

Partial Generalized Correlation For Hyperspectral Data

Marc Strickert
Institute of Vision and Graphics
University of Siegen, Germany

Björn Labitzke
Computer Graphics Group
University of Siegen, Germany

Volker Blanz
Media Systems Lab
University of Siegen, Germany

Abstract—A variational approach is proposed for the unsupervised assessment of attribute variability of high-dimensional data given a differentiable similarity measure. The key question addressed is how much each data attribute contributes to an optimum transformation of vectors for reaching maximum similarity. This question is formalized and solved in a mathematically rigorous optimization framework for each data pair of interest. Trivially, for the Euclidean metric minimization to zero distance induces highest vector similarity, but in case of the linear Pearson correlation measure the highest similarity of one is desired. During optimization the not necessarily symmetric trajectories between two vectors are recorded and analyzed in terms of attribute changes and line integral. The proposed formalism allows to assess partial covariance and correlation characteristics of data attributes for vectors being compared by any differentiable similarity measure. Its potential for generating alternative and localized views such as for contrast enhancement is demonstrated for hyperspectral images from the remote sensing domain.

Keywords: distance pursuit, partial generalized correlation.

I. INTRODUCTION

Progress in opto-electrical engineering, multi-channel spectrometers, and highly accurate time-of-flight technologies stimulates new demands on computational data analysis. One important direction is the real-time management, interactive data exploration and rapid analysis, while another direction is the development of data-specific algorithms, exploiting spatial arrangements and spectral characteristics. Mass spectrometry imaging and remote sensing are popular technologies that deliver massive spatially connected data cubes for which the main task is the highlighting of application-specific unknown or expected features [1], [2]. Image segmentation techniques for the generation of false color images help to visually emphasize such features. Examples are the assignment of protein compositions in biological tissue images or the identification of geological resources in remote sensing multi-spectral images. These are already high-level goals that require prior knowledge, i.e. label information, about candidate materials. In this case, supervised methods like learning vector quantization, k-nearest neighbor, or linear discriminant analysis can be used to map the spectra onto class-related spectral prototypes and into meaningful subspaces. If additional label information is not available, faithful unsupervised characterization of the data is needed. In this case, data-intrinsic features like mathematical moments such as the mean, variance, or skewness are essential ingredients for many data operations, among which normalization, e.g. z-score transformation; clustering, e.g. k-means; and data-driven projections and analyses, e.g. principal component analysis [3].

Maybe the most important aspect of unsupervised data analysis is the choice of distance or similarity measure. Euclidean distance and, more generally, Minkowski distances belong to one popular class of metrics focusing on differences of individual data attributes and pairwise attribute mixtures, respectively. More relaxed comparisons result from dissimilarity measures like the Pearson correlation [4] or its descendant, the spectral angle [5], while density vectors can be faithfully compared by information-related divergence measures [6].

Depending ultimately on data measures, the nearest neighbor search is one of the most important operations in data processing models, such as hierarchical clustering [7], vector quantization [8], endmember detection [9], [10] or morphological operations [11]. In supervised scenarios the selection of the 'right' measure can be alleviated by data-driven adaptive metrics. For example, weighting factors of the adaptive Euclidean distance can be learned from the data for scaling individual data attributes according to optimizing features such as class discrimination [12]. Other examples include learning a matrix metric [13], or adaptive Pearson correlation [14].

In practice, analytic pipelines tend to be inconsistent, for example, if a k-means algorithm is used, but with other than Minkowski distances. Then calculated Euclidean centers of gravity are not necessarily in minimum displacement to data compared by other measures like Pearson correlation. For example, the Euclidean center $(1, 1, 1)$ of three vectors $(0, 0, 3)$, $(0, 3, 0)$, and $(3, 0, 0)$ even leads to singular values if taken as argument to Pearson correlation [15]. Analog reasoning holds true for measures of attribute variability. Principal component analysis relies on the eigen decomposition of the data covariance matrix, but covariance, measuring attribute variability, is an inherently Euclidean concept and thus conflicting with other similarity measures used in the analysis pipeline. The attribute assessment proposed in the following is as general as possible, only requiring differentiability of the data measure.

II. DISTANCE PURSUIT METHOD

A variational point-of-view is taken by collecting all information about attribute involvement and path length during the transformation of source vectors into target vectors. Thereby, the underlying data measure imposes constraints on the optimum way between pairwise vectors, as illustrated in panel A of Figure 1. As a consequence, asymmetric relationships between the vectors can occur for non-Euclidean as depicted in panel B of Figure 1. By traversing such measure-specific

paths partial generalized covariance and correlation quantities can be computed, which is the goal of this work.

A pragmatic solution to getting along the optimum path from a start vector $\mathbf{w} \in \mathbb{R}^d$ to a target $\mathbf{x} \in \mathbb{R}^d$ can be best formulated in an iterative manner. The general procedure is summarized in Algorithm 1. Therein, t denotes the number of steps, i.e. the resolution to transform \mathbf{w} into \mathbf{x} ; for this, a monotonic sequence $s = \{s_0 = d(\mathbf{x}, \mathbf{w}), \dots, s_t = d_{target}\}$ from the initial vector distance to the maximum possible degree of similarity d_{target} is created with $d_{target} = 0$ for metrics and $d_{target} = 1$ for Pearson correlation. This sampling density is set to $t = 16$ here, because an even higher density did not change the final values significantly in the applications. During execution, the current position of \mathbf{w} is continuously updated and the following two quantities are collected:

- g is the overall line integral summing up path fragment lengths $\sqrt{\langle \mathbf{v} \rangle_2}$, with self dot-product $\langle \mathbf{v} \rangle_2 := \langle \mathbf{v}, \mathbf{v} \rangle$,
- \mathbf{b} is the integrated differential attribute vector.

This approach works under the assumption that vector transformations can be expressed as piecewise linear approximations in d -dimensional space, i.e. target measures are differentiable.

