Detection and classification of heterogeneous materials as well as small particles using NIR-spectroscopy by validation of algorithms

Xiaozheng Chen and Alexander Feil

RWTH Aachen University,
Department of Processing and Recycling
Wüllnerstr. 2, 52062 Aachen

Abstract Detection, characterization and sorting of plastics using Near-Infrared (NIR) Spectroscopy is State of the Art for plastic recycling processes. NIR active materials could be characterized according to the specific spectra of each material in the NIR spectrum. This works well for homogeneous materials, as they have known uniform spectra. For the detection of heterogeneous material and particles with smaller size than the resolution of NIR camera, the analysis becomes difficult due to mixed spectra. In this paper, the capturing of spectra information with a NIR camera takes place simultaneously with the optical detection with a VIS (Visible light) camera. The NIR spectra and VIS information are combined for the analysis of the mixed spectra, because the higher resolution of VIS camera contributes to a clear definition between the plastic materials and the background as well as between the selected materials. In order to determine the material composition, different kinds of mixed spectra of plastics with background as well as of plastic composition were detected and analyzed. The background is a conveyer belt made of black plastic, and the studied types of plastics are Polypropylene (PP), Polystyrene (PS), High-Density-Polyethylene (HDPE) and Polyvinylchloride (PVC). For the analysis, several algorithms will be developed and tested. In the end, a universal algorithm which performs well for all kinds of mixed spectra will be selected and improved.

Keywords: NIR-spectroscopy, resolution, analysis, sorting, heterogeneous material, Classification.

DOI: 10.58895/ksp/1000087509-7 erschienen in:

OCM 2019 - 4th International Conference on Optical Characterization of Materials, March 13th – 14th, 2019, Karlsruhe, Germany

DOI: 10.58895/ksp/1000087509 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000087509/

1 Introduction

Plastic is one of the most widely used materials in the industry. Detection, classification and sorting of all types of plastic materials are important processes in the recycling chain of plastics. The state of the art for the sorting of plastics is sensor-based sorting with Near-Infrared (NIR) spectra technology [1,2]. Different kinds of plastic materials show distinct spectra in the NIR wavelength area. According to a known specific reflection of different wavelength, the material could be detected and classified. The accuracy of NIR based sorting of plastics can be higher than 98 % with a throughput up to 6 tons/h [3]. However, the application of NIR classification is limited. In some cases, the detection of heterogonous material can be a challenge due to the mixed spectra, mainly because the pixels in the contact zones and at the object edges are compositions of two or more types of materials [4]. The spectra of these pixels are not unique anymore as pure material. Especially when processing small heterogonous particles, the recognition of these particles can result in a wrong classification due to low spatial resolution. In this context, the particle size for the detection with NIR technology is also limited. For example, during the quality-management-process of the Polyethylene terephthalate (PET) flakes, the flakes can be smaller than the resolution of NIR camera, the spectra are, accordingly, mixing of PET and background. Thus, the classification is not possible. In this study, different types of mixed spectra were analyzed. The pixels with mixed spectra are confirmed with the information from the VIS camera due to its higher resolution. The mixed spectra of various plastic materials with background has been captured and analyzed. Besides, the mixed spectra of contact zones of different kinds of materials have been investigated. Several algorithms were tested for the analysis in order to compare with each other and find out the most suitable one for the classification of mixed spectra. The aim of this research is to investigate the possibility of the detection and classification of heterogeneous material and particles with smaller size than the resolution of NIR camera through analysis of mixed spectra with different kinds of algorithms.

2 Material and methodology

The test rig used in this investigation was the NIR sensor-based sorting system at the Department of Processing and Recycling (I.A.R.), RWTH Aachen University. A simplified scheme of the test rig is shown in figure 7.1.

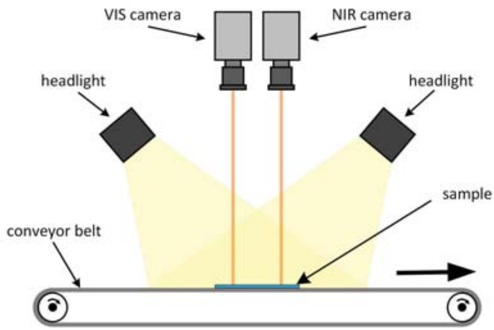

Figure 7.1: scheme of the test rig at I.A.R.

