ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Processing Hyperspectral Data in Machine Learning

T. Villmann[1], M. Kästner[1]*, A. Backhaus[2] and U. Seiffert[2]

1- University of Appl. Sciences Mittweida - Dept. of Mathematics
Mittweida, Saxonia - Germany

2- Fraunhofer IFF Magdeburg - Dept. Biosystems Engineering
Magdeburg, Germany

Abstract. The adaptive and automated analysis of hyperspectral data is mandatory in many areas of research such as physics, astronomy and geophysics, chemistry, bioinformatics, medicine, biochemistry, engineering, and others. Hyperspectra differ from other spectral data that a large frequency range is uniformly sampled. The resulting discretized spectra have a huge number of spectral bands and can be seen as good approximations of the underlying continuous spectra. The large dimensionality causes numerical difficulties in efficient data analysis. Another aspect to deal with is that the amount of data may range from several billion samples in geophysics to only a few in medical applications. In consequence, dedicated machine learning algorithms and approaches are required for precise while efficient processing of hyperspectral data, which should include also expert knowledge of the application domain as well as mathematical properties of the hyperspectral data.

1 Introduction

Spectral data play a key role in many areas of theoretical and applied research. Among them are physics, earth sciences, biochemistry, life-sciences and medicine, where the analysis of hyperspectra are essential [16, 58, 70]. During the last years, the resolution of measurement equipment and scanners has drastically improved [17, 42, 84]. Thus, scanners with a wide range of spectral information obtained for a single measurement are available. Multispectral scanners sample the frequency range using a few spectral channels with wide bandpasses. Hyperspectral data differ from multispectral data by narrowly spaced and uniformly sampled bandpasses with a huge number of bands. The typical vectorial representation of the spectra causes serious numerical problems: Theoretically, because of the large data dimension a huge number of data samples is required for representative data space sampling. The amount of available data may range from up to several millions samples in geophysics and astronomy to only a few in medical applications. For analysis of such data, standard techniques like multivariate statistical data analysis [22, 60], support vector machines and statistical learning [21, 67, 95], as well as neural network methods [28, 74] have been used. In this paper we give an overview about recent developments and challenges in hyperspectral data analysis (HDA) in the context of machine learning approaches emphasizing the particular characteristics of hyperspectral data.

*M.K. is supported by a grant of the ESF, Saxony, Germany.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

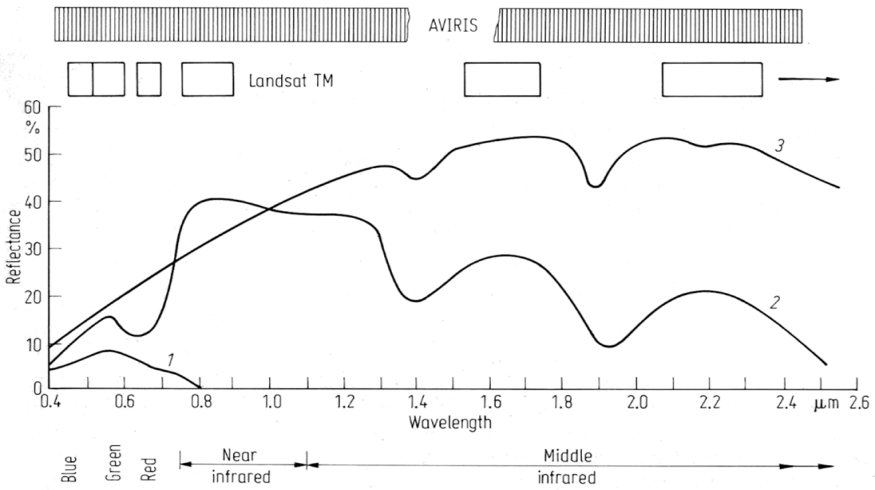

Figure 1: Comparison of multispectral and hyperspectral data sensors. Visualized examples are from satellite remote sensing (adapted from [14]).

2 Characteristics of Hyperspectral Data

Frequently, hyperspectral data are given in vectorial form $\mathbf{v} = (v_1, \ldots, v_n)^\top$ with typically large n, i.e. a hyperspectral vector may contain up to thousands of dimensions. The dimensions are also denoted as spectral bands in this context. Hyperspectral data can be distinguished from multispectral data in a way that hyperspectral scanners provide a uniform representation of the spectral range, for example see Fig. 1.

