Generic paper and plastic recognition by fusion of NIR and VIS data and redundancy-aware feature ranking

Alla Serebryanyk[1], Matthias Zisler[2], Claudius Schnörr[1]

[1] University of Applied Sciences Munich, Germany,
alla.serebryanyk@hm.edu, schnoerr@cs.hm.edu,
home page: http://schnoerr.userweb.mwn.de/
[2] University of Heidelberg, Inst. f. Applied Mathematics, Germany,
zisler@math.uni-heidelberg.de

Abstract. Near infrared (NIR) spectroscopy is used in many applications to gather information about chemical composition of materials. For paper waste sorting, a small number of scores computed from NIR-spectra and assuming more or less unimodal clustered data, a pixel classifier can still be crafted by hand using knowledge about chemical properties and a reasonable amount of intuition. Additional information can be gained by visual data (VIS). However it is not obvious what features, e.g based on color, saturation, textured areas, are finally important for successfully separating the paper classes in feature space. Hence, a rigorous feature analysis becomes inevitable. We have chosen a generic machine-learning approach to successfully fuse NIR and VIS information. By exploiting a classification tree and a variety of additional visual features, we could increase the recognition rate to 78% for 11 classes, compared to 63% only using NIR scores. A modified feature ranking measure, which takes redundancies of features into account, allows us to analyze the importance of features and reduce them effectively. While some visual features like color saturation and hue showed to be important, some NIR scores could even be dropped. Finally, we generalize this approach to analyze raw NIR-spectra instead of score values and apply it to plastic waste sorting.

Keywords: near infrared (NIR) spectroscopy, waste sorting, visual features (VIS), CART, feature ranking, machine-learning

1 Introduction

More than 16 million tons of waste paper are processed each year in Germany [4]. At our partner facility around 130,000 tons per year are handled. A high sorting quality of the waste paper is critical to achieve a high grade of recycled paper while keeping the environmental footprint to a minimum. In [10], a general overview of many methods in the field of paper waste sorting is given, and the impact is emphasized these methods can have on the conservation of natural resources in terms of energy and water consumption, CO_2-footprint, and environmental pollution. Ultimately, good knowledge about the input material may be used to optimize the parameters of the sorting facility, e.g. the conveyor belt speed.

We address this paper sorting problem by using near infrared (NIR) and additional RGB (red-green-blue) visual data. From the visual data, we use the RGB and HSV (hue-saturation-value) color components and compute a huge variety of features consisting of classical and statistical texture sensitive features (VIS-features).

There is also a strong need for optimizing the parameters of sorting facilities for plastic waste based on the composition of the input material in order to improve the throughput and the sorting quality. In the European Community alone there are 26 million tons of plastic waste to be sorted, only 30% of them are recycled [3]. This is all the more important since China has denied to take the plastic waste from Europe any longer. The quality of the sorted output in terms of purity and attainable constant properties of sorts is crucial for the usability in many applications and thus for the price of the recycled materials.

Our classifier implementation of a Classification and Regression Tree (CART) allows a ranking of the features by importance and thus can be used to select only the most important features.

[3] According to a recent newspaper report

Furthermore, the complexity of the classifier can be parameterized to create simpler decision trees which has proven to be more robust in case of high measuring errors and partly non-representative data. The optimal decision tree ultimately results by a cross-validation training scheme.

For paper waste, we compare the classification performance in three experiments: First, only NIR scores are used for training, then RGB and HSV data is added, and finally a whole variety of visual (VIS) features is combined. Based on the set of NIR and VIS features we were able to show the power of an importance ranking for an effective feature selection.

For plastic waste, we have direct access to the raw spectra, so we can analyse the raw spectra of a NIR camera instead of pre-processed score values, as we were limited to do in the paper waste case. In this case the improved feature ranking is able to identify the wavelengths with most discriminative power for the trained plastic sorts.

The rest of the paper is arranged as follow: in section 2 the setting for the recording of the paper and plastic waste material is sketched and the characteristics of the available sensor data is described. Section 3 briefly mentions classic approaches to analyse and classify waste material, and a list of feature ranking approaches is given, one of them based on the CART is pursued further and discussed in more detail in section 4. In particular, in section 4.2, our modification of the CART feature ranking is given to adequately regard the redundancy of features. This modification is empirically verified by a synthetic data example. Section 4.3 states a modification to the pruning of the CART to improve its robustness. The preprocessing of the paper data and plastic spectra is stated in section 4.4. Section 5 describes how the recognition rate could be increased from 63% to 78% by fusing NIR and VIS data, and the effectiveness of our feature ranking and reduction method is proved on the used paper features and on the plastic spectra. Finally, section 6 summarizes the main results and states ideas for future work.

2 Characteristics of waste data

2.1 Paper data

Line scan cameras for NIR and RGB were used to image the conveyor belt transporting the waste paper. The system used in a real paper sorting plant recorded 172 NIR tracks and 1204 RGB tracks at 175 scans per second and a belt speed of around 0.5 m/sec and covered a width of circa 90 cm (see top at figure 1).