During the capturing process, the samples passed through the NIR light and the cameras simultaneously on a conveyer belt. The halogen lamps were placed with an angle to the capturing area of the two cameras to ensure enough light for the process. The conveyer belt of this test rig is made of black PVC, which reflects much less light comparing to other colors. The NIR-camera used was the model N17E from Specim®, Spectral Imaging Ltd, with a frame rate up to 30 fps. It captures the reflection of the light in the wavelength area between 900 nm and 1700 nm. As the capturing of the wavelength area between 900 – 1000 nm was not stable, only the wavelength area of 1000 – 1700 nm was analyzed in this study. The spatial resolution was 320 pixels in the capturing width with 400 mm and the spectral resolution was approximately 5 nm. In case of the VIS camera, the manufacturer is IDS Imaging Development Systems GmbH. The camera has a spatial resolution of 1280 pixels in the capturing width. The frame rate is up to 1550 fps as line camera with an automated optical inspection height of 4. Both NIR and VIS camera are applied as line camera. For the VIS camera, the line-photos are merged together as a two-dimensional (2D) picture. The NIR camera, however, provides each line a 2D picture, in which one direction is the spatial resolution of 320 pixels and the other direction with 256 pixels is the spectral resolution. The reflectance is reflected from the picture with different grey scales of each pixel. In the following, the grey scales are shown as a line in diagram to be intuitive.

As the frame rate of VIS camera is more than 50 times higher, and the spatial resolution 4 times higher than of NIR camera, a definition of the contact zones of NIR picture was achieved with the help of VIS camera. To clearly define the contact zones of the materials, the pictures of NIR and VIS camera were combined. The capturing with NIR camera was firstly merged to a picture to show the position and the size of the samples.

2.1 Algorithms

Different kinds of algorithms are tested for the identification of the mixed spectra. The most important character to determine the performance of the algorithms in prediction is the accuracy of the classification. Besides, the computation time was compared, since the results of the classification are often needed immediately.

Manual classification

There are different studies in which various manual classification algorithms for plastics detection with NIR spectra were developed [1,5]. The idea is to classify the materials according to the position and/or ratio of the spectra peaks. The spectra can be raw spectra or the spectra processed with first derivative. One of the examples is to distinguish among PET, Polypropylene (PP), Polystyrene (PS), High-Density-Polyethylene (HDPE) and Polyvinylchloride (PVC) through calculating the ratio of the reflectance in two different wavelength areas, 1656 and 1724 nm. The algorithm was capable to classify the materials according to the known ratio range of each plastic type, but was limited to the five types mentioned above [1]. The mixed spectra were firstly analyzed manually based on the classification algorithms developed. The reflectance of the contact zones with different compositions are shown and the characteristic peaks were found manually. The performance of the peaks, e.g., the position, the level and the ratio of positive and negative peaks were analyzed and the most relevant characteristic performances are chosen as classification factor for the mixed spectra.

Machine learning

Machine learning provides automated methods for data analysis to make computers modify or adapt their activities. In other words, with machine learning, the patterns in data are automatically detected and used to predict future data, for example, to classify the data [1, 3]. There are different types of machine learning and numerous algorithms within them. The algorithms which are tested in this investigation are decision trees, k Nearest Neighbors algorithm (KNN) and Support Vector Machines (SVM).

Decision tree Decision tree for classification has been more popular over recent years due to low computational cost. It is important for machine learning that the algorithm works as fast as possible since the results are often needed immediately for sorting or online quality control. The idea of classification tree is to start at the root of the tree and progress down to the leaves, until the features match and the decision is made. In other words, the classification is broken down into a set of logical disjunctions about each feature in turn. [5, 6] The constructing of decision tree algorithms is based on heuristics starting at the root and choosing the most informative feature at each step to construct the tree gradually [5, 6]. Firstly, the root is assigned a label according to a major vote among all labels over the training data. For a new node, a series of iteration is carried out and the effect of splitting a single leaf is examined on each iteration. The split which performs best among all possible splits is chosen to be a new node and local optimization is made at the construction of each node [6].

