TRANSFORMATION OF HYPERSPECTRAL DATA TO IMPROVE CLASSIFICATION BY MITIGATING NONLINEAR EFFECTS

Wolfgang Gross, Sebastian Wuttke, Wolfgang Middelmann

Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB

ABSTRACT

Non-linear effects in hyperspectral data are caused by varying illumination conditions, different viewing angles or multiple scattering of the incident light. These effects interfere with commonly used data analysis procedures. Manifold learning procedures are slow and require certain assumptions about the data structure that do not necessarily hold in real hyperspectral data. In this paper, a transformation is proposed that uses neighborhood distances to track the nonlinear structures of multiple classes simultaneously. The transformation is evaluated using a hyperspectral data set containing nonlinearities. A classification is performed and the results on the original and the transformed data are compared.

***Index Terms*—** Hyperspectral, data transformation, mitigating nonlinear effects, supervised classification

1. INTRODUCTION

Non-linear effects in hyperspectral data are a significant source of errors in data analysis. They are caused by varying illumination conditions, different viewing angles [1] or multiple scattering of the incident light [2]. Modeling these effects in airborne hyperspectral data is difficult; as specific material parameters, a geometric model and detailed information about the incident irradiance are required to perform a full correction with the bidirectional reflectance distribution function (BRDF). Recently, practical application of BRDF correction to airborne hyperspectral data is performed [3]. However, they are usually restricted to specific materials like certain vegetation classes or minerals.

The recent advance of hyperspectral full frame videos could also benefit from a fast nonlinear transform to support real time classification. Ground-based sensors are even more affected by BRDF effects, shadows etc. than airborne scanners. By assuming locally low-dimensional structures in the data, nonlinear effects can be modeled with ISOMAP [4] or Locally Linear Embedding (LLE) [5]. Each method has its limitations. The LLE assumes a locally linear structure, which is not always the case, observable in Figure 1b. Here, the class spectra follow a dominant linear curve, but it also has a certain width that is not the result of noise. ISOMAP requires a convex parameter space, otherwise the geodesic estimation will be flawed [4]. Also, the commonly used approach to parametrize the data by geodesics becomes complicated when the underlying manifold has more than two dimensions. When the data contains noise, the geodesics tend to go through outliers and skip over nonlinearities.

In this paper, a transformation is proposed that contracts sample spectra towards given reference spectra based on the samples distances to their k-Nearest Neighbors (k-NN) in each reference class. The transformation of each spectrum is a linear combination of vectors pointing in the direction of the class references. The coefficients of the linear combination are calculated to penalize larger distances. For samples with smaller distances to the training spectra, this penalizing results in a translation towards the reference spectra. If the training data contains nonlinear effects, the transformation achieves a linearization by identifying samples with the reference spectra, while the contraction of spectra reduces the intraclass variance. A further result is increased interclass distances, which can lead to improved data analysis.

Section 2 contains the description of the proposed method. Evaluation of the transformation using a classification algorithm on airborne hyperspectral data is performed in Section 3. The results and additional information about selecting suitable reference data are discussed in Section 4. An outlook about practical application and possible improvements can be found in Section 5.

2. METHOD

The scatter plots of hyperspectral data sets usually appear as a big cluster with multiple peaks, as can be seen in Figure 1a. Each peak stands for one material and is usually caused by spectra that are brightly illuminated or have a specific orientation relative to the sensor. Both can increase the measured radiance [3]. The resulting cluster contains varying radiance measurements for different samples of the same material, with a difficulty to separate classes when the radiance is too low. The aim is to identify each spectrum of a material with a single reference spectrum.

The nonlinear effects result in varying radiance measurements for different samples of the same material. The goal is to identify each measurement of a material with a single reference spectrum.

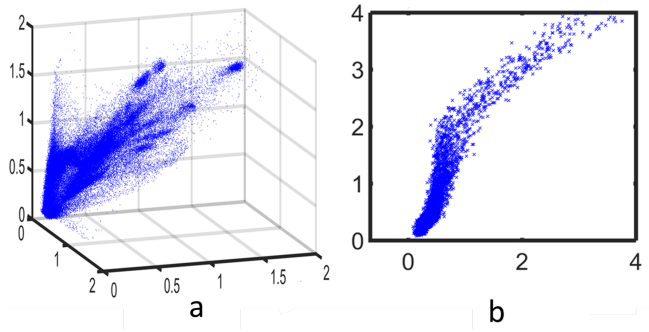

Fig. 1: a) Example scatter plot of hyperspectral radiance data for three bands, structured like a cluster with several peaks that can be identified with different materials. b) Scatter plot of the nonlinear effect from the metal dome in Figure 2 in two bands.

