A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing

Hilda Deborah, Noël Richard, and Jon Yngve Hardeberg

***Abstract*—Distance functions are at the core of important data analysis and processing tools, e.g., PCA, classification, vector median filter, and mathematical morphology. Despite its key role, a distance function is often used without careful consideration of its underlying assumptions and mathematical construction. With the objective of identifying a suitable distance function for hyperspectral images so as to maintain the accuracy of hyperspectral image processing results, we compare existing distance functions and define a suitable set of selection criteria. Bearing in mind that the selection of distance functions is highly related to the actual definition of the spectrum, we also classify the existing distance functions based on how they inherently define a spectrum. Theoretical constraints and behavior, as well as numerical tests are proposed for the evaluation of distance functions. With regards to the evaluation criteria, Euclidean distance of cumulative spectrum (ECS) was found to be the most suitable distance function.**

***Index Terms*—Image processing, multidimensional signal processing.**

I. Introduction

QUANTITATIVELY measuring the difference between two objects is the core of many important image processing tasks, e.g., PCA, vector median filtering, mathematical morphology, etc. While for binary and grayscale images, the task of measuring distance or dissimilarity is rather straightforward, it is not the case for images of higher dimension such as color and hyperspectral images. Vector median filter by Astola *et al*. [1] was constructed based on the notion of aggregate distance which enables the ordering of multivariate data, which soon after was followed by other constructions of multivariate nonlinear filters using other different distance functions [2], [3]. Mathematical morphology, a nonlinear image processing framework, relies on the ordering of its input data. The core of mathematical morphology is therefore similar to that of nonlinear filter, and one way of ordering multivariate data is using distance functions [4], [5]. Different PCA algorithms and implementations are available, varying by the distance function behind its classifier [6], [7]. A comparative study on dissimilarity measures for the use in content-based image retrieval is available, trying to identify which among the existing dissimilarity or distance functions are suitable for such a task [8]. In hyperspectral image domain, several distance-based method to relate two data sets or image regions are available [9]. All the previously mentioned works which use distance function illustrates that a distance function is indeed at the core of image processing especially for multivariate data.

Distance functions or metrics are particularly essential for image quality (IQ) assessment. IQ metrics typically compute quality values locally, either pixel-wise or spatially, resulting in a quality map. The information given by a quality map will depend on the IQ metric behind, e.g., simple distortion measure using Euclidean distance, structural similarity (SSIM) index [10], blur index [11], and IQ according to human judgment [12], [13]. Several spectral quality metrics that are in use in remote sensing are spectral similarity value (SSV) [14] and general image-quality equation (GIQE) [15]; Kerekes *et al.* have provided the comparative evaluation of them with some other spectral quality metrics [16]. In general, IQ metrics can be categorized into no-reference, reduced reference, and full reference quality metrics. Shresta *et al.* [17] evaluated several IQ metrics for the purpose of spectral imaging systems quality assessment. In their work, existing IQ metrics were categorized into basically full reference metrics and some task specific metrics. In most of these existing IQ metrics, the mathematical formulations are typically variations of basic distance or similarity function, whether with incorporating weighting factors or by combining basic distance functions.

Before proceeding into the use of distance functions in intermediate or high-level image processing or in IQ assessment, a careful consideration has to be made regarding which distance or metric to use and the characteristics of the problem at hand. Each of the existing distance functions has its inherent assumption that is due to the mathematical construction. Take the distance functions from the Minkowski family, e.g., Manhattan and Euclidean distance functions, as an example. In many cases, these functions are thought to be general and useful for any kind of input data. However, Minkowski distance functions are defined on Euclidean space, and thus are only valid when the input data is a Euclidean vector. In addition to that, many distance functions were originally developed

Manuscript received September 15, 2014; revised December 10, 2014; accepted January 19, 2015. Date of publication March 17, 2015; date of current version July 30, 2015.

H. Deborah is with The Norwegian Colour and Visual Computing Laboratory, Gjøvik University College, Gjøvik 2815, Norway and also with SIC Laboratory, XLIM Research Institute UMR CNRS 7252, Futuroscope Chasseneuil Cedex 86962, France (e-mail: hildad@hig.no).

N. Richard is with SIC Laboratory, XLIM Research Institute UMR CNRS 7252, Futuroscope Chasseneuil Cedex 86962, France (e-mail: richard@sic.univ-poitiers.fr).

J. Y. Hardeberg is with Colorlab, Gjøvik University College, Gjøvik 2815, Norway (e-mail: jon.hardeberg@hig.no).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2015.2403257

TABLE I
NOTATIONS

$d(x,y)$	Distance value between two objects, x and y.
η_p	Parameter evolution, i.e., standard deviation for translation and magnification factor for magnitude change case, $p \subset [0, np-1]$, $\eta_{\min} = \eta_0$, $\eta_{\max} = \eta_{np-1}$
η_{sat}	Saturation point
dnr	Dynamic range of a distance response
S_1, S_2	Two spectral reflectance signals, $S_1 = \{s_{1,\lambda}\}$
λ	Wavelength or spectral band, $\lambda \subset [0, nb-1]$
nb	Number of wavelengths

for some specific tasks, e.g., spectral angle mapper [18] and Levenshtein distance [19]. Last but not least, strictly speaking, a distance function is not necessarily a metric for there are theoretical requirements for a metric which are different from those of a distance function.

This paper is organized as follows. The objective criteria needed to evaluate the existing distance functions are provided in Section II. As we classify the existing distance functions based on how they consider a spectrum, each of the sections starting from Sections III to VI is dedicated for each of the distance categories or families. Finally, the result of this comparative study is concluded in Section VII. Table I is provided to summarize all the mathematical notations used throughout this article.

II. Objective Evaluation Criteria

Before proceeding with each of the criteria, note that the evaluation will be carried out for reflectance spectra in hyperspectral imaging setting. Here, we define hyperspectral image as having many *spectrally contiguous bands*.

A. Simulation of Basic Spectra

To evaluate the performance of a distance function with regard to its theoretical properties, simple artificial reflectance spectra will be used. A basic reflectance spectrum is simulated with a theoretical Gaussian distribution function. Gaussian distribution function is selected with the purpose of showing the limitations of the existing distance functions and not to model the nature of a spectrum. More complex distribution will yield more complex distance response and therefore is not able to explain the behavior of the distance function or metric under evaluation.

Based on our previous observation on the spectral reflectances of pigments, we simulate hue change and brightness or illumination change with basic mathematical transformations, i.e., translation and magnitude change, respectively. Simulated spectral reflectances for each case of translation and magnitude change are shown in Fig. 1. The two sets of simulated spectral reflectances are going to be used to obtain a distance response which in turn will be used to evaluate the corresponding distance function.

B. Theoretical Properties and Numerical Constraints

There are several means to quantify similarity or dissimilarity between two objects, i.e., distance, similarity, or divergence measures. In this work, we strictly evaluate the existing

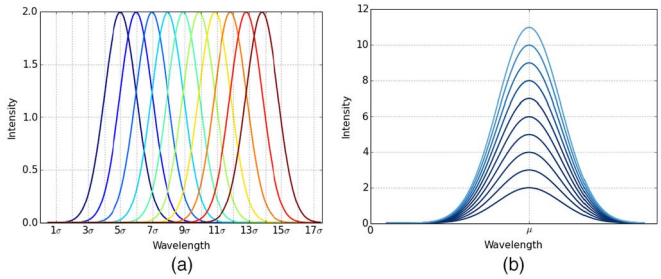

Fig. 1. Simulation of a basic spectrum using theoretical Gaussian function. Hue change and illumination change in perceptual level is simulated by (a) translation and (b) magnitude change applied to the basic spectrum, respectively.

measures to theoretical requirements of a distance and metric. Below we can see several theoretical properties. A distance function must possess the properties of reflexivity, nonnegativity, and symmetry. A metric is stricter than a distance, it has to satisfy nonnegativity, symmetry, identity of indiscernibles, and triangular inequality properties.