Algorithm 1 distance pursuit $DP_{\{\mathbf{b}, g\}}(\mathbf{x}, \mathbf{w}, t)$

- 1: {input \mathbf{x} target vector; \mathbf{w} source vector; t number steps}
 - 2: $g \leftarrow 0$; $\mathbf{b} \leftarrow \mathbf{0}_d$ {to be filled and returned}
 - 3: **for** $s \leftarrow s_1$ to s_t {traverse distance sequence w/o s_0 } **do**
 - 4: $\mathbf{v} \leftarrow \alpha(s) \cdot \partial d(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w}$ {optimum scaled gradient}
 - 5: $\mathbf{w} \leftarrow \mathbf{w} + \mathbf{v}$ {move based on gradient ascent/descent}
 - 6: $\mathbf{b} \leftarrow \mathbf{b} + \mathbf{v}$ {integrate changes per attribute}
 - 7: $g \leftarrow g + \sqrt{\langle \mathbf{v} \rangle_2}$ {line integral}
 - 8: **end for**
-

In Algorithm 1, the optimum scaling factor $\alpha(s)$ can be either computed by a line search approach, or in some cases analytically, as will be illustrated for Euclidean distance and Pearson correlation. The sign of $\alpha(s)$ depends on the optimization approach: for Euclidean distance, the vectors start at a certain distance that should be minimized to zero by gradient descent (negative gradient), while for Pearson correlation the quantity needs to be maximized to one by gradient ascent (positive gradient).

The general idea is to sample the interval starting at $d(\mathbf{x}, \mathbf{w})$ and going down to zero for regular distances or going up to one for correlation measures. Using the current sample position s , the relationship

$$s = d(\mathbf{x}, \mathbf{w} + \alpha(s) \cdot \partial d(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w}) \quad (1)$$

can be used for isolating the gradient scaling term $\alpha(s)$. In Algorithm 1, line 4, a very general update of \mathbf{v} is possible without $\alpha(s)$ by finding the least norm solution of \mathbf{w}^* starting at $\mathbf{w}^* = \mathbf{w}$ for $\mathbf{v} \leftarrow (\arg \min_{\mathbf{w}^*} |s - d(\mathbf{x}, \mathbf{w}^*)|) - \mathbf{w}$.

Algorithm 1 is formulated for only one pair of vectors, one being fixed, the other being moved. Depending on the application many pairs of data vectors need to be analysed by this algorithm in a bidirectional manner, as will be discussed.

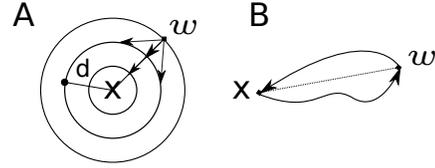


Fig. 1. **Sub-figure A:** graphical illustration of the distance pursuit Algorithm 1. Vector \mathbf{x} is fixed, and vector \mathbf{w} is adapted to minimize its dissimilarity with \mathbf{x} . After each small step on that way, at fixed displacements d to \mathbf{x} indicated by concentric circles, the new optimum direction towards \mathbf{x} is calculated. In the depicted Euclidean plane this finally results in a straight line between \mathbf{x} and \mathbf{w} . **Sub-figure B:** Illustration of non-Euclidean relationships of vectors \mathbf{x} and \mathbf{w} . Arrow end points refer to maximum similarity, i.e. not necessarily identity. The optimum paths for the transformations $\mathbf{x} \rightarrow \mathbf{w}$ and $\mathbf{w} \rightarrow \mathbf{x}$ in a given space can be different, and their line integrals may be equal or different. The dotted line denotes the displacement of both vectors.

A. Euclidean Distance

The Euclidean distance $d(\mathbf{x}, \mathbf{w}) = \sqrt{\langle \mathbf{x} - \mathbf{w} \rangle_2}$ is a good starting point for showing the interplay with the distance pursuit procedure. Using Equation 1 the optimum gradient scaling term $\alpha(s)$ can be isolated as

$$\alpha(s) = q_1 + \sqrt{q_1^2 - q_2(s)} \quad \text{with} \quad (2)$$

$$q_2(s) = \frac{\langle \mathbf{x} - \mathbf{w} \rangle_2 - s^2}{\langle \partial d(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} \rangle_2} \quad (3)$$

$$q_1 = \frac{\langle \mathbf{x} - \mathbf{w}, \partial d(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} \rangle}{\langle \partial d(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} \rangle_2} \quad (4)$$

Therein, the derivative of the Euclidean distance is given by:

$$\frac{\partial d(\mathbf{x}, \mathbf{w})}{\partial \mathbf{w}} = \frac{\mathbf{w} - \mathbf{x}}{\sqrt{\langle \mathbf{x} - \mathbf{w} \rangle_2}} \quad (5)$$

After all, the results from the distance pursuit algorithm are exactly according to the expectations: g is just the undirected Euclidean distance between \mathbf{x} and \mathbf{w} , and \mathbf{b} is the difference vector $\mathbf{x} - \mathbf{w}$.

Still, Euclidean distance will become interesting for the introduction of partial variance and for creating structural analogies to other measures.

B. Pearson Correlation

Correlation measures are widely used in scientific data analysis, because they help finding associations between data vectors [16], [17]. These relationships might be linear or non-linear, or symmetric or asymmetric [18]. A maximum value of 1 indicates perfect correlation, a value of 0 disconnection, and -1 denotes perfect anti-correlation in terms of inverse patterns.

