An endmember-based distance for content based hyperspectral image retrieval

Manuel Graña*, Miguel A. Veganzones

Grupo de Inteligencia Computacional, Universidad del País Vasco (UPV/EHU), Spain

A R T I C L E   I N F O

Article history:
Received 9 April 2011
Received in revised form 17 March 2012
Accepted 21 March 2012

Keywords:
Hyperspectral images
Content based image retrieval
Endmember induction algorithm
Image similarity

A B S T R A C T

We propose a specific content-based image retrieval (CBIR) system for hyperspectral images exploiting its rich spectral information. The CBIR image features are the endmember signatures obtained from the image data by endmember induction algorithms (EIAs). Endmembers correspond to the elementary materials in the scene, so that the pixel spectra can be decomposed into a linear combination of endmember signatures. EIA search for points in the high dimensional space of pixel spectra defining a convex polytope, often a simplex, covering the image data. This paper introduces a dissimilarity measure between hyperspectral images computed over the image induced endmembers, proving that it complies with the axioms of a distance. We provide a comparative discussion of dissimilarity functions, and quantitative evaluation of their relative performances on a large collection of synthetic hyperspectral images, and on a dataset extracted from a real hyperspectral image. Alternative dissimilarity functions considered are the Hausdorff distance and robust variations of it. We assess the CBIR performance sensitivity to changes in the distance between endmembers, the EIA employed, and some other conditions. The proposed hyperspectral image distance improves over the alternative dissimilarities in all quantitative performance measures. The visual results of the CBIR on the real image data demonstrate its usefulness for practical applications.

© 2012 Published by Elsevier Ltd.

1. Introduction

Mining the increasing amount of Earth Observation data has been recognized as a key problem even for panchromatic images or low spatial resolution multispectral images [8–10,13]. The increasing amount of hyperspectral images provided by new deployed hyperspectral sensors introduces new technical problems. Specifically, searching through these increasingly big databases using content based image retrieval (CBIR) techniques [12,43] has not been properly addressed for the case of hyperspectral images. Approaches to CBIR in remote sensing images proposed up to now have been focused on panchromatic or low dimension multispectral images [8–11,13,14,40,41], but not on hyperspectral images.

There are two main elements in the definition of a CBIR system [43]: (a) the image features and the corresponding feature extraction process and (b) the similarity measure defined on the feature space guiding the database search. Hyperspectral images contain rich spectral information, therefore it is natural to define spectral image features. As far as we know the only attempts to propose such a spectral image characterization to guide CBIR search in hyperspectral image databases are [33,49]. Assuming the image formation framework given by the linear mixing model [25], the image pixel spectra are a linear combination of elementary material signatures, called endmembers. Therefore, endmembers are appropriate spectral features providing a global characterization of the image. Endmember induction algorithms (EIAs) induce the endmember signatures from the image data, providing an autonomous processes for feature extraction which can be applied to each image independently. Thus, we are interested in this paper in assessing the descriptive power of endmembers for database search, and on the value of endmember induction algorithms (EIAs) as feature extraction processes.

The definition of an appropriate dissimilarity function must take into account that the endmembers are global image features, so that metrics or semimetrics defined for pattern matching in images can be less appropriate for this kind of data. Specifically the notion of outlier due to occlusions in shape detection [42], does not apply well to sets of endmembers. We need a dissimilarity measure which decreases when the hyperspectral images have more similar elementary materials, as characterized by their spectral signatures, regardless of their spatial distribution on the image. This paper introduces a formal distance between hyperspectral images based on spectral information. This hyperspectral image distance is a function of the distances between individual...
image endmembers. Therefore, the whole approach may be sensitive to the endmember distance and the applied EIA. We compare the proposed distance with the Hausdorff distance and a robust least trimmed squares (LTS) modification of the Hausdorff distance [42].

The issue of validation, or evaluation, of the CBIR approach is not trivial. CBIR systems are desired to handle large amounts of data, however, the available labeled information is scarce. Methodological frameworks for validation under conditions of little ground truth knowledge, such as the hybrid methodology in [48] are needed. However, it is preferable to use as much available information as possible. One useful approach is to provide an extensive collection of synthetic images whose meaning is well known, i.e. we have perfect knowledge of the ground truth allowing exact computation of the performance measures. Other approach is to perform bootstrapping of image blocks of a large image containing well identified regions that can provide a convenient image block labeling. We have followed both approaches in the experimental validation of the CBIR system constituted by the endmember induction and the proposed distance.

1.1. Contributions and structure of the paper

The contributions of the paper are the following ones.

- The definition of image endmembers induced by a given EIA as the hyperspectral image features for CBIR.
- The definition of a dissimilarity function over the image endmembers, proving that this image dissimilarity is a distance, to guide the CBIR search.
- The test on CBIR performance of the effect of the distance between endmembers (Euclidean versus Angular), and of the EIA (N-FINDER [50] versus EIHA [20,21]).
- The realization of an extensive experimental validation on a collection of synthetic hyperspectral images, and on an image dataset extracted from a real hyperspectral image.

In summary, the results of the computational experiments show little sensitivity of the CBIR system to EIA and endmember distance. The proposed distance improves over the comparing distances.

