a decade, near-infrared sensors are also used for sorting but still play a minor role compared to the huge amount of sensors working in the visible range. The advantage of the near-infrared range is that light is reflected which physically interacts differently with the material than electromagnetic waves do in the visible range. Thus, the material can be characterized differently in addition to an RGB-sensor. The disadvantage of a near-infrared sensor is that it is expensive, has got less pixels, and no multi-line sensor exists so far, to mention only a few aspects.

State-of-the-art near-infrared sensors are sensitive over a wide spectral range. For discriminating desired product and foreign material it is usually necessary to filter this spectral range.

Hyperspectral imaging systems are ideal tools for analysis of material that has to be sorted as they acquire a complete spectrum for each single pixel instead of only three spectral values as an RGB-sensor does. Although it is preferable for laboratory analysis it is the exception to the rule to put a hyperspectral imaging system into work for, e.g., mineral sorting, recycling, or food sorting, due to its slowness, expensiveness, and insensitivity.

2 State-of-the-art

State-of-the-art sorting systems are based mainly on grayscale or color, i.e., red, green, blue images. Nowadays, the sorting systems are well established in different industry branches producing high resolution images while performing with a high throughput. Nevertheless, the list of unsolved tasks is still quite long, thus, new technology is required to tackle them. Hyperspectral imaging has been well known in remote sensing of the earth's surface for more than 30 years [1], whereas it entered into quality related checks about 10 years ago[2].

Examples for hyperspectral imaging applications for quality tests or sorting related tasks are apple, cucumber, orange quality checks [3, 4, 5], or for example recycling [6, 7]. Depending on the material that either has to be identified or defects that have to be detected, various ranges of the near-infrared spectrum are selected and applied. When only a few different wavelength ranges are used we talk about multispectral application, whereas an application with a huge amount of wavelength ranges is designated as hyperspectral.

Of course, spectral analysis for laboratory application exists for many decades. The drawback of this application is the restriction of acquiring the data only from one single point or region of interest. In contrast, the hyperspectral imaging takes into account the complete spectral information of each single pixel that is acquired. Hyperspectral imaging systems scan the objects line by line and project the line into a lens-diffraction combination to split up the broad band information into well-defined spectral intervals onto a 2D sensor being sensitive in the required range, e.g., from 1000 to 2500 nm. Fig. 1 depicts the principle of hyperspectral imaging.

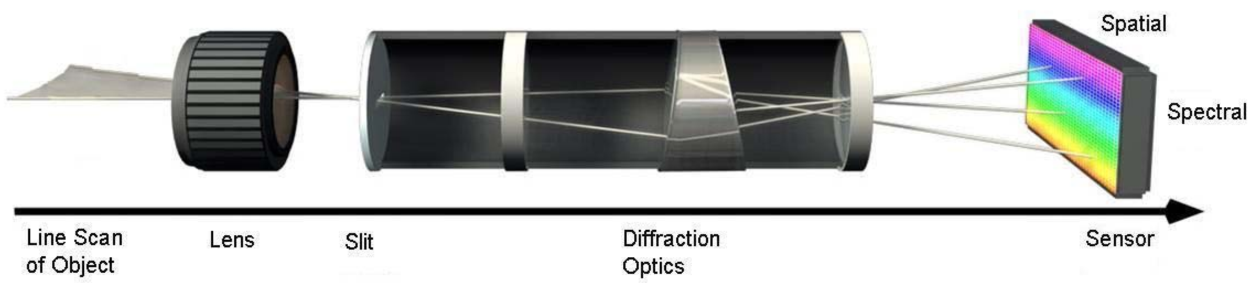

Fig. 1 Principle of hyperspectral imaging.

Hyperspectral imaging can also be applied to other types of quality checks. For about 10 years, the IOSB has worked with hyperspectral imaging systems in the visible range in order to determine precisely the color of plastic granules [8].