Due to this characteristics, hyperspectral vectors can be seen as discrete representations (approximations) of continuous spectra $v_i = \varphi(\omega_i)$ with φ being a continuous function of a frequency/wavelength value ω_i. Even through hyperspectral data have a large data dimension, the intrinsic dimension (Hausdorff dimension) is usually much lower than the number of data dimensions, because the spectral bands are highly correlated. The Hausdorff dimension can be estimated with several methods [12, 13, 25, 38, 79, 6]. Hence, functional data analysis (FDA), specifically optimized for hyperspectral data analysis can be applied [27, 77, 57]. One key point in FDA is that the functional data vectors may be linearly represented by a convex linear combination

$$\varphi(\omega) \sim \sum_{k=1}^{N} \gamma_k \cdot \psi_k(\omega, \alpha_k) \tag{1}$$

of so-called basis functions $\psi_k(\omega, \alpha_k)$ specified by a parameter vector α_k. In this linear view, data analysis can be done in the coefficient data space formed by the vectors $\gamma = (\gamma_1, \ldots, \gamma_N)^\top$. Examples of basis functions are radial basis functions (rbf), the logistic function, Gaussians or Lorentzians. Piecewise linear functions and related norms are considered in [10, 29, 77]

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Further assumptions on properties of spectral data can easily be verified: The spectra can be mathematically treated as *positive* functions $\varphi(\omega) \geq 0$, which are, at least for its discrete realizations, bounded. These functions are denoted as positive measures. A special subset of positive measures are probability functions. For both function types, appropriate functional dissimilarity measures can be considered: Information theoretic learning is based on divergences [18, 56, 86]. Many divergences can be extended such that they are applicable also for positive measures [87]. Functional norms like Sobolev norms can be applied if the spectral functions $\varphi(\omega)$ are additionally supposed to be differentiable [85]. The Sobolev norm of a spectral function can be written in the form $\|\varphi\|_{K,p}^{S} = \|\varphi\|_p + \sum_{j=1}^{K} \|D_j[\varphi]\|_p$ where $\|\bullet\|_p$ is the $\mathcal{L}_p$-Lebesgue-norm and D_j is the differential operator of order j [32]. Hence, Sobolev norms pay attention to spatial correlations in the frequency domain. Correlation measures also give an alternative to the frequently inappropriate Euclidean distance [94, 78, 35]. Correlation measures can also be applied to line or peak spectra because they do not require any spatial information.

Principal component analysis (PCA, [31]) is a standard technique for dimensionality reduction and data compressing. It can be extended to functional PCA which is a standard PCA in the coefficient data space using the linear decomposition (1) [77]. Standard PCA in the original data space may become computationally crucial due to the high data dimensionality blowing up the required covariance matrix. If only a few principal components are sufficient for data description, adaptive Hebbian learning is an alternative, which only implicitly uses the information of the covariance matrix [54, 63]. Originally, the method was developed for Euclidean space, Sobolev-Hilbert spaces are considered in [88]. Recently, this method was further extended to be applicable for non-Euclidean spaces, such as $\mathcal{L}_{p=1}$-Lebesgue-normed spaces, kernel spaces, and Sobolev spaces [8].

3 Machine Learning Approaches for Hyperspectral Data Analysis in Astronomy and Geosciences

One of the most promising fields of HDA applications are in satellite or airborne remote sensing image analysis [16, 41, 58, 70]. Spectra are collected by remote sensing, from telescopes, air- or spacecrafts, and by robots. Visualization of this information as well as knowledge extraction and data mining are challenging primary tasks, that become increasingly attractive for machine learning approaches with improved and accelerated hardware for fast processing [75]. Applications can be found in all kind of geosciences, agriculture, etc.

Depending on the considered object and region as well as the spectral measurement method (near-infrared - NIR, and thermal infrared -TIR), subtle differences in local wavelengths (bands) may provide substantial information to distinguish the organic or anorganic material of the observed surface area. The underlying physical process that determines the spectral shape is the preferential interaction of light with different materials at different wavelengths. Materials can have multiple absorption features, each of which may be very narrow or

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

quite wide. Therefore, investigations based on principal component analysis frequently fail [92]. Neural networks and machine learning methods may offer alternatives [7, 23, 48, 51, 81]. However, the mentioned features contribute to difficulties for adequate data modeling and processing also in machine learning and computational intelligence. Otherwise, precise analysis of synthetic spectral data may help to develop successful variants of known algorithms specifically designed for a given problem [11].