kNN Nearest Neighbor (NN) algorithms belong to the simplest of all machine learning algorithms for classification. The idea of kNN is to look at similar data and choose to be in the same class as them without searching for a predictor within some predicted class of functions. The training set is memorized, the label of any new instance is predicted on the basis of the labels of the closest neighbors in the training data. In other words, NN is a learning-by-memorization type of rule. [5,6]

The datapoints which positioned within input space are classified according to the nearest neighbors. K nearest neighbors could be identified and the class is set to the most common one out of those for the nearest neighbors. The choice of k is important. If k is too large, the consideration of points which are too far away reduces the accuracy. In case k is too small, the methods are sensitive to noise. Computing the distance to the datapoints in the training set is required for the algorithms, which can cost relatively much time. The computational costs are higher as the number of dimensions grows. [5]

SVM The SVM is one of the most popular and widely used algorithms in modern machine learning due to the computational advantages over probabilistic methods. It provides impressive classification performance on relatively large datasets. [5,6]

SVM algorithm was originally designed for learning linear predictors in high dimensional feature spaces for binary classification. The data are classified through searching for largest margin as large as possible. Figure 7.2 shows the principle of the SVM binary classification. The red and black lines are possible separators to classify the data and the separator in black should be chosen as separator due to largest margin. [6]

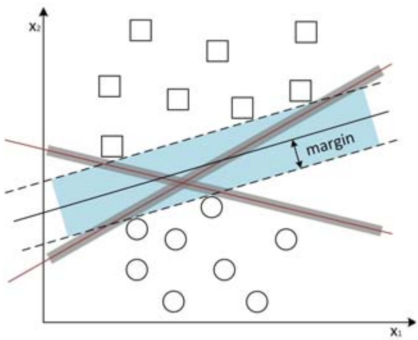

Figure 7.2: The principle of SVM binary classification.

The SVM is also capable to classify data which are not linearly separable. The solution is to introduce some slack variables so that the separator is a combination of various linear separators. Besides, SVM is not limited in two-class classification. For the problem of N-class classification, the SVM can be trained to learn how to classify one from all other classes, then a new SVM which classifies a new class from all others. This means, for N-class classification, there are N SVMs. [5]

2.2 Materials

For the analysis of mixed spectra, samples made from PP, HDPE, PS and PVC were collected. The materials were divided equally into two groups. The first group is for learning information. The mixed spectra of the first group were analyzed manually or learned with machine learning algorithms. The second group contributed to test the accuracy. For all the algorithms, the information for learning was the same and the data for testing purpose were identical.

3 Results and discussion

According to the known frame rate ratio and the resolution of both cameras, the relationship between the same pixels in NIR picture and in VIS picture could be confirmed. In the picture of VIS camera, the contact zones were chosen and the coordinates thereof were determined, the corresponding coordinates of the pixels in NIR picture can be accordingly calculated and clearly defined. The spectra data of the pixels are then analyzed.

3.1 Mixed spectra of plastics and background

For the analysis of the mixed spectra of PP, HDPE, PS and PVC with background, there are 9 classes for the classification: Background, pure PP, pure HDPE, pure PS, pure PVC, PP with background, HDPE with background, PS with background and PVC with background.

Manual classification

To analyze the data manually, the data for different classes were indicated in diagrams to find out the characters which are unique that could be used as classification factor. Figure 7.3 shows the raw spectra and the spectra process with the first derivative of learning pixels for PP, mixed pixels with PP mixed with background as well as background as examples. For the spectra processed with the first derivative, the thick lines are the average of all the values for each wavelength area.