Mean-centered Pearson correlation

$$d_r(\mathbf{x}, \mathbf{w}) = \frac{\langle \mathbf{x} - \mu_{\mathbf{x}}, \mathbf{w} - \mu_{\mathbf{w}} \rangle}{\sqrt{\langle \mathbf{x} - \mu_{\mathbf{x}} \rangle_2 \cdot \langle \mathbf{w} - \mu_{\mathbf{w}} \rangle_2}} \quad (6)$$

is used for characterizing the degree of linear dependence, invariant to the argument vectors' mean and variance. Its application for detecting common patterns and relationships has a long tradition in biosciences, technical analyses, and econometrics. In spectral data, the spectral angle which is the arcus cosine \cos^{-1} of uncentered correlation is one of the

standard analysis measures [5]. The optimum gradient scaling factor for the Pearson correlation d_r is computed by

$$\alpha(s) = \frac{-(q_1 \cdot q_2) + q_3 \cdot q_5 \cdot s^2 + s \cdot \sqrt{q_3 \cdot q_7(s)}}{q_2^2 - q_3 \cdot q_6 \cdot s^2} \quad (7)$$

$$q_7(s) = q_1^2 \cdot q_6 + q_2^2 \cdot q_4 - 2 \cdot q_1 \cdot q_2 \cdot q_5 + s^2 \cdot q_3 \cdot (q_5^2 - q_4 \cdot q_6) \quad (8)$$

$$q_6 = \langle \partial d_r(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} - \mu_{\partial} \rangle_2 \quad (9)$$

$$q_5 = \langle \mathbf{w} - \mu_{\mathbf{w}}, d_r(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} - \mu_{\partial} \rangle \quad (10)$$

$$q_4 = \langle \mathbf{w} - \mu_{\mathbf{w}} \rangle_2 \quad (11)$$

$$q_3 = \langle \mathbf{x} - \mu_{\mathbf{x}} \rangle_2 \quad (12)$$

$$q_2 = \langle \mathbf{x} - \mu_{\mathbf{x}}, d_r(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w} - \mu_{\partial} \rangle \quad (13)$$

$$q_1 = \langle \mathbf{x} - \mu_{\mathbf{x}}, \mathbf{w} - \mu_{\mathbf{w}} \rangle \quad (14)$$

In these equations, $\mu_{\mathbf{x}}$ denotes the arithmetic mean of the components in vector \mathbf{x} , $\mu_{\mathbf{w}}$ is the mean of vector \mathbf{w} , and μ_{∂} is the mean value of the gradient components of $\partial d_r(\mathbf{x}, \mathbf{w}) / \partial \mathbf{w}$. The derivative of the Pearson correlation is given by

$$\frac{\partial d_r(\mathbf{x}, \mathbf{w})}{\partial \mathbf{w}} = d_r(\mathbf{x}, \mathbf{w}) \cdot \left(\frac{\mathbf{x} - \mu_{\mathbf{x}}}{\langle \mathbf{x} - \mu_{\mathbf{x}}, \mathbf{w} - \mu_{\mathbf{w}} \rangle} - \frac{\mathbf{w} - \mu_{\mathbf{w}}}{\langle \mathbf{w} - \mu_{\mathbf{w}} \rangle_2} \right). \quad (15)$$

Finally, two notes in favor of the Pearson correlation are given. First, if components in \mathbf{x} and \mathbf{w} are turned into their rank order values Equation 6 yields the Spearman rank correlation coefficient which is very robust against outliers. Second, when the logarithm is applied to Equation 6 and if entries in \mathbf{x} and \mathbf{w} are non-negative and summing up to one, then $\log d_r(\mathbf{x}, \mathbf{w}) = \log \langle \mathbf{x} \rangle_2 \cdot \log \langle \mathbf{w} \rangle_2$ becomes the negative Cauchy-Schwarz divergence [19], which links the uncentered Pearson correlation to information-theoretic processing of density distributions.

III. GENERALIZED PARTIAL COVARIANCE

Variance is one of the most basic measures to assess the variability of data attributes. This helps to characterize the interestingness of the j -th attribute in a set \mathbf{X} of d -dimensional data vectors. One of the obvious properties is that variances of the j -th and i -th attribute are usually computed independently of each other as $\sigma_k^2 = 1/(n-1) \cdot \sum_{l=1}^n (x_k^l - \mu_k)^2$ according to the textbook form, μ_k being the mean value of the k -th attribute and x_k^l the entry of the k -th attribute in the l -th data vector \mathbf{x}^l . This common notation of variance quantifies the quadratic deviation of the k -th attribute from its mean. Alternatively, the expansion of $\mu_k = 1/n \cdot \sum_{j=1}^n x_k^j$ could be used for deriving a double sum notation of the variance visiting all pairs of data. Inversely, one can consider all data vector pairs of interest and compute the contribution of their attributes to the minimization of their common distance, or, more generally, to the maximization of their total similarity. A double sum expansion of σ_k^2 avoids the explicit calculation of the Euclidean mean values μ_k .

Generally, pairwise attribute contributions, i.e. covariance, can be quantified via the partial derivatives for data relationships defined by either all pairs of data vectors or by a subset of pairs. Let \mathcal{I}^i be the set of indices of data vectors connected to

data vector \mathbf{x}^i then the partial generalized covariance matrix \mathbf{V} with elements $k, l = 1 \dots d$ is calculated by

$$V_{kl} = \frac{1}{G} \cdot \sum_{i=1}^n \sum_{j=1}^{|\mathcal{I}^i|} \text{DP}_{b_k}(\mathbf{x}^i, \mathbf{x}^{\mathcal{I}_j^i}, t) \cdot \text{DP}_{b_l}(\mathbf{x}^i, \mathbf{x}^{\mathcal{I}_j^i}, t). \quad (16)$$

Here, the DP Algorithm 1 is called with a resolution of t steps to calculate the integrated attribute derivative vectors \mathbf{b} of which the k -th and l -th components are taken. The normalization constant $G = 2 \cdot (-z + \sum_{i=1}^n |\mathcal{I}^i|)$ is twice the number of pairwise comparisons, excluding the number z of non-contributing self-comparisons where $i = \mathcal{I}_j^i$. This normalization constant makes sure that an unbiased estimation of variance ($k = l$) and covariance ($k \neq l$) is obtained for the Euclidean distance when all pairs of data are considered, i.e. $\mathcal{I}^i = \{1 \dots n\}$ for $i = 1 \dots n$ and $G = 2 \cdot (-n + n^2)$. In any case the matrix \mathbf{V} will be symmetric, because the summed product in Equation 16 is commutative for index swaps k, l and l, k . The variance is given in the diagonal elements of \mathbf{V} .