The structure of the paper is as follows: Section 2 introduces the CBIR for hyperspectral images relating it to the general field of CBIR. This section introduces also the formal definition of endmembers, the EIAs, and a discussion of relevant validation issues. In Section 3 we prove that the proposed hyperspectral image dissimilarity is a distance. Section 4 introduces the dissimilarity functions used for comparison. In Section 5 we give comparative and sensitivity results computed on the synthetic hyperspectral image databases. In Section 6 we give results on dataset composed of image blocks extracted from a large real hyperspectral image. Finally, we give some conclusions in Section 7. Appendix A contains miscellaneous results on the comparing dissimilarity functions. Appendix B describes the EIAs used for the experiments: the N-FINDER and EIHA algorithms.

2. CBIR in hyperspectral image databases

The acknowledgment of the semantic gap in general CBIR approaches [12,29,43] has diverted significant efforts in the CBIR literature into two avenues. One is focused on specific domain systems, such as medical images [15,24,35,22], where there is a reduced semantic uncertainty, and similarity based search is always meaningful. The other is the development of relevance feedback approaches for broad domain image databases [7,51]. Relevance feedback approaches include feature weighting [26], probabilistic combination of image distances [2], genetic algorithms [3], kernel machines [4] among a multitude of other techniques. Relevance feedback has also been applied to remote sensing images [14,16]. In fact [4] applies a CBIR-like approach to image segmentation, instead of database search. This is an indication of the terminological confusion found in the literature. In [1] there is an attempt to define a relevance feedback CBIR approach using several features, among them the average of the image pixel spectra. The main handicaps when trying to design a retrieval feedback on hyperspectral image databases based on spectral information are the design of an appropriate user interface, and the difficulty to interpret the spectral information by untrained people. Hyperspectral image database CBIR is at an early stage of development, when appropriate dissimilarity measures need to be designed for the adequate exploitation of the rich spectral information contained in the images. An additional difficulty is the lack of available public databases for CBIR system validation and evaluation. Most reported validation experiments of CBIR systems over remote sensing images rely on one image or a relatively small collection of images [1,14,16,33]. For hyperspectral images this problem is even stronger, because there are few public images with a known ground truth which can provide supporting evidence for CBIR approaches.

In this paper, CBIR on hyperspectral images is treated as a specific domain CBIR problem where a key technical issue is to find a domain-dependent dissimilarity measure providing enhanced database search results. We follow a query-by-example approach, where the database interrogation is done through the presentation of a query image, and the most similar images in the database are returned as the answer [43]. The image features are the endmembers and the feature extraction processes are the EIA, both explained below. Previous works using endmembers to obtain image features [33] did not provide a formal distance between endmember-based image representations, but a heuristic search based on corresponding fractional abundances obtained from a greedy spectrum matching. The emphasis in [33] was the design of distributed computing systems to approach real-time response, not CBIR performance.

The structure of the proposed Spectral CBIR system is illustrated in Fig. 1. Users interrogate the system providing a sample query image. The EIA calculates the image endmembers which will be used as feature vectors. The same EIA is applied to both the query image and all and each of the hyperspectral images in the database, and will remain the same along the database exploitation. A dissimilarity measure between the query and the database images is computed on the endmember-based features. Finally, the k most similar images on the database are returned as the answer to the query.

2.1. Spectral features

We assume the linear mixing model [25] of hyperspectral image formation. Each pixel is the result of the linear combination

![Fig. 1. Structure of a CBIR system based on the proposed hyperspectral image distance.](image-url)
of the pure spectral signatures of elementary materials, the endmembers, with fractional coefficients, the abundance coefficients. Let \( E = \{e_1, \ldots, e_m\} \) be the pure endmember signatures, where each \( e_i \in \mathbb{R}^q \) is a \( q \)-dimensional vector. Then, the hyperspectral signature \( r \) at each pixel in the image is defined by the expression

\[
r = s + n = \sum_{i=1}^{m} e_i \phi_i + n,
\]

where \( r \) is formed by the sum of the pixel's signal \( s \) and an independent additive noise component \( n \); and, \( \phi \) is the \( m \)-dimensional vector of fractional abundances at the given pixel subject to constraints: \( \phi_i \geq 0, \forall i = 1, \ldots, m \) and \( \sum_{i=1}^{m} \phi_i = 1 \). This equation can be extended to the full image as \( H = F \Phi + n \), where \( H \) is the hyperspectral image, \( \Phi \) is a matrix of fractional abundances and \( n \) is independent additive noise. For image analysis applications outside the scope of this paper, the fractional abundances of each pixel are obtained solving a constrained least squares problem which ensure abundances are well defined.

We characterize hyperspectral image \( H \) by the set of endmembers \( E = \{e_1, e_2, \ldots, e_m\} \) induced from \( H \) by an EIA. EIAs, introduced below, are autonomous processes in the sense that they do not rely in information sources outside the hyperspectral image, such as laboratory libraries of material spectral signatures. Therefore, endmembers are independent content based spectral features of the image.

### 2.2. Endmember induction algorithms

EIAs are unsupervised approaches determining the image endmembers from the image data. The geometrical interpretation of the linear mixing model of Eq. (1) with non-negative and sum to one constraints is that the endmember signatures are affinely independent points in the high dimensional space of pixel spectra defining a convex polytope, often a simplex, that covers all the image pixels represented as points in this space. Fractional abundances are the convex coordinates of a pixel signature inside this polytope. The EIAs aim to induce the endmember signatures from the image data without any \textit{a priori} knowledge, therefore, they are unsupervised algorithms. The induced endmembers can be extrapolations from the image data, therefore, not corresponding to any real pixel spectra, or pixel spectra selected from the image. In this latter case, the corresponding polytope often will not cover all pixel data points and the constraints on the fractional abundances need to be relaxed.