For the proposed transformation the following assumptions are necessary. We assume that the dimension p of the spectral signatures is sufficiently high to distinguish between two different materials or classes, even if they can not be linearly separated. This means, that nonlinear effects do not cause the spectra of two different materials to be equal. Also, we assume the signal-to-noise ratio to be sufficiently high that neighboring classes do not interpenetrate each other.

Finally, a set of training samples for each class is required, which should contain labeled samples of each class with different illumination conditions. These serve as an approximation to the class structure.

Expressed mathematically, let S be the given data set with n samples $s_1, \ldots, s_n \in \mathbb{R}^p$. From the above assumptions it follows, that each of the t different classes $M_1, \ldots, M_t \subseteq \mathbb{R}^p$ occupies a unique space in $\mathbb{R}^p$. Then, for two different materials M_i and M_j, $M_i \cap M_j = \emptyset$ for $i \neq j$ holds. If there exist two samples s_1, s_2 with $s_1 = s_2$ and $s_1 \in M_i$, then it automatically follows that $s_2 \in M_i$.

Let $\widetilde{M}_i \subset M_i$ be the set of training spectra for class i, $i = 1, \ldots, t$. Approaches to select suitable training data $\widetilde{M}_i$ are discussed in Section 4.

Let $\widetilde{s}_i \in \mathbb{R}^p$ be a chosen reference spectrum for class i. To compute the transformation, the vectors $v_{s,i} \in \mathbb{R}^p$ from a sample $s \in S$ to the class references $\widetilde{s}_i, i = 1, \ldots, t$ are required. The transformed spectrum $\widetilde{s} \in \mathbb{R}^p$ is computed by translating the original sample s with a convex combination of all $v_{s,i}$. The coefficients $\widetilde{c}_{s,i}$ for the convex combination are calculated from the k-NN distances to the training spectra $\widetilde{M}_i$ for all classes i.

With the data set S, the reference spectra $\widetilde{s}_i$ and the training samples $\widetilde{M}_i$ for each class i, the transformation can be computed with Algorithm 1.

For each sample $s \in S$, the first step is to calculate its k-Nearest Neighbors among the training spectra $\widetilde{M}_i$ of every class M_i. The k-dimensional distance vector $d_k(s, \widetilde{M}_i)$ can be saved to apply different penalty functions during transformation. In Step 3 the directional vector $v_{s,i}$ between the sample s and the reference spectra $\widetilde{s}_i$ are calculated. These are used in Step 10 to calculate the translation of s by a convex combination of the vectors $v_{s,i}$. Steps 5 and 6 guarantee, that any sample $s \in \widetilde{M}_i$ is identified with the correct reference $\widetilde{s}_i$. In Steps 8 and 9 the coefficients $\widetilde{c}_{s,i}$ for the convex combination of the $v_{s,i}$ are calculated and applied to s in Step 10 to calculate the transformed spectrum $\widetilde{s}$. The convex combination ensures that the transformed data $\widetilde{S}$ lies on a hyperplane containing all the reference spectra $\widetilde{s}_i$.

Algorithm 1 Collapsing point cloud by clusters

Inputs

1: Hyperspectral data set $S \subseteq \mathbb{R}^p$
2: Training data $\widetilde{M}_i \subseteq M_i$ for $i = 1, \ldots, t$
3: Reference spectra $\widetilde{s}_i \in \mathbb{R}^p$

Steps

1: **for all** $s \in S$ **do**
2: Calculate k-Nearest Neighbor distance vectors $d_k(s, \widetilde{M}_i)$ for $i = 1, \ldots, t$.
3: $v_{s,i} \leftarrow \widetilde{s}_i - s$
4: Define $f : \mathbb{R} \to \mathbb{R}$ to be a suitable decay function to penalize larger distances $d_k(s, \widetilde{M}_i)$
5: **if** At least one element of $d_k(s, \widetilde{M}_i) = 0$ for one class $i = 1, \ldots, t$ **then**
6: $\widetilde{s}_i \leftarrow s$
7: **else**
8: $\forall i = 1, \ldots, t$: $c_{s,i} \leftarrow f(d(s, \widetilde{M}_i))$
9: $\widetilde{c}_{s,i} \leftarrow c_{s,i} / \left\| \sum_{i=1}^{t} c_{s,i} \right\|_2$
10: $\widetilde{s} \leftarrow s + \sum_{i=1}^{t} c_{s,i} v_{s,i}$
11: **end if**
12: **end for**

Output $\widetilde{S}$

3. EXPERIMENTS

To evaluate the proposed transformation, a hyperspectral data set with 400×400 samples and $p = 127$ bands was selected. The data was recorded with an AISA Eagle in the wavelength range of $400 - 990$ nm. The scene contains several flat roof buildings, tree lines, roads and a big metallic dome. Figure 2 depicts the scene as a color composite.