Reflexive $d(x,x) = 0$
Nonnegative $d(x,y) \geq 0$
Symmetric $d(x,y) = d(y,x)$
Identity of indiscernibles $d(x,y) = 0 \Leftrightarrow x = y$
Triangular inequality $d(x,y) \leq d(x,z) + d(z,y)$

As a spectrum is translated or magnified further from a referenced spectrum, see Fig. 1, we expect an increase in its distance response. This then leads to an additional constraint which is *monotonically increasing* (nondecreasing is included), see Fig. 2(b). The relation shown in (1) is induced by three requirements, i.e., triangular inequality, identity of indiscernibles, and monotonically increasing. η_p is translation or magnification factor evolution and $d_k(\eta_p)$ is the response of distance function k at η_p. As we will see later, many of the existing distance functions are not able to respond proportionally to the given parameter evolution, they saturate after a certain number of evolution. To measure this, saturation point η_{sat} information as shown in (2) will be given. Another additional information is dynamic range dnr of a distance function, see (3).

$$\frac{d_k(\eta_p)}{d_k(\eta_{\max}) - d_k(\eta_{\min})} \geq \frac{\eta_p}{\eta_{\max} - \eta_{\min}} \tag{1}$$

$$\eta_{sat} = \bigwedge \{\eta_p : \arg\max d_k(\eta_p)\} \tag{2}$$

$$dnr = d_k(\eta_{\max}) - d_k(\eta_{\min}) \tag{3}$$

C. Real Pigment Reflectance Spectra

To evaluate the performance of distance measures on real hyperspectral data with all its spectral noise or variations, hyperspectral images of two pigment patches as shown in Fig. 3 will be used. Defining a theoretical absolute white reflectance spectrum as the reference, distance responses to each of the pixels located under the yellow horizontal lines will be obtained for each distance function. This assessment is intended to qualitatively measure the performance of a distance function in classifying the pixels into their respective classes of color shades, in the presence of spectral noise and variations. An ideal

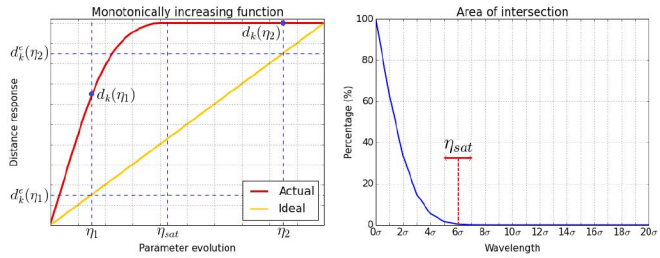

Fig. 2. (a) Saturation point η_{sat} as the location where the distance response reaches its maximum. (b) Saturation of a distance function is in fact related to the amount of intersection region between two spectra, if it saturates it will be around the first point of zero intersection.

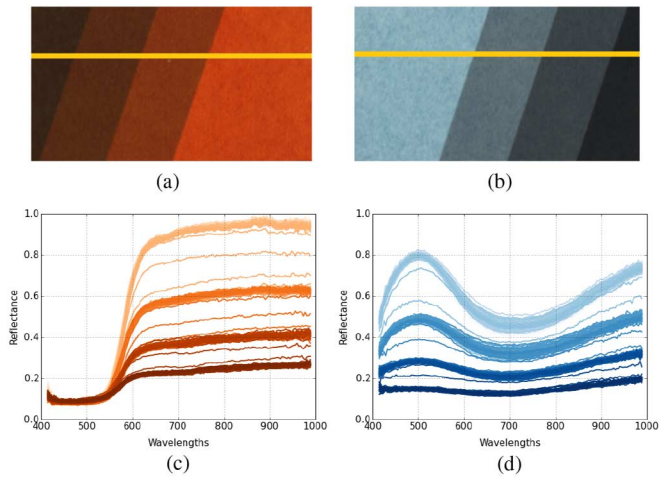

Fig. 3. Two spectrally acquired pigment patches to be used to evaluate distance measures performances. The pixels under the yellow horizontal lines will be used to see whether a distance measure is able to separate pixels into groups of different color shades. (a–c) 23800, patch and spectra; (b–d) 44450, patch and spectra.

distance function is expected to give a response such that is shown in Fig. 4, i.e., a step function. As the hyperspectral images of both pigment patches have noise and variations, the step function is expected to fluctuate as well. However, the fluctuations must not be significant such that the intraclass variations are smaller than the interclass distance.

III. Spectrum as Vector in Euclidean Space

A. Hypothesis of Validity

In Euclidean space, a vector is represented as n-tuples where each of the axes or dimensions is orthogonal to each other. Thus, it inherently assumes that the spectral bands or channels are orthogonal and do not correlate. Furthermore, in Euclidean space, the bands are said to have no particular order, or in other words, distance between two spectra will remain unchanged whatever spectral bands ordering is used. Such assumptions are, however, false, because it is known that in hyperspectral imaging the neighboring bands are highly correlated [20]. Using distance functions that fall into this category would then ignore the higher order relationship or correlation between neighboring spectral bands and will result in inaccurate measurement.

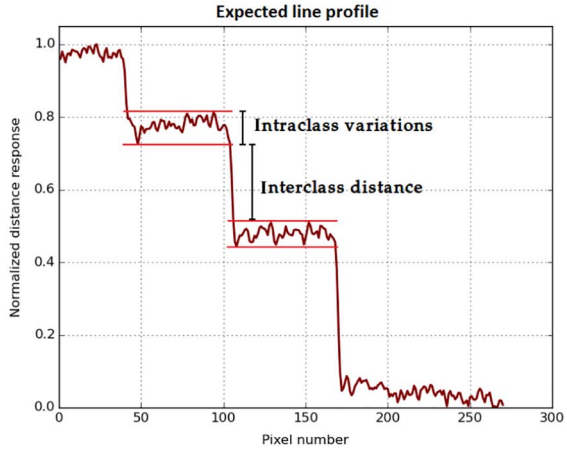

Fig. 4. Distance function is expected to give this response, i.e., a noisy step function, when given a theoretical absolute white reflectance spectrum as reference and spectral reflectance signals from the pigments patches in Fig. 3.

B. Distance Functions

Many of distance functions in this category can be generalized as variations of Minkowski formula, see (4). By varying the Minkowski order p, we obtain Manhattan ($p = 1$), Euclidean ($p = 2$), and Chebyshev ($p = \infty$) distance functions, which are also metrics. Chebyshev distance function can also be written as in (5). Another variation is fractional Minkowski $p < 1$, which was shown by Aggarwal *et al.* [21] to significantly improve the effectiveness of clustering algorithms for high-dimensional data; fractional Minkowksi is, however, not a metric as it violates triangular inequality [22].