3 Identification of relevant wavelength ranges

Hyperspectral imaging systems provide a huge amount of information which characterizes the inspected objects. Focusing on the near-infrared wavelength range, the interaction of the light with the material is due to the bonding of the molecules and thus the reflection and absorption is the result of the overtones of molecule oscillation. For the discrimination of different types of material, usually the hyperspectral data cube contains more information than necessary. Hence, it is possible to decrease computation time, and therefore to simplify analysis and identification, if redundant information is removed and only relevant information is processed. To identify wavelength ranges that have a major impact on the classification, multivariate analysis is applied which is well known from remote sensing and chemometric analysis [9,10].

To extract features from the feature space often the principle component analysis (PCA) or linear discriminant analysis (LDA) is performed. Once the features have been extracted, it has to be tested if these features allow to solve the problem optimally [11]. We perform these tests by computing the distance between classes which have been gained due to spectral band selection. If the test yields good results, the user can continue, e.g., to design a fast sorting system based on only very few wavelength ranges.

Because the design is based on the results of the distance measure of classes, it is important to analyze the results of different existing distance measures which are applied to the same data set in order to understand which distance measure is applicable. This article is based on a diploma thesis by T.-T. Le [12].

4 Distance Measures

Here, we compare three different distance measures, the Mahalanobis Distance [13], Variable Bin Size Distance (VBSD) [14], and Diffusion Distance [15]. The Mahalanobis Distance is a parametric distance measure which describes the distance of two classes by means of mean value and covariance. The Variable Bin Size Distance and the Diffusion Distance are histogram based measures. Histogram based distance measures can be divided into bin-by-bin distance measures which compare bins at identical positions whereas cross-bin methods compare bins of the complete defined range. The VBSD belongs to the bin-by-bin methods and the Diffusion Distance to the cross-bin methods.

A histogram is computed for objects of two different classes as they appear due to the use of an optical filter with a certain central wavelength and a well-defined spectral width. This is repeated for a given range of central wavelength and each time for a variable spectral width.

As the name of the VSBD implicitly describes, the bin size of the histograms changes. The bin-by-bin distance is calculated several times while the bin size is different during each run. It is initialized by calculating the distance with a small bin size. Then the overlap of two histograms is subtracted and the remaining part of the histogram is transformed to a new histogram with bigger bins. This process is repeated several times until a threshold is reached. All distances that have been computed for each bin size are weighted and added in order to receive the Variable Bin Size Distance.

The Diffusion Distance method for histograms is based on the idea of deforming the difference of two histograms as a temperature distribution where a diffusion compensation is performed. We describe it rather by example than by mathematics. Fig. 2 shows three different histograms on the left side, its bin-by-bin difference in the middle and the deformed differences at different times are depicted on the right side.