Five classes, asphalt, metal, grass, roof and tree, were manually extracted as ground truth reference. The training samples $\widetilde{M}_i$ were specifically chosen to contain different illumination conditions. Especially the tree class poses many variations. The samples from treetops have different orientations between illumination and sensor, spectral variation due to nu-

Fig. 2: RGB composite of three hyperspectral bands. The depicted scene was used for the experiments in section 3. The classes for transformation and classification were asphalt, metal, grass, roof and tree.

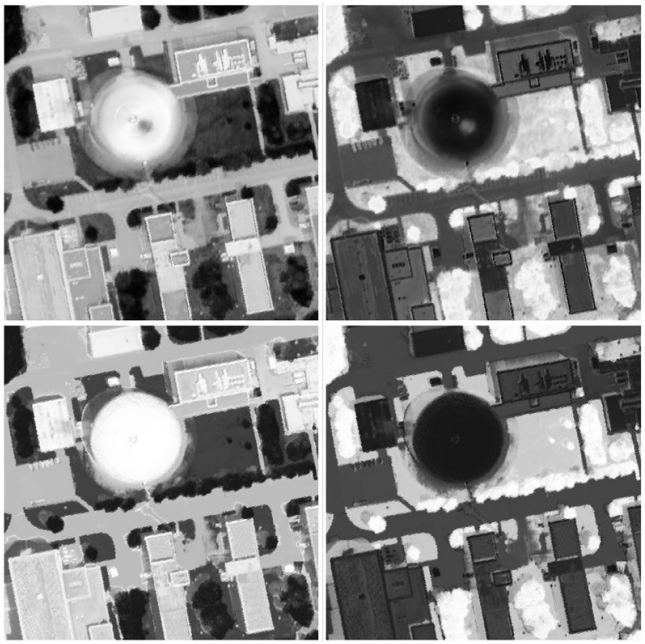

Fig. 3: Results of SAM for the classes metal in the left column and tree on the right. The upper images were computed from the original data, the lower from the transformed data. Looking at the dome and the trees, the SAM results look more homogeneous in the lower line, especially the area on the dome where the total reflection occurs.

tritional variances and shadowing. Reference spectra $\widetilde{s}_i$ of each class were directly taken from field spectroscopy data.
The transformation was calculated using random sets of training data from the selected reference data. The decay function was chosen to be $f(d_5(s, \widetilde{M}_i)) = 1/d_5(s, \widetilde{M}_i)$ to increase the distance between the transformed sample and the classes that are far away.
To classify the original and the transformed data, the Spectral Angle Mapper (SAM) was chosen. The SAM calculates the inner product between a sample and the reference spectrum and takes values between 0 (no similarity) to 1 (sample and reference are identical). Table 1 contains the reliability and overall accuracy of the classification results for the original and the transformed data. The reliability is defined as the fraction of correctly classified pixels with regard to all pixels classified as this class. The overall accuracy is the fraction of correctly classified pixels with regard to all pixels of that ground truth class.

4. DISCUSSION

From Table 1 it is apparent, that the classification on the original data set has high rate of false classifications for the labels metal, tree and grass. The SAM can not handle nonlinear data structures, e.g. samples on some parts of the metal dome, and confuses them with roof spectra. In the original data no sample was incorrectly classified as asphalt or metal.

While a confusion of the vegetation classes is understandable, the SAM classification in the transformed data is much better. The transformation significantly improves the classification of the metal dome and vegetation classes.
Two rule images of the SAM classification are depicted in Figure 3 for the original (upper row) and the transformed (lower row) data. The SAM results are more homogeneous in the vegetation areas and on the metal dome in the transformed data. As there was not much bare soil in the scene, the grass mask also contained areas where grass was very thin. Thus, the selected reference data also contained mixtures of grass and bare soil. In the transformed data an unmixing of these two materials would no longer be possible.
There are several approaches to select the training data $\widetilde{M}_i \subseteq$

	$\mathrm{REL}_{\mathrm{orig}}$	$\mathrm{ACC}_{\mathrm{orig}}$	$\mathrm{REL}_{\mathrm{trans}}$	$\mathrm{ACC}_{\mathrm{trans}}$
asphalt	1	0.9750	0.9976	0.9840
metal	1	0.7115	0.9641	0.9975
grass	0.9391	0.9351	0.9795	0.9979
roof	0.7330	1	0.9751	0.9544
tree	0.9118	0.9222	0.9762	0.9762

Table 1: Reliability (REL) and overall accuracy (ACC) for the classification results of original and transformed data, computed from the confusion matrices.