One of the most prominent vector quantizers is the self-organizing map (SOM, [37]). Beside its vector quantization abilities, the property of topographic mapping makes SOMs an appropriate tool for visualization in remote sensing [6, 49, 91]. Precise data analysis by means of SOMs requires additional efforts like magnification control as well as SOM-based cluster and separation analysis [50, 45, 47, 80, 83]. An alternative vector quantizer to SOMs is neural gas (NG, [43]), which frequently yields better quantization results and is, therefore, well-suited also for hyperspectral data analysis [71]. An unsupervised multi-view feature extraction for dimensionality reduction using the specific data structure in image cubes is proposed in [93]. Another successful alternative to SOMs are ART maps [15] for clustering and novelty detection.

Supervised classification can be realized using multi-layer perceptrons or support vector machine as powerful adaptive classifiers [22, 28, 67, 76]. An overview in the context of hyperspectral imaging can be found in [17, 84]. An alternative to these approaches are variants of learning vector quantization (LVQ, [37]), which extend the basis algorithm and are specifically adapted to high-dimensional problems with subtle features in spectra to distinguish the classes. These extensions include relevance learning for weighting and extracting those bands that are important for classification [26, 44]. A further extension of relevance learning is matrix learning taking into account the correlations between bands [68, 69]. Related to classification problems is spectral unmixing of components, which are comprised in the spectral signature of a single pixel covering a spatial area in the image. Unmixing is also known as the problem of automatic endmember detection. Unmixing allows the estimation of physical parameters of the observed material from their complex spectral shapes. Commonly, this problem is savaged using convex optimization tools, however, restricted to only a few components due to numerical instabilities [46, 48]. Neural network alternatives, also applicable for a serious number of components, are proposed in [24, 55, 90].

4 Machine Learning Techniques for Hyperspectral Data Computational Biology, Medicine and Related Fields

Hyperspectral data in biology, medicine, and life sciences become more and more attractive for precise non-invasive analysis of objects, plants/vegetation, crops, meat, etc. [3, 4, 2, 72]. Hyperspectral data are obtained from several measurement techniques like mass spectrometry (MS), nuclear magnetic resonance spectroscopy (MNR), NIR and other. The amount of data in typical applications tends to be still quite large but limited compared to astrophysical applications. Particular challenges arise in medicine, where often only a few

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

samples are available. Further challenges in biomedicine include topics like diversity and inconsistency of biological data, unresolved functional relationships within the data, imbalanced data, or large variability. Yet, machine learning methods seem to be successfully contributing to the solution of these problems in hyperspectral data analysis [73] as well.

In comparison to remote sensing, here additional difficulties occur due to the application area: the raw spectra are usually contaminated with high-frequency noise and systematic baseline disturbances. Additionally, the alignment of spectra, i.e., a frequency shifting, is necessary to remove the inaccuracy of the instruments [1, 65]. Even preprocessed hyperspectra often still remain high-dimensional such that dimensionality reduction is required. Although the spectra are functional vectors, an underlying smooth differentiable function cannot always be assumed, such that the above outlined decomposition methods, see eq. (1), are not applicable for dimensionality reduction. Here, specific procedures and heuristics have to be applied for dimensionality reduction including detailed knowledge about the data. For example, generating informative peak lists for MS-spectra is highly non-trivial [66, 64]. Information theoretic methods for feature extraction are investigated in [29, 40, 62]. Regularization techniques may help to achieve sparseness in data representation [10, 36, 59, 82] Denoising using wavelet decomposition together with PCA for hyperspectral data is studied [9] to reflect the spatial as well as the spectral relations in the data within the denoising procedure. Independent component analysis (ICA, [19, 20, 30]) is considered in [39, 61].