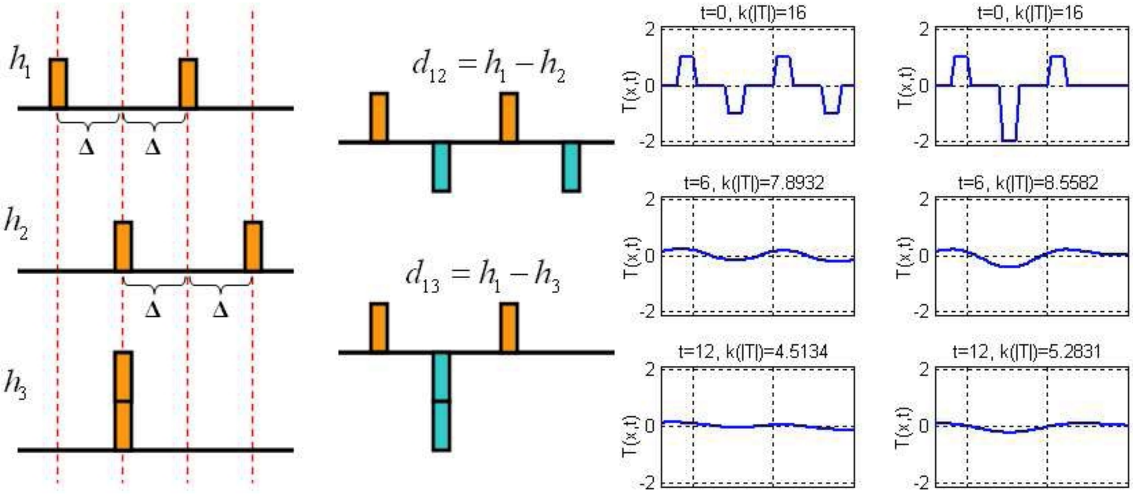

Fig. 2 Histograms and its differences [16].

The diffusion results are identical for the initial point of time but they change with ongoing time due to the difference of the histograms. Hence, the results differ by about 10% after 6 time steps and about 16% after 12 time steps. The number of time steps is restricted by a threshold of the diffusion result. The results of all diffusion steps are added in order to compute the distance between histograms.

5 Evaluation and Application of Distance Measures

On the one hand, the distance measures were evaluated with simulated normal distribution whose parameters were changed and on the other hand the distance measures were applied to measured data. For both cases, each single data value corresponds to the intensity measured by a pixel after the light of the source has been reflected by the object and transmitted by the lens and diffraction optics or by a filter. The distribution of the data can be varied according to the relative midpoint, the scattering of the distribution, and the orientation of the distribution within the range of definition.

Fig. 3 shows intensity values for a bi-chromatic camera, i.e., for two colors due to the use of “Filter 1” and “Filter 2”. The shift of the distribution, depicted in Fig. 3, has been done by several steps. I.e., the initial position of the distribution is shown on the left and the final position on the right. For each step the distances have been computed. The result can be seen in Fig. 4, where it obvious that the Variable Bin Size Distance fluctuates depending on the position of the distribution, whereas the Mahalanobis Distance and the Diffusion Distance yield position-independent results.

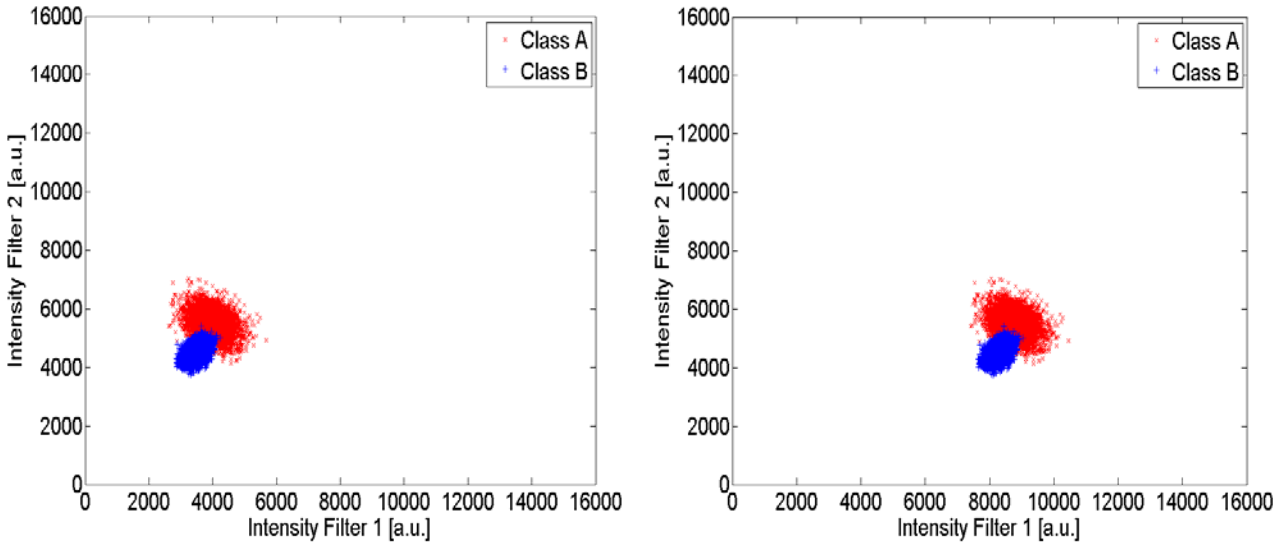

Fig. 3 Initial and final position of histograms.