We distinguish three fundamental kinds of EIAs [46]: geometric approaches, that try to find a simplex that covers the image data [6,50]; lattice computing approaches, that use some kind of lattice theoretic formalism or mathematical morphology approach [21,32]; heuristic approaches, that are ad hoc solutions guided by the intuition about the problem [5]. Examples of lattice computing [17] EIAs are the works in [19,21,39,18] based on the notion of Strong Lattice Independence (SLI), following the conjecture in [38] that SLI vectors are Affine Independent vectors and, thus, their convex hull defines an approximation to the simplex covering the data. The EIHA algorithm described in Appendix B corresponds this kind of EIA. Geometric EIA, such as the NFINDR described in Appendix B try to build the simplex encompassing the image data directly.

### 2.3. Validation issues

CBIR systems are envisaged to deal with large databases, however, the quality and quantity of the data available for system evaluation is low. The problem of the validation of CBIR in the case of lack of ground truth information has been considered in [47,48]. In this paper we perform two kinds of validation processes. One is performed over a dataset of synthetic images, the second is performed upon a partition of a large real hyperspectral image into image blocks labeled by visual inspection.

Validation on synthetic images benefits from the complete knowledge of the ground truth, so very accurate performance measures can be given over a large dataset which simulates the general CBIR setting. However, it has some disadvantages. First, there are no classes defined, so that it is not feasible to define generality plots or performance measures based on class size normalized query scopes. Second, the set of relevant images must be defined on the basis of the dissimilarity computed on the ground truth, which do not allow a class definition. The recall and precision measures must be interpreted as the degree of agreement between the image orderings induced by the dissimilarity measures on the ground truth and the results of image feature extraction processes, respectively.

Validation on the real data allows the manual labeling of the image blocks, therefore, the ground truth is completely independent from the feature extraction processes. The labeling allows the computation of standard recall and precision performance measures, scope normalization and the generation of generality plots [23]. The main disadvantage is that the size of the database is not large. Results however have a natural intuitive interpretation.

### 3. Endmember based hyperspectral image distance

Let it be \( E = \{e_1^2, e_2^2, \ldots, e_m^2\} \) the set of endmembers induced from the hyperspectral image \( H \) in the database, where \( p_i \) is the number of induced endmembers from the \( i \)-th image. A dissimilarity function between two hyperspectral images, \( s(H_a, H_b) \) is defined on the basis of the distances between their corresponding set of endmembers \( E_a \) and \( E_b \). We compute the matrix whose elements are the distances between each pair of endmembers from each image

\[
D_{a,b} = [d_{ij}; \ i = 1, \ldots, p_a; \ j = 1, \ldots, p_b],
\]

where \( d_{ij} \) is the distance between the endmembers \( e_i^2 \in \mathbb{R}^q \) and \( e_j^2 \). In this paper we consider the Euclidean distance, \( d_E \)

\[
d_E(e_1, e_2) = \|e_1 - e_2\|,
\]

and the angular distance, aka Spectral Angle Mapper (SAM) distance in remote sensing applications, \( d_S \)

\[
d_S(e_1, e_2) = \cos^{-1} \left( \frac{e_1 \cdot e_2}{\|e_1\| \cdot \|e_2\|} \right),
\]

where \( e_1 \cdot e_2 \) denotes the vector inner product.

We compute the vectors of row and column minimal values of \( D_{a,b} \), respectively, denoted as \( m_a^r = [m_a^{r,1}, \ldots, m_a^{r,p_b}] \) and \( m_b^c = [m_b^{c,1}, \ldots, m_b^{c,p_a}] \). Their components are computed as

\[
m_a^{r,j} = \min_{i=1}^{p_a} \{d_{ij}\},
\]

\[
m_b^{c,j} = \min_{i=1}^{p_b} \{d_{ij}\}.
\]

**Definition 1.** The dissimilarity between two hyperspectral images, \( H_a, H_b \), is given by the following expression:

\[
s(H_a, H_b) = (\|m_a^r\| + \|m_b^c\|).
\]
Lemma 2. The dissimilarity function of Definition 1 is a semimetric.

Proof. The dissimilarity function of Definition 1 complies with the axioms required for a semimetric for any hyperspectral images $H_a, H_b$:

- **Non-negativity:** $s(H_a, H_b) \geq 0$ because both norms are non-negative $\|m_a\| \geq 0$, $\|m_b\| \geq 0$.

- **Identity of indiscernibles:** it is restricted to the equivalence of images whose endmembers obtained by the EIA are the same.
  - If $E_a = E_b$ then both row and column minimal vectors will be null $\|m_a\| = \|m_b\| = 0$, because the diagonal of $D_{a,b}$ will be zero, $d_{ij} = 0$, then $s(H_a, H_b) = 0$. This is true even if the EIA finds a permutation of the endmembers when applied to the same or different images.
  - On the other sense, if $s(H_a, H_b) = 0$ then all the components of the row and column minimal vectors must be zero, $m_{a,i} = m_{b,i} = 0$ for $i \in 1, \ldots, p$. Therefore, for each endmember in $E_a$ there is another identical to it in $E_b$ and, conversely, for each endmember in $E_b$ there is another identical to it in $E_a$. Thus, $E_a = E_b$.