Overall, 29 NIR-based features or scores were used for the classification problem and were processed from the raw NIR spectra similarly to [9]. A third party project partner, a NIR camera manufacturer, provided these scores. These consist of 11 scores discriminating plastic versus paper, 15 scores sensitive to different paper classes, and 3 values measuring the content of characteristic chemicals: talcum, kaolin, and lignin. Plastic content may result from coated paper classes, adhesive tapes or foils, for example.

Based on the visual RGB data a huge variety of features is computed consisting of co-occurrence features, histogram moments, Haar wavelet filters, anisotropic Gaussian filters, and first and second order spatial derivatives for various mask widths and orientation angles (VIS features).

The NIR-scores and VIS-features are then combined in a feature vector of dimension d: $\mathbf{x} \in \mathbb{R}^d$ for each pixel of a track. The set of feature vectors $\mathcal{X} = \{\mathbf{x}_i\},\ i \in \{1, \ldots, N\}$ along with a class label from labeled data form the training data set we operate on. Thus, NIR- and VIS-features are fused in these vectors and treated in a common sense by the classifier and feature ranking procedure.

We discriminate 10 paper classes which were defined by a third party project partner. The conveyor belt is treated as a separate background class. Thus, a total number of 11 classes are discriminated for the results in this paper (see table 1).

2.2 Plastic data

To test the recognition of plastic waste only one bottle per plastic class was available. The bottles were cleaned, and labels or markers were removed. This is only a small data set, and the preparation had to lead to too optimistic results in terms of recognition rates, but we wanted to check two aspects:

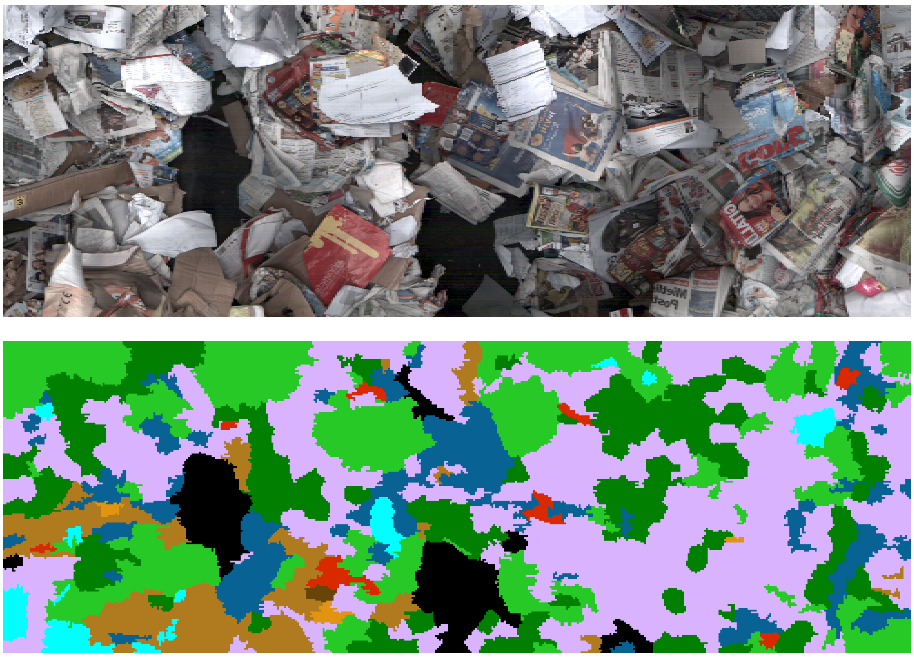

Fig. 1. Example visualization of the classification results on real world paper data. The upper image shows the RGB data of a section of the conveyor belt. Each color in the lower image represents the recognized paper class. The background is colored in black.

– does our generic approach have a chance to be successfully transferred to the treatment of plastic waste?
– can the feature selection analysis be successfully applied to raw NIR-spectra as well to overcome the need of experts experience to compute application dependent score values?

For plastic objects, the NIR-camera recorded 320 tracks perpendicular to the belt movement in the range of $900 - 1200\eta m$ and a wavelength resolution of 256 values. The background was suppressed by an intensity threshold. For the training of the background as a separate class some additional measurements were taken from an empty belt. The background data were reduced as in the paper data experiments so that the background does not dominate the other classes and hence the determined recognition rate. Based on these data a labeled training set was built up.

Note that some PET-classes only differ in color. Table 2 lists all defined plastic classes.

3 Related Work

NIR spectroscopy is a well established technique for material identification in general and paper sorting in particular [9, 10, 11]. Besides characteristic absorption bands, also first and second order derivatives are used to preprocess the raw reflectance spectra. Smoothing filters like Savitzky-Golay are used to reduce noise in the derivatives [9]. Furthermore, Principal Component Analysis (PCA) is used to reduce the dimension of the feature space [7]. Classification is then carried out by evaluating several subsequent binary decision rules, for which Partial Least Squares (PLS) regression is applied. The order of these substeps is based on a sequence of manual analysis steps or on rather intuitive decisions.