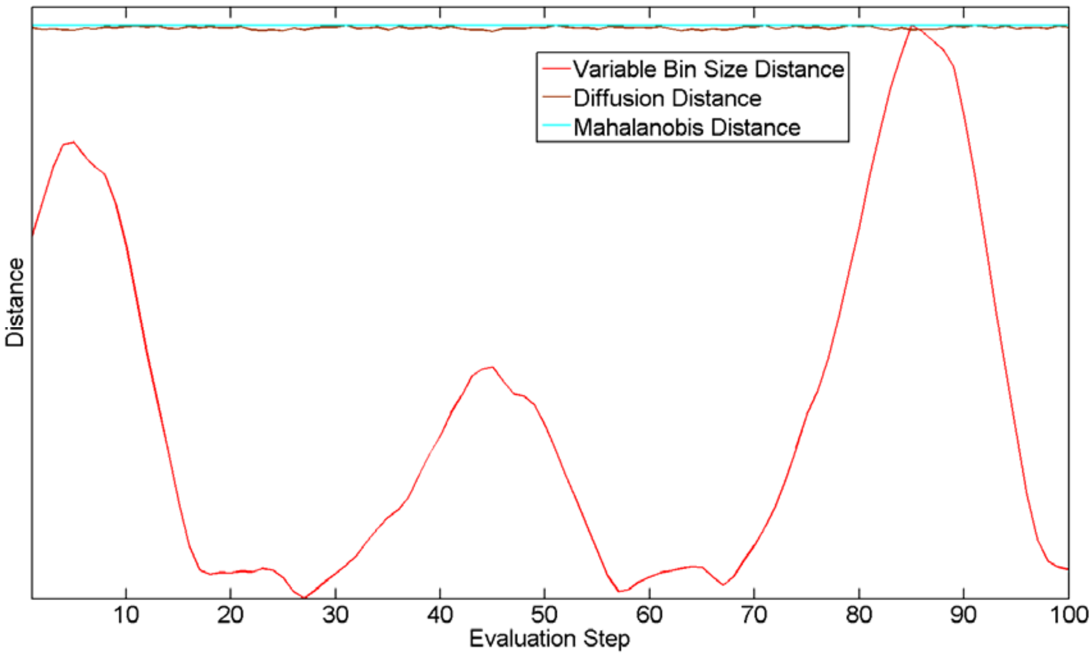

Fig. 4 Evaluation result of shifted histograms.

Of course, also combinations of the variation of parameters are possible and were examined. Fig. 5 depicts the distributions having large covariances on the left and small covariances on the right. In addition to changing them, the expected value has been changed.