If the resolution of the spectral data is not too high, i.e., if the dimension of the functional data vector is moderate, then a processing without complexity reduction may become feasible. Otherwise, parallelization of algorithms may become attractive and promising [5]. Of course, the functional aspect of the data should be kept in mind, although standard techniques like the above mentioned neural networks may be successfully applied. Functional metrics like Sobolev norms or divergences for dissimilarity estimation and vector quantization of hyperspectral data are applied in [52, 53]. Relevance learning in classification by LVQ methods taking into account the differentiability of the hyperspectra to obtain smooth relevance profiles are considered in [33, 34, 89].

Generally, the integration of expert knowledge beside the functional aspect of hyperspectra is still underestimated although promising in other fields of research [34]. This could include knowledge about hierarchies in data and data classes or asymmetric classification costs, the latter one to reflect the problem of sensitivity and specificity. Classification learning including those assumptions and restrictions would provide new perspectives. These may become particularly attractive in bio-/life-sciences applications, because of the limited number of available data samples in comparison to the huge dimensionality of hyperspectra.

5 Conclusion

In this tutorial paper we discussed some new trends and developments in machine learning of hyperspectral data. We emphasize that hyperspectra should

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

be treated as functional data taking into account their specific characteristics. In particular, the machine learning methods should deal with the inherent correlations in the vectors directly or use adequate preprocessing to achieve a faithful analysis. Additionally, structural expert knowledge should be integrated to reduce the complexity of the problems such that more precise results can be obtained. This becomes especially important, if only a few spectra are available due to the application area like in biomedicine, for example. Without any claim of completeness, we highlighted several new possibilities developed during the last years for an appropriate handling of hyperspectral data.

References

[1] J. B. Adams, M. O. Smith, and A. R. Gillespie. Imaging spectroscopy: Interpretation based on spectral mixture analysis. In C. Peters and P. Englert, editors, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pages 145–166. Cambridge University Press, New York, 1993.

[2] A. Backhaus, P. C. Ashok, B. B. Praveen, K. Dholakia, and U. Seiffert. Classifying Scotch Whisky from near-infrared Raman spectra with Radial Basis Function Network with Relevance Learning. In M. Verleysen, editor, *Proceedings of the 20. European Symposium on Artificial Neural Networks ESANN 2012*, pages 411–416, Evere, Belgium, 2012. D-Side Publications.

[3] A. Backhaus, F. Bollenbeck, and U. Seiffert. High-throughput quality control of coffee varieties and blends by artificial neural networks and hyperspectral imaging. In *Proceedings of the 1st International Congress on Cocoa, Coffee and Tea, CoCoTea 2011*, pages 88–92, 2011.

[4] A. Backhaus, F. Bollenbeck, and U. Seiffert. Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. In *Proceedings of the 3rd IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing WHISPERS 2011*, page 9. IEEE Press, 2011.

[5] A. Backhaus, J. Lachmair, U. Rückert, and U. Seiffert. Hardware accelerated real time classification of hyperspectral imaging data for coffee sorting. In M. Verleysen, editor, *Proceedings of the 20. European Symposium on Artificial Neural Networks ESANN 2012*, pages 627–632, Evere, Belgium, 2012. D-Side Publications.

[6] H.-U. Bauer and T. Villmann. Growing a Hypercubical Output Space in a Self–Organizing Feature Map. *IEEE Transactions on Neural Networks*, 8(2):218–226, 1997.

[7] J. A. Benediktsson, P. H. Swain, and et al. Classification of very high dimensional data using neural networks. In *IGARSS'90 10th Annual International Geoscience and Remote Sensing Symp.*, volume 2, page 1269, 1990.

[8] M. Biehl, M. Kästner, M. Lange, and T. Villmann. Non-euclidean principal component analysis and Oja's learning rule – theoretical aspects. In P. Estevez, J. Principe, and P. Zegers, editors, *Advances in Self-Organizing Maps: 9th International Workshop WSOM 2012 Santiage de Chile*, volume 198 of *Advances in Intelligent Systems and Computing*, pages 23–34, Berlin, 2012. Springer.

[9] F. Bollenbeck, A. Backhaus, and U. Seiffert. A multivariate Wavelet-PCA denoising-filter for hyperspectral images. In *Proceedings of the 3rd IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing WHISPERS 2011*, page 25. IEEE Press, 2011.

[10] M. Boullé. Functional data clustering via piecewise constant nonparametric density estimation. *Pattern Recognition*, 45:4389–4401, 2012.