- **Symmetry:** $s(H_a, H_b) = s(H_b, H_a)$ is immediate if we note that $D_{a,b} = (D_{b,a})^T$. The component of $D_{a,b}$ will be identical to the column minimal vector of $(D_{b,a})^T$, and the column minimal vector of $D_{b,a}$ will be identical to the row minimal vector of $(D_{b,a})^T$. Therefore, the expressions of $s(H_a, H_b)$ and $s(H_b, H_a)$ are identical by the commutativity of addition. \( \square \)

To prove that the dissimilarity of Definition 1 is a distance we need to prove that it satisfies the triangle inequality. We proceed proving first that the norms of the vectors of minimum values satisfy the triangle inequality. Then, it follows immediately from these results that the dissimilarity satisfies the triangle inequality.

Lemma 3. Given three hyperspectral images $H_a, H_b, H_c$, with corresponding induced sets of endmembers $E_a, E_b, E_c$, the following inequality holds

$$\|m_a\| \leq \|m_b\| + \|m_c\|.$$  

Proof. For each $e^a_i \in E_a$ there is a $e^b_j \in E_b$ such that $d(e^a_i, e^b_j) = \min_{1 \leq j \leq p} d(e^a_i, e^b_j)$. In the same way, there is a $e^c_k \in E_c$ such that $d(e^b_j, e^c_k) = \min_{1 \leq k \leq p} d(e^b_j, e^c_k)$. By the triangle inequality of the distance between endmembers we have that:

$$d(e^a_i, e^b_j) \leq d(e^a_i, e^c_k) + d(e^c_k, e^b_j).$$

If $d(e^a_i, e^c_k) = \min_{1 \leq k \leq p} d(e^b_j, e^c_k)$ then the proposition is true. Suppose that there is another $e^c_k \in E_c$ such that $d(e^b_j, e^c_k) < d(e^a_i, e^c_k)$. This implies that

$$d(e^a_i, e^b_j) < d(e^a_i, e^c_k) + d(e^c_k, e^b_j).$$

We have two cases in which this could be true

- $j^* = j$ implies that $d(e^b_j, e^c_k) < d(e^b_j, e^c_k)$ contradicting the fact that we have selected $e^c_k$ as the minimum distance endmember.

- $j^* \neq j$ implies that $d(e^b_j, e^c_k) < d(e^a_i, e^b_j)$ contradicting the fact that we have selected $e^b_j$ as the minimum distance endmember.

The triangle inequality $\|m_a\| \leq \|m_b\| + \|m_c\|$ follows directly from the triangle inequalities of all the entries of the distance matrix. \( \square \)

Theorem 4. The dissimilarity function of Definition 1 is a distance

Proof. According to Lemma 2 the dissimilarity function of Eq. (5) is a semimetric. To prove that, given three hyperspectral images $H_a, H_b, H_c$, it complies with the triangle inequality,

$$s(H_a, H_c) \leq s(H_a, H_b) + s(H_b, H_c).$$

we refer to Lemma 3 which shows that the triangle inequality holds for the norm $\|m_a\|$ of the vector of row minimal values. An identical reasoning proves that the triangle inequality holds for the norm $\|m_a\|$ of the vector of column minimal values. Furthermore, the elementary triangle inequalities of the endmembers are completely determined and the addition $\|m_a\| + \|m_b\| = s(H_a, H_b)$ complies with the triangle inequality, thus, the proposed dissimilarity measure complies with the axioms of a distance. \( \square \)

Remark 5. We have used in previous works [48,49] the following definition for the dissimilarity between hyperspectral images represented by their corresponding sets of endmembers:

$$s(u, v) = (\|m_u\| + \|m_v\|)(|p_u - p_v| + 1).$$

It is easy to check that this dissimilarity meets the conditions of a semimetric following the reasoning of Lemma 2, however, it does not comply with the triangular inequality, as proved in Lemma 15 of Appendix A. Therefore, it is not a distance and we have discarded it for the works presented in this paper.

4. Other dissimilarity functions

We compare the proposed distance with state of the art dissimilarity measures. This section gives their definitions and some discussion of their usefulness in our image domain. We maintain the notation introduced in the previous section specializing our discussion of dissimilarity functions to the comparison of hyperspectral images characterized by their endmembers.

Definition 6. The Hausdorff distance between the endmembers of a pair of hyperspectral images is defined as follows:

$$s(u, v) = \max(\max(m_u), \max(m_v)),$$

where $H_u$, $H_v$ are hyperspectral images; and $m_u$, $m_v$ are the vectors of row and column minimal values of $D_{a,b}$ as it was explained in Section 3.

Hausdorff distance is used for matching of point clouds in arbitrary spaces, with applications in pattern recognition. For CBIR it is applied on the image features, allowing for comparison among images with different feature set sizes. Compared with the distance of Definition 1, the Hausdorff distance disregards information that can be relevant in our CBIR application in a way that we cannot formalize easily, but that can be illustrated with an example.

Example 7. Assume that the endmember distances are normalized in $[0, 1]$ so that 1 is extreme dissimilarity. Let us have hyperspectral images $H_a$, $H_b$, and $H_c$, such that $m_a = [0.9, 0.9, 0.9]$, $m_b = [0.9, 0.9, 0.9]$, $m_c = [0.9, 0.1, 0.1]$. Then the similarity between images $H_a$ and $H_b$ relative to $H_c$ will be the same according to the Hausdorff distance $s_{H_a}(H_b, H_c) = s_{H_b}(H_a, H_c)$, that is 0.9. However, the spectral content of $H_a$ is more similar to $H_b$ than to $H_c$, because they share two endmembers, while $H_a$ and $H_b$ do not share anyone. The proposed distance of Definition 1 gives a result according to our intuition $s(H_a, H_b) = 2.8316 > s(H_b, H_a) = 1.0525$.