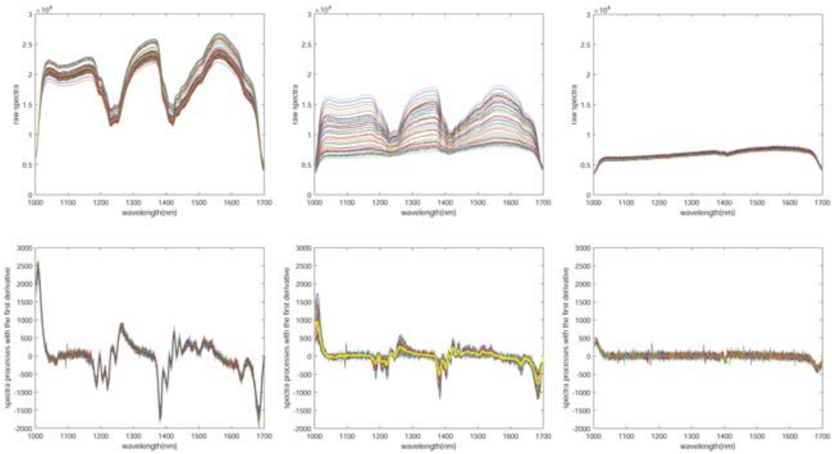

Figure 7.3: Raw spectra and first derivative for the manual analysis (from top left to bottom right: Raw spectra of PP, PP with background, background; first derivative of the same sequence).

According to the raw spectra, the differences between PP, background and PP with background were considerably big. The reflectance of PP was stronger than PP with background and the background reflects, as mentioned above, much less light. A general classification of PP, PP with background and background can be achieved with raw spectra values. However, in order to classify different kinds of mixed spectra, the classification with average values of reflectance is not sufficient. Moreover, the spectra have a large range in the reflectance of certain wavelength, see raw spectra of PP with background.

Since the form is characteristic for the classification, the spectra processed with first derivative were further analyzed and trained to classify the mixed spectra, along with the average value of raw spectra. The average value contributes to the classification between plastic, plastic with background and background, the first derivative is responsible for the classification of types of material. Looking at the first derivative of PP and PP with background, the peaks have the same positions, only the values are different. Thus, the values of the peaks are the classification factor between pure materials and the same type of material with background. Figure 7.4 shows the mixed spectra of PS with background, PVC with background as examples and the average values of all kinds of mixed spectra in one diagram.

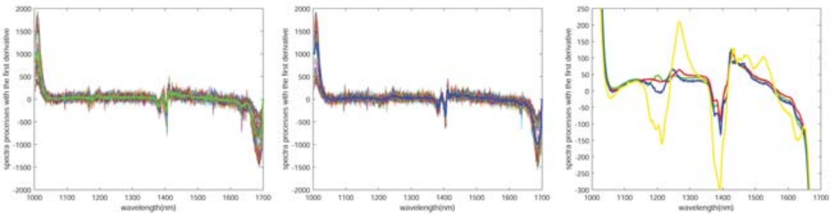

Figure 7.4: Mixed spectra processed with first derivative of PS with background, PVC with background and the average values (Red: HDPE with Background; Yellow: PP with background; Green: PS with background; Blue: PVC with background).

In the first two figures, the spectra are very similar to each other and it is difficult to classify the data. In the diagram at the right side, only the wavelength area from about 1050 to 1670 nm is shown to make the difference clearer. From the results, it can be seen that the distinct differences of HDPE with background, PS with background and PVC with background are in the wavelength area of approximately 1150 to 1300 nm. The characteristic peaks in this area are located in different wavelength, some of which are positive peaks and some are negative. Based on the information above, the algorithm for manual classification is defined as following:

- Pre-classification of the background: find the background with the average values of reflectance using a relatively low threshold, as the average value of some of the pixels with mixed spectra are similar to the value of background.

- Find the important peaks of the spectra processed with first derivative for each material, background included, and classify the pixels in the same class if they have the similar peaks
- According to the average values of the raw spectra and the values of peaks, the pixels are classified to pure material or material with background

The classification results of 4 samples with different materials as examples are shown in figure 7.5. The samples in the picture are HDPE, PP, PS, PVC from left to right. The colors which represent the mixed spectra of the materials are the same as the thick lines in figure 7.3, the black represent the background. The greyscales from low to high are HDPE, PP, PS, PVC.

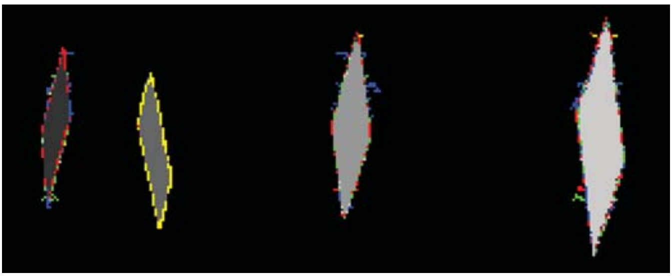

Figure 7.5: Classification results with manual classification.