Along with PCA also other techniques for feature analysis like Fisher Linear Discriminant Analysis (LDA) or the divergence measure based on Kullback-Leibler distance for probability distributions, besides others, have been used for similar problems in pattern recognition [3]. Generally, the linear techniques PCA and LDA will be only optimal if the class distributions are well separated and Gaussian in feature space.

Well known classifiers include Classification and Regression Trees (CART) [2], Randomized Trees or Random Forests [1] and Support Vector Machines (SVM), besides many others [3]. Feature

class index	abbreviation	description	samples N_i
0	BG	background	853573
1	ZD	newspaper	473144
2	MGWD	magazine/advertising print	854485
3	BP	bureau paper	540297
4	WPb	corrugated paper brown	196494
5	WPw-u	corrugated paper white covered and uncoated	217558
6	WP-g	corrugated paper coated	118834
7	KA-u	carton package uncoated	90218
8	KA-g	carton package coated	538842
9	SV	other packages	152433
10	UN	unassigned objects	139243

Table 1. Paper classes to be discriminated, with $N = \sum_i N_i = 4175121$ samples in total.

class index	abbreviation	description
0	BG	background
1	PET_raw	Polyethylene Terephthalate raw material
2	PET_bottles	PET bottles
3	PET_blue	PET blue
4	PET_brown	PET brown
5	PET_green	PET green
6	PET_transp	PET transparent
7	ABS	Acrylnitril-Butadien-Styrol
8	PE	Polyethylene
9	PE_UHMW	PE ultra high-molecular
10	PEUHMWTG_1.2	PE ultra high-molecular TG 1.2
11	PE_hard	Polyethylene hard
12	Polyester resin	Polyester resin
13	PA	Polyamide
14	PC	Polycarbonate
15	PP	Polypropylene
16	PVC_hard	Polyvilylchloride hard
17	PAK	Polyacrylate

Table 2. Plastic classes to be discriminated

ranking can be done, e.g. by using a CART with surrogates [2], Randomized Trees [5], or Recursive Feature Elimination (RFE) using weight parameters of trained SVMs [6].

We decided to use a CART classifier, since it is a rule-based and parameter free technique which can handle a large number of features and performs well on arbitrary distributions, provided a large number of training samples is available, which is clearly the case in our application [2].

In [8], the approach of a generic data fusion of VIS and NIR data using a classifier and a Machine-Learning approach was first described. In the following sections, we describe the progress of this work and the first step towards an application of the methods to the task of plastic waste sorting by analyzing whole raw NIR spectra.

4 Methodology

4.1 Classifier

We use our own C++ implementation of the CART algorithm which is based on the principles presented in [2]. The CART algorithm trains a binary decision tree. In each node the pattern set is split at a threshold for a feature which minimizes the impurity in the following subsets. As impurity metric we use the Gini diversity index for a node t as proposed by [2]:

$$i(t) = \sum_{j \neq k} p(j|t)p(k|t) \quad , \tag{1}$$

where the indices j and k represent different classes. A splitter s is defined by the feature which is used to split and the corresponding threshold. The decrease of impurity from one node to the left and right child nodes t_L and t_R by a splitter s is described by the delta impurity

$$\Delta i(s,t) = i(t) - p_R i(t_R) - p_L i(t_L) \quad , \tag{2}$$

where p_L and p_R are the proportions of data in t_L and t_R respectively. The splitter s which maximizes $\Delta i(s,t)$ is then used as primary splitter. Each leaf of the tree finally represents a class. To use a trained classification tree, the tree is traversed for a given pattern according to the splits in each node and the class of the reached leaf node is returned.

4.2 Feature Ranking and Selection

In order to rate the importance of features, surrogates are chosen in each node of the tree. Therefore, splitting thresholds for the other features not used in the primary splitter are sought so that the resulting child trees would be most similar to the trees created by the original primary splitter. For each surrogate s^* and the primary splitter s, the delta impurity measure from (2) is calculated. Finally these delta impurities are summed up over all nodes for each feature, which gives a measure $M(x_m)$ for the importance of each feature x_m:

$$M(x_m) = \sum_{t \in T} \left(\Delta i(s_m^*, t) + \Delta i(s_m, t) \right) \quad , \tag{3}$$

where $m \in \{1, \ldots, d\}$ denotes the index of the specific feature, T is the set of all nodes representing the decision tree and s_m^* and s_m denote the surrogates and the primary splitter which involve feature x_m. As opposed to the importance measure found in [2], which ignores the delta impurity for the primary splitter, we deliberately included it, since we think the feature actually used in the primary splitter is important by definition. Tests with an artificially designed test dataset also yielded more realistic importance measures when the primary splitter was included.

Moreover, we defined an importance measure $M'(x_m)$ which only sums up the delta impurities of the primary splitter of each node, thus leaving out these of the surrogate splitters. This means that only features actually used by the classifier gain importance. This has the effect, that the importance ranking selects between similar important but redundant features, thus dropping unnecessary features, as we observed in the selection of characteristic wavelengths in raw NIR-spectra of plastic waste (see later in section 5.2).