[11] B. Bue, E. Merényi, and B. Csathó. An evaluation of class knowledge transfer from synthetic to real hyperspectral imagery. In *Proceedings of the 3rd IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing WHISPERS 2011*, page 9. IEEE Press, 2011.

[12] F. Camastra and A. Colla. Data dimensionality estimation methods: a survey. *Pattern Recognition*, 36:2945–2954, 2003.

[13] F. Camastra and A. Vinciarelli. Intrinsic dimension estimation of data: an approach based Grassberger-Procaccia's algorithm. *Neural Processing Letters*, 14:27–34, 2001.
[14] N. W. Campbell, B. T. Thomas, and T. Troscianko. Neural networks for the segmentation of outdoor images. In *Solving Engineering Problems with Neural Networks. Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN'96). Syst. Eng. Assoc, Turku, Finland*, volume 1, pages 343–6, 1996.
[15] G. A. Carpenter, M. N. Gjaja, and et al. ART neural networks for remote sensing: Vegetation classification from Landsat TM and terrain data. *IEEE. Trans. Geosci. and Remote Sens.*, 35(2):308–325, 1997.
[16] C. Chang. *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer, 2003.
[17] C. Chang. *Hyperspectral Data Processing: Algorithm Design and Analysis*. Whiley, 2013.
[18] A. Cichocki and S.-I. Amari. Families of alpha- beta- and gamma- divergences: Flexible and robust measures of similarities. *Entropy*, 12:1532–1568, 2010.
[19] A. Cichocki, R. Zdunek, A. Phan, and S.-I. Amari. *Nonnegative Matrix and Tensor Factorizations*. Wiley, Chichester, 2009.
[20] P. Comon and C. Jutten. *Handbook of Blind Source Separation*. Academic Press, 2010.
[21] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.
[22] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.
[23] M. Grana and R. Duro, editors. *Computational Intelligence for Remote Sensing*. Springer, 2008.
[24] M. Grana and J. Gallego. Associative morphological memories for spectral unmixing. In *Proc. 12th European Symposiumon Artificial Neural Networks (ESANN 2003), Bruges, Belgium*, Brussels, April 28-30 2003.
[25] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica*, 9D:189–208, 1983.
[26] B. Hammer and T. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15(8-9):1059–1068, 2002.
[27] T. Hastie and W. Stuetzle. Principal curves. *J. Am. Stat. Assn.*, 84:502–516, 1989.
[28] S. Haykin. *Neural Networks. A Comprehensive Foundation*. Macmillan, New York, 1994.
[29] G. Hébrail, B. Hugueney, Y. Lechevallier, and F. Rossi. Exploratory analysis of functional data via clustering and optimal segmentation. *Neurocomputing*, 73:1125–1141, 2010.
[30] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. J. Wiley & Sons, 2001.
[31] I. Joliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.
[32] I. Kantorowitsch and G. Akilow. *Funktionalanalysis in normierten Räumen*. Akademie-Verlag, Berlin, 2nd, revised edition, 1978.
[33] M. Kästner, B. Hammer, M. Biehl, and T. Villmann. Functional relevance learning in generalized learning vector quantization. *Neurocomputing*, 90(9):85–95, 2012.
[34] M. Kästner, W. Hermann, and T. Villmann. Integration of structural expert knowledge about classes for classification using the fuzzy supervised neural gas. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2012)*, pages 209–214, Louvain-La-Neuve, Belgium, 2012. i6doc.com.
[35] M. Kästner, M. Strickert, D. Labudde, M. L. L. Haase, and T. Villmann. Utilization of correlation measures in vector quantization for analysis of gene expression data - a review of recent developments. *Machine Learning Reports*, 6(MLR-04-2012):5–22, 2012. ISSN:1865-3960, http://www.techfak.uni-bielefeld.de/~fschleif/mlr/mlr_04_2012.pdf.
[36] M. Kästner, T. Villmann, and M. Biehl. About sparsity in functional relevance learning in generalized learning vector quantization. *Machine Learning Reports*, 5(MLR-03-2011):1–12, 2011. ISSN:1865-3960, http://www.techfak.uni-bielefeld.de/~fschleif/mlr/mlr_03_2011.pdf.
[37] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).
[38] A. Kraskov, H. Stogbauer, and P. Grassberger. Estimating mutual information. *Physical Review E*, 69(6):66–138, 2004.
[39] C. Krier, F. Rossi, D. François, and M. Verleysen. A data-driven functional projection approach for the selection of feature ranges in spectra with ICA or cluster analysis.