Variations of the Hausdorff distance are suggested for robust pattern matching [42] in images. We reformulate them in the terms of Section 3, discussing their properties and usefulness for our problem.
Definition 8. The robust M-estimation Hausdorff dissimilarity between the endmembers of a pair of hyperspectral images is defined as follows:

\[ s_r(H_x, H_y) = \max(h_x(m_a), h_y(m_b)) \]

where the robust M-estimation of the distance between endmembers is given by

\[ h_i(m_a) = \frac{1}{p} \sum_{i=1}^{p} \rho_m(m_{a,i}) \]

the thresholding function removing outliers is specified as follows:

\[ \rho_m(x) = \begin{cases} x, & x \leq \tau, \\ \tau, & x > \tau. \end{cases} \]

Notice that \( m_{a,i} \geq 0 \), therefore, we do not need to compute \(|x|\) as specified in [42]. The parameter \( \tau \) specifies the minimum distance for an endmember to be considered as an outlier. We prove in Lemmas 11 and 13 of Appendix A that this dissimilarity is a semimetric.

Definition 9. The robust Hausdorff dissimilarity between the endmembers of a pair of hyperspectral images based on the least trimmed square (LTS) is defined as follows:

\[ s_L(H_x, H_y) = \max(h_L(m_a), h_L(m_b)) \]

where the robust LTS estimation of the distance between endmembers is given by

\[ h_L(m_a) = \frac{1}{l_p} \sum_{i=1}^{l_p} \alpha(m_{a,i}) \]

in this expression \( \alpha(m) \), denotes the \( i \)-th element of vector \( m \) after ordering its components in ascending order, i.e. \( \alpha(m_1) \leq \alpha(m_2) \leq \cdots \leq \alpha(m_n) \). The parameter \( L \in [0, 1] \) specifies the order statistics to be taken into account. In the experimental works we will use the value \( L = 0.6 \) recommended in [42]. We prove in Lemmas 12 and 14 in Appendix A that this dissimilarity is a semimetric.

Both robust M-estimation and LTS Hausdorff semimetrics are designed to cope with outlier points in image pattern matching applications looking for point correspondences [42], where the outliers come from the occurrence of occlusions inducing confusion between different shapes or background noise. Therefore, these dissimilarities may be expected to be biased to find partial matchings of the data points. From the point of view of the kind of images and features in this paper, these dissimilarities have two inconveniences. First, endmembers are global descriptors of the image content, not strictly associated with an image coordinate or point. Therefore, the concept of outlier as such does not apply well in our CBIR system: outlier endmembers represent a wide variation of a holistic image feature, i.e. the appearance of a different material, regardless of image region distributions. Diminishing or bounding the effect of widely different endmembers forces the similarity between images containing very disparate materials. Second, the estimation and validation of the appropriate parameter settings, either the threshold \( \tau \) or the order statistic \( L \), implies a delicate and complex computational process, which does not add general value to the approach because any setting will often be specific of the particular data domain. We will consider the robust LTS Hausdorff semimetric for comparison over the real image data in Section 6.

Example 10. Assume that the endmember distances are normalized in [0, 1] so that 1 is extreme dissimilarity. Let us have hyperspectral images \( H_x, H_y, \) and \( H_z \), such that \( m_{x}^{2} = [0.9, 0.1, 0.1], m_{y}^{2} = [0.1, 0.1, 0.1], m_{z}^{2} = [0.1, 1.0, 1.0] \). The interpretation of these values is as follows: \( H_z \) and \( H_y \) share two endmembers, there is a third endmember in \( H_z \) that is not present in \( H_x \) and \( H_y \) and \( H_z \) have the same endmembers. The robust LTS Hausdorff dissimilarity with \( L = 0.6 \) does not distinguish these cases \( s_L(H_x, H_y) = s_L(H_x, H_z) = 0.1 \). On the other hand, the ability of the robust M-estimation Hausdorff to detect this situation depends on the choice of the threshold parameter, assuming \( \tau > 0.1 \): \( s_r(H_x, H_y) = \frac{1}{2}(\tau + 0.2) > s_r(H_x, H_z) = 0.1 \).

5. Experiments on synthetic data

In this section we will first describe how we construct the synthetic hyperspectral image collections used for the validation computational experiments. Then, we describe the experimental design, giving the definition of the computed performance measures. Finally, we provide the results of the experiments.

5.1. Synthetic hyperspectral images

Synthetic hyperspectral images are generated as linear mixtures of a set of spectra (the ground-truth endmembers) according to simulated abundance coefficients for each pixel.\(^1\) The image ground-truth endmembers were selected from a subset of the USGS spectral library.\(^2\) We have structured the experimental data in datasets, where each dataset is characterized by the underlying set of candidate ground-truth endmember spectra and the image size. We defined three sets of candidate ground-truth endmembers, with 5, 10 and 20 endmembers each, representing an increasing diversity of the materials that can be found in the scene. We denote the datasets generated from each set of candidate ground-truth endmembers as 5-E, 10-E and 20-E datasets, respectively. Fig. 2 shows the collection of candidates for the generation of the 10-E dataset. We also defined three categories of image spatial size, with images having 64 × 64, 128 × 128 and 256 × 256 pixels. Increasing image size, we increase the available information and the computational cost. We have synthesized a total of 18,000 hyperspectral images divided in nine datasets of 2000 images each. Spectral resolution is 269 bands per pixel. In each dataset, images are built from a number of randomly picked endmembers that varies from 2 up to 5. The procedure for the generation of each image, once fixed the set of candidate endmember spectra and the image size, is as follows:

1. Select the number of endmembers, \( n_e \in [2, 5] \).
2. Select randomly and without repetition the \( n_e \) ground-truth endmember spectra from the corresponding set of candidates.
3. Compute the simulated abundance image for each ground-truth endmember as a Gaussian random field with Matern correlation function of parameters \( \theta_1 = 10 \) and \( \theta_2 = 1 \). We applied the procedure proposed by [28] for the efficient generation of Gaussian random fields with large domains.
4. Normalize the abundances of each pixel. We preserve the greatest abundance coefficient value, normalizing the remaining coefficients to ensure that the abundance coefficients sum up to one. This procedure does not guarantee the existence of pure pixels in the resulting hyperspectral image.

Fig. 3 shows an example of randomly selected ground truth endmembers and corresponding generated abundance images used to synthesize a hyperspectral image from the 10E-256 × 256 pixels dataset.

---

\(^1\) The code to generate such synthetic images can be found at http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB

\(^2\) http://specalab.cr.usgs.gov/spectral-lib.html
5.2. Experimental design

We have performed independent query computational experiments over each of the nine synthetic hyperspectral datasets using the distance of Definition 1, hereafter denoted as “Grana” in the plots and tables. The experiments are of two kinds. First we test the sensitivity of the approach to the distance between individual endmembers, and the sensitivity to the EIA
used to induce the endmembers. We test the Euclidean distance of Eq. (3), and the SAM distance of Eq. (4). The EAs tested are the N-FINDER and the EIHA algorithms presented in Appendix B. Second, we compare the results of the proposed “Grana” distance with those obtained computing the Hausdorff distance \([12, 27]\) on the same image features (the endmembers).

For each image \(H_k\) in a dataset we calculate the dissimilarity between \(H_k\) and each of the remaining images in the dataset. These dissimilarities are represented as a vector \(s_k = [s_{k1}, \ldots, s_{kN}]\), where \(N\) is the number of images in the dataset (2000 in our experiments) and \(s_{kj}\) is the dissimilarity between the images \(H_k\) and \(H_j\). We define the ranking of the images in a dataset relative to image \(H_k\), \(\Omega_k = \{o_{kj} \in \{1, \ldots, N\}; p = 1, \ldots, N\}\), as the set of image indexes ordered according to increasing values of their corresponding entries in the dissimilarity vector \(s_k\). That is, we sort in increasing order the components of \(s_k\) and the resulting permutation of image indices constitute \(\Omega_k\), so that

\[ s_{k, o_{kj}} \leq s_{k, o_{kj+1}}. \]

The validation process compares the ground-truth induced image ordering with the ordering induced by the dissimilarity function. The vector of ground-truth dissimilarities computed using the knowledge of the ground image truth endmembers is denoted \(s_k^{GT}\). The ground-truth dissimilarity \(s_{k, o_{kj}}^{GT}\) between the images \(H_k\) and \(H_j\) is the number of different endmembers in the images. The vector of dissimilarities computed using the endmembers induced by one of the EAs (either N-FINDER or EIHA) is denoted \(s_k^{IND}\). The ground-truth dissimilarity \(s_{k, o_{kj}}^{GT}\) between the images \(H_k\) and \(H_j\) is computed using either the semimetric of Definition 1 or the Hausdorff distance of Definition 6. We distinguish rankings \(\Omega_k^{GT}\) and \(\Omega_k^{IND}\) corresponding to the dissimilarities computed on the ground-truth and EIA induced endmember, respectively. Some limited experimental circularity is unavoidable, because we must provide an image ordering based on ground-truth information, which consists of endmember spectra and spatial abundances. However, if the ground-truth dissimilarity function is different from the CBIR dissimilarities experimental circularity is reduced.

5.3. Performance measures on the synthetic data

In the collection of synthetic data we do not have a priori labeled classes of hyperspectral images, therefore, we need to adapt the definition of precision measures to this situation. We cannot define the normalized score, nor the class embedding to compute generality plots [23,30] because there is no defined class size. The lack of a priori defined classes implies that the relevant images must be defined on the basis of their dissimilarity. Setting a threshold independent of the query scope is arbitrary, therefore, we have defined the relevant images on the basis of the image ordering induced by the ground truth endmembers, as the ideal response to the query.

A query \(Q_k(H_a)\) is a search for the \(k\) most similar (less dissimilar) images \(H_p\) in the dataset respect to the image \(H_a\), with \(1 \leq k \leq N\). The set of returned images \(T_k(H_a)\) and the set of relevant images \(V_k(H_a)\) for a query \(Q_k(H_a)\) are defined according to \(s_k^{GT}\) and \(s_k^{IND}\) as follows:

\[ T_k(H_a) = \{a_{kj} \in \Omega_k; s_{k, o_{kj}} \leq s_{k, o_{kj}}^{GT}\}. \]

\[ V_k(H_a) = \{a_{kj} \in \Omega_k; s_{k, o_{kj}}^{IND} \leq s_{k, o_{kj}}^{GT}\}. \]