To validate this observation we created an artificial dataset comprising 1000 samples of 11 overlapping Gaussian distributions with identity covariance matrices each, that is they scatter isotropically. One distribution is centered at the origin, and the others are placed at the coordinate axes at increasing distances from the origin. These distributions overlap mostly with the distribution around the origin and not with each other. A sketch is given in figure 2 for $d = 2$ features.

A CART classifier can easily separate the centered distribution around the origin from an apart distribution by one threshold on the corresponding coordinate axis, that means the corresponding feature. The farther apart a distribution is the less is the overlap and thus the more important is that feature. When applying the CART, the measure $M(x_m)$ leads to an increasing feature ranking of features $1, 2, \ldots, 10$, as expected.

In a next step, we replicated the feature 5 in the data set as feature 11. Thus, these two features are completely redundant. As expected, these features are assigned the same importance by $M(x_m)$, as shown in table 3. By the way, an Randomized-Tree classifier leads to the same ranking result.

In contrast, when using the measure $M'(x_m)$ the classifier decides to use feature 5 and rates the completely redundant feature 11 worthless, as shown in table 4. This is the sort of feature ranking we need to strongly reduce the feature count while retaining most information about material classes.

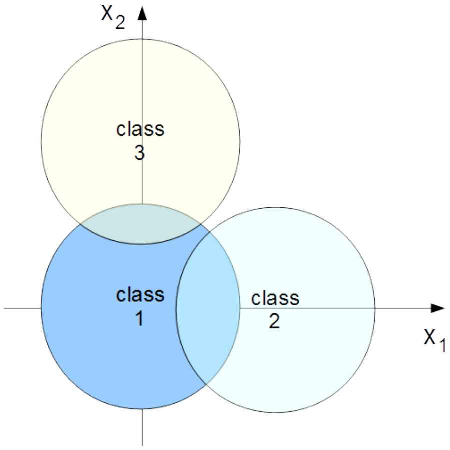

Fig. 2. A sketch of two isotropic Gaussian distributions overlapping at a different degree with the distribution centered at the origin. The circles represent the contour lines of the distributions. Feature x_2 can better separate class 1 and 3 by a threshold than x_1 can with class 1 and 2, thus feature x_2 is regarded more important than x_1 by the ranking measure.

feature	importance
10	1
9	0.90932413
8	0.86688438
7	0.76397420
6	0.66307053
5	0.65805597
11	0.65805597
4	0.47340730
3	0.18303054
2	0.11822442
1	0

Table 3. Normalized feature ranking by $M(x_m)$ with two redundant features 5 and 11 ranked equally

feature	importance
10	1
9	0.90932413
8	0.86688438
7	0.76397420
6	0.66307053
5	0.65805597
4	0.47340730
3	0.18303054
2	0.11822442
1	0
11	0

Table 4. Normalized feature ranking by $M'(x_m)$ with two redundant features 5 and 11. Note that feature 11 is ranked 0 in this case.

4.3 Robustness Improvement

If the classifier is trained until each leaf contains one single training pattern the classifier will likely be overfitted, since also outliers are 'learned by heart' and might be confused with representative data from other classes. This problem is addressed by an internal cross-validation scheme that prunes back the fully trained tree to some degree until it generalizes well on the given dataset.

However, in a real-world scenario with changing side conditions, feature measurements might be slightly influenced by additional effects not covered by the original training dataset. We address this problem by continuing the pruning process of the trained tree to make it more robust against small changing measurement effects. By the way, this leads to simpler trees as well.

4.4 Data Preprocessing

Paper Data The training data is compiled from mono-fraction recordings for each class. As a preprocessing step the paper objects were separated from the background by using a threshold on the intensity of the visual data.

For the results in this paper, the visual resolution of 1204 pixels per scan was scaled down to the resolution of 172 pixels of the NIR data, by a simple data reduction.

Since the background class of the conveyor belt showed to be quite dominant and very well distinguishable from the paper classes, the background data was resampled to roughly the same amount as the next bigger classes. This avoids the overall recognition rate to be too optimistic just because of a good background recognition.

Plastic Data According to [12], varying intensities from scan line to scan line were caused by varying distances between camera and the objects and by diffuse scattering effects. Following the norming procedure described in [12], all spectra are normed so that

$$\sum_{i=1}^{d} |x_i| = \text{const} = 256 \,,$$

where x_i is a component of the feature vector $\mathbf{x} \in \mathbb{R}^d$, in this case the intensity value at a particular wavelength of the spectrum at a pixel of the scan track. Essentially, this normalization removes a constant bias. The constant value 256 is chosen to avoid inaccuracies due to floating point errors for big or small spectral values. Imposed PP-spectra, normalized and smoothed, are shown in figure 3 as an example. These spectra match quite well, they don't spread much vertically.

Since the spectra don't show sharp peaks, no peak retaining smoothing filter is necessary. We used simple Gaussian smoothing filters, and calculated the first and second derivatives by derivated gaussian filters as additional spectral features used in the material classification.

