DOI: 10.58895/ksp/1000032143-18 erschienen in:
OCM 2013 - Optical Characterization of Materials - conference proceedings
DOI: 10.58895/ksp/1000032143 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000032143/

Understanding multi-spectral images of wood particles with matrix factorization

Mark Asbach[1], Dirk Mauruschat[2] and Burkhard Plinke[2]

[1] Fraunhofer IAIS,
Schloss Birlinghoven, 53754 Sankt Augustin
[2] Fraunhofer WKI,
Bienroder Weg 54 E, 38108 Braunschweig

Abstract Multispectral image data can be used to quantify the concentrations of chemical substances in material compounds by differential spectroscopy. In this paper, we describe Simplex Volume Maximization (SiVM), a matrix factorization method derived from Archetypal Analysis (AA), that is well suited to separate spectra. Exemplarily, we apply the technique to multispectral images of wood strands partially covered with adhesives and wood-polymer composites and show how to determine the concentration of the adhesives and how to distinguish the polymer types.

In the multispectral domain, our objective is to separate the spectral characteristics of the adhesives and polymers from those spectral components caused by variation in the natural wood, including differences in moisture.

Our experiments show that wood particles with different concentrations of adhesives or different polymer components can be distinguished after applying SiVM-based factorization to NIR spectral imaging. We therefore conclude that this technique has great potential for quality control applications that rely on multispectral imaging.

1 Introduction

Wood is an important raw material for the enterprises producing particle boards and other wood-based material like e.g. wood polymer composites (WPC). Because wood is an eco-friendly renewable material many efforts in research and development take place to reduce production costs [1] and to increase possibilities for recycling [2].

Spectral imaging in the near-infrared range (NIR) is one of many measurement techniques with great potential for classification and sorting processes [3]. However, the requirements are much higher than e.g. in recycling of plastic packages, because better resolution is needed and because the signals acquired by NIR cameras are superposed by optical scattering due to rough surfaces and statistical/temperature noise in the detector. The "classical" method to classify spectra using chemometric methods like linear filtering and principal component analysis (PCA) works but has its limitations [4]. Especially for wood particles improved methods would help to optimize wood products and to increase the recycling rate.

In addition to improved classification performance, a second motivation to use alternative methods results from the fact, that classical subspace transformation methods like PCA result in numerical representations of the data that have no physical meaning and are hard to interpret. In contrast, non-negative matrix factorization (NMF) has been shown to provide meaningful results, if the data are inherently non-negative [5]. But because the underlying problem is NP hard [6], optimal solutions are costly to obtain for real-world problems. Instead, recent extensions to NMF introduce additional constraints on the basis vectors to reduce the search space. Several of these extensions have been demonstrated successfully on hyperspectral image data for remote-sensing applications [7, 8].

In order to obtain a meaningful decomposition of the multispectral NIR imagery with low algorithmical complexity, we apply Archetypal Analysis (AA) [9] or rather its approximative implementation Simplex Volume Maximization (SiVM) [10]. SiVM requires efforts of only $O(kn)$ to derive basis functions and was shown to provide highly accurate reconstructions [11].

2 Wood particles

Two tasks are presented here as examples where the evaluation of multispectral image data could be improved by new approaches for factorization and classification:

2.1 Adhesive coverage of wood strands

Oriented strand boards (OSB) are made from big wood particles (e.g. 120mm $\times$ 25mm $\times$ 0.8mm) in automatic production lines. Emulgated adhesive (or resin) is sprayed onto the surface of the strands while they pass a rotating drum, then the strands are oriented and formed to a mat on a conveyor, pass a continuous hot press and leave it as particle boards. The board quality depends on many manufacturing conditions. An important one is the adhesive distribution on the strands before the mat enters the press.

But the adhesive is "visible" only in the NIR range by using spectral cameras and detection methods which are suitable also for the surface of an OSB mat. Figure 18.3 shows a scene with strands made from aspen wood and partially covered with urea-formaldehyde (UF) resin. The adhesive concentrations, based on the dry mass of wood, were 0% (definitely too low), 6% (good concentration) and 12% (too high because resin is an important cost factor). These concentrations are estimated from the amount of glue added to the rotary drum, but cannot be deduced from the visible light image.