This definition allows for the inclusion in the query answer of images whose dissimilarity is equal to the maximum one, thus, the cardinality of both returned and relevant sets may be bigger than \(k\). The precision \(P_k(H_a)\) and recall \(R_k(H_a)\) for a query \(Q_k(H_a)\) are standard performance measures in CBIR literature [12,31],

\[ \text{precision} = \frac{|V_k(H_a)|}{|T_k(H_a)|}, \]

\[ \text{recall} = \frac{|V_k(H_a)|}{|Q_k(H_a)|}. \]

they are defined as

\[ P_k(H_a) = \frac{|V_k(H_a) \cap T_k(H_a)|}{|T_k(H_a)|}. \]

\[ R_k(H_a) = \frac{|V_k(H_a) \cap T_k(H_a)|}{|V_k(H_a)|}. \]

the average precision and recall of the system for a query of size \(k\) are defined as

\[ P_k = \frac{1}{N} \sum_{a=1}^{N} P_k(H_a). \]

\[ R_k = \frac{1}{N} \sum_{a=1}^{N} R_k(H_a). \]

The precision–recall curve is obtained plotting the precision versus recall obtained varying the query size from 1 up to the entire database. Usually, increasing recall implies decreasing precision.

We compute two scalar performance measures for summary comparison. One is the area under the precision–recall curve computed using the trapezoidal method. This value is similar to the area under the Receiver Operating Characteristic used for model comparison in classification system design. Increasing values correspond to increasing performance. The optimal value is 1, corresponding to perfect precision and recall.

The other scalar performance measure is the normalized average rank of relevant images [31]. The normalized rank for a given image ranking \(\Omega_k\), denoted as \(\text{Rank}(H_a)\), is defined as

\[ \text{Rank}(H_a) = \frac{1}{N_s} \left( \sum_{a=1}^{N_s} \phi_i \frac{N_s(N_s+1)}{2} \right). \]

where \(N_s\) is the number of relevant images for the query, i.e. \(N_s = |V_k(H_a)|\), and \(\phi_i\) is the rank at which the \(ith\) relevant image is retrieved, i.e. \(\phi_i = o_{ij}^{IND}\). Because there are no classes defined for the synthetic data, we employ \(k=1\) to compute Eq. (16), thus, the relevant set is the set of images with exactly the same ground truth endmembers. This measure is 0 for perfect performance, and approaches 1 as the performance worsens. Perfect performance means that the system is able to recover the relevant images in the normalized scope given by the relevant set size. The average normalized rank (ANR) for the whole dataset is given by

\[ \text{ANR} = \frac{1}{N} \sum_{a=1}^{N} \text{Rank}(H_a). \]

where \(N\) is the number of images in the dataset (\(N=2000\) in synthetic image experiments).

5.4. Results

In the figures and tables below we call “Grana” the distance of Definition 1. We performed the following computational experiments: for each dataset of 2000 synthetic images we computed independently the precision and recall obtained using the endmembers induced by the alternative EIA (N-FINDER and EIHA) and alternative distances (Euclidean and SAM), for the Grana and Hausdorff distances. We present these results in Figs. 4–9. Each figure corresponds to a dataset and EIA. Each figure shows twelve precision–recall curves corresponding to each combination of the dissimilarity function (Grana or Hausdorff), spatial image size \((64 \times 64, 128 \times 128\) or \(256 \times 256\)) and endmember distance (Euclidean or SAM). Tables 1–3 show the area under the precision–recall curves and the average normalized rank for each of the 5-D, 10-D and 20-D datasets.
5.5. Discussion of results

We have structured the discussion of the results into a series of questions that we feel relevant.

Classification results. Can the proposed CBIR approach discover the underlying classification of the images induced by the ground truth endmembers? Artificial classes could be defined by images having a zero distance on the ground truth endmembers. These classes will be small sized, therefore, to determine if the CBIR really uncovers the ground-truth classification we need to look at the values of precision for low recall values, which are pretty high in all the plots. Increasing the query scope many irrelevant
images are forced into the answer of the query, reducing precision and increasing recall. For all combinations of the CBIR system parameter values corresponding to plots in Figs. 4–9 the system is able to discover quite efficiently the underlying classification.

**Endmember distance robustness.** Is the CBIR system robust to the choice of the distance computed between individual endmembers? From the examination of Tables 1–3 the SAM distance consistently gives better results than the Euclidean distance:

**Fig. 7.** Precision and recall results for the 10-E dataset using the N-FINDER endmember induction algorithm for the Grana distance and the Hausdorff distance.

**Fig. 8.** Precision and recall results for the 20-E dataset using the EIHA endmember induction algorithm for the Grana distance and the Hausdorff distance.

**Fig. 9.** Precision and recall results for the 20-E dataset using the N-FINDER endmember induction algorithm for the Grana distance and the Hausdorff distance.

Please cite this article as: M. Graña, M.A. Veganzones, An endmember-based distance for content based hyperspectral image retrieval, *Pattern Recognition* (2012), http://dx.doi.org/10.1016/j.patcog.2012.03.015
M. Graña, M.A. Veganzones / Pattern Recognition 000 (2012) 1 – 10

64. Experiments of real data

We have designed a CBIR experiment over a large real hyperspectral image where we have manually identified several classes of terrain. The computed performance measures follow standard evaluations in the literature [23,30]. We have included also for comparison results using the robust LTS Hausdorff dissimilarity of Eq. (9).