Other than varying the order of Minkowski, other distance functions vary on the weighting function. Canberra in (6) [23], Sørensen [24], Gower [25], Kulczynski [23], and Lorentzian [23] are variations of weighted Manhattan distance functions. Examples of weighted Euclidean are root mean square error (RMSE) in (7) and the two χ^2 distances in (8) and (9). Geman-McClure function in (12) [26] can be considered as squared Euclidean distance but normalized individually. Among all these distance functions, only several will be evaluated as mathematically they are very similar.

$$d_p(S_1, S_2) = \left(\sum_{\lambda} |s_{1,\lambda} - s_{2,\lambda}|^p \right)^{\frac{1}{p}} \tag{4}$$

$$d_{Che}(S_1, S_2) = \max_{\lambda}(|s_{1,\lambda} - s_{2,\lambda}|) \tag{5}$$

$$d_{Can}(S_1, S_2) = \sum_{\lambda} \frac{|s_{1,\lambda} - s_{2,\lambda}|}{s_{1,\lambda} + s_{2,\lambda}} \tag{6}$$

$$d_{RMS}(S_1, S_2) = \sqrt{\frac{1}{nb} \sum_{\lambda} (s_{1,\lambda} - s_{2,\lambda})^2} \tag{7}$$

$$d_{\chi_1^2}(S_1, S_2) = \sum_{\lambda} \frac{(s_{1,\lambda} - s_{2,\lambda})^2}{(s_{1,\lambda} + s_{2,\lambda})^2} \tag{8}$$

$$d_{\chi_2^2}(S_1, S_2) = \frac{1}{2} \sum_{\lambda} \frac{(s_{1,\lambda} - s_{2,\lambda})^2}{s_{1,\lambda} + s_{2,\lambda}} \tag{9}$$

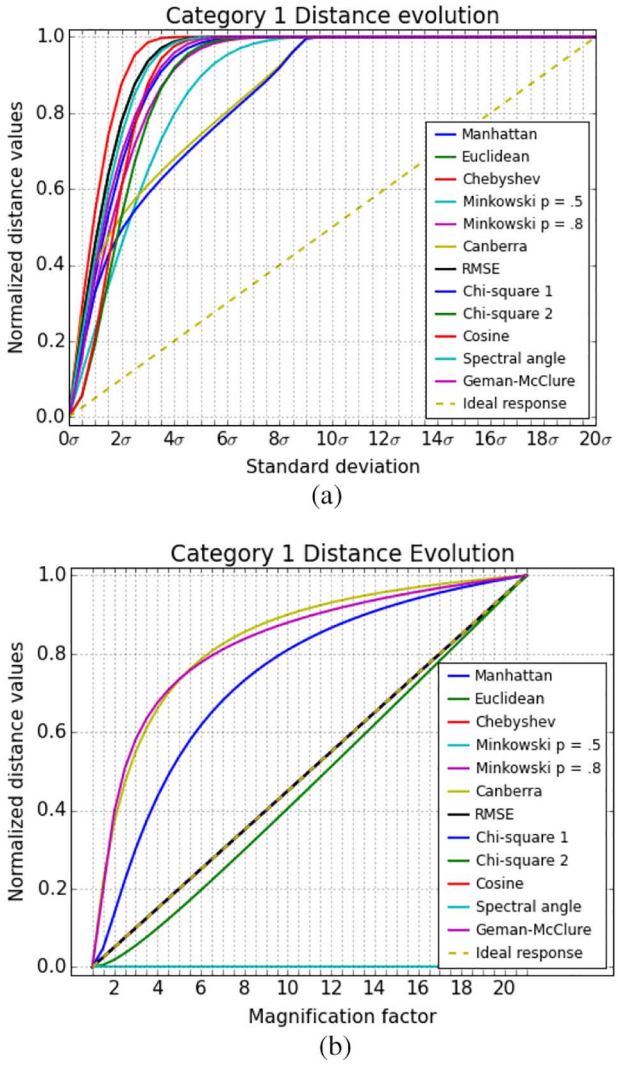

Fig. 5. Distance responses of distance functions that consider a spectrum as a vector in Euclidean space, shown in normalized values for (a) translation and (b) magnitude change cases. All distance function saturates for the translation case. Spectral angle, cosine, and χ_2^2 distance responses lie below the ideal linear response and therefore are not metrics.

$$d_{Cos}(S_1, S_2) = 1 - \frac{\sum_\lambda s_{1,\lambda} \cdot s_{2,\lambda}}{\sqrt{\sum_\lambda s_{1,\lambda}^2}\sqrt{\sum_\lambda s_{2,\lambda}^2}} \tag{10}$$

$$d_{SAM}(S_1, S_2) = \cos^{-1}\left(\frac{\sum_\lambda s_{1,\lambda} \cdot s_{2,\lambda}}{\sqrt{\sum_\lambda s_{1,\lambda}^2}\sqrt{\sum_\lambda s_{2,\lambda}^2}}\right) \tag{11}$$

$$d_{GMC}(S_1, S_2) = \sum_\lambda \frac{(s_{1,\lambda} - s_{2,\lambda})^2}{1 + (s_{1,\lambda} - s_{2,\lambda})^2} \tag{12}$$

Some other distance functions calculate the difference between two spectra based on angular information. Two of such distance functions are cosine distance in (10) and spectral angle in (11). Spectral angle was initially used in spectral angle mapper [18] and has since been used extensively in remote sensing field.

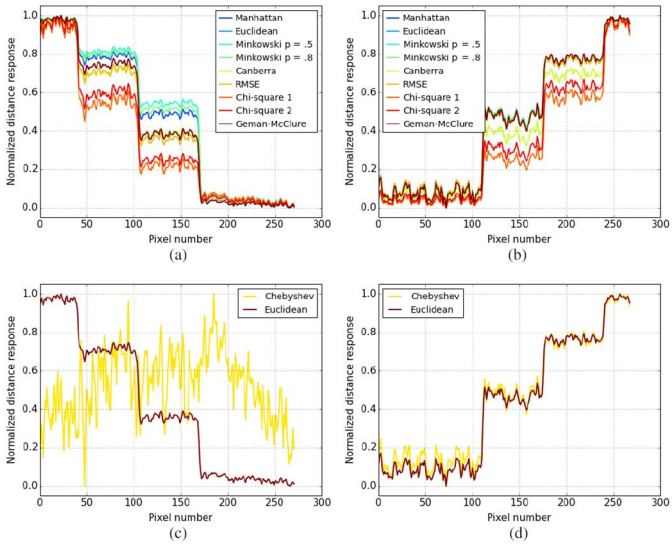

Fig. 6. Responses of category 1 distance functions on two pigment patches in normalized values, i.e. (a–b) Minkowski on 23800 and 44450, and (c–d) Chebyshev on 23800 and 44450. Different variations of Minkowski distances behave identically, with magnitude being the only difference. Chebyshev is, however, shown to be rather unstable.

C. Evaluation

Two sets of simulated spectral reflectance signals as shown in Fig. 1 are used to evaluate the theoretical properties of distance functions in this category. Distance responses for both cases of translation and magnitude change are provided in Fig. 5. For an increasing standard deviation or magnitude change, we are expecting a proportionally increasing distance response. However, as shown in Fig. 5(a) none of the distance functions is giving the desired behavior as all of them saturates, which as mentioned before is related to the intersection between two spectra, see Fig. 2(b). From observing distance responses for magnitude change in Fig. 5(b), we know that spectral angle, cosine, and χ_2^2 distances are not metrics as their responses are below the ideal linear response. To be specific, χ_2^2 violates triangular inequality, while spectral angle and cosine distance violate identity of indiscernibles due to their zero responses.

Regarding many variations of Minkowski distances, several points can be concluded. Even though all these distance functions saturate for translation case, functions with larger Minkowski order p will saturate at an earlier point than those with smaller p. Variations on the weighting function would not only affect the saturation but also curve shape of the distance response and therefore it will affect whether the function is a metric or not, e.g., χ_2^2 distance in Fig. 5(b). If we observe Fig. 6(a) and (b), which are distance responses to the two sets of pigment spectral reflectances, it is clear that variations of Minkowski either in its order or weighting function result in similar behavior, differing only by its magnitude. Chebyshev distance is, however, not stable as compared to, e.g., Euclidean distance, as seen in Fig. 6(c) and (d) where in 23 800 it fails in correctly classifying the spectral reflectances.