The partial generalized covariance matrix \mathbf{V} can be used further to calculate partial generalized correlation matrix \mathbf{R} according to the formula:

$$R_{kl} = V_{kl} / \sqrt{V_{kk} \cdot V_{ll}} \quad (17)$$

If only \mathbf{R} is of interest, the factor G in Equation 16 is canceled out and does not need to be calculated. Again, for data living in Euclidean space, these values are exactly the values much more efficiently calculated by Equation 6. The benefit of this approach is that the arithmetic mean does not need to be computed explicitly and is thus suitable for non-Euclidean measures too. Incidentally, the Pearson correlation measure can be plugged in here to compute a generalized partial correlation of data compared by Pearson correlation. For brevity, this will be called meta-correlation in the following.

IV. PROCESSING OF HYPERSPECTRAL IMAGES

The generalized measure of attribute variability has been applied to a hyperspectral data set from the remote sensing domain. The cuprite data set, acquired in a mission of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [20], was made available for free download at the website <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>; it contains geological features recorded in the visible and infrared wavelength range between 380 to 2500 nm. The present study concentrates on the structurally rich third patch named f970619t01p02_r02_sc03 which is geographically located at 117°11'W, 37°32'N. The patch size is 614 × 512 pixels, with each pixel covering an approximate area of 20 × 20 m². In the available 224 bands, the 19 channels 1–3, 108–111, and 155–166 do only contain zero data, and they are thus excluded from the analysis. The remaining valid channels have been renumbered in consecutive order.

A first aerial overview is obtained by k-means clustering with a number of 32 centers for each of the two images displayed in Figure 2. The top panel shows the quantized 205-dimensional spectra without any transformation of the

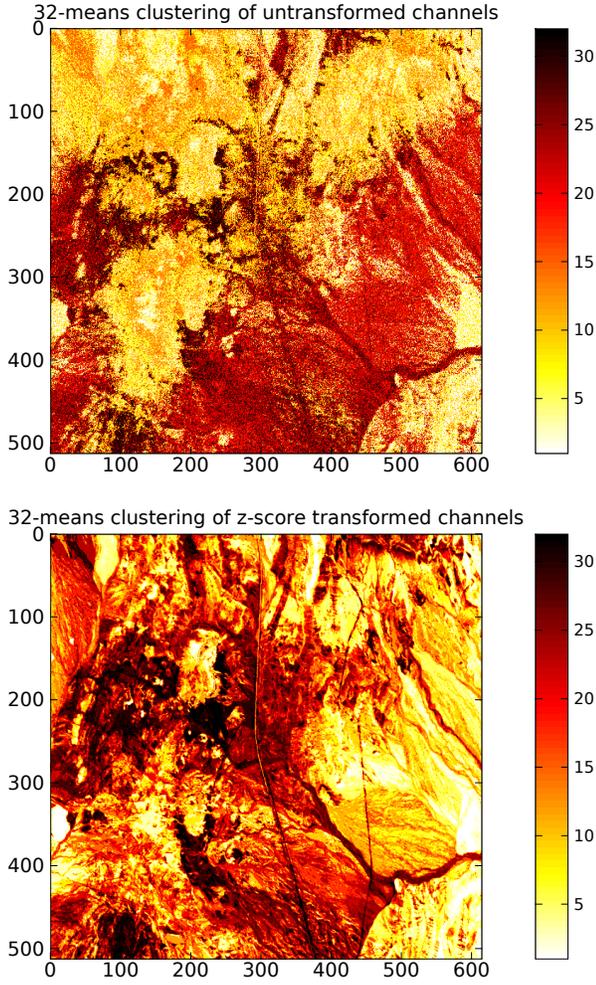


Fig. 2. Quantization of 205-dimensional spectra using k-means clustering. **Top:** non-transformed data. **Bottom:** the 205 channels z-score transformed.

data set, the bottom panel gives the result for a data set in which each of the 205 channels have been separately z-score transformed, i.e. by removing the mean and dividing by the standard deviation. Comparing both images, more consistent patterns can be found in the transformed case. For this reason all following operations were carried out on the transformed data set. This transformation would make queries against spectral fingerprint data bases difficult, because each channel is shifted and scaled independently for each spectrum, but since such queries are not made here, this is not a problem.

Two cases are considered for the partial generalized variance measures: one refers to localized views on channels, the other refers to spatially localized spectra.

A. Channel Neighborhood

In hyperspectral data one of the often observed characteristics is a strong correlation between adjacent channels, especially if the spectral gap between the bands is small [21]. The standard variance of a spectrum in Euclidean distance can be calculated by Equation 16, basically by looking at the sum

of squared differences of all pairs of spectral channels, each channel containing 614×512 values. In this case, structurally very different channels of far apart frequency bands contribute to their variance value, which leads to a dominance of large differences over subtle characteristics. A more localized view is obtained by comparing only adjacent channels, i.e. channel 2 with channels 1 and 3, 3 with 2 and 4, and so forth. Thus, if the very boundaries are ignored, the number of effective comparisons is $2 \cdot (d - 2)$, leading to the normalization term $G = 4 \cdot (d - 2)$. As a consequence of the localized view, the analysis gets more sensitive to changes in else highly correlated channels. The differences between standard variance and the partial variance for Euclidean distance is shown in Fig 3. Note that for each pixel the corresponding spectrum variance can be calculated independently. While the standard variance gets dominated by one spectrum around pixel position $(x = 490, y = 280)$ with a variance of about 8, the partial version provides more equalized solutions across all spectra. As a result, more variational features are displayed for the partial version, although its total range of values only reaches a level of about 0.35.

Due to z-score transformed channels no noticeable difference between the shown partial Euclidean variance and its analogon for Pearson correlation (not shown) can be found [15]. Yet, the Euclidean example indicates that it might be worthwhile to consider partial rather than global Euclidean variance for focusing on localized spatially arranged features.

B. Pixel Neighborhood

The spatial connectivity in images can be exploited for local analyses. Rather than being interested in the relationship of far apart spectra, i.e. distant 205-dimensional image pixels, the neighborhood perspective is of crucial importance, especially if lateral blurring can be expected. In image operations, local filters like sharpening and edge enhancement do rely on the proper definition of the pixel context. Here the standard Moore environment of 8-neighborhood is considered for the characterization of spectrum contexts.