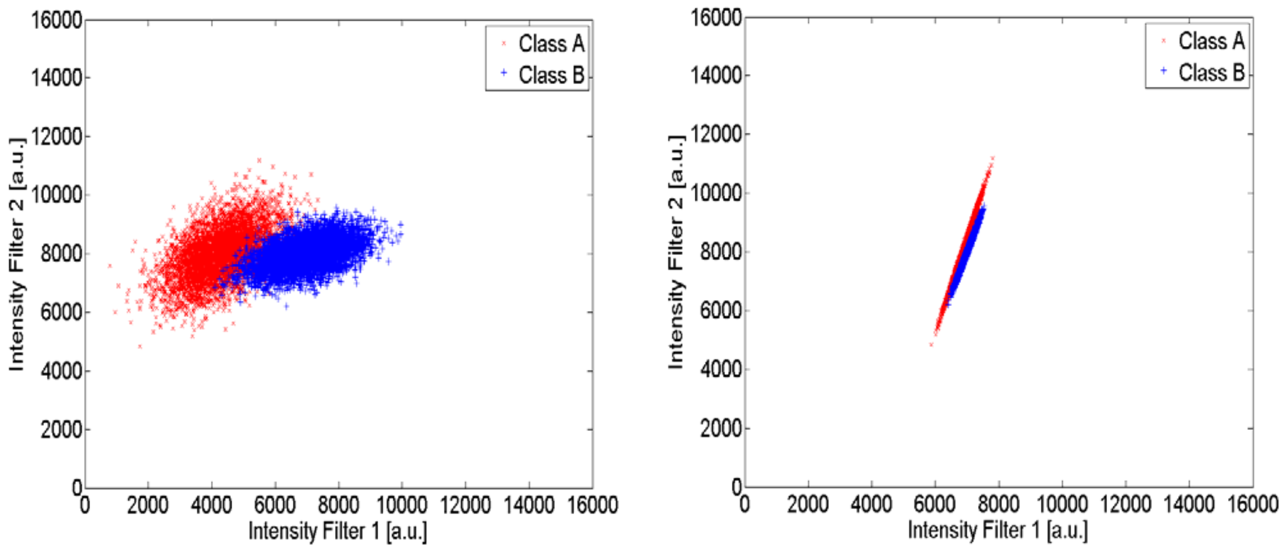

Fig. 5 Covariance and expected values are changed.

The impact of the concurrent change of the covariance and the expected value is documented in Fig. 6. The Mahalanobis Distance increases, the Variable Bin Size Distance decreases, and the Diffusion Distance is nearly constant. One would expect that the distance decreases as the overlap of the final distribution is quite large.

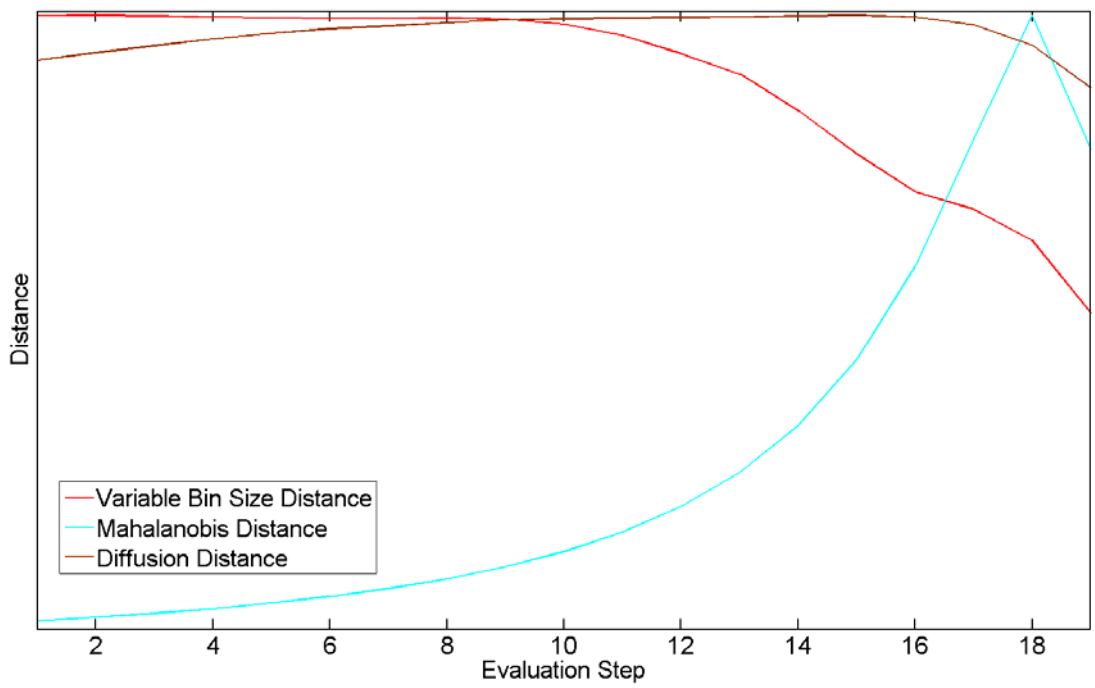

Fig. 6 Results of distance measures when covariance and expected value is changed.