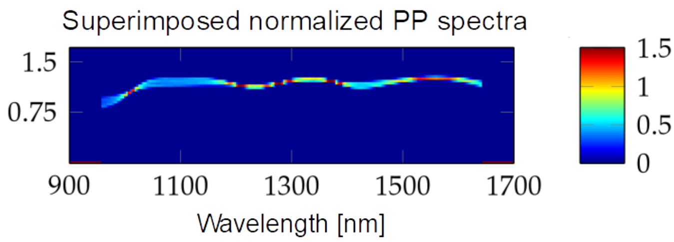

Fig. 3. Example of superimposed spectra for plastic sort Polypropylene (PP) after normalization and smoothing to show the variation in the spectra. The spectra don't spread much vertically after normalization (The color scale represent frequency of overlapping spectra and can be ignored here).

5 Experimental Results

5.1 Paper Data

The dataset used for the following results consisted of almost 4 million samples of which 80% were used as training set and 20% as validation set in a 3-fold cross-validation scheme. To be clear, the purpose of this cross-validation is to get a most accurate estimation of the real recognition rate. We emphasize that this dataset originates from a real sorting facility with all dirty effects like probe contamination, light scattering, changing detector-probe distances, shadow effects, etc.

Solely using the given NIR features as described in section 2.1, our classifier achieved an overall recognition rate of 63%. The classification statistics are given in table 5, and the corresponding error matrix or confusion matrix F is visualized in figure 4. N_i/N is the fraction of data belonging to class i. The elements F_{ij} of F are the number of samples from class i which are classified as class j, where i is the row index and j the column index. The diagonal elements of F represent the frequency of correct classification decisions, while the off-diagonals show false-positive and false-negative decision rates. From F the diagonal elements $diag(F)$ are extracted and the F_1 measure is computed. The F_1 measure is the harmonic mean of precision and recall and thus also considers false positives and false negatives. The overall recognition rate is calculated as $1 - P(F)$, where $P(F)$ is the error probability.

class index i	0	1	2	3	4	5	6	7	8	9	10
class abbrev.	BG	ZD	MGWD	BP	WPb	WPw-u	WP-g	KA-u	KA-g	SV	UN
N_i/N	16.65	11.87	21.44	13.56	4.93	5.46	2.98	2.26	13.52	3.83	3.49
F_1 measure	95.09	54.68	60.35	65.75	43.68	36.32	36.03	19.23	68.98	30.82	34.39
$\mathrm{diag}(F)$	16.169	7.120	14.346	9.618	2.284	1.702	0.736	0.276	9.060	0.789	0.858
$1 - P(F) = 62.958$											

Table 5. Classification statistics for all NIR features ($d = 29$).

Adding the RGB and HSV channels the recognition rate could be raised to 69%. In a first attempt to include other features, a variety of 386 additional visual features were computed consisting of co-occurrence features, histogram moments, Haar wavelet filters, anisotropic Gaussian filters, and first and second order spatial derivatives for various mask widths and orientation angles. The total of 419 features resulted in a recognition rate of around 77%.

As a remark, the trained CART classifier consists of 484054 decision nodes and 33371 leaves in this case. Two reasons led us to the decision not to use a Randomized Tree (RT) instead of a CART: first a RT ranks the features like a CART with surrogate rules accorting to $M(x_m)$. Second, the time of a couple of minutes needed to read in a trained RT consisting of e.g. 100 CART classifiers is a bit prohibitive in a real facility environment.

class index i	0	1	2	3	4	5	6	7	8	9	10
class abbrev.	BG	ZD	MGWD	BP	WPb	WPw-u	WP-g	KA-u	KA-g	SV	UN
N_i/N	16.65	11.87	21.44	13.56	4.93	5.46	2.98	2.26	13.52	3.83	3.49
F_1 measure	96.49	72.60	75.19	80.84	82.79	70.18	63.42	69.81	75.57	62.53	61.99
$\mathrm{diag}(F)$	16.026	8.704	17.086	11.074	4.079	3.629	1.641	1.457	10.242	2.172	1.973
$1 - P(F) = 78.082$											

Table 6. Classification statistics for the best $d = 59$ features selected among NIR, RGB, HSV and a mixture of visual features.

By iteratively deleting the most unimportant features (according to the measure described in section 4.2), the number of features could be reduced to just 59, while even improving the recognition rate slightly to 78%. The error statistics are listed in table 6, and the corresponding error matrix F is visualized in figure 5.

It is worth to be noted, that the increase in recognition rate from 63% to 78% contributed mainly to the paper classes and not to the background class (compare F_1 measures in tables 5 and 6). An example of classified paper waste is shown at the bottom of figure 1 where the paper classes are labeled by different colors.

To further illustrate the feature selection process and its relevance to the achievable recognition rate, figure 6 shows the recognition rate versus the number of selected features among the 419 total features. At the far right, when all NIR and VIS features are used, 77% recognition rate is achieved. Surprisingly, when moving to the left in this plot, a further deletion of features results in a slight increase of the recognition rate, because the classifier is no longer worried about useless and redundant information in the data set. However, the CART classifier is a parameter free approach and deals robustly with useless information. The most important result is, however, that the features can be reduced down to 59 with no loss in the recognition rate, which leads to 78%. Only when reducing the features further, a significant decrease of the recognition rate results (see far left in figure 6). Thus, with appropriate feature selection, the computational cost can be reduced, since only the best visual features need to be computed.