An indexing scheme compatible with the notation in the partial generalized covariance Equation 16 can be easily established by serializing the two-dimensional image coordinates (i, j) into column-major order using the number of rows n_r and the serialized index $\iota = i + j \cdot n_r$. Then index sets $\mathcal{I}^\iota = \iota + \{-n_r - 1, -n_r, -n_r + 1, -1, 1, n_r - 1, n_r, n_r + 1\}$ are defining the Moore neighbors of index ι .

1) *Partial generalized correlation matrix and eigenstructure:* With the defined indexing scheme Equation 16 is used to calculate the partial generalized correlation matrix \mathbf{R} defined in Equation 17 for the Pearson correlation measure, thereby conducting a total number of $(614 - 2) \times (512 - 2) \times 8$ pairwise spectrum comparisons within the 8-neighborhood.

A comparison of correlation matrices resulting from the proposed generalization of covariance is shown in Figure 4. From top to bottom, standard correlation, partial correlation computed from the 8-neighborhood, and partial meta-correlation are shown as color matrices on the left with their

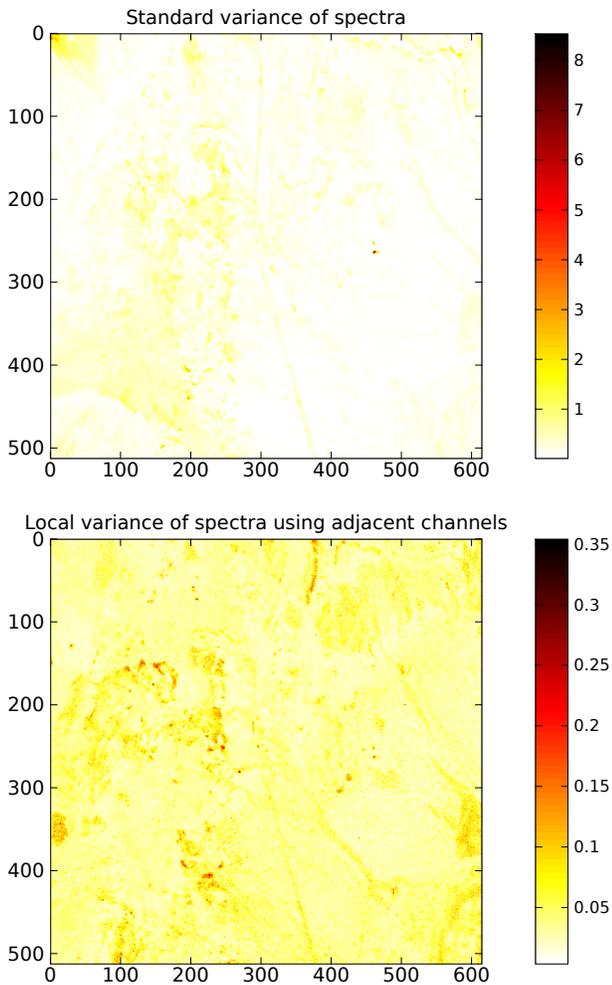


Fig. 3. **Top**: standard variance. **Bottom**: partial variance of the 205 spectral bands per pixel.

corresponding scatter plots of channel dissimilarities on the right. It can be observed that a transition from global to local correlation, i.e. from top to middle row, helps to better resolve the dominating extremal channels around 105, 147, and 204. Furthermore, the change from Euclidean distance to Pearson correlation, i.e. from middle to bottom row, both computed from the Moore neighborhood, changes the relationships between the channels strongly. Particularly, distant channels are affected, and the range of values for the discrimination of channels is increased from -0.2 to 1 for the Euclidean distance to -0.6 to 1 for the meta-correlation. The global meta-correlation matrix is not shown here because of excessive computing times for the roughly $100 \cdot 10^9$ pairwise vector comparisons using Algorithm 1.

Principal component analysis (PCA) is commonly based on eigen decomposition of either covariance or correlation matrix. Correlation matrices, reflect standardized variables and are therefore preferred in most applications. Each of the three displayed correlation matrices could be used for creating false color image representations by mapping the 205-dimensional

TABLE I
PERCENT VARIANCE EXPLAINED BY FIRST THREE EIGENVECTORS

Standard correlation	80%	9%	5%	$\sum=94\%$
Partial correlation	78%	8%	3%	$\sum=89\%$
Partial meta-correlation	36%	20%	9%	$\sum=65\%$

spectra per pixel to RGB space, i.e. to the three major eigenvectors, or to only the most prominent eigenvector for addressing an indexed color set [21]. For two reasons these standard displays of hyperspectral data are not provided here: Firstly, linear mapping operations of spectra to eigenvectors are fully justified only in Euclidean spaces. Secondly, even if such mapping was valid, the eigenstructures of the three correlation matrices, summarized in Table I, would be difficult to compare.

Although the matrices of standard and partial correlation look similar, the decrease from 94% to 89% of explained total variation indicates that the structural complexity of the eigenstructure of the partial correlation is higher than for standard correlation. This becomes even much more evident for the meta-correlation with only 65% of total variance represented by the first three eigenvectors. In many applications standard PCA exhibits an empirically rapid decrease of eigenvalues, often being examined for thresholding using Catell, Horn, Kaiser, or other criteria [22]. One line of argumentation indicates that since a high amount of 94% total variability can be explained by standard correlation, a faithful RGB color image could be obtained for representing the underlying spectra. A more critical reflection leads to realizing outliers as possible reason for 'explaining' major directions in high-dimensional spaces, or apparent redundancies may dominate the principal directions. Thus, the richer eigenvalue structure of partial correlation and meta-correlation bares the potential of more diverse representations of the spectra while being at the same time more difficult to visualize properly. That's why a comparison of generated false color images would be misleading here, because the different color spaces carry different amounts of represented information. Furthermore, keeping in mind that a projection onto the principal coordinates is a linear Euclidean data transformation, more investigation is needed on the faithful mappings for partial and generalized correlation measures.

