

# Using hybrid information of colour image analysis and SWIR-spectrum for high-precision analysis of construction and demolition waste

Petr Kuritcyn<sup>1</sup>, Katharina Anding<sup>1</sup>, Elske Linß<sup>2</sup>, and Gunther Notni<sup>1</sup>

<sup>1</sup> Ilmenau University of Technology, Faculty of Mechanical Engineering,  
Department of Quality Assurance and Industrial Image Processing,  
Gustav-Kirchhoff-Straße 2, 98693 Ilmenau

<sup>2</sup> Bauhaus-University Weimar, Faculty Civil Engineering, F. A. Finger  
Institute of Building Materials Science,  
Geschwister-Scholl-Straße 8, 99423 Weimar

**Abstract** This paper discusses the accuracy improvement of automatic analysis of construction and demolition waste (CDW) by using the combination of image analysis and spectral information. This means using the combination of methods of image processing, methods of spectral analysis and methods of supervised learning. The classification performances in colour images and also in SWIR-spectrums showed, that we have to use a combination of these two components in a combined feature vector to improve the accuracy of analysis. Investigations on hybrid information from colour images and SWIR-spectrums were done and compared with the separate usage of these information sources.

**Keywords:** Construction and demolition waste, machine learning, image processing, spectral analysis, hybrid information.

DOI: 10.58895/ksp/1000087509-6 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

Construction and demolition waste (CDW) are one of the biggest waste flow in Germany with around 68 million tons per year [1]. At the same time, the rate of the recyclable amount in producing of high-quality materials like recycling concrete and recycling asphalt is still relatively low, with 14 million tons per year (20.7%). For recycled masonry aggregates and recycled mixed aggregates, the lowest recycling rates are found because of the high heterogeneity and the mineral admixtures. Therefore, the desired reuse of these materials is very difficult. An important limitation factor is here the composition and heterogeneity of the material. The materials from the dismantling of buildings can be very heterogeneous in its material composition, depending on the origin, the method of dismantling and the processing. Recycled aggregates are characterized in the context of the quality assurance with respect to various structural parameters as well as the material composition. The latter is determined by a manual sorting analysis at relatively small sample sizes and is on the one hand very time-consuming and on the other hand strongly subjective. On the basis of the parameters and the material composition, the possibilities of utilization or possible fields of application are defined. The quality control is the prerequisite for obtaining the product status. Quality-controlled products have better chances on the market and thus lead to an image gain for the building material recycling industry and to a better utilization of given resources. This requires processes for the analysis of recycled aggregates, which provide precise, rapid and, above all, representative results. There are no method or device is developed or in common use, which allows a high performance recognition of all possible ingredients of CDW to realise an automatically and exactly analysis of CDW samples in context of the guidelines to reach a high quality product for building industry.

The results of own investigations by using image processing techniques in colour images [2], [3] and spectral analysis methods in SWIR-spectrums [4], [5] showed, that there is a possibility to improve the accuracy of analysis by combining these two information sources in one feature vector as basis for classification. So investigations in colour images, SWIR spectrums and combined hybrid-information were done and then compared with each other.

We used pre-engineered system from project “Autopetrographie” with a 3CCD-line-scan camera to capture colour images and the SWIR spectrometer Polytec PSS 2120 with an InGaAs detector and a range of 1100 to 2100 nm to record spectrums of the CDW classes. In our classification routine we used supervised machine learning classifiers [6], feature selection methods and principal component analysis (PCA).

Previous investigations on a separate colour image dataset showed a total recognition rate of 96.9% and on a separate spectral dataset of 99.2% by using an optimized support vector machine with polynomial kernel. We did know that we have to combine the image information with the spectral information to enhance the recognition accuracy, because the best achieved total recognition rate of 99.2% is not good enough for an automatically analysis of the material composition of CDW. So we combined the separate feature vectors to a hybrid dataset for training and testing our supervised classifier.