We tested the distance measures using boletus and corresponding foreign material, see Fig. 7. Hyperspectral data of good boletus was acquired and also from the foreign material. Then the filter response for the materials was computed by integrating the data for a given band width at a defined central wavelength. For the central wavelength, the range from 1000 – 2500 nm was split into 30 intervals and the band width varied 15 times between 10 and 710 nm. Altogether 450 combinations of central wavelength and bandwidth were computed for the simulation of one NIR-filter. Of course, it can also be simulated for two or more NIR-filters.

Fig. 7a Foreign material of boletus.

Fig. 7b Good material of boletus.

The computation of the distances for the different methods for two NIR-filters is shown in Fig. 8 and Fig. 9. The distances are normalized to one. Thus, red color means large distance, blue means low distance. Only the upper triangle is colored as the result for the lower triangle is mirrored along the diagonal. Each of the 900 squares represents the distance calculated at two fixed central wavelengths. For each square, the spectral bandwidth is changed separately 15 times for both central wavelengths. The result of the Mahalanobis Distance is rather poor as it is restricted locally, whereas the results of the VBSD and the Diffusion Distance are quite similar at a first glance. The distance of the VBSD method is large for filters with central wavelengths of around 2100 and 1100 nm and the Diffusion Distance is large around 1600 and 1200 nm.

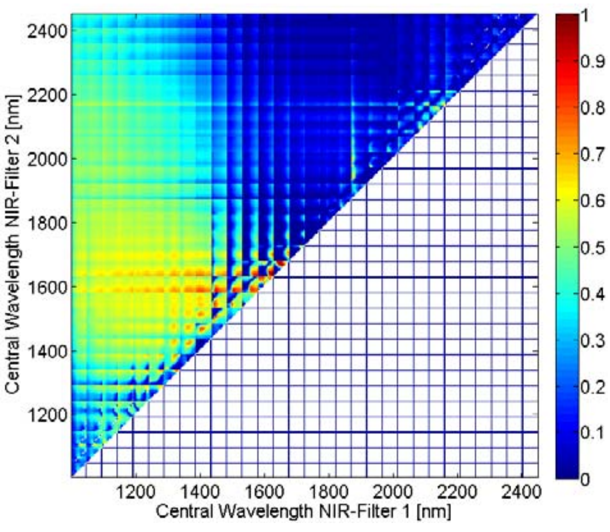

Fig. 8 Mahalanobis Distance for two NIR-filters, ranging from 1000 – 2500 nm.