2.2 Detection of different polymer types in WPC granulate

Wood polymer composites (WPC) consist of approx. 50 to 70 mass percent wood fibers and a polymer component, e.g. polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC), or a bio-based polymer like poly-L-lactic acid (PLLA). They are produced in a compounding/extrusion process as profiles and can substitute solid wood in applications as terrace deckings, windows, and door frames or car interior parts. Recycling of WPC has proven to be possible. However, methods for material management and especially for grading and sorting are not yet available, and therefore most of the material is only reused as fuel [12].

Developments for WPC sorting are ongoing but require a sorting or grading method for WPC granulate to make sure that the material stream is free of impurities and contains only one polymer component. The second example in Fig. 18.4 shows WPC granulate with four types of polymer components (from top to bottom: PLLA, PVC, PP, PE) which can not be distinguished in the visible light range.

3 Multispectral image decomposition

We interpret multispectral imaging as a discrete form of reflectance spectroscopy: a sample reflects light to a sensor that measures a discrete reflectance spectrum per pixel. A multispectral image with F spectral bands is then represented as a matrix $X = [x_1, x_2, \ldots, x_N]$ of N pixels $x_n \in \mathbb{R}^F$.

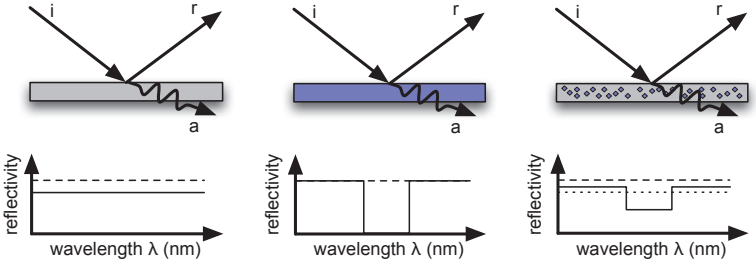
The run of the spectrum is characteristic of the chemical substance under investigation. For a mixture of different substances, the measured spectrum is a weighted sum of the individual spectra. If a pixel x_n shows a mixture of K chemical components with reflectance spectra $s_k \in \mathbb{R}^F$ and mixture weights w_{nk}, the pixel can be expressed as a linear combination

$$x_n = w_{n1}s_1 + w_{n2}s_2 + \cdots + w_{nK}s_K = SW_n, \tag{18.1}$$

with

$$w_{nk} \geqslant 0, \sum_K w_{nk} = 1.$$

This is schematically depicted in Fig. 18.1: A pure substance A shall have a flat spectral reflectivity of 0.8, thus reflecting 80 % of spectrally white light, while a second, pure substance B shall have a spectral reflectivity of 1.0 over all but a given spectral band, where it has a reflectivity of 0. When measuring the spectral reflectivity of a compound material consisting of 60% of substance A plus 40% of substance B, we expect a combined spectral reflectivity that is the linear combination of the pure spectra, weighted with the respective lots of the substances. For our assumed compound substance, we would therefore expect a spectral reflectivity of $0.88 = 0.6 \cdot 0.8 + 0.4 \cdot 1.0$ over most of the spectrum and



(a) Pure substance A (b) Pure substance B (c) Mixture of A and B

Figure 18.1: We consider three different, thick materials with diffuse surface scattering, where incident light i is split into a fraction a that is absorbed and a second part r that is reflected. The spectral reflectivity of a material mixture with 60% of substance A and 40% of substance B is considered to be the weighted sum of the spectral reflectivities of the pure substances.

$0.48 = 0.6 \cdot 0.8 + 0.4 \cdot 0$ over the band where substance B is fully absorbent.