## 2 Analysis of image information of several CDW classes

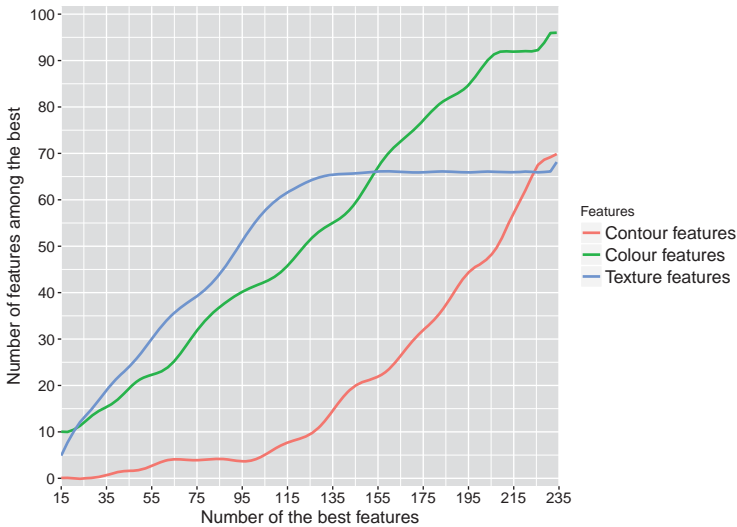
Like described in [3], we used the machine vision software HALCON to extract features from colour images. It results in feature vectors with more than 200 features (several contour, colour and texture features). Our previous investigations on colour images showed, that not all these features are relevant for classification and it's necessary to reduce the amount of redundant information by using feature selection methods. Several feature selection algorithms were tested and compared with each other.

All three feature selection methods (InfoGain, chiSquare, ReliefF) showed similar results. The application of ReliefF results in a smaller number of significant features in comparison to the two other methods, but results are better only for one class out of five. The InfoGain feature selection algorithm provide a good balance between accuracy and speed.

The classifiers svmPoly and MLP showed an underfitting by using a low number of features and an overfitting by using a high number of features. Recognition rates of classifiers are showing a plateau from 115 features. Further increase of the number of features resulted in a degradation of classification performance. It means that the first 115

features with the highest relevance are playing a crucial role for the recognition task.

Fig. 6.1 shows the number of specific features among the best. Texture and colour features are most frequently usable until 115 features (62 texture and 46 colour features among 115 best features based on InfoGain—filter, i.e. 94% of these features). It means that texture and colour feature are the most important features for recognition of CDW in images.



**Figure 6.1:** Number of specific features among the best, selected by InfoGain.

The application of principal component analysis on image dataset showed a degradation of results. Therefore, this method is not suitable for this task and leads to overcomplication of features.

### 3 Analysis of hybrid information of several CDW classes

In our investigations we used a dataset with similar size like in [5] (nearly 1100 samples out of the 8 superordinated classes: lightweight concrete, concrete, aerated concrete, sand-lime brick, dense and porous

brick, gypsum and granite) (see Table 6.1). The image acquisition and spectrum record were done parallel and as a result, the feature vector of objects consists of both, the image and the spectral information for the same object.

The dataset was used for training and testing of different classifiers: SVM with different kernels, neural networks, decision trees and logistic regression classifier.

The recognition rates of all classifiers on hybrid dataset are increased between 1.6% (for Random Forest) and 11% (for MLP) in comparison to separate usage of spectral information from this dataset. On the contrary, the recognition rates decreased between 0.6% (for LogitBoost) and 6.6% (for Random Forest) in comparison to separate usage of image information from this dataset. A similar tendency was found on dataset with 8 material classes: increasing between 0.4% (for svmPoly) and 10% (for MLP) in comparison to usage of spectral information and decreasing between 0.3% (for LogitBoost) and 7.7% (for Random Forest) in comparison to usage of image information. It shows, that spectral part of datasets consists irrelevant information, which leads to degradation of results.

**Table 6.1:** Number of objects in hybrid dataset.

Material	Number of objects in dataset
Concrete	155
Granite	25
Gypsum	57
Lightweight concrete	200
Aerated concrete	225
Sand-lime brick	199
Dense brick	75
Porous brick	105

Feature selection and feature transformation methods are required to solve this problem. Investigations in [5] showed, that it is necessary to use PCA for the transformation of spectrum to achieve higher accuracy of analysis. The application of 16 first principal components for classification results in reduction of computation time in comparison