M_i for each class in the data set. The straightforward but most time-consuming method is to select them manually from the data. This approach is subject to the accuracy of the analyst and errors in class labeling may occur. Another approach is to use segmentation algorithms with significantly more classes than expected and identifying segments of different classes with one material, e.g. grass in the sun and in the shadow. Active classification approaches can also help to describe the data better by giving hints to regions that can not be classified properly [6].

The transformation on the test scene was performed on a standard computer, the implementation was carried out in MATLAB. Computation depends on the number of reference spectra in $\widetilde{M}_i$ and the number of samples in S. With 2000 reference spectra per class it took approximately 2 minutes to compute the results for 400×400 samples. Run time optimization through parallelization is possible.

During selection of reference data $\widetilde{M}_i$ it is important to cover the nonlinear effects of each class M_i. The method can not transform e.g. shadow regions properly, if no material information about any samples in the shadow is available.

It has not yet been analyzed whether transferring the training data to another data set is possible without correction. Technically, the training data could still be useful, as the assumptions from section 2 about the uniqueness of the present classes still holds. However, if the illumination conditions are completely different from the original, the previously extracted reference spectra $\widetilde{M}_i$ might not be good representatives of the class structures in the new image. Also, no general approach exists to determine whether the training data sufficiently describes the class structures. Analyzing the distribution of training spectra in $\mathbb{R}^p$ could give additional information.

5. OUTLOOK

The first experimental results from section 3 look very promising. The next step is to apply the transformation to commonly used data like the RIT SHARE 2012 data set [7]. Considering the recorded data and the different experiments, generating a subset to test classification of targets under varying illumination conditions is possible.

The proposed method suffers from the fact that it relies on the completeness of the training data. If a nonlinear effect has no sample in the training data, the transformation will not resolve it. A further task is to research data-driven point cloud analysis in high-dimensional data to select suitable training samples, thus reducing or completely skipping manual ground truth generation [8].

Applying a suitable noise model before data transformation could also improve the results. Especially, when samples have such a low radiance that they violate the uniqueness assumption of the occupied space in $\mathbb{R}^p$ of each class from Section 2.

The next step is to compare results of nonlinear classifiers on the original data with linear classifiers on the transformed data.

6. REFERENCES

[1] S.R. Sandmeier, E.M. Middleton, D.W. Deering, and W. Qin, "The potential of hyperspectral bidirectional reflectance distribution function data for grass canopy characterization," vol. 104, no. D8, pp. 9547–9560, 1999.

[2] D.A. Roberts, M.O. Smith, and J.B. Adams, "Green vegetation, nonphotosynthetic vegetation, and soils in aviris data," vol. 44, no. 2/3, pp. 255–269, 1993.

[3] J. Weyermann, A. Damm, M. Kneubuhler, and M.E. Schaepman, "Correction of reflectance anisotropy effects of vegetation on airborne spectroscopy data and derived products," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 616–627, January 2014.

[4] J.B. Tenenbaum, V. deSilva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," in *Science*, 2000, vol. 290, pp. 2319–2323.

[5] D.H. Kim and L.H. Finkel, "Hyperspectral image processing using locally linear embedding," in *Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on*, March 2003, pp. 316–319.

[6] Sebastian Wuttke, Hendrik Schilling, and Wolfgang Middelmann, "Reduction of training costs using active classification in fused hyperspectral and lidar data," in *Image and Signal Processing for Remote Sensing XVIII*, 2012, vol. 8537 of *SPIE Proceedings*.

[7] A. Giannandrea, N.G. Raqueno, D.W. Messinger, J. Faulring, J.P. Kerekes, J. van Aardt, K. Canham, S. Hagstrom, E. Ontiveros, A. Gerace, J. Kaufman, K.M. Vongsy, H. Griffith, and B.D. Bartlett, "The share 2012 data collection campaign," in *Proc. SPIE, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX*, 2013, vol. 8743.

[8] W. Gross, G. Keskin, H. Schilling, A. Lenz, and W. Middelmann, "Automatic modeling of nonlinear signal source variations in hyperspectral data," 2014, pp. 2965–2968.