Chemometrics and Intelligent Laboratory Systems, 91:43–53, 2008.

[40] C. Krier, M. Verleysen, F. Rossi, and D. François. Supervised variable clustering for classification of NIR spectra. In M. Verleysen, editor, *Proceedings of 16th European Symposium on Artificial Neural Networks (ESANN)*, pages 263–268, Bruges, Belgique, 2009.

[41] D. Landgrebe. *Signal Theory Methods in Multispectral Remote Sensing*. Wiley, Hoboken, New Jersey, 2003.

[42] T. Lillesand, R. Kiefer, and J. Chipman. *Remote Sensing and Image Interpretation*. Wiley, 6th edition edition, 2008.

[43] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten. 'Neural-gas' network for vector quantization and its application to time-series prediction. *IEEE Trans. on Neural Networks*, 4(4):558–569, 1993.

[44] M. Mendenhall and E. Merényi. Relevance-based feature extraction for hyperspectral images. *IEEE Transactions on Neural Networks*, 19(4):658–672, 2008.

[45] E. Merenyi. "precision mining" of high-dimensional patterns with self-organizing maps: Interpretation of hyperspectral images. In P. Sincak and J. Vascak, editors, *Quo Vadis Computational Intelligence? New Trends and Approaches in Computational Intelligence (Studies in Fuzziness and Soft Computing, Vol. 54*. Physica-Verlag, ?, 2000.

[46] E. Merényi, W. Farrand, J. Taranik, and T. Minor. Classification of hyperspectral imagery with neural networks: Comparison to conventional tools. In T. Villmann and F.-M. Schleif, editors, *Machine Learning Reports*, volume 5, pages 1–15, 2011. ISSN:1865-3960. on-line http://www.techfak.uni-bielefeld.de/fschleif/mlr/mlr_04_2011.pdf. Also submitted to EURASIP Journal on Advances in Signal Processing.

[47] E. Merenyi, A. Jain, and W. H. Farrand. Applications of SOM magnification to data mining. *Wseas Transactions on Systems. July 2004; 3(5): 2122-7*, 2004.

[48] E. Mereńyi, J. V. Taranik, T. B. Minor, and W. H. Farrand. Quantitative comparison of neural network and conventional classifiers for hyperspectral imagery. In R. O. Green, editor, *Summaries of the Sixth Annual JPL Airborne Earth Science Workshop, Pasadena, CA, March 4–8*, volume 1: AVIRIS Workshop. 1996.

[49] E. Merényi and T. Villmann. Self-organizing neural network approaches for hyperspectral images. In M. Tolba and A. Salem, editors, *Intelligent Computing and Information Systems*, pages 33–42. Ain Shams University Cairo, Fac. of Computer and Information Science, 2002. ISBN 977-237-172-3.

[50] E. Merényi. Self-organizing ANNs for planetary surface composition research. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'98)*, pages 197–202, Brussels, Belgium, 1998. D facto publications.

[51] E. Merényi. The challenges in spectral image analysis: An introduction and review of ANN approaches. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 93–98, Brussels, Belgium, 1999. D facto publications.

[52] E. Mwebaze, P. Schneider, F.-M. Schleif, S. Haase, T. Villmann, and M. Biehl. Divergence based learning vector quantization. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks (ESANN'2010)*, pages 247–252, Evere, Belgium, 2010. d-side publications.

[53] D. Nebel and M. Riedel. Generalized functional matrix learning vector quantization. Master's thesis, University of Applied Sciences Mittweida, Germany, 2012.

[54] E. Oja. Neural networks, principle components and subspaces. *International Journal of Neural Systems*, 1:61–68, 1989.

[55] N. Pendock. A simple associative neural network for producing spatially homogeneous spectral abundance interpretations of hyperspectral imagery. In *Proc. Of European Symposium on Artificial Neural Networks (ESANN'99)*, pages 99–104, Brussels, Belgium, 1999. D facto publications.

