Improvement of dry paper waste sorting through data fusion of visual and NIR data

P. Klippel, M. Zisler, F. Schröder, S. Schleich, A. Serebryanyk, Cl. Schnörr [1]

{pklippel,mzisler,aserebry}@hm.edu, schnoerr@cs.hm.edu

Research group CORSNAV (Computer Vision, Remote Sensing, Navigation)
University of Applied Sciences Munich
Lothstraße 64
80335 Munich

Keywords: paper waste sorting, NIR, visual features, data fusion, CART, feature ranking

Abstract

Near Infrared (NIR) spectroscopy is a well-known sensor technology which is used in many applications to gather information about chemical composition of materials. For paper waste sorting this information can be used to classify different paper classes which enables better sorting and higher recycling quality. With a small number of NIR scores 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. However it is not obvious how this information can be well captured by describing features, and what features are finally important for successfully separating the paper classes in feature space. Due to the huge variety of possible visual features, e.g. based on color, saturation, textured areas with different structure size, etc., a rigorous feature analysis becomes inevitable. We therefore have chosen a pattern recognition approach to deal with the curse of dimensionality. By exploiting a classification tree and a variety of additional visual features, followed by a forceful feature selection, we achieve a recognition rate of 78% for 11 classes, compared to 63% only using NIR features. The feature reduction shrinks the otherwise high computational burden to compute all features and furthermore even increases recognition rate slightly. While some visual features like color saturation and hue showed to be important, some NIR scores could even be dropped.

1 Introduction

More than 16 million tons of waste paper are processed each year in Germany [9]. 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 [7] 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

[1] The authors appreciate financial support of this project by the BMWi - Bundesministerium für Wirtschaft und Energie, Germany

material may be used to optimize the parameters of the sorting facility, e.g. the conveyor belt speed.

We address this sorting problem by using NIR and additional RGB visual data. From the visual data, a 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).

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. Furthermore, the complexity of the classifier can be parametrized 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.

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 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.

2 Characteristics of 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 figure 3).

Overall, 29 features were used for the classification problem which were processed from the raw NIR spectra similarly to [6] and were provided by a third party project partner. 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.

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.

3 Related Work

NIR spectroscopy is a well established technique for material identification in general and paper sorting in particular [6, 7, 8]. 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 [6]. Furthermore, Principal Component Analysis (PCA) is used to reduce the dimension of the feature space [5]. 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 are only optimal if the class distributions are well separated and normal 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 ranking can be done using CART with surrogates [2] or Recursive Feature Elimination (RFE) using weight parameters of trained SVMs [4].

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].

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]:

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

where i and j are different classes. The decrease of impurity from one node to the left and right child nodes t_L and t_R by the 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. A splitter s is defined by the feature which is used to split and the corresponding threshold. Each leaf of the tree finally represents one 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 than used in the primary splitter are sought, such 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 theses 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 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 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.

4.3 Robustness Improvements

If the classifier is trained until each leaf contains one single training pattern the classifier is likely to 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 crossvalidation scheme that prunes the fully trained tree to a certain 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

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 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.

5 Experimental Results

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 crossvalidation is to get a most accurate estimation of the real recognition rate. We emphasise that this dataset originates from a real sorting facility with all dirty effects.

Solely using the given NIR features as described in section 2, our classifier achieved an overall recognition rate of 63%. 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% (see far right in figure 1). 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% (see peak in figure 1). Further deletion of features would result in a significant decrease of the recognition rate (see far left in figure 1). Thus, with appropriate feature selection, the computational cost can also be reduced, since only the best visual features need to be computed.

For each classifier, error statistics are computed (see table 1 and 2). N_i/N is the fraction of data belonging to class i. F is called error matrix or confusion matrix and is visualized in figure 2. The elements 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. From F the diagonal elements $diag(F)$ are extracted and the F_1 measure is computed. The F_1 measure is the harmonic

class i	1	2	3	4	5	6	7	8	9	10	11
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
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 1: Classification statistics for NIR features.

class i	1	2	3	4	5	6	7	8	9	10	11
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
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 2: Classification statistics for best 59 features of NIR, RGB, HSV and a mixture of visual features.

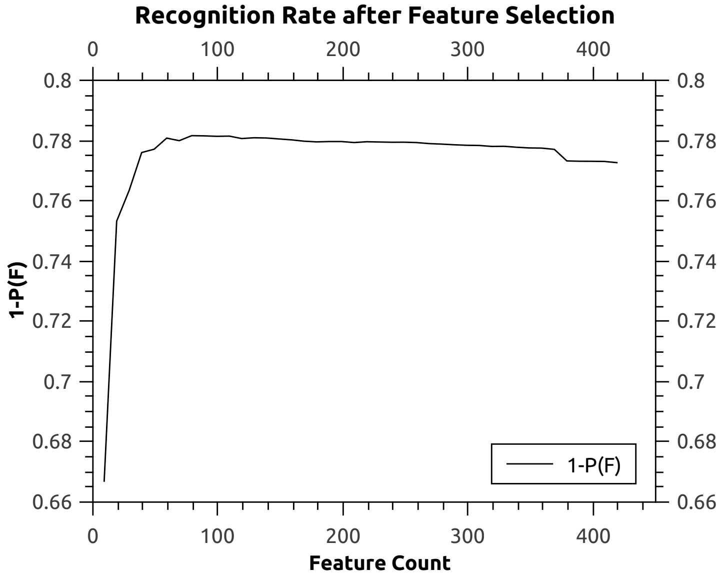

Figure 1: Recognition rate over selected features. Best trade-off with 59 features and recognition rate of 78%

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.

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 1 and 2).

Interestingly, our feature ranking also showed, that the H and S channel of the HSV data are quite important, which is also stated by [6]. More surprisingly, almost half of the original NIR features could be dropped in the remaining set of 59 features – even the values for talcum and lignin.

While [7] 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 parallelize further by hardware. When, however, exploiting many hundreds of visual features, more sophisticated data preprocessing steps need to be applied.

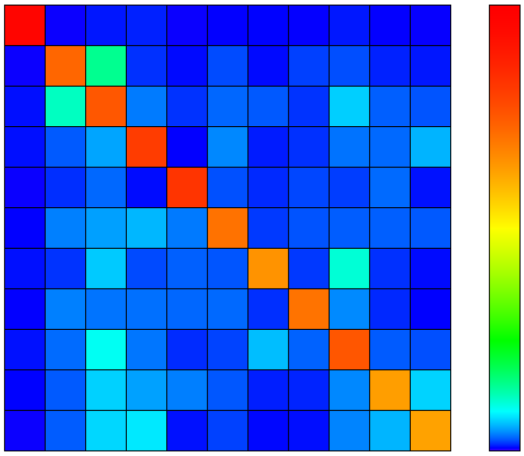

Figure 2: Visualization of the class error matrix F for best 59 features (see peak in figure 1). With i being the row index and j the column index, the elements F_{ij} are the number of samples from class i which are classified as class j. Low values are colored in blue, high values in red.

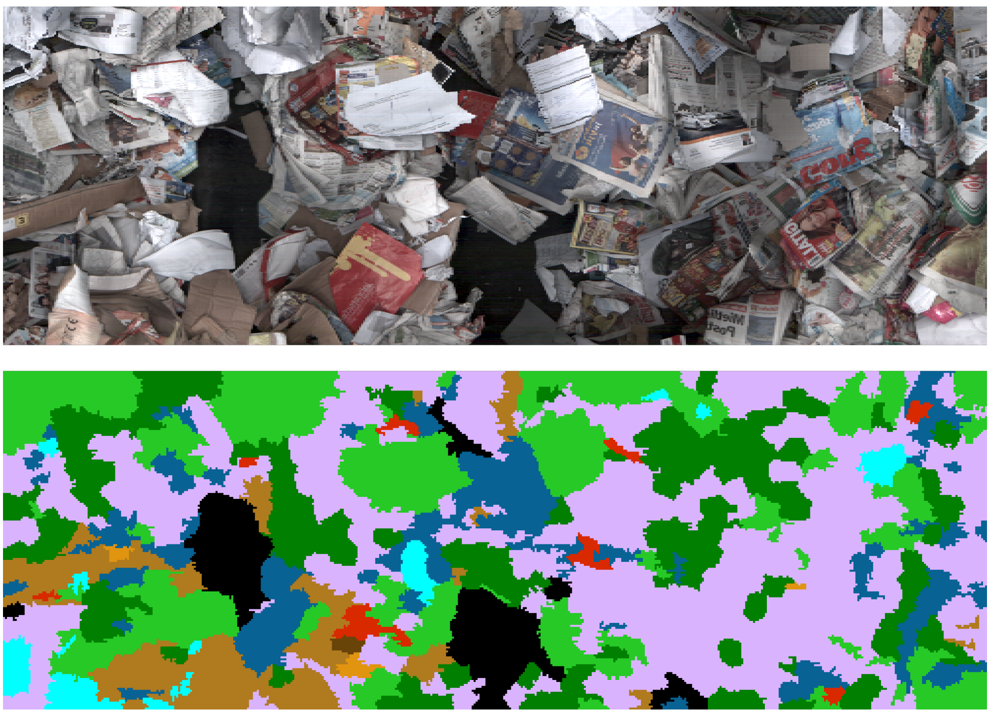

Figure 3: Example visualization of the classification results on real world data. The upper image shows the RGB data of a section of the conveyor belt. Each color in the lower image represents the recognized paper class. The background is colored in black.

6 Conclusion and Outlook

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

For the future, we plan to exploit the full visual resolution in order to capture finer structure details. 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 cost 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. 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 inflammables 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. Furthermore, we aim to extend our methods to other tasks like recognition of plastic materials and such.

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] 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.

[5] I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. New York: Springer–Verlag, 1986. ISBN: 0-387-96269-7.

[6] 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.

[7] 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.

[8] M. O. Rahman et al. “Waste paper grade identification system using window features”. In: *Journal of Computational Information Systems* 6.7 (July 2010), pp. 2077–2091. ISSN: 1553-9105.

[9] Verband Deutscher Papierfabriken e.V. *Facts about Paper*. Brochure. Accessed: 2015-11-30. 2015. URL: `http://www.vdp-online.de/en/papierindustrie/statistik`.