(a) the area under the precision–recall curve is greater for SAM than for the Euclidean distance, all other parameters equal and (b) the ANR is consistently lower for SAM than for Euclidean distance. There is, thus, some sensitivity of the CBIR system to the endmember distance, but with less impact than other experimental factors. This effect can be due to the scale robustness of SAM which allow to cope better with the imprecise recovery of the endmembers by the EIA.

EIA sensitivity. Is the CBIR system sensitive to the EIA used for endmember induction? Figs. 4 and 8 show the precision–recall curves obtained on the endmembers induced by the EIHA, while Figs. 5–9 show the precision–recall curves obtained on the endmembers induced by N-FINDER. Pairing corresponding plots with the same set of candidate endmembers, image size, endmember distance and image distance gives a qualitative impression of the CBIR sensitivity to EIA. We find them very similar regardless of the EIA used, therefore, a qualitative insensitivity of the CBIR approach on the EIA can be declared. For a more quantitative assessment, examining Tables 1–3 we find that the N-FINDER performs better for smaller images and smaller sets of candidate endmembers. The EIHA improves N-FINDER for the largest spatial size images and the biggest set of candidate endmembers. However, we do not find these effects strong enough to claim that the CBIR must employ a specific EIA.

Effect of image size. Is there any effect of the image size on the CBIR system? The examination of Tables 1–3 shows that results improve with the size of the image for both scalar performance measures. Change in image size induces changing the relative results of the EIA considered. The effect of image size is very clearly appreciated in the precision–recall curves in Figs. 4–9 gathering plots corresponding to the same image size regardless of all remaining experimental parameters. Smaller images give lower precision–recall curves. Increasing the image size the curves grow, but the effect does saturate. This effect decreases as the size of the endmember pool grows.

Ground truth diversity. Is there any effect in the CBIR system performance due to the ground-truth diversity measured by the size of the underlying set of candidate endmembers? We expect some effect because if the diversity of candidate endmembers is higher, the image semantic domain may be considered as broader and the expected size of the query relevant set is smaller. We find a strong qualitative effect in Figs. 4–9, regardless of other parameters. Minimum recall values found in the experiments grow with the size of the set of candidate endmembers. In Tables 1–3 the area under the precision–recall curves grows accordingly. The effect is less strong for the ANR than for the area under the precision–recall curve.

Comparing distances. Does the proposed distance of Definition 1 improve over the conventional Hausdorff distance? If we examine Figs. 4–9 we can appreciate in all of them that the precision–recall curve corresponding to the Hausdorff distance falls below the one corresponding to the Grana distance. If we consider the scalar performance results given in Tables 1–3 we find a clearer confirmation of this effect. The area under the precision–recall curve is consistently greater for the Grana distance than for the Hausdorff distance, under all combinations of experimental factors. At the same time, the ANR is consistently smaller for the Grana distance than for the Hausdorff distance, with no sensitivity to the size of the set of candidate endmembers, while the area under the precision–recall curve shows less improvement for the large set of candidate endmembers.

Please cite this article as: M. Graña, M.A. Veganzones, An endmember-based distance for content based hyperspectral image retrieval, Pattern Recognition (2012), http://dx.doi.org/10.1016/j.patcog.2012.03.015
6.1. HyMAP dataset

The hyperspectral HyMAP data was made available from HyVista Corp. and German Aerospace Center’s (DLR) optical Airborne Remote Sensing and Calibration Facility service.\(^3\) The sensed scene corresponds to the radiance captured by the sensor in a flight line over the facilities of the DLR center in Oberpfaffenhofen (Germany) and its surroundings, mostly fields, forests and small towns. Fig. 10 shows the scene captured by the HyMAP sensor. The data cube has 2878 lines, 512 samples and 125 bands; and the pixel values are represented by 2-B signed integers.

We cut the scene in blocks of 64 × 64 pixels size for a total of 360 image blocks forming the hyperspectral database used in the experiments. We label the image blocks by visual inspection in five classes. There are three image classes with a definite meaning: Forests, Fields and Urban, representing image blocks where the corresponding kind of terrain covers almost all the surface. Image blocks containing two or three of these terrains are labeled as Mixed. Finally, image blocks which are not easily categorized by visual inspection are labeled as Others.

6.2. Performance measures on the real data

Image blocks in the real dataset are labeled, therefore, there are some small but substantial differences in the definition of the performance measures that need some detailed definition. For each image \(H_i\) in the database we calculate the dissimilarity between \(H_i\) and each of the remaining images in the dataset. These dissimilarities are represented as a vector \(s_i = [s_{i1}, \ldots, s_{iN}]\), where \(N\) is the number of images in the dataset and \(s_{ip}\) is the dissimilarity between the images \(H_i\) and \(H_p\), with \(z, \beta = 1, \ldots, N\). We define the ranking of the images in a dataset relative to image \(H_i\), \(\Omega_i = \{s_{ip} \in \{1, \ldots, N\} ; p = 1, \ldots, N\}\), as the set of image indexes ordered according to increasing values of their corresponding entries in the dissimilarity vector \(s_i\). That is, we sort in increasing order the components of \(s_i\), and the resulting permutation of image indices constitute \(\Omega_i\), so that \(s_{i1} \leq s_{i2} \leq \ldots \leq s_{iN}\). In this real dataset, the ground-truth is given by the manual labeling of the image blocks described above.

A query \(Q_k(H_j)\) of scope \(k\) is a search for the \(k\) most similar (less dissimilar) images \(H_p\) in the database respect to the image \(H_j\), with \(1 \leq k \leq N\). The set of returned images \(T_k(H_j)\) and the set of relevant images \(