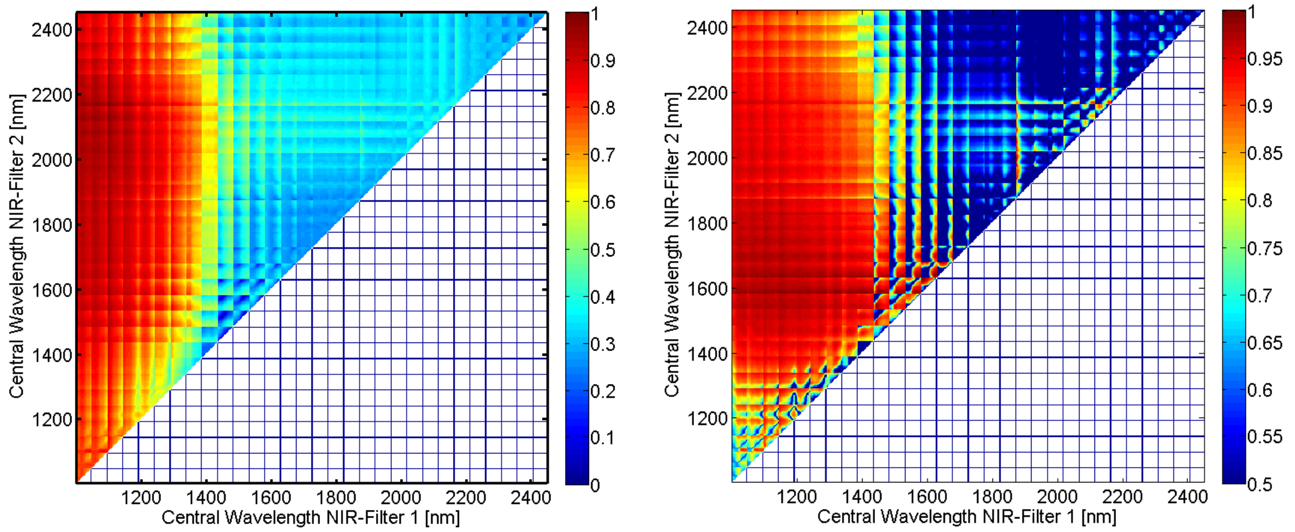

Fig. 9 VBSD (left) and Diffusion Distance (right) for two NIR-filters, ranging from 1000 – 2500 nm.

Compared to the Mahalanobis Distance, the region of large distance values is in Fig. 9 much greater for the VBSD method and Diffusion Distance. The conclusion is that the respective filter would produce a result which is more robust.

When the best filters are selected and applied to the two classes, boletus and its foreign material, it can be seen that the distribution for the Mahalanobis distance is close together, Fig. 10. The distribution of the VBSD is scattered much more and the classes are further apart, Fig. 10 right side. The result of the data distribution by means of the selection of the NIR-filter by Diffusion Distance is similar to the VBSD results.

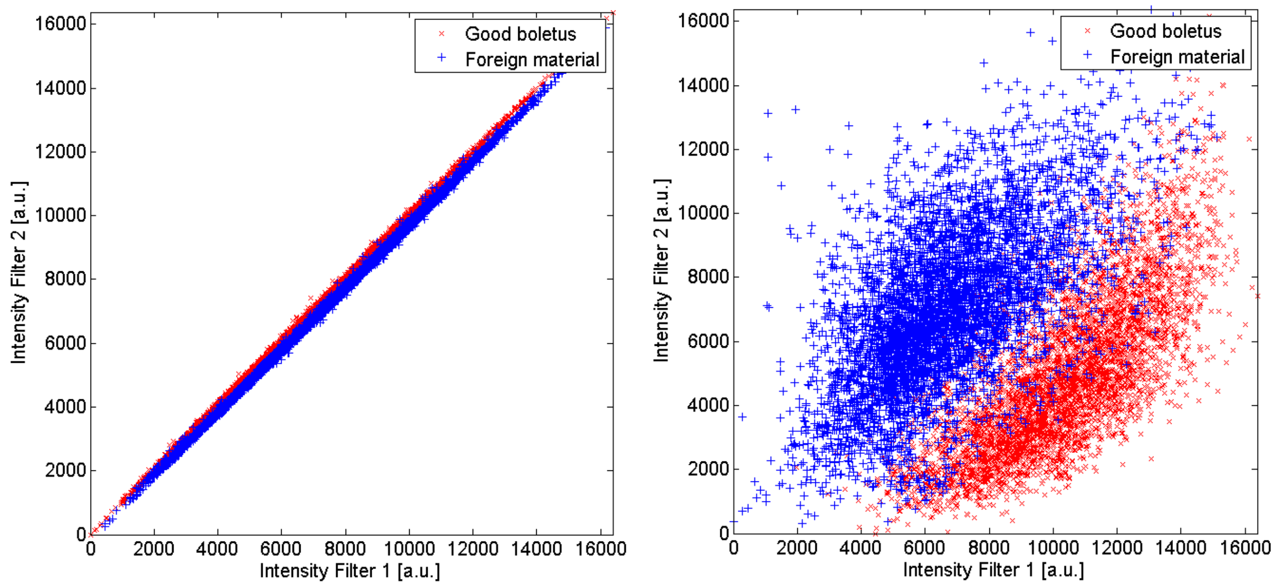

Fig. 10 Data distribution due to NIR-filter selection of Mahalanobis Distance (left), and due to selection of VBSD (right).