As a result, the classification worked properly for most of the pixels for pure material. For the classification of mixed spectra, only pixels of PP with background were correctly classified. The accuracy for classifying other mixed spectra was low. In some cases, the background pixels were classified as mixed materials. The accuracy of the classification was about 90.7 % and the computational time of the classification was 5.18 seconds.

Decision tree

The information used for the analysis of the manual classification was learned with decision tree algorithms. Figure 7.6 demonstrates the best classification resulted from decision tree.

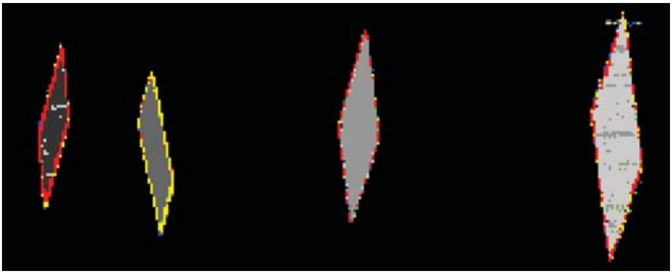

Figure 7.6: Classification results with decision tree algorithm.

With decision tree algorithm, the mixed spectra of HDPE with background, PP with background could be mostly classified to the right group, as well as pure PP and PS. The classification of pure HDPE and PVC was not very successful, as some of the pixels were classified as PS. The accuracy of classification for mixed spectra of PS with background was lower comparing to HDPE and PP with background. Almost all the mixed spectra of PVC with background were not classified correctly. The accuracy for training and classification was 91.5 % and the computational time was 2.85 seconds.

kNN

Figure 7.7 shows the classification results with kNN algorithm.

With kNN algorithm, almost all pixels of pure HDPE, PP and PS were classified properly. The accuracy of the classification of PVC was lower than the others, similar to the decision tree algorithm, a part of the pixels were classified as HDPE. For the classification of the mixed spectra, the accuracy decreased, since for HDPE with background and PP with background, some of the pixels were classified as background.

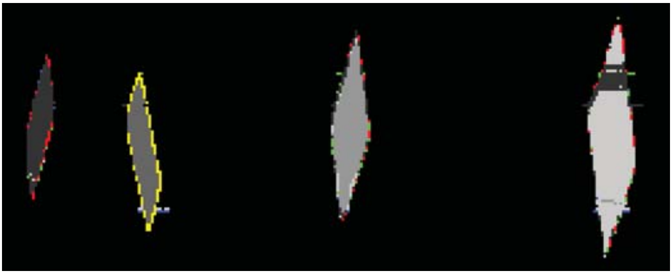

Figure 7.7: Classification results with kNN algorithm.

The classification of PS with background and PVC with background was not successful, only few pixels were right classified. The accuracy for training and classification was 89.4 % with a computational time of 17.4 seconds.

SVM

Figure 7.8 shows the best classification achieved of all the SVM algorithms.

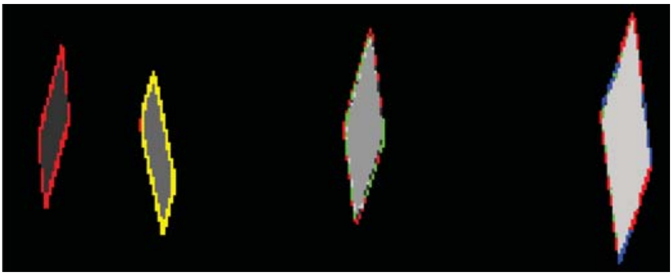

Figure 7.8: Classification results with SVM algorithm.

With this algorithm, the classification of both the pure material and mixed spectra was successful, comparing to other algorithms. However, the accuracy of classification of PS with background and PVC with background was lower than of other classes. The accuracy for the classification was 96.1 % with a computational time of 27.1 seconds.