[56] J. Principe. *Information Theoretic Learning*. Springer, Heidelberg, 2010.

[57] J. Ramsay and B. Silverman. *Functional Data Analysis*. Springer Science+Media, New York, 2nd edition, 2006.

[58] J. Richards and X. Jia. *Remote Sensing Digital Image Analysis*. Springer-Verlag, Berlin, Heidelberg, New York, third, revised and enlarged edition edition, 1999.

[59] M. Riedel, F. Rossi, M. Kästner, and T. Villmann. Regularization in relevance learning vector quantization using l_1-norms. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*

(ESANN'2013), page in this volume, Louvain-La-Neuve, Belgium, 2013. i6doc.com.

[60] B. Ripley. *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.

[61] F. Rossi, N. Delannay, B. Conan-Gueza, and M. Verleysen. Representation of functional data in neural networks. *Neurocomputing*, 64:183–210, 2005.

[62] F. Rossi, A. Lendasse, D. François, V. Wertz, and M. Verleysen. Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and Intelligent Laboratory Systems*, 80:215–226, 2006.

[63] T. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 12:459–473, 1989.

[64] F.-M. Schleif, A. Hasenfuss, and T. Villmann. Aggregation of multiple peaklists by use of an improved neural gas network. *Machine Learning Reports*, 1(MLR-02-2007):1–14, 2007. ISSN:1865-3960, http://www.uni-leipzig.de/~compint/mlr/mlr_01_2007.pdf.

[65] F.-M. Schleif, T. Villmann, T. Elssner, J. Decker, and M. Kostrzewa. Machine learning and soft-computing in bioinformatics - a short jorney. In D. Ruan, P. D'hondt, P. Fantoni, M. D. Cock, M. Nachtegael, and E. Kerre, editors, *Applied Artificial Intelligence – Proceedings FLINS 2007*, pages 541–548, Singapore, 2006. World Scientific. ISBN:981-256-690-2.

[66] F.-M. Schleif, T. Villmann, and B. Hammer. Analysis of proteomic spectra by multiresolution analysis and self-organizing maps. In F. Masulli, S. Mitra, and G. Pasi, editors, *Applications of Fuzzy Sets Theory – Proceeding of the 7th International workshop on Fuzzy Logic and Applications*, LNAI 4578, pages 563–570. Springer, Camogli, Italy, 2007. ISBN 978-3-540-73399-7.

[67] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2002.

[68] P. Schneider, B. Hammer, and M. Biehl. Adaptive relevance matrices in learning vector quantization. *Neural Computation*, 21:3532–3561, 2009.

[69] P. Schneider, F.-M. Schleif, T. Villmann, and M. Biehl. Generalized matrix learning vector quantizer for the analysis of spectral data. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2008)*, pages 451–456, Evere, Belgium, 2008. d-side publications.

[70] R. Schowengerdt. *Remote Sensing*. Academic Press, second edition edition, 1997.

[71] U. Seiffert and F. Bollenbeck. Clustering of hyperspectral image signatures using neural gas. *Machine Learning Reports*, 4(4):49–59, 2010.

[72] U. Seiffert, F. Bollenbeck, H.-P. Mock, and A. Matros. Clustering of crop phenotypes by means of hyperspectral signatures using artificial neural networks. In *Proceedings of the 2nd IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing WHISPERS 2010*, pages 31–34. IEEE Press, 2010.

[73] U. Seiffert, B. Hammer, S. Kaski, and T. Villmann. Neural networks and machine learning in bioinformatics - theory and applications. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2006)*, pages 521–532, Brussels, Belgium, 2006. d-side publications.

[74] U. Seiffert, L. C. Jain, and P. Schweizer. *Bioinformatics using Computational Intelligence Paradigms*. Springer-Verlag, 2004.

[75] U. Seiffert and B. Michaelis. Multi-dimensional self-organizing maps on massively parallel hardware. In N. Allinson, H. Yin, L. Allinson, and J. Slack, editors, *Advances in Self-Organising Maps*, pages 160–166. Springer-Verlag, London, 2001.

[76] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis and Discovery*. Cambridge University Press, 2004.

[77] B. Silverman. Smoothed functional principal components analysis by the choice of norm. *The Annals of Statistics*, 24(1):1–24, 1996.