4 Archetypal analysis

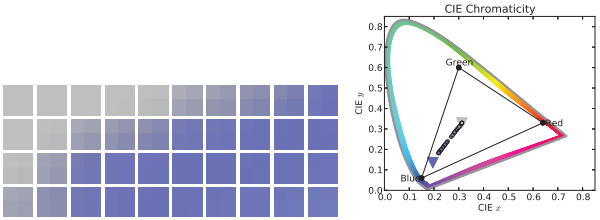
The spectra s_k of simple chemical substances are known, but natural materials like wood exhibit mixtures of a high number of components that can be learned from sample data only. Analogous to Equation 18.1, we approximate a multispectral image X with N pixels from K spectral components $s_k \in \mathbb{R}^F$ and K weights $w_k \in \mathbb{R}^F$ as

$$X \approx S \cdot W, \tag{18.2}$$

with

$$S = [s_1, s_2, \ldots, s_K], W = [w_1, w_2, \ldots, w_K]^T$$

with the approximation error $E = \min ||X - SW||^2$. While non-negative matrix factorization (NMF) provides a solution to Equation 18.2 satisfying the physical requirements for the spectra s_k, convex-NMF [13] and



(a) Exemplary color image (b) Color gamut

Figure 18.2: Color mixing example: All pixel colors in (a) can be composed as linear combination of two archetypal colors light blue and light gray (triangular color entries in (b)). They form the convex hull (here: a line) enclosing all other pixel colors.

convex-hull-NMF [14] further guarantee meaningful weights w_k satisfying the requirements for w_{nk} from Equation 18.1.

The resulting spectra s_k typically coincide with actual data points x_n, which makes convex-hull-NMF representations readily interpretable: If an individual image location n' exist, at which only a single substance k is present, the discrete spectrum of this substance $s_k = x_{n'}$ resides on the convex hull of all data points x_n.

For illustrational purposes, we can interpret RGB colors as a colorspace with three spectral bands. Figure 18.2 depicts an image containing various colors mixed from light blue and light gray. As can be seen from the gamut diagram, these colors represent data points in a subspace of the full RGB spectrum. The "pure" colors light blue and light gray reside on the convex hull. Given an image like the one depicted in Fig. 18.2(a), our goal is to find the "pure" colors and to unmix the colors of all pixel spectra. Archetypal Analysis (AA) is a method that selects suitable data points as basis functions for the above mentioned convexity-constrained matrix factorization techniques. We use its speeded up derivate—Simplex Volume Maximization (SiVM)—to quickly identify archetypal datapoints in our multispectral image data. Then, all other pixels of the image can be approximated by linear combinations of the archetypes. By definition, all resulting coefficients are of the range $[0, 1]$ and we can interpret them as relative amount of "pure" ingredients (archetypes) used to "mix" a certain pixel's spectrum.

5 Results

We apply the Archetypal Analysis to two material characterization problems from the woodworking industry.

5.1 Adhesive coverage of wood strands

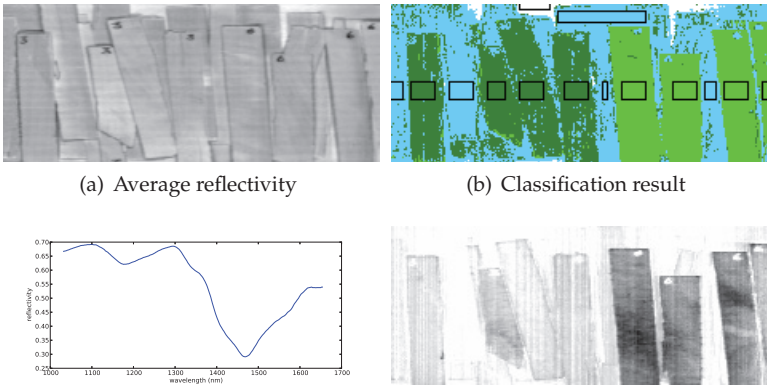
In our first experiment, we seek to analyze the glue coverage of wood strands. Data were acquired with an InGaAs line scan camera that records 316 spectral bands in the near infrared (NIR) range between 1032 nm and 1656 nm. The image depicted in Fig. 18.3 shows about twenty strands of aspen wood. Four strands are covered with $\approx$ 6%, four other strands are covered with $\approx$ 12% of adhesive.