Interestingly, our feature ranking also showed, that the H and S channel of the HSV data are quite important, which is also stated by [9]. More surprisingly, almost half of the original NIR

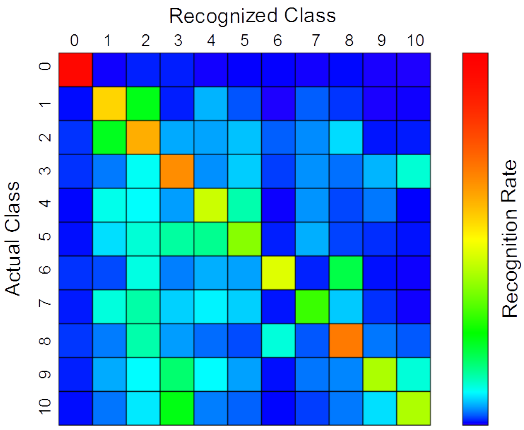

Fig. 4. Visualization of the class error matrix F for 29 NIR features. With i being the row index and j the column index, the elements F_{ij} are the number of samples from class i which are classified as class j. Low values are colored in blue, high values in red.

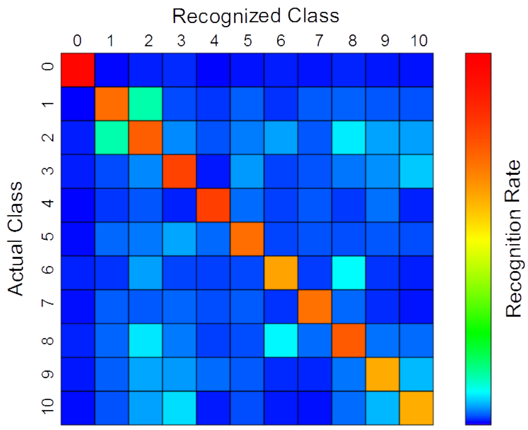

Fig. 5. Visualization of the class error matrix F for best 59 NIR+VIS features (see peak in figure 6). The recognition rate is improved much compared to figure 4.

features could be dropped in the remaining set of 59 features – even the values for talcum and lignin.

While [10] states, that rule-based classifiers like CART are generally too slow for real-time applications, we would be able to process at a conveyor speed of 4m/sec on a standard 4-core computer based on 29 NIR, 3 RGB and 3 HSV features without the need to further parallelize by hardware. This would be eight times the actual conveyor speed. When, however, exploiting many hundreds of visual features, more sophisticated data preprocessing steps need to be applied.

5.2 Plastic Data

In the first experiment, a CART-classifier was trained for all 17 classes with 768 features. The size of the training data is big enough, and the classifier uses an internal cross-validation so that overfitting is avoided. The class error matrix in figure 7 however shows an almost perfect recognition of all classes with $1 - P(F) = 89.57\%$. Even the five PET-classes, that only differ in color and cause the most recognition errors, are recognized quite well. This is an overly optimistic result, of course, but it shows it's worth to proceed with our generic approach.

In the next experiment, only the most important classes from an application point of view are considered further by merging all PET-classes (1-6) and all PE-classes (8-11) to one PET and PE class respectively, and dropping classes 7, 12, and 17, see table 7 and compare with table 2.

class index i	abbreviation	class	pattern samples N_i
0	BG	background	192678
1	PET	Polyethylene Terephthalate	192676
2	PE	Polyethylene	105113
3	PA	Polyamide	12078
4	PC	Polycarbonate	2641
5	PP	Polypropylene	15059
6	PVC_hard	Polyvilylchloride hard	17022

Table 7. Most important plastic classes to be discriminated, with $N = 537267$ samples in total. The class index runs from $0, \ldots, c$ with $c = 6$ classes plus background.

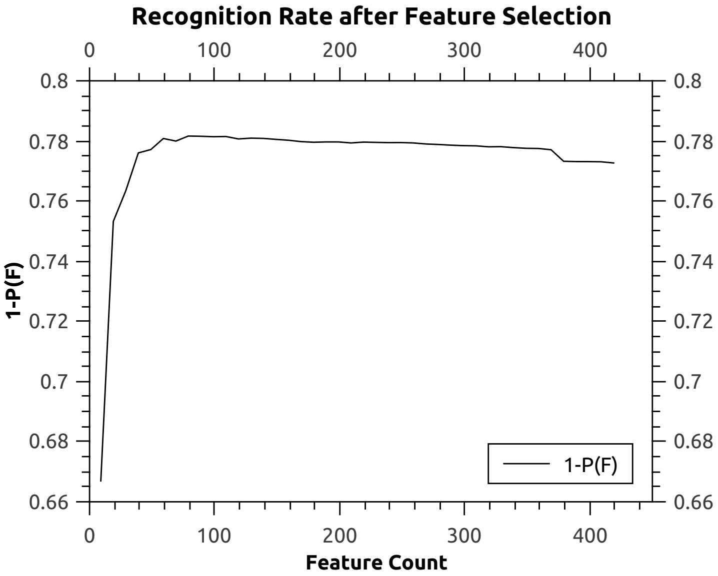

Fig. 6. Recognition rate over selected features. Best trade-off with 59 features and recognition rate of 78%