2) *Reconstruction of channel relationships*: A method to visualize the dependencies between the spectral channels is the reconstruction of their dissimilarity matrix in the Euclidean space. For correlation measures such a matrix is easily obtained by subtracting the respective correlation value from 1. This leads to minimum values of zero for maximum correlation and to maximum values of two for anti-correlation. Reconstructions are canonically obtained by multi-dimensional scaling (MDS), here, maximizing reconstruction quality r , the correlation of the given dissimilarity matrix and the distance matrix of optimized 2D points [23].

Embedding results for the correlation matrices are displayed as scatter plots in the right column of Figure 4. Reconstruction

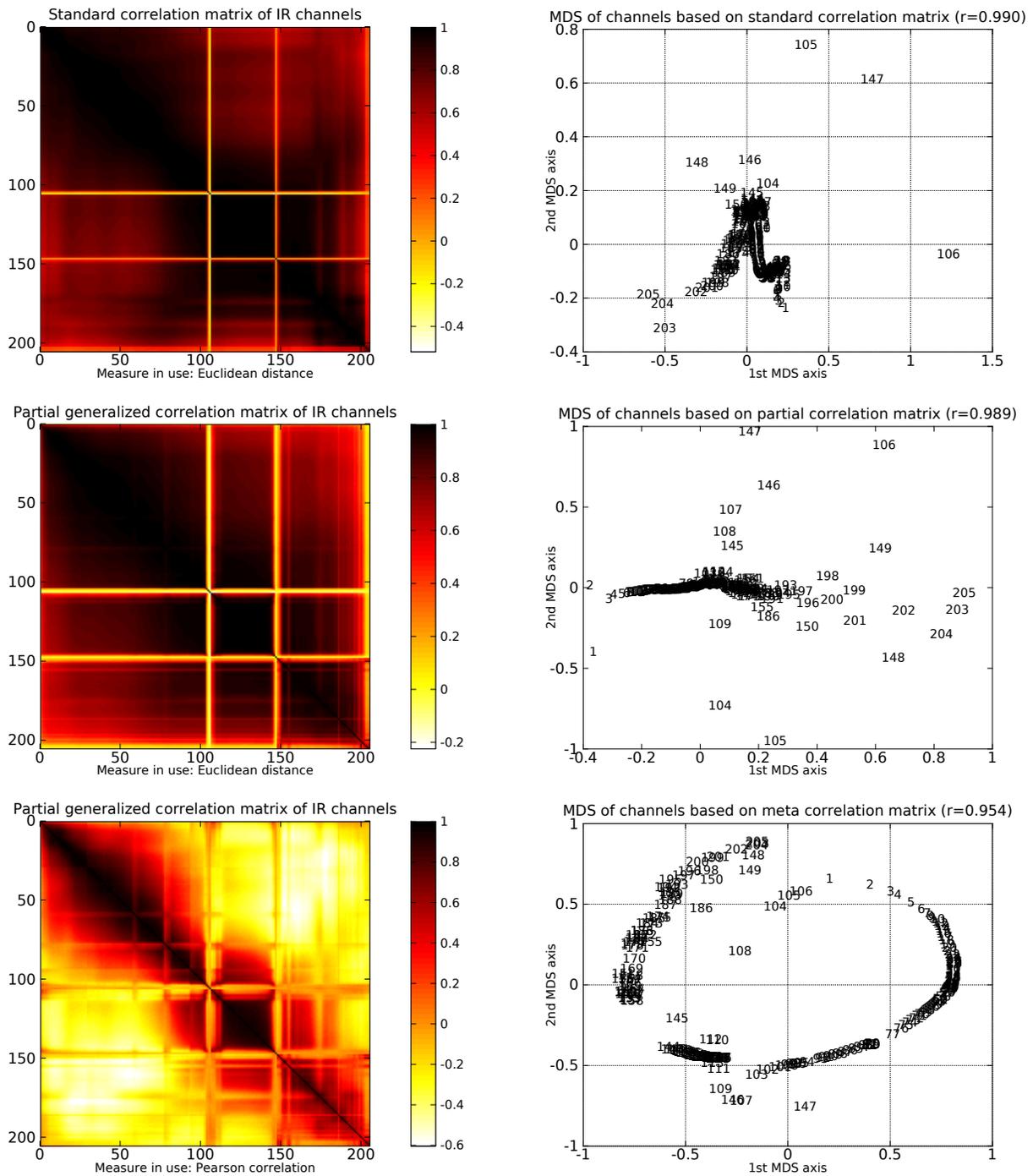


Fig. 4. Data visualization based on correlation matrices. **Left:** correlation matrices. **Right:** channel scatter plots obtained by MDS, reflecting dissimilarities one minus matrix of top row. **Top row:** standard correlation. **Center row:** partial correlation based on Euclidean distance. **Bottom row:** partial correlation based on Pearson correlation.

qualities of $r = 0.990$ for standard correlation (top) and $r = 0.989$ for partial correlation (center) are almost equally excellent, and both scatter plots show some common, but also quite different characteristics. Things in common are the smooth trajectories of adjacent channels 5–190 with the exception of some characteristic channels 104–106, 146–148, and 203–205. Different features are the trajectory bendings, and

more characteristic channels are found by partial correlation in addition to those already identified by standard correlation. Since these additional channel indices are adjacent to the ones of standard correlation, this indicates a semantically meaningful reason, especially, partial correlation of the 8-neighborhood tends to better discriminate between strongly correlated channels. Partial meta-correlation (bottom), like