[78] M. Strickert, F.-M. Schleif, T. Villmann, and U. Seiffert. Unleashing pearson correlation for faithful analysis of biomedical data. In M. Biehl, B. Hammer, M. Verleysen, and T. Villmann, editors, *Similarity-based Clustering*, volume 5400 of *LNAI*, pages 70–91. Springer, Berlin, 2009.

[79] F. Takens. On the numerical determination of the dimension of an attractor. In B. Braaksma, H. Broer, and F. Takens, editors, *Dynamical Systems and Bifurcations*, pages 99–106, Berlin, 1985. Lecture Notes in Mathematics, N0 1125, Springer-Verlag Berlin.

[80] K. Tasdemir and E. Merényi. A validity index for prototype-based clustering of data sets with complex cluster structures. *IEEE Transactions on Systems, Man, and Cybernetics*,

41(4):1039 – –1053, 2011.

[81] M. Tellechea and M. Grana. On the application of competitive neural networks for unsupervised analysis of hyperspectral remote sensing images. In S. B. Serpico, editor, *Proceedings of SPIE—The International Society for Optical Engineering*, volume 4170, pages 65–72, 2001.

[82] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B*, 67:91–108, 2005.

[83] A. Ultsch. Data mining and knowledge discovery with emergent selg-organizing feature maps for multivariate time series. In E. Oja and S. Kaski, editors, *Kohonen Maps*, pages 33–46. Elsevier, Amsterdam, 1999.

[84] P. Varshnev and M. Arora. *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*. Springer, 2010.

[85] T. Villmann. Sobolev metrics for learning of functional data - mathematical and theoretical aspects. *Machine Learning Reports*, 1(MLR-03-2007):1–15, 2007. ISSN:1865-3960, http://www.uni-leipzig.de/~compint/mlr/mlr_01_2007.pdf.

[86] T. Villmann, A. Cichocki, and J. Principe. Information theory related learning. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks (ESANN'2011)*, pages 1–10, Louvain-La-Neuve, Belgium, 2011. i6doc.com.

[87] T. Villmann and S. Haase. Divergence based vector quantization. *Neural Computation*, 23(5):1343–1392, 2011.

[88] T. Villmann and B. Hammer. Functional principal component learning using Ojas method and Sobolev norms. In J. Principe and R. Miikkulainen, editors, *Advances in Self-Organizing Maps - Proceeding of the Workshop on Self-Oranizing Maps (WSOM)*, pages 325–333. Springer, 2009.

[89] T. Villmann, M. Kästner, D. Nebel, and M. Riedel. Enhancement learning in functional relevance learning vector quantization. *Machine Learning Reports*, 6(MLR-03-2012):46–57, 2012. ISSN:1865-3960, http://www.techfak.uni-bielefeld.de/~fschleif/mlr/mlr_03_2012.pdf.

[90] T. Villmann, E. Merényi, and W. Farrand. Unmixing hyperspectral images with fuzzy supervised self-organizing maps. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2012)*, pages 185–190, Louvain-La-Neuve, Belgium, 2012. i6doc.com.

[91] T. Villmann, E. Merényi, and B. Hammer. Neural maps in remote sensing image analysis. *Neural Networks*, 16(3-4):389–403, 2003.

[92] T. Villmann, E. Merényi, and U. Seiffert. Machine learning approaches and pattern recognition for spectral data. In M. Verleysen, editor, *Proc. Of European Symposium on Artificial Neural Networks (ESANN'2008)*, pages 433–444, Evere, Belgium, 2008. d-side publications.

[93] M. Volpi, G. Matasci, M. Kanevski, and D. Tuia. Multi-view feature extraction for hyperspectral image classification. In M. Verleysen, editor, *Proc. of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN'2013)*, page in this volume, Louvain-La-Neuve, Belgium, 2013. i6doc.com.

[94] D. Wei, L. Hualiang, L. Xi-Lin, V. Calhoun, and T. Adali. ICA of fMRI data: Performance of three ICA algorithms and the importance of taking correlation information into account. In *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011*, 2011.

[95] Z. Zhang, J. Kwok, and D.-Y. Yeung. Parametric distance metric learning with label information. Technical Report HKUST-CS-03-02, The Hong Kong University of Science and Technology, Acapulco, Mexico, 2003.