6 Summary

The requirements on the inspection of products still increase and require to extend the inspection technology outside the visible range. Because one does not have the experience for those ranges, the selection of filters, e.g., NIR-filters has to be supported by means of distance measures. Three different distance measures, Mahalanobis Distance, Variable Bin Size Distance, and Diffusion Distance were evaluated and applied to real data. Although it seems that Mahalanobis Distance yields clearly separated classes and is quite independent with respect to shifted distributions, the result of the Variable Bin Size Distance is to be preferred as a variation in intensity always exists. Thus, the filter combination of the VBSD result is more robust.

References

[1] Goetz, A. (2009): Three decades of hyperspectral remote sensing of the Earth: A personal view," Remote Sensing of Environment, 113, pp. 5-16.

[2] Zude, M. (Ed.) 2008: Optical Monitoring of Fresh and Processed Agricultural Crops, CRC Press.

[3] Mehl, P; Chen, Y.; Kim, M.; Chan, D. (2004): Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations, Journal of Food Engineering, 61(1), pp. 67-81.

[4] Ariana, D.; Lu, R.; Guyer, D. E. (2006): Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers, Computers and Electronics in Agriculture, 53(1), pp. 60-70

[5] s Cayuela, J. A., Weiland, C. (2010): Intact orange quality prediction with two portable NIR spectrometers, Postharvest Biology and Technology, 58 (2), pp. 113-120

[6] Balthasar, D; Rehrmann, V. (2010): PET flake sorting based on visual and infrared spectroscopy, Conference, Sensor based Sorting, Aachen, 9-11th March 2010.

[7] Feierabend, A. (2010): Application of NIR measurement technique for the identification and sorting of mineral rocks, Conference, Sensor based Sorting, Aachen, 9-11th March 2010.

[8] Vieth, K.-U.; Enderle, E.; Burkhard, M (2009): Online Farbmessung mittels bildgebendem Spektrometer, 15. Workshop Farbbildverarbeitung, Gesellschaft zur Förderung angewandter Informatik, Berlin: GFaI, pp.71-76.

[9] Serpico, S.; G. Moser, G. (2007): Extraction of spectral channels from hyperspectral images for classifcation purposes, IEEE Transactions on Geoscience and Remote Sensing, 45(2), pp. 484-495.

[10] Brereton, R. G. (2009): Chemometrics for pattern recognition, John Wiley & Sons Ltd.

[11] Gruna, R.; Vieth, K.-U.; Michelsburg, M.; Puente León, F. (2010): Hyperspectral Imaging – From Laboratory to In-line Food Sorting, Proceeding of the Second International Workshop on Image Analysis in Agriculture, pp. 79-90.

[12] Le, T.-T. (2011): Evaluation verschiedener Distanzmaße zur Merkmalsselektion aus hyperspektralen NIR-Daten, Diploma theses, Karlsruher Institut für Technologie, Institut für Industrielle Informationstechnik.

[13] Mahalanobis, P.C. (1936): On the generalized distance in statistics, Proceedings of the National Institute of Science, Calcutta, Band 12, Seite 49.

[14] Ma, Y.; Xiaodong, G.; Yuanyuan, W. (2010): Histogram similarity measure using variable bin size distance. Computer Vision and Image Understanding, Vol. 114, pp. 981– 989.

[15] Ling, H.; Okada, K. (2006): Diffusion Distance for Histogram Comparison, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 246 – 253.