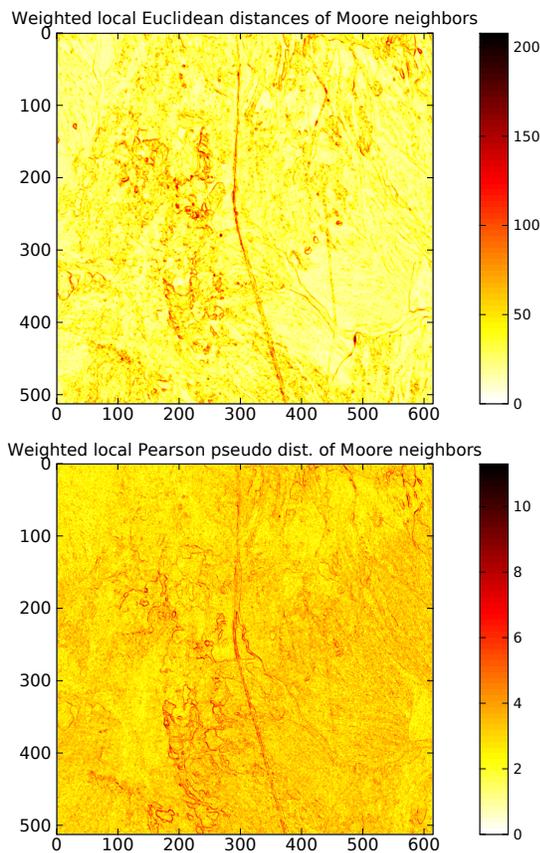


Fig. 5. Images of weighted Moore neighbor line integrals. **Top:** Euclidean distance being at the same time the line integral between spectra. **Bottom:** pseudo-distance in terms of line integrals for partial Pearson correlation between neighbors.

Euclidean correlation, shows a consecutive indexing, pointing out consistent relationships of adjacent channels. Yet, the drop of embedding quality to $r = 0.954$ also indicates a slightly less faithful representation in 2D Euclidean space. The channel index gaps in the circular arrangement correspond to the vertical lines in the meta-correlation matrix, the outlier indices coincide with those of Euclidean distance. It seems that the closing of the endpoints in the channel index circle is forced by the channels 104, 105, 106, 148, and 149 which are structurally located between the endpoints. After all, meta-correlation shows a quite distinct scattering of the channel relationships compared to the Euclidean distance.

3) *Line integrals:* Finally, the novel feature of line integrals available as scalar value g by the distance pursuit Algorithm 1 is investigated. The matrix of relationships of the Moore neighborhood with entries $g_{(i,j) \rightarrow (i,j+1)}$, and so forth, is summarized to the weighted sum with the diagonal corners weighted by factors $1/[(4 + 4/\sqrt{2}) \cdot \sqrt{2}]$, and horizontal and vertical neighbors weighted by $1/[4 + 4/\sqrt{2}]$.

The resulting images are displayed in Figure 5. The Euclidean case exhibits a good contrast, for example, for highlighting the vertically oriented highway. In case of line integrals of Pearson correlation, more curly structures get empha-

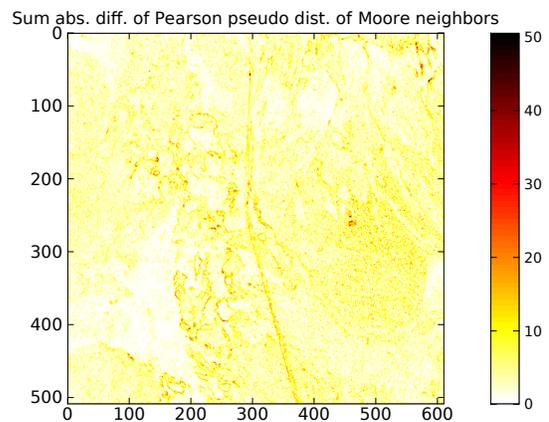


Fig. 6. Terrain image using weighted sums of absolute differences of line integrals of adjacent channels for the Pearson correlation measure.

sized. Yet, the image is less homogeneously colored, because neighbored spectra contribute asymmetrically to neighbored pixel colors, while Euclidean neighbors share the influence of their symmetric edge. Note the very different ranges of the color bars. This shows how little efforts are needed to rotate, i.e. transform, neighbored spectrum vectors from one into the other direction in order to reach maximum correlation between them, while their Euclidean distance might get rather large.

The spatial distribution of asymmetry in line integrals of Pearson correlation is visualized in Figure 6. A high degree of correspondence with the two previous images in Figure 5 can be observed. As shown, the range of differences between commuted line integrals can be quite substantial. Yet, the mechanism underlying the observed and spatially coherent structure and its connection with asymmetric efforts needed to rotate spectrum \mathbf{x}^i into the direction of spectrum \mathbf{x}^j and vice versa needs more systematic investigations. Especially the identification of spatially directed features is essential for morphological image operations for which the presented approach contributes a new direction.

V. CONCLUSIONS

A generalized way has been introduced for the calculation of partial covariance and correlation of attributes in multidimensional data. It contains standard covariance and correlation of Euclidean data as special case. The new approach is thus considered to be of interest for the unsupervised assessment of data attribute variability and attribute correlatedness in localized or sparsely connected data relationships. In image processing, such pixel neighborhood operations open a wide field of morphological operations, such as edge detection, and local feature enhancement.

For large data sets sparse connectivity is an essential prerequisite for being computationally feasible, because of the otherwise bi-quadratic runtime of $\mathcal{O}(d^2 \cdot n^2)$ for the generalized covariance matrix calculation. Although the distance pursuit algorithm is computationally quite demanding it is designed for being computed in parallel on graphics hardware,

because each thread requires exactly the same number of t steps composed of structurally identical operations which can be efficiently warped on hardware, cf. [24]. For Euclidean distance, the \mathbf{b} in the distance pursuit algorithm can be just replaced by the difference vectors of the arguments, and for Pearson correlation, a number of $t = 16$ steps provided very close results to those involving steps.

Partial generalized variance has been implemented for two measures, the Euclidean distance and Pearson correlation, but the framework allows to integrate other measures than these two. In the experiments with hyperspectral images both spectrum and channel relationships have been analyzed. As a main result, the eigenstructure of the partial correlation matrix for Pearson correlation turns out to be more complex than the one of standard correlation, but also localized, i.e. partial, variance helps to equalize the eigenspectrum. This means that Euclidean data at a global level of comparison is dominated by almost collinear 'redundant' vectors along the main eigenvector, that is, by very large pairwise differences, but the data vectors get spread out for better discrimination by using partial Euclidean or even partial Pearson correlation. Thus, instead of forcing basis orthogonalization in Euclidean space, alternative and localized measures may help to achieve the same goal for a better discrimination of data vectors. Another result stimulating upcoming research is the observation of asymmetric relationships between spectra that seem to be useful for creating contrast maps.