Spectral angle and cosine distance also show unstable behaviors on the two pigment patches. In Fig. 7(a), both distance

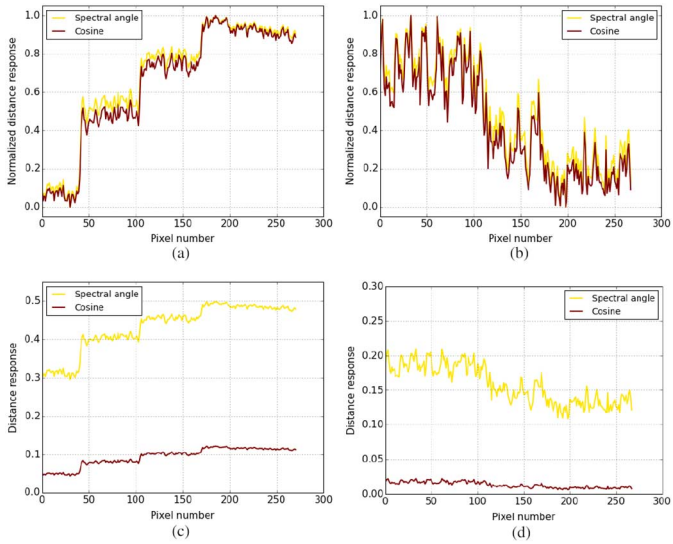

Fig. 7. Responses of angular distance functions (category 1) on pigment (a–c) 23 800 and (b–d) 44 450, shown in normalized and real response values.

functions are able to identify the different classes, with some expected spikes due to the variations in the signals. However in Fig. 7(b) both functions are not able to tell the difference between classes. Spectral angle itself was originally used such that the magnitude differences between spectra caused by illumination change or shading effect have little significance on the resulting distance response [18]. With magnitude difference being the most prominent difference between spectra in pigment 23 800 and 44 450 [see Fig. 3(c) and (d)] and knowing that cosine distance and spectral angle represent distance in values between 0 and 1, it is expected that the dynamic ranges of distance responses to 23 800 and 44 450 are rather small (see Fig. 7(c) and (d). The reason why both cosine distance and spectal angle completely fail at 44 450 is because the spectral reflectances of 44 450 contain two Gaussian curves and angular distances are proven to be limited in such case.

To summarize, so far in this category we are not able to find a single distance function that responds proportionally to the increasing parameter evolution, mainly for translation case; they always saturate when there is no overlapping between two spectra. The summary of theoretical and numerical properties of the distance functions is provided in Table II.

IV. Spectrum as n-Dimensional Data in Manifold

A. Hypothesis of Validity

While assuming that a spectrum is a vector in Euclidean space is incorrect, it can be assumed as an n-dimensional data whose spectral variations form a manifold. With this assumption, two different approaches in computing a distance function can be explored. The first approach addresses a distance function as a direct measure between two spectra, e.g., Goodness-of-fit coefficient (GFC) [27]. On the other hand, the second approach explores this question within the dimensionality reduction purpose [28]–[30].

TABLE II
Summary of the Theoretical Properties of Several Distance Functions in Category 1, I.E., Those That Consider a Spectrum as a Vector in Euclidean Space

Fun	Var.	Triangular inequality	Identity of indiscernibles	Mon. increasing
Category 1: Vector in Euclidean space				
Minkowski	p = 1	✓	✓	✓
	p = 2	✓	✓	✓
	p = ∞	✓	✓	✓
	p = 0.5	x	✓	✓
	p = 0.8	x	✓	✓
Canberra	–	✓	✓	✓
RMSE	–	✓	✓	✓
χ^2	χ^2_1	✓	✓	✓
	χ^2_2	x	✓	✓
Angular	Cosine	✓	x	✓
	Spectral angle	✓	x	✓
Geman-McClure	–	✓	✓	✓

*Reflexivity, nonnegativity, and symmetry are all satisfied.

B. Distance Functions

GFC is one example of the straightforward distance expressions in this category. It was originally used to test the accuracy of reconstructed daylight spectra [27]. It is based on Schwartz's inequality and in Euclidean space it is basically the cosine distance or the cosine of spectral angle. Since GFC is a similarity function, the distance function is as shown in (13).

$$d_{GFC}(S_1, S_2) = 1 - \frac{\left|\sum_\lambda s_{1,\lambda} \cdot s_{2,\lambda}\right|}{\sqrt{\left|\sum_\lambda s^2_{1,\lambda}\right|}\sqrt{\left|\sum_\lambda s^2_{2,\lambda}\right|}} \tag{13}$$

In parallel, Tenenbaum *et al.* [28] and Roweis *et al.* [29] proposed a new way to consider a set of spectra, i.e., as manifolds where each spectrum is represented by a node in a graph. With this new point of view, the dimensionality of spectral variations can be reduced using the proposed algorithms, i.e., isometric feature mapping (Isomap) [28] and locally linear embedding (LLE) [29]. Several more advanced algorithms are locally linear coordination (LLC) [30] and manifold charting [31].

In Isomap, a distance between two spectra can be expressed as a sum of local distances between these two spectra and their immediate neighbors in a trusted set of spectra; local distances are typically computed with Euclidean distance or dot product. Isomap is considered as a global method as it preserves the global geometric features of the initial spectral set. Several applications of Isomap for hyperspectral images are available [31], [32]. The LLE algorithm considers that each spectrum is a weighted combination of a trusted set of nearest neighbors. It then computes the minimally distorting low-dimensional barycentric embedding [29], [31]. The main idea behind the previous algorithms is that the n-dimensional values mostly consist of noise or inaccurate features and that the lattice or neighborhood structure is more important than the values of the data.

$$d(S_a, S_b) = \min \sum_{i=0}^{P-1} d_2(S_i, S_{i+1}) \tag{14}$$

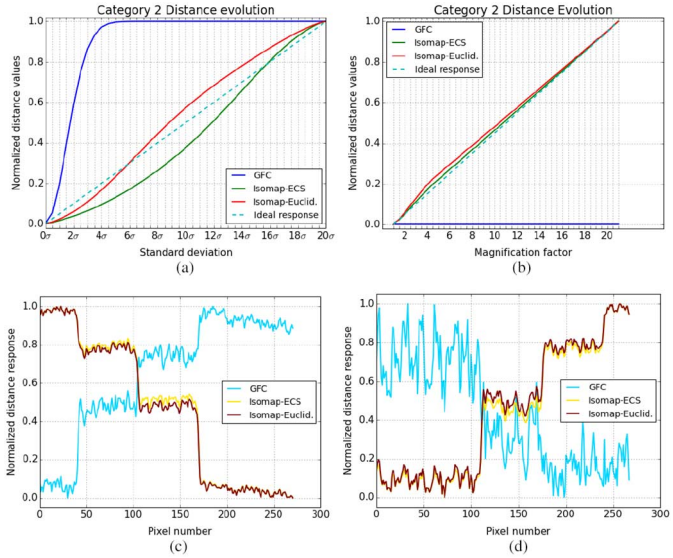

Fig. 8. Theoretical behavior evaluation of category 2 distance functions with regard to (a) translation and (b) magnitude change, shown in normalized values. Their numerical responses to spectral reflectances of pigment 23 800 and 44 450 are shown in (c) and (d), respectively.

As the second approach of distance function in manifold is centered around the assessment of local distance between neighboring spectra, we are only going to evaluate and consider as valid measure the Isomap construction as it preserves the global structure of the lattice of spectra. Let $\mathcal{S} = \{S_i, i = [0, L-1]\}$ be a set of spectra, the distance $d(S_a, S_b)$ between two spectra S_a and S_b is expressed as the sum of local distances between S_i and S_j, where the list of S_i forms the shortest path $\mathcal{P}$ between S_a and S_b. Or in other words, the shortest path between S_a and S_b is formed by a chain of P spectra, see (14), where $S_0 = S_a$, $S_L = S_b$, $P \leq L$, and $S_i \neq S_j, \forall i \neq j$.