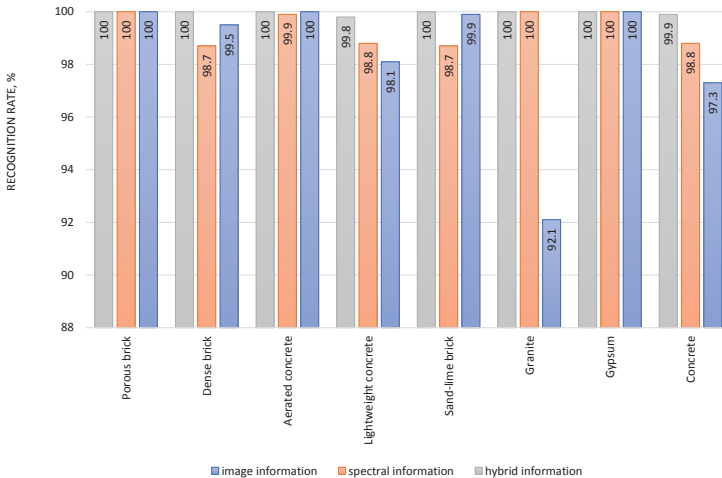
to using of all 501 wavelength specific features and leads to increase of classifier performance.

The classifier svmPoly showed the highest recognition rate of 99.7% on hybrid dataset with 8 material classes. The classifier MLP is close with total recognition rate of 99.6%.

The current investigations on the spectral part of hybrid dataset have confirmed previous results – the optimal number of principal components is between 14 and 18 PC's.

After optimization of spectral part, there is one more possibility to improve results. Image part of hybrid dataset can be optimized as well. Feature selection methods were used for this task.

Some classifiers like MLP and svmLinear showed weak dependency of performance from number of used features. Other classifiers like svmPoly and LogitBosst showed worse results by using low number of features (underfitting) and by using high number of features (overfitting). Classifier Random Forest showed worse results by using high number of features.



**Figure 6.2:** Comparison of the accuracy of the SVM with polynomial kernel for image, spectral and hybrid dataset (consists of 8 different material classes).

The two algorithms InfoGain and PCA were applied on the part of image information (120 features were selected) and spectral information (16 PC were selected) of hybrid dataset respectively. The best results showed the SVM classifier with polynomial kernel (Figure 6.2). In Figure 6.2 the comparison of the achieved accuracy of the SVM with polynomial kernel is shown for the image dataset, the spectral dataset and also for the combined hybrid dataset.

The achieved individual recognition rates based on the use of feature selection methods on the image feature vector and the principal component analysis on the spectral feature vector. In summary the individual recognition rates of the 8 classes achieved by SVMPoly for the hybrid dataset showed a very high recognition performance with nearly 100% (only for the classes lightweight concrete of 99.8% and for concrete of 99.9%).

## 4 Summary

The investigations on combined information (hybrid information) from colour images and SWIR-spectrums showed, that it's possible to improve the accuracy of analysis for the classes concrete, lightweight concrete, sand-lime brick and dense brick in comparison to the independent usage of the separate data. With application of the InfoGain feature selection on colour image part and PCA on spectral part of the data and using of a SVM classifier with polynomial kernel it's possible to achieve recognition rates of nearly 100% for all reviewed material classes. Most of them showed a 100% recognition rate due to relatively small dataset size. The results showed that a method for automatic recognition of all common CDW classes could developed, which allows a high performance recognition of all CDW classes to realise an automatically analysis of the material composition of CDW in context of the guidelines. It will be a reliable method for ensuring the quality of secondary building materials produced from construction and demolition waste.

Further investigations are planned for testing the method on a bigger dataset.

## References

1. "Mineral construction waste – Monitoring 2014. Kreislaufwirtschaft Bau," Berlin, Tech. Rep., 2013.
2. K. Anding, E. Linß, H. Träger, M. Rückwardt, and A. Göpfert, "Optical Identification of Construction and Demolition Waste by Using Image Processing and Machine Learning Methods," in *14th Joint Int. IMEKO Symp.*, Jena, Germany, 2011.
3. P. Kuritcyn, K. Anding, E. Linß, and S. M. Latyev, "Increasing the safety in recycling of construction and demolition waste by using supervised machine learning," in *IMEKO Joint Symposium 2014 TC1 - TC7 - TC13 „Measurement Science Behind Safety and Security”*, Madeira, Funchal, September 2014.
4. "NIR-Sorting of CDW. Final report of AiF scientific project. AIF-KF 2155712," Bauhaus-University Weimar, Tech. Rep., 2013.
5. K. Anding, P. Kuritcyn, E. Linß, and S. M. Latyev, "Significant characteristics in VIS- and IR-spectrum of construction and demolition waste for high-precision supervised classification," in *OCM 2015 - 2nd International Conference on Optical Characterization of Materials*, Karlsruhe, Germany, March 2015.
6. M. Kuhn, "The caret Package," 2009.