Objective quality assessment of unsupervised methods is a principal challenge. Here such an assessment is even more problematic, because no alternative method for the characterization of measure-specific attribute variability is known so far. If you find variance useful, you may find the generalized distance pursuit-based approach useful too. Rather than being interested in the 'Euclidean' attribute variation of strongly preprocessed and transformed data, making a combinatorial choice of methods in the data processing pipeline necessary, the ultimate goal here is the integrated analysis of the original data given a specified data measure. Besides contributions of formal solutions for other data measures, providing a real-time data exploratory analysis with different measures and neighborhood definitions is a great challenge for the presented approach, but a goal to be pursued in future.

ACKNOWLEDGMENTS

We thank Gordon Strickert, Institute of Flight Systems at the German Aerospace Center, and three anonymous reviewers for their helpful comments. This research is kindly supported by the DFG Graduiertenkolleg 1564 'Imaging new Modalities'.

REFERENCES

- [1] L. McDonnell and R. Heeren, "Imaging mass spectrometry," *Mass Spectrometry Reviews*, vol. 26, no. 4, pp. 606–643, 2007.
- [2] C.-I. Chang, Ed., *Hyperspectral Data Exploitation: Theory and Applications*. John Wiley & Sons, 2007.
- [3] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988. [Online]. Available: <http://www.jstor.org/stable/2685263>

- [5] N. Keshava, "A survey of spectral unmixing algorithms," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 55–78, 2003.
- [6] T. Villmann, S. Haase, F.-M. Schleif, B. Hammer, and M. Biehl, "The mathematics of divergence based online learning in vector quantization," in *Artificial Neural Networks in Pattern Recognition*, ser. Lecture Notes in Computer Science, F. Schwenker and N. El Gayar, Eds. Springer Berlin / Heidelberg, 2010, vol. 5998. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-12159-3_10
- [7] S. C. Johnson, "Hierarchical Clustering Schemes," *Psychometrika*, vol. 2, pp. 241–254, 1967.
- [8] R. M. Gray, "Vector quantization," *IEEE ASSP Magazine*, vol. 1, pp. 4–29, 1984.
- [9] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/spectral endmember extraction by multidimensional morphological operations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 9, pp. 2025–2041, 2002.
- [10] M. A. Veganzones and M. G. na, *Endmember Extraction Methods: A Short Review*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008, no. Part 3, ch. Knowledge-Based Intelligent Information and Engineering Systems, pp. 400–407. [Online]. Available: <http://www.springerlink.com/content/nm2132725k134663/>
- [11] H. J. Heijmans and J. B. Roerdink, Eds., *Mathematical Morphology and its Applications to Image and Signal Processing*, ser. Proceedings of the 4th International Symposium on Mathematical Morphology (ISMM'98), 1998.
- [12] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization," *Neural Networks*, vol. 15, pp. 1059–1068, 2002.
- [13] M. Strickert, J. Keilwagen, F. M. Schleif, T. Villmann, and M. Biehl, "Matrix metric adaptation for linear discriminant analysis of biomedical data," in *Bio-Inspired Systems: Computational and Ambient Intelligence*, ser. Lecture Notes in Computer Science, J. Cabestany, F. Sandoval, A. Prieto, and J. Corchado, Eds., vol. 5517. Springer, 2009, pp. 933–940.
- [14] M. Strickert, U. Seiffert, N. Sreenivasulu, W. Weschke, T. Villmann, and B. Hammer, "Generalized relevance LVQ (GRLVQ) with correlation measures for gene expression data," *Neurocomputing*, vol. 69, pp. 651–659, 2006.
- [15] M. Strickert, F.-M. Schleif, T. Villmann, and U. Seiffert, *Similarity-Based Clustering - Recent Developments and Biomedical Applications*, ser. Lecture Notes in Computer Science. Springer, 2009, vol. 5400, ch. Unleashing Pearson Correlation for Faithful Analysis of Biomedical Data, pp. 70–91.
- [16] N. Nielsen, J. Carstensen, and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping," *Journal of Chromatography*, vol. 805, pp. 17–35, 1998.
- [17] I. Scarminio and M. Kubista, "Analysis of correlated spectral data," *Analytical Chemistry*, vol. 65, no. 4, pp. 409–416, 1993. [Online]. Available: <http://pubs.acs.org/doi/abs/10.1021/ac00052a017>
- [18] W. Xu, C. Chang, Y. Hung, S. Kwan, and P. Chin Wan Fung, "Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5552–5563, dec. 2007.
- [19] E. Mwebaze, P. Schneider, F. M. Schleif, S. Haase, T. Villmann, and M. Biehl, "Divergence based classification and learning vector quantization," *Neurocomputing*, to appear April 2011.
- [20] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sensing of Environment*, vol. 44, no. 2–3, pp. 127–143, 1993, airborne Imaging Spectrometry. [Online]. Available: <http://www.sciencedirect.com/science/article/B6V6V-489S4HM-11/2/01726a3463e121148c95c08b2ab2b54a>
- [21] W. Wang, Z. Zhao, and H. Zhu, "Hyperspectral image compression method based on spectral statistical correlation," 2009, pp. 1–5.
- [22] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.
- [23] M. Strickert, N. Sreenivasulu, B. Usadel, and U. Seiffert, "Correlation-maximizing surrogate gene space for visual mining of gene expression patterns in developing barley endosperm tissue," *BMC Bioinformatics*, vol. 8, no. 165, 2007. [Online]. Available: <http://www.biomedcentral.com/1471-2105/8/165>
- [24] S. Rosario-Torres and M. Velez-Reyes, "Speeding up the MATLAB Hyperspectral Image Analysis Toolbox using GPUs and the Jacket Toolbox," 2009, pp. 1–4.