C. Evaluation

Theoretical evaluation on GFC with regards to translation and magnitude change is as shown in Fig. 8(a) and (b), respectively. It is shown that for translation GFC is not able to proportionally reflect the input parameter evolution; it saturates. For magnitude change, it completely fails at measuring distance between two spectra that only differ in magnitude. Such stationary response for magnitude change implies that it violates the property of identity of indiscernibles, and therefore GFC is not a metric (see Table III).

Regarding Isomap, due to the particular nature of the metrics on lattice, distance between two spectra must be assessed inside a lattice. Consequently, all the possible spectrum transformations including translation and magnitude change need to be combined to construct the lattice support. Only after this, the distance between two spectra can be computed using the Isomap algorithm. Isomap relies on other distance functions to measure local distances, which in this case are Euclidean distance and Euclidean distance of cumulative spectrum (ECS), see (17). Fig. 8(a) and (b) shows the responses of distances in lattice using Isomap. Using the lattice structure allows assessing

TABLE III
SUMMARY OF THE THEORETICAL PROPERTIES OF SEVERAL DISTANCE FUNCTIONS IN CATEGORY 2, I.E., THOSE THAT CONSIDER A SPECTRUM AS AN N-DIMENSIONAL DATA IN MANIFOLD

Fun	Var.	Triangular inequality	Identity of indiscernibles	Mon. increasing
Category 2: N-dimensional data in manifold				
GFC	–	✓	x	✓
Isomap	Euclidean	x	✓	✓
	ECS	x	✓	✓

*Reflexivity, nonnegativity, and symmetry are all satisfied.

spectral distance with Euclidean expression when the translation transformations are applied. As the the lattice is relatively dense, due to the closeness of the transformed spectra, there are only few differences between the behavior of the two local distances, i.e., Euclidean distance and ECS.

Using spectral reflectance signals of pigment 23 800 and 44 450, the behaviors of GFC and Isomap are evaluated on real spectra with known properties. As mentioned before, the expected response is a step function as shown in Fig. 4. The actual responses are as shown in Fig. 8(c) and (d), where for GFC only the response to 23 800 matches the expected one. Having identical behavior to that of angular distances due to GFC's similar mathematical construction, the failure of classifying signals in 44 450 is due to its bimodal construction, including the two Gaussian functions which reduce the sense of angle measurement. For Isomap, the responses to both pigment patches are as expected. Note that due to the lattice density, the distance responses are more directly related to the local distances rather than to Isomap algorithm. A more complete study should be developed to assess the relationship between the lattice density and the distance accuracy, and evidently, the computational cost that is due to the shortest path computation in Isomap algorithm.

V. SPECTRUM AS DISTRIBUTION

A. Hypothesis of Validity

By assuming that a spectrum is a distribution, one assumes that spectral bands are ordered and that closer spectral bands have more correlation than those which are further away. Such assumptions do agree with the common knowledge about spectral data and its high correlation between neighboring spectral bands [20]. As a distribution, a spectrum could then be represented as either a probability density function (pdf) or cumulative distribution function (cdf).

$$R(S_1, S_2) = \frac{\sum_\lambda (s_{1,\lambda} - \overline{s_{1,\lambda}})(s_{2,\lambda} - \overline{s_{2,\lambda}})}{\sqrt{\sum_\lambda (s_{1,\lambda} - \overline{s_{1,\lambda}})^2} \sqrt{\sum_\lambda (s_{2,\lambda} - \overline{s_{2,\lambda}})^2}}$$

$$d_{Cor}(S_1, S_2) = 1 - \frac{1 + R(S_1, S_2)}{2} \tag{15}$$

$$d_{Smi}(S_1, S_2) = 1 - \frac{\sum_\lambda \min(s_{1,\lambda}, s_{2,\lambda})}{\min(\sum_\lambda s_{1,\lambda}, \sum_\lambda s_{2,\lambda})} \tag{16}$$

$$d_{ECS}(S_1, S_2) = \left(\sum_\lambda \left| \int s_{1,\lambda} d\lambda - \int s_{2,\lambda} d\lambda \right|^2 \right)^{\frac{1}{2}} \tag{17}$$

$$d_{Jef}(S_1,S_2)=\sum_{\lambda}\left(s_{1,\lambda}\log\frac{2.s_{1,\lambda}}{s_{1,\lambda}+s_{2,\lambda}}+s_{2,\lambda}\log\frac{2.s_{2,\lambda}}{s_{1,\lambda}+s_{2,\lambda}}\right) \quad (18)$$

$$d_{Pea}(S_1,S_2)=\sum_{\lambda}\frac{(s_{1,\lambda}-m_{\lambda})^2}{m_{\lambda}};\quad m_{\lambda}=\frac{s_{1,\lambda}+s_{2,\lambda}}{2} \quad (19)$$

$$d_{SqC}(S_1,S_2)=\sum_{\lambda}\left(\sqrt{s_{1,\lambda}}-\sqrt{s_{2,\lambda}}\right)^2 \quad (20)$$

B. Distance Functions

Spectral correlation, initially used for spectral correlation mapper (SCM) [33], is a similarity function that is used in remote sensing. Aiming at suppressing the effect of shading in distance computation of remote sensing data, spectral correlation does not take into account the magnification difference between spectra. Although spectral angle also aims the same purpose, spectral correlation was said to be more accurate to spectral angle. To fit our context of distance, the mathematical formulation of spectral correlation is changed into that is shown in (15). The most common distance functions for distribution are those that measure the amount of intersection between two pdfs or histograms, e.g., Smith distance in (16). Other dissimilarity measures that will be taken into account are Jeffrey divergence in (18) [34], Pearson χ^2 in (19) [35], squared chord distance in (20) [36], and ECS in (17). Among all the previous measures, ECS uses cdf and other measures use pdf. A different approach to measure dissimilarity between spectra is by using transformation cost based distance, i.e., Earth Mover's distance (EMD) [37] and combined EMD [38].

C. Evaluation

The response of distance functions in this category to theoretical evaluation is as shown in Fig. 9 and Table IV. With regards to translation, we obtain three distance functions that do not saturate, i.e., EMD, combined EMD, and ECS. In the case of magnitude change, only combined EMD and ECS satisfy the triangular inequality and identity of indiscernibles, as the responses are exactly on top of the ideal response, while the other functions are below it. Spectral correlation and Smith distance are giving zero responses, which violate the identity of indiscernibles property. For spectral correlation, this result is expected as it was designed to remove the effect of shading or magnitude difference. For Smith distance, if two spectra are only different by magnitude, due to the mathematical formulation the similarity would be 1 and the distance 0.