class index i	0	1	2	3	4	5	6
class abbrev.	BG	PET	PE	PA	PC	PP	PVC
N_i/N	35.86	35.86	19.56	2.25	0.49	2.80	3.17
F_1 measure	99.79	99.62	99.74	99.94	94.48	99.47	97.15
diag(F)	35.809	35.734	19.507	2.247	0.454	2.793	3.061
$1 - P(F) = 99.604\%$							

Table 8. Classifying statistics for 6 important classes with $d = 768$ features

Figure 8 shows the related class error matrix, and table 8 the classification statistics. As before the recognition rate is very good, almost 100% now.

The effect, only to consider the primary splitter in the feature ranking is shown in figure 9. The recognition rate drops at less features compared to the feature selection based on the original ranking criterion. That's because now the ranking selects between equally important, but redundant features, thus dropping high ranked but unnecessary features as well.

Figure 10 shows the second derivative of spectra of various plastic materials. The grey bars indicate the importance assigned to wavelengths according to this feature by the importance measure $M'(x_m)$. Wavelengths where this feature shows great diversity are rated high.

As mentioned above, these recognition rates are overly optimistic due to a) the careful probe preparation and b) the data set being far from realistic for all possible appearances of plastic waste in a real facility. But the results show, that even identic PET probes, only differently colored, can be recognized well, and that the feature selection scheme can be applied to whole raw NIR-spectra too. This is all the more important as

- it is a generic approach without the need of any expert knowledge, and
- the amount of data of a raw spectrum is about eight times that of preprocessed score values, hence the need for a data reduction increases much.

6 Conclusion and Outlook

The experimental results including additional visual features show a significant improvement over NIR scores alone. Our results on the real world paper data approve the preliminary results

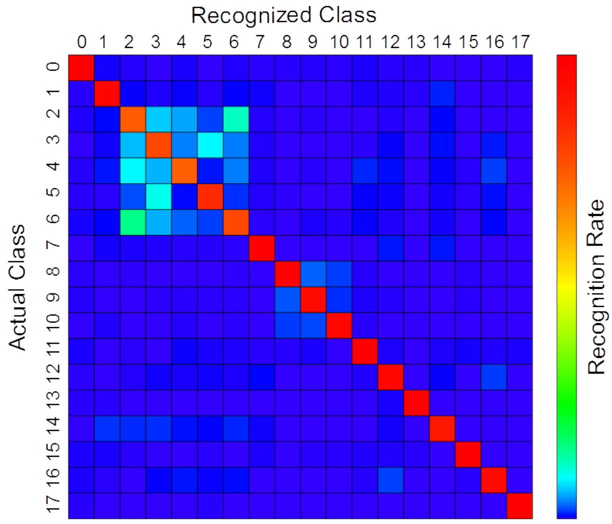

Fig. 7. Class error matrix for all plastic classes. The overall recognition rate is $1 - P(F) = 89.57\%$. Mostly the differently colored PET-classes contribute to the recognition error.

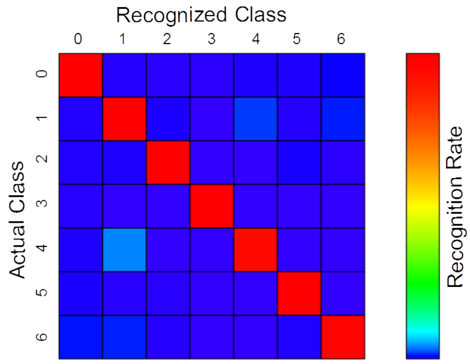

Fig. 8. Class error matrix for 6 important plastic classes. The overall recognition rate is $1 - P(F) = 99.604\%$

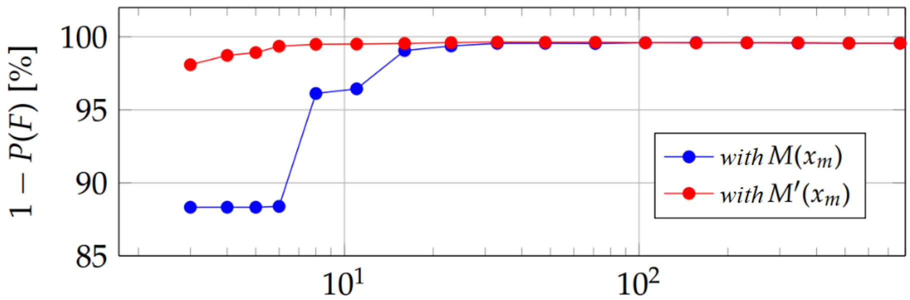

Fig. 9. Recognition rate versus selected features count for Breiman-measure (blue) and only primary splitter measure (red). Less features are needed in the red case.

attained on a laboratory-dataset with 14 different paper classes. The feature ranking of the CART classifier enables us to use many potential features at first and automatically select only the best subset for a productive environment.

The application of the material recognition methods on raw NIR-spectra of plastic waste reveals that wavelengths can be selected in an generic way, where material classes exhibit characteristic diversity, thus preprocessed scores dependent on the experience of a particular camera manufacturer are no longer necessary. This way, the amount of data of raw spectra can be successfully reduced as well while retaining the crucial information.

For the future, we plan to exploit the full visual resolution in order to capture finer structure details in paper waste. At the same time, intelligent data fusion of multivariate data of different resolutions is needed to avoid resubstitution error due to partially replicated data. With a sevenfold higher resolution, the computational costs will also be a critical factor. Therefore, we want to investigate the applicability of a regional pre-clustering procedure and other data reduction techniques. We also intend to compare the feature ranking technique used in our CART classifier to other possible techniques, like e.g. l_1-regularized data reduction. Compared to a simple RGB camera a NIR sensor is rather expensive. Thus, it is also of interest, if visual features alone suffice to achieve an at least acceptable recognition rate for a lower price. Since real world paper waste is not guaranteed to only contain paper, detection of problematic material like inflammable materials or rigid objects which might damage the sorting plant would be much appreciated. For these classes it is generally hard to gather much training data, as the variety of possible objects is huge.

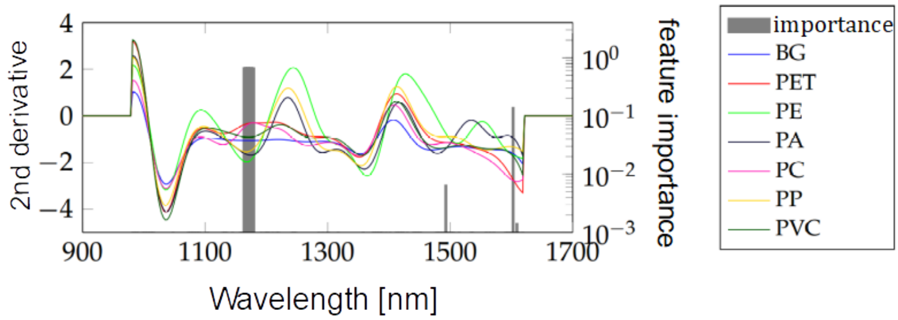

Fig. 10. Importance (grey bars) of the 2nd derivative of spectra versus wavelength.

The recognition results for plastics on a small data set of raw NIR-spectra are quite promising and advice us to determine the recognition rates on a large scale in a real sorting facility for plastic materials as well.

References

[1] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 0885-6125.
[2] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall/CRC, 1984. ISBN: 978-0-412-04841-8.
[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. 2nd. New York: Wiley & Sons, 2000. ISBN: 0-471-05669-3.
[4] Verband Deutscher Papierfabriken e.V. *Facts about Paper*. Brochure. Accessed: 2015-11-30. 2015. URL: `http://www.vdp-online.de/en/papierindustrie/statistik`.
[5] R. Genuer, J-M. Poggi, and C. Tuleau-Malot. "Variable Selection using Random Forests". In: *Pattern Recognition Letters* 31.14 (2010), pp. 2225–2236.
[6] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. "Gene Selection for Cancer Classification using Support Vector Machines". In: *Machine Learning* 46.1-3 (Jan. 2002), pp. 389–422. ISSN: 0885-6125.
[7] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. New York: Springer–Verlag, 1986. ISBN: 0-387-96269-7.
[8] P. Klippel, M. Zisler, F. Schröder, S. Schleich, A. Serebryanyk, and C. Schnörr. "Improvement of dry paper waste sorting through data fusion of visual and NIR data". In: *7th Sensor-Based Sorting & Control 2016*. Ed. by T. Pretz and H. Wotruba. Shaker, 2016.
[9] R. Leitner and S. Rosskopf. "Identification of Flexographic-printed Newspapers with NIR Spectral Imaging". In: *International Journal of Computer, Information, Systems and Control Engineering* 2.8 (2008), pp. 68–73. ISSN: 1307-6892.
[10] M. O. Rahman, A. Hussain, and H. Basri. "A critical review on waste paper sorting techniques". English. In: *International Journal of Environmental Science and Technology* 11.2 (2014), pp. 551–564. ISSN: 1735-1472.
[11] M. O. Rahman, A. Hussain, E. Scavino, N. E. A. Basri, H. Basri, and M. A. Hannan. "Waste paper grade identification system using window features". In: *Journal of Computational Information Systems* 6.7 (July 2010), pp. 2077–2091. ISSN: 1553-9105.
[12] H.W. Siesler, S. Ozaki Y. ans Kawata, and H. M. Heise. *Near-Infrared Spectroscopy. Principles, Instruments, Applications*. Wiley-VCH Verlag GmbH, 2002.