With Simplex Volume Maximization, 15 archetypes were extracted from the pixel spectra. We expect to need several different archetypes to model the spectra of the wooden texture, the spectrum of the adhesives, moisture, image background, and the spectrum of the graphite used to mark the strands. Using a slightly higher number of archetypes allows to model shadows, specular highlights, and noise as well. Figure 18.3(c) and Fig. 18.3(d) show archetype s_{13} and the corresponding mixture weights w_{13}, that seem to model the absorption spectrum of wood covered with $\approx$ 12% adhesives.

The 15-dimensional SiVM space can be used to classify image pixels based on the adhesives coverage. Exemplarily, we have marked small regions with known (compare section 2) glue coverage as training data for a simple nearest-neighbor classifier (light green is used for 12% and dark green for 6%). In addition, areas showing strands not covered with adhesives were marked with light blue and visible background marked in light grey. The trained classifier was then used to predict the glue coverage of all other image pixels. Classification result and training data (marked by boxes) are depicted in Fig. 18.3(b).

5.2 Detection of different polymer types in WPC granulate

In a second experiment, we use the same method to disambiguate pellets made from wood fibres and different sorts of polymers. WPC pellets with four different polymer components (PLLA, PVC, PP, and PE



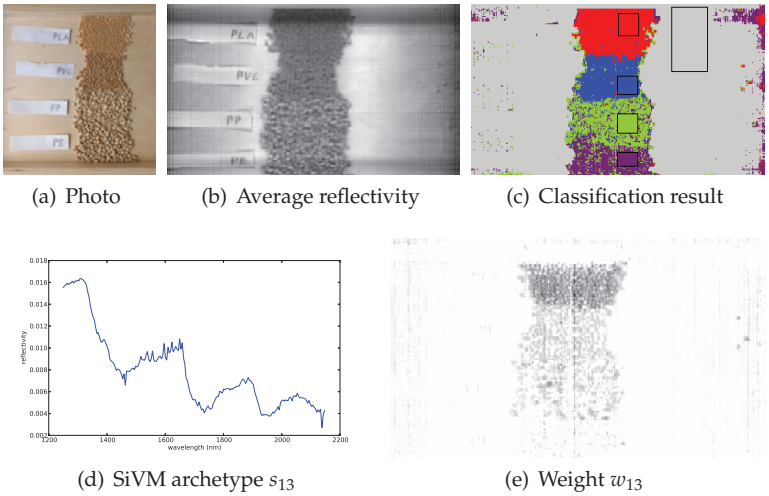
(a) Average reflectivity (b) Classification result

(c) SiVM archetype s_{13} (d) Weight w_{13}

Figure 18.3: Wood strands, partially covered with adhesives. The adhesive is transparent in visible light as well as in the averaged NIR reflectivity (a). After decomposition with SiVM into 15 basis functions and training a nearest-neighbor classifier (training regions marked with black boxes), it can be predicted for the whole image (b). An exemplary basis function (c) with corresponding weight (d) seems to model adhesive concentration quite well.

respectively) have been placed in a wooden box together with paper labels. The wooden box has walls that cast shadows on part of the scenery and the whole setup is far from an ideal laboratory environment but nevertheless closer to industrial conditions. We chose it to illustrate the resilience of our method against adverse data acquisition conditions. The image data was captured with an Extended-InGaAs line scan camera with a spectral resolution of 248 bands in the range of 1161 nm to 2262 nm, of which the 20 lowest and 25 highest wavelengths were discarded because of extremely low signal-to-noise ratio.

Again, SiVM with 15 archetypes was used in combination with a nearest-neighbor classifier to learn the characteristics of the different plastics from some image pixels and predict it for the remainder of the image. Figure 18.4 depicts the wood-plastic-compound (WPC) dataset, including classification result and an exemplary archetype. Archetype 13—depicted in Fig. 18.4(d)—shows contributions in the area of PLLA pellets.



(a) Photo (b) Average reflectivity (c) Classification result

(d) SiVM archetype s_{13} (e) Weight w_{13}

Figure 18.4: Composite pellets made from PLLA, PVC, PP, and PE, in a wooden box. From a conventional color photo (a) or the averaged NIR reflectivity (b), polymer components can not be estimated. The decomposition of the multispectral NIR image with SiVM into 15 basis functions, however, allows classification even with a simple nearest-neighbor classifier. Training data (black boxes) and result are shown in (c). An exemplary basis function is depicted together with its corresponding weights in (d) and (e).

The nearest-neighbor classification result was obtained by selecting per class one rectangular image region as training samples. The wooden box was learned from a slightly larger region. Overall, this—really simple—attempt already results in acceptable classification performance. Errors are only visible in shadowed areas. As the pellets themselves generate small shadows, classification errors looking like speckle noise are observed especially in the area of PP and PE pellets.

6 Summary

In this paper, we have demonstrated the application of Archetypal Analysis (AA) / Simplex Volume Maximization (SiVM), a matrix fac-

torization method, to multi-spectral image analysis for wooden materials. It was shown that adhesive concentrations on wood strands and polymer types in WPC granulates can be distinguished with simple classification algorithms using the image decompositions derived from SiVM/AA.

The proposed method allows for a better understanding of the multispectral image decomposition than standard methods from chemometry. And while the latter are usually sensitive to outliers, surface and acquisition conditions rather than to the chemical composition of the material, the proposed approach appears to deliver predictable results, invariant to small changes in the initialization.

An objective for future research will be to quantitatively evaluate the algorithms—shown here as a prove-of-concept—on a larger set of image data and to optimize the computational performance.

References

1. J. Aderhold and B. Plinke, "Innovative methods for quality control in the wood-based panel industry," in *Wood-based Panels: An Introduction for Specialists*, H. Thoemen, M. Irle, and M. Sernek, Eds. London: Brunel Univ. Press, 2010, pp. 225–249.
2. P. Meinlschmidt and B. Plinke, "Optical infrared detection of contaminations in recovered wood," in *Sensor Based Sorting*, Aachen, 2012.
3. S. Wendel and B. Plinke, in *7. Kolloquium Prozessanalytik*, Linz.
4. B. Plinke and D. Ben Yacov, "Detection of adhesives on wood surfaces: Spatially resolved monitoring of adhesive application," in *Adhesion Adhesives & Sealants*, vol. 4, 2010, pp. 25–29.
5. D. Lee and S. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
6. S. Vavasis, "On the Complexity of Nonnegative Matrix Factorization," *SIAM J. on Optimization*, vol. 20, no. 3, pp. 1364–1377, 2009.
7. L. Miao and H. Qi, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.
8. J. Nascimento and J. B. Dias, "Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 4, pp. 898–910, 2005.

9. A. Cutler and L. Breiman, "Archetypal Analysis," *Technometrics*, vol. 36, no. 4, pp. 338–347, 1994.
10. C. Bauckhage and C. Thurau, "Making Archetypal Analysis Practical," in *Pattern Recognition*, ser. LNCS, J. Denzler and G. Notni, Eds., vol. 5748. Springer, 2009, pp. 272–281.
11. C. Thurau, K. Kersting, and C. Bauckhage, "Yes We Can – Simplex Volume Maximization for Descriptive Web-Scale Matrix Factorization," in *Proc. ACM CIKM*, 2010.
12. C. Gahle, "Wpc: Recyclingfrage noch nicht abschließend geklärt," *Holz-Zentralblatt*, pp. 253–254, 2009.
13. C. Ding, T. Li, and M. Jordan, "Convex and Semi-Nonnegative Matrix Factorizations," *IEEE Trans. on Pattern Analalysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2010.
14. C. Thurau, K. Kersting, and C. Bauckhage, "Convex Non-Negative Matrix Factorization in the Wild," in *Proc. IEEE ICDM*, 2009.