To further evaluate the performances of distance functions in this category, as it was carried out for the previous categories, we test them on pigment 23 800 and 44 450 and the corresponding results are as shown in Fig. 10. In this results, however, EMD and Combined EMD were not evaluated, for a reason that will be explained later. With regard to the two pigments, most of the evaluated distance functions are able to classify the signals to their respective color shades, except for

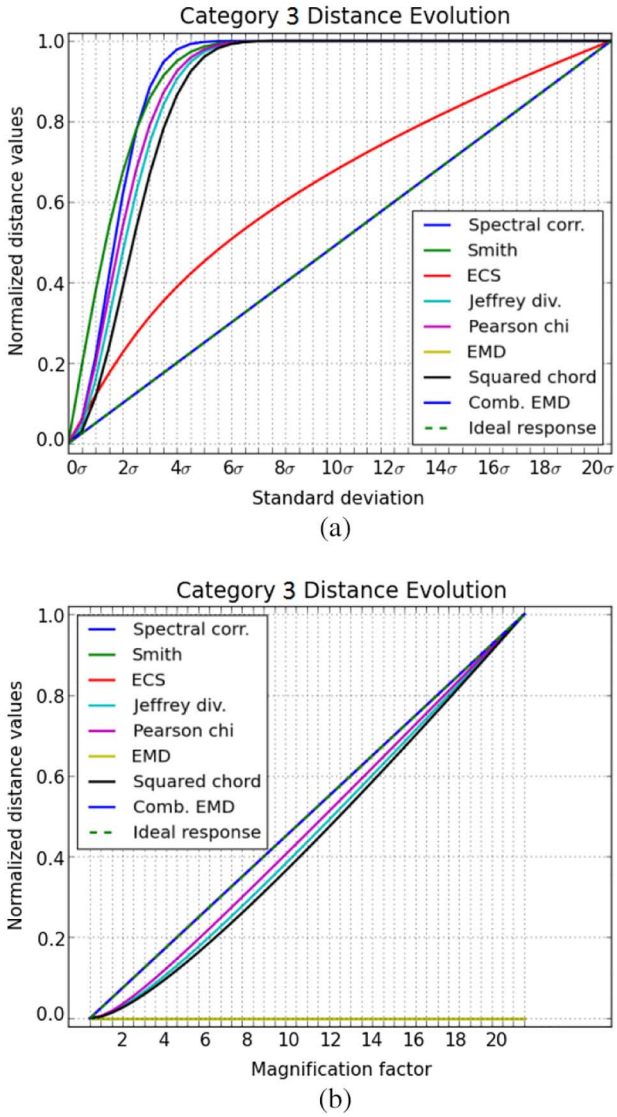

Fig. 9. Responses of distance functions that consider a spectrum as a distribution with regard to theoretical evaluation. For both (a) translation and (b) magnitude change cases, only Combined EMD and ECS that are able to give responses that do not saturate and lie on top of or above the ideal response.

TABLE IV
SUMMARY OF THE THEORETICAL PROPERTIES OF SEVERAL DISTANCE FUNCTIONS IN CATEGORY 3, I.E., THOSE THAT CONSIDER A SPECTRUM AS A DISTRIBUTION

Fun	Var.	Triangular inequality	Identity of indiscernibles	Mon. increasing
Category 3: Distribution				
Spectral-correlation	–	✓	x	✓
Smith	–	✓	x	✓
ECS	–	✓	✓	✓
Jeffrey div.	–	x	✓	✓
EMD	– Comb.	✓ ✓	x ✓	✓ ✓
Pearson χ^2	–	x	✓	✓
Squared-chord	–	x	✓	✓

*Reflexivity, nonnegativity, and symmetry are all satisfied.

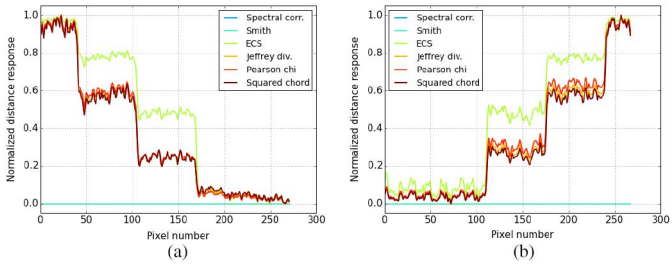

Fig. 10. Responses of distance functions in category 3, except EMD and Combined EMD, to real spectral reflectance signals taken from pigment (a) 23 800 and (b) 44 450.

spectral correlation and Smith distance. The results for these two distances agree with their corresponding theoretical evaluation as the spectra of the two pigments are predominantly different only by magnitude.

In the implementation of EMD, an optimization loop is used to obtain a distance value. With simulated spectra using Gaussian functions, this distance worked as expected. When the real spectral data were used to test it; however, the function failed to give any result because it reached the limit of optimization loop. Such behavior is due to the nature of real hyperspectral data that has a lot of noise and variations. To support this statement, we simulate several spectral reflectance signals using Gaussian functions with additive uniform noise. As our pigment spectral data are essentially different by their magnitude, we simulate the spectra only for the case of magnitude change [see Fig. 11(a)], in which a stationary or all-zero response is expected from EMD. The response is, however, showing an unpredictable behavior when noise or variations are present, see Fig. 11(b). Such response is due to the noise or variations that were added to the simulated spectra.

As a conclusion, in this category we find that according to the theoretical evaluation, two distance functions are giving the desired response, i.e., they do not saturate for the case of translation and magnitude change. The two functions are combined EMD and ECS. EMD was found to be unstable when dealing with noisy data. Since the core of combined EMD is EMD, in this category it is only ECS that is the most suitable distance function for hyperspectral data. Nevertheless, regarding the definition of distribution function, a remark must be added to the fact that a spectrum cannot be defined as a distribution because the integral of its energy or intensity across the wavelength is not equal to 1.

VI. Spectrum as Sequence

A. Hypothesis of Validity

This category is inspired by distance functions that are used in DNA matching. If a spectrum is considered as a sequence, it means that spectral bands or channels are ordered by its position in the sequence. Another inherent assumption that is taken is that the values of the signals come from a finite set. The latter assumption is incorrect as spectral data are represented by real numbers which is an infinite set of values.

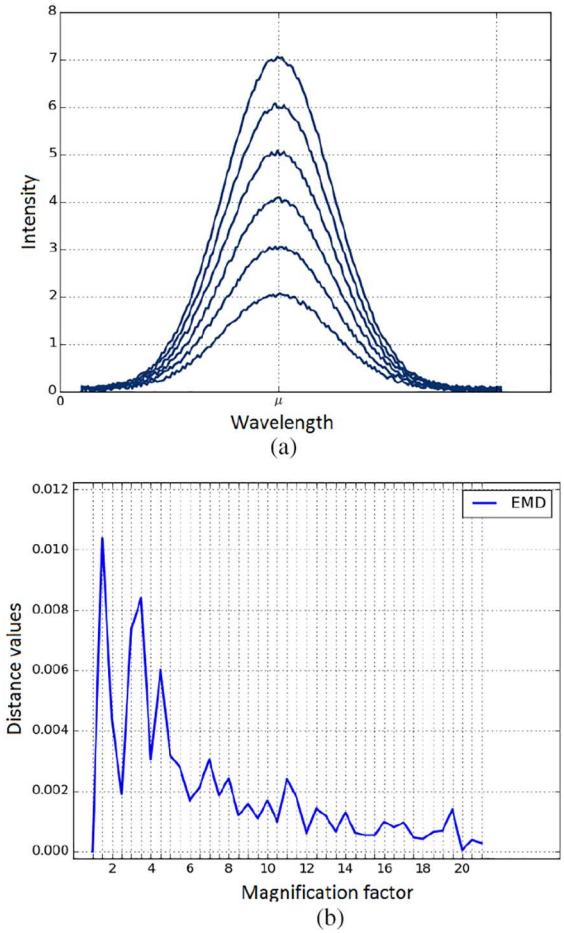

Fig. 11. Simulated spectra with additive uniform noise used to test the behavior of EMD when noise or variations are present. (a) Noisy simulated spectra. (b) EMD response.

B. Distance Functions

Distance functions that fall into this category are those that are used mainly for string matching. Hamming distance [39] measures the difference between two sequences based on the minimum number of character substitutions required to transform one string into another. This distance is only applicable if the two sequences are of the same length. In Levenshtein distance [19], instead of only taking into account character substitution, it considers all the single character edit, i.e., insertion, deletion, and substitution. Levenshtein distance is therefore suitable for spectra or sequences of unequal length. Damerau–Levenshtein [40] add another operation to those considered in Levenshtein, i.e., transposition of two adjacent characters, saying that it corresponds to human misspelling.

C. Evaluation

The responses of distance functions in this category to theoretical evaluation are as shown in Fig. 12 and Table V. For translation case, all three distance functions saturate. For magnitude change case, all the distance functions saturate as soon as the magnification factor is bigger than 1. This behavior is not unexpected. All three distance functions measure the difference

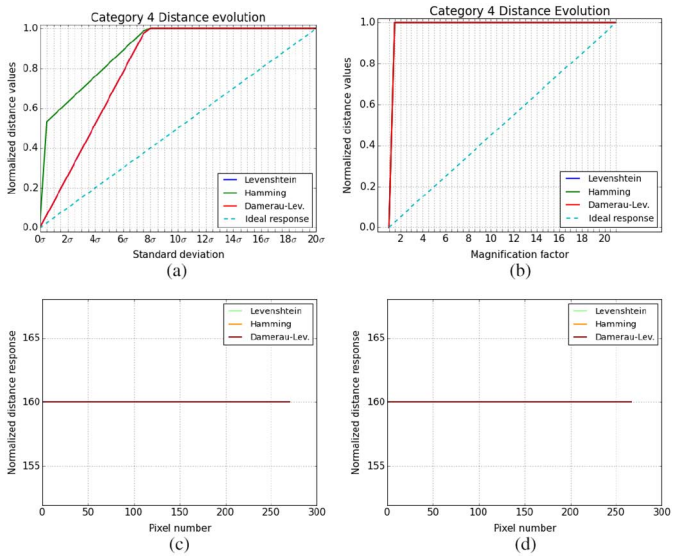

Fig. 12. Responses of distance functions in category 4 to theoretical evaluations for the case of (a) translation and (b) magnitude change, as well as to real spectral data from pigment (c) 23 800 and (d) 44 450. The first two figures are shown in normalized values.

TABLE V
SUMMARY OF THE THEORETICAL PROPERTIES OF SEVERAL DISTANCE FUNCTIONS IN CATEGORY 4, I.E., THOSE THAT CONSIDER A SPECTRUM AS A SEQUENCE

Fun	Var.	Triangular inequality	Identity of indiscernibles	Mon. increasing
Category 4: Sequence				
Hamming	–	✓	✓	✓
Levenshtein	–	✓	✓	✓
	Damerau	✓	✓	✓

*Reflexivity, nonnegativity, and symmetry are all satisfied.

between two sequences by means of character operations. For the magnitude change case, there is no shifting of the curve involved and therefore only values coming from the first and last spectral bands or channels remained having the same values. For this reason, the three distance functions are giving the results as shown in Fig. 12(a) and (b). As the theoretical behavior has been explained, the reason to why testing these functions with real spectral data results in Fig. 12(c) and (d) becomes clear. Spectral data are represented by real numbers. Furthermore, noise and variations are present in real spectral data. Such is the limitation of distance functions in this category as they assume that there are repetitions in the occurring values, which are not the case due to real numbers being an infinite set of values and noise that are present in the data making repetition of values even more unlikely.

VII. CONCLUSION

Distance functions are at the core of many image processing tasks, e.g., PCA and nonlinear filtering. However, in many cases, the use of a distance function is not carefully considered and adapted to the problem at hand. This would result in inaccurate measurement as the underlying assumptions of a distance function are ignored. Aiming at identifying the most suitable distance function for hyperspectral images, existing distance functions were classified into several categories based on their spectrum definition. Several evaluation criteria were defined, from theoretical evaluation with simulated spectra to numerical evaluation using real spectra of known properties.

TABLE VI
SUMMARY

Fun	Var.	D	M	Translation		Mag. change	
				η_{sat}	dnr	η_{sat}	dnr
Category 1: Vector in Euclidean space							
Minkowski	p = 1	✓	✓	7σ	272.07	–	2k
	p = 2	✓	✓	5.5σ	19.59	–	277.07
	p = ∞	✓	✓	4σ	1.99	–	39.89
	p = 0.5	✓	x	8σ	72k	–	364k
	p = 0.8	✓	x	7.5σ	1k	–	8k
Canberra	–	✓	✓	8σ	399.97	–	181.82
RMSE	–	✓	✓	5.5σ	3.06	–	43.27
χ^2	χ^2_1	✓	✓	8σ	399.94	–	165.29
	χ^2_2	✓	x	7σ	136.03	–	1k
Angular	Cosine	✓	x	5.5σ	1.0	–	0.0
	Spectral angle	✓	x	5.5σ	1.0	–	0.0
Geman-McClure	–	✓	✓	6σ	115.46	–	146.52
Category 2: N-dimensional data in manifold							
GFC	–	✓	x	5.5σ	1.0	–	0.0
Isomap	Euclid.	✓	x	–	276.06	–	249.94
	ECS	✓	x	–	195.07	–	155.0
Category 3: Distribution							
Spectral correlation	–	✓	x	6σ	0.58	–	0.0
Smith	–	✓	x	7σ	0.99	–	0.0
ECS	–	✓	✓	–	3k	–	2k
Jeffrey div.	–	✓	x	7σ	82.36	–	664.34
Pearson χ^2	–	✓	x	7σ	136.81	–	1k
EMD	–	✓	x	–	520	–	0.0
	Comb.	✓	✓	–	260	–	138.94
Squared-Chord	–	✓	x	7σ	273.07	–	1k
Category 4: Sequence							
Hamming	–	✓	✓	8σ	400	1.5	200
Leven-shtein	–	✓	✓	8σ	400	1.5	200
	Damerau	✓	✓	8σ	400	1.5	200

Legend

Fun	Function name	**M**	Metric
Var	Variation	η_{sat}	Saturation point
D	Distance	dnr	Dynamic range

The summary of theoretical properties of the evaluated distance functions is provided in Table VI. From this table, we can see that many of the distance functions are in fact not metrics. Even though a metric is not always necessary, it allows us to estimate the distance of a target to a reference point using distances obtained from other points (triangular inequality property). If we are looking for a metric, quite many candidates are available. However, information given in this table is only with respect to theoretical spectral data. After testing the metrics on real spectral data, we could eliminate Chebyshev, EMD, and combined EMD from the candidates as they were incapable of dealing with noise or variations.

Finally, the most suitable distance function for hyperspectral data, one that is a metric and responds proportionally without saturating in both cases of translation and magnitude change is ECS. Putting a less strict constraint on the selection, we could still select those metrics that saturate with preferably bigger saturation point. Therefore, we obtain Canberra and χ^2_1 distances.

Even though the metrics in category 4 saturate at the same point for translation case, they were not selected as they saturate as well for magnitude change case. We are aware that in applications in remote sensing, there is often a need to ignore magnitude change information. In such case, distance functions that give all-zero responses for magnitude change case can be used, i.e., Smith distance and spectral correlation. Angular-based distance functions, including GFC, are not chosen as they were shown to be unstable when given real spectral data as input. Manifold distances appear to be a possible choice. Nevertheless, their computational costs are a big limit for the use at a low-level image processing.

REFERENCES

[1] J. Astola, P. Haavisto, and Y. Neuvo, "Vector median filters," *Proc. IEEE*, vol. 78, no. 4, pp. 678–689, Apr. 1990.

[2] P. Trahanias and A. Venetsanopoulos, "Vector directional filters–A new class of multichannel image processing filters," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 528–534, Oct. 1993.

[3] M. E. Celebi, H. A. Kingravi, and Y. A. Aslandogan, "Nonlinear vector filtering for impulsive noise removal from color images," *J. Electron. Imaging*, vol. 16, no. 3, pp. 033 008–033 008–21, 2007.

[4] X. Guo and B. Guo, "Color image morphology based on distances in the HSI color space," in *Proc. ISECS Int. Colloquium Comput. Commun. Control. Manage. (CCCM)*, 2009, vol. 3, pp. 264–267.

[5] A. Ledoux, N. Richard, A.-S. Capelle-Laizé, and C. Fernandez-Maloigne, "Perceptual color hit-or-miss transform: Application to dermatological image processing," in *Signal Image Video Process.*, 2013, pp. 1–11.

[6] D. Chakrabarti and D. Dutta, "Facial expression recognition using PCA and various distance classifiers," in *Emerging Trends in Computing and Communication*. New York, NY, USA: Springer, 2014, vol. 298, pp. 79–85.

[7] K. Mahantesh, V. N. Manjunath Aradhya, and C. Naveena, "An impact of PCA-mixture models and different similarity distance measure techniques to identify latent image features for object categorization," in *Advances in Signal Processing and Intelligent Recognition Systems*. New York, NY, USA: Springer, 2014, vol. 264, pp. 371–378.

[8] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing dissimilarity measures for content-based image retrieval," in *Proc. 4th AIRS Inf. Retrieval Technol.*, Berlin, Germany: Springer-Verlag, 2008, pp. 44–50.

[9] C. M. Cuadras, S. Valero, D. Cuadras, P. Salembierd, and J. Chanussot, "Distance-based measures of association with applications in relating hyperspectral images," *Commun. Statist. Theory Methods*, vol. 41, no. 13–14, pp. 2342–2355, 2012.

[10] Z. Wang, A. C. Bovik, H. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[11] S. Chen and H. Jiao, "No-reference video monitoring image blur metric based on local gradient structure similarity," in *Artificial Intelligence and Computational Intelligence*. Berlin, Germany: Springer, 2011, vol. 7003, pp. 328–335.

[12] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barbba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric," in *Proc. IEEE Int. Conf. Image Process.*, 2007, vol. 2, pp. 169–172.

[13] M. Pedersen and J. Y. Hardeberg, "A new spatial filtering based image difference metric based on hue angle weighting," *J. Imaging Sci. Technol.*, vol. 56, pp. 50 501-1–50 501-12(12), 2012.

[14] J. Sweet, J. Granahah, and M. H. Sharp, "An objective standard for hyperspectral image quality," in *Proc. 9th AVIRIS Earth Sci. Workshop*, Pasadena, CA, USA, 2000.

[15] J. C. Leachtenauer, W. Malila, J. Irvine, L. Colburn, and N. Salvaggio, "General image-quality equation: GIQE," *Appl. Opt.*, vol. 36, no. 32, pp. 8322–8328, 1997.

[16] J. P. Kerekes, A. P. Cisz, and R. E. Simmons, "A comparative evaluation of spectral quality metrics for hyperspectral imagery," in *Proc. SPIE*, 2005, vol. 5806, pp. 469–480.

[17] R. Shrestha, R. Pillay, S. George, and J. Y. Hardeberg, "Quality evaluation in spectral imaging–Quality factors and metrics," *J. Int. Colour Assoc.*, vol. 12, pp. 22–35, 2014.

[18] F. A. Kruse *et al.*, "The spectral image processing system (SIPS)–Interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2–3, pp. 145–163, 1993.

[19] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Phys. Dokl.*, vol. 10, no. 8, pp. 707–710, 1966.

[20] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surv. Land Inf. Syst.*, vol. 62, no. 2, pp. 115–122, 2002.

[21] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional spaces," in *Proc. 8th Int. Conf. Database Theory*, New York, NY, USA: Springer-Verlag, 2001, pp. 420–434.

[22] P. Howarth and S. Rüger, "Fractional distance measures for content-based image retrieval," in *Proc. 27th Eur. Conf. Inf. Retrieval*, New York, NY, USA: Springer, 2005, pp. 447–456.

[23] M.-M. Deza and E. Deza, *Dictionary of Distances*. Amsterdam, The Netherlands: Elsevier Science, 2006.

[24] T. Sørensen, *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*, vol. 5. Kopenhagen: I kommission hos E. Munksgaard, 1948.

[25] J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, vol. 27, no. 4, pp. 857–871, Dec. 1971.

[26] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Statist. Inst.*, 1987, vol. 2, pp. 4–5.

[27] J. Hernández-Andrés, J. Romero, A. García-Beltrán, and J. L. Nieves, "Testing linear models on spectral daylight measurements," *Appl. Opt.*, vol. 37, no. 6, pp. 971–977, 1998.

[28] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[29] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[30] Y. W. Teh and S. T. Roweis, "Automatic alignment of hidden representations," in *Advances in Neural Information Processing Systems*, vol. 15. Cambridge, MA, USA: MIT Press, 2002, pp. 841–848.

[31] M. Brand, "Charting a manifold," in *Advances in Neural Information Processing Systems*, vol. 15. Cambridge, MA, USA: MIT Press, 2003, pp. 961–968.

[32] M. Crawford, L. Ma, and W. Kim, "Exploring nonlinear manifold learning for classification of hyperspectral data," in *Optical Remote Sensing*, vol. 3, S. Prasad, L. M. Bruce and J. Chanussot, Eds. Berlin, Germany: Springer, 2011, pp. 207–234.

[33] O. de Carvalho Jr. *et al.*, "Spectral multiple correlation mapper," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. Symp.*, 2006, pp. 2773–2776.

[34] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. Roy. Soc. London A Math. Phys. Eng. Sci.*, vol. 186, pp. 453–461, 1946.

[35] Y. Rubner, "Perceptual metrics for image database navigation," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, May 1999.

[36] D. Gavin, W. Oswald, E. R. Wahl, and J. W. Williams, "A statistical approach to evaluating distance metrics and analog assignments for pollen records," *Quat. Res.*, vol. 60, no. 3, pp. 356–367, 2003.

[37] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[38] A. Ledoux, N. Richard, and A. Capelle-Laizé, "How to specify or identify the most accurate multispectral distance function for non-linear image processing?" *Proc. Colour Visual Comput. Symp.*, 2013, pp. 1–7.

[39] R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 26, no. 2, pp. 147–160, 1950.

[40] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Commun. ACM*, vol. 7, no. 3, pp. 171–176, 1964.

Hilda Deborah received the Bachelor's degree in computer science from the University of Indonesia, Depok, Indonesia and the Master's degree from Erasmus Mundus Color in Informatics and Media Technology (EM CIMET). She is currently working toward the Ph.D. degree at the University of Poitiers, Poitiers, France, and Gjøvik University College, Gjøvik, Norway.

Her research interests include hyperspectral image processing and cultural heritage applications.

Noël Richard received the M.Sc. and Ph.D. degrees from the University of Poitiers, Poitiers, France, in 1988 and 1993, respectively.

Currently, he is an Associate Professor with the University of Poitiers, Poitiers, France. He is a Researcher with XLIM Laboratory, UMR CNRS SIC (Signals, Images and Communications) Department since 1992 and has been working on color image processing for 20 years. Facing the lack of stability and accuracy of the existing approaches in color image processing, he developed a new image processing paradigm based on distance functions, by which a vector computation of color texture features was applied. In 2013, he extended the paradigm to hyperspectral images, where a full-band vector processing was achieved.

Jon Yngve Hardeberg received the M.Sc. (Eng.) degree in signal processing from the Norwegian Institute of Technology, Trondheim, Norway, in 1995, and the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1999.

He is currently a Professor of Colour Imaging with GUC's Faculty of Computer Science and Media Technology, and member of the Norwegian Colour and Visual Computing Laboratory, where he teaches, supervises M.Sc. and Ph.D. students, and researches in the field of color imaging. His research interests include multispectral color imaging, print and image quality, colorimetric device characterization, color management, and cultural heritage imaging, and he has co-authored more than 150 publications within the field.