The results of the classification with all kinds of algorithms show that the classification of mixed spectra of PS with background and PVC with background was more difficult than other classes. The reason for this is that the mixed spectra of HDPE with background, PS with background and PVC with background are very similar, like figure 7.4 shows. It must be noticed that most of the wrong classification was PS with background and PVC with background as HDPE with background, due to more data of mixed spectra of HDPE with background than the other two in the learning data. The pixels for the classification were more likely to be classified into the class with more data. Another reason why the classification accuracy is not high enough is the small size of the database. There were enough spectra for the learning of pure materials, for the mixed spectra, however, much less. A larger size of database brings, in the other side, a problem of increased computational time, especially for the kNN-Algorithms, as the distance to the datapoints must be calculated and compared.

3.2 Mixed spectra of different compositions of plastics

As the conveyer belt of the test rig was also plastic, the analysis of the mixed spectra of plastics with background is a special case of the analysis of different kinds of compositions. In that case, one of the components was certain, the conveyer belt. In order to classify the mixed spectra of the contact zones and the pure materials, there are 11 classes: Background, HDPE, PP, PS, PVC, HDPE with PP, HDPE with PS, HDPE with PVC, PP with PS, PP with PVC, PS with PVC. The analysis methods were the same as for the analysis of plastics with background. The classification results are shown below in table 7.1.

Table 7.1: Accuracy and computational time for the analysis of mixed spectra of different plastic compositions.

Algorithms	manual	Decision Tree	kNN	SVM
Accuracy (%)	89.3	91.0	92.5	96.5
Computational time (seconds)	9.8	10.4	28.7	48.8

The classification results show that the accuracy with SVM algorithm was higher than others, like the analysis of plastics with background.

However, the computation costed significantly more time. It must be mentioned that the accuracy and the computational cost are dependent on the size of learning information and the pixels chosen as learning information. In this case, the analysis and classification of 6 samples in one process results in a higher computational cost. One of the solutions to reduce the computational time of SVM is to analyze based on lines instead of objects. The classification can be more than 300 times faster, since the computational time listed above is for over 300 lines.

4 Conclusions

The analysis of mixed spectra of different kinds of plastics with black PVC as background, as well as of different kinds of compositions have been implemented. The classification of each of them was achieved through manual classification, decision tree algorithms, kNN algorithms and SVM algorithms. The results demonstrate that the classification of mixed spectra is generally possible and, the accuracy together with the computational time are depending on the used algorithms. For the classification of the mixed spectra of plastics with background and composition of plastics, the SVM algorithms proved to be the more accurate one, although the computation costs more time than the other algorithms. Classification manually is possible for pure materials, but it does not work properly for the mixed spectra. However, it must be mentioned that for the algorithms for manual classification, the way of programming is for each person different and the accuracy and computational time are depending on the program. In general, machine leaning algorithms are more universal, as the only effort is to pre-processing of learning information as input for the computer, the classification can be done automatically. This is easier for especially numerous types of materials, the data can be added and learned faster than manual analysis, since for manual analysis, data of all types of material must be compared.

With the successful analysis of mixed spectra of plastics and background, it is possible to detect and classify particles which are smaller than the resolution of NIR camera with NIR technology. Besides, the success of classification of mixed spectra of different compositions makes the detection and classification of heterogeneous material theoretically possible, which offers an additional option for plastic sorting.

References

1. S. M. Safavi and Z. Khani, "Identification and classification of plastic resins using near infrared reflectance p spectroscopy," 2012.
2. T. Pretz and J. Julius, "Stand der technik und entwicklung bei der berührungslosen sortierung von abfällen," *Österreichische Wasser- und Abfallwirtschaft*, vol. 60, no. 7, pp. 105–112, Aug 2008.
3. I. 4R Sustainability, "Demingling the mix: An assessment of commercially available automated sorting technology," 2011.
4. K. Nienhaus, Ed., *Sensor technologies : impulses for the raw materials industry*, ser. Schriftenreihe Zur Aufbereitung und Veredlung ; Bd. 50. Aachen: Shaker Verl., 2014. [Online]. Available: http://d-nb.info/1046907050/04
5. S. Marsland, *Machine learning : an algorithmic perspective*, 2nd ed., ser. Chapman and Hall/CRC machine learning and pattern recognition series. Boca Raton[u.a.]: CRC Press, 2015, erscheinungsjahr in Vorlageform:c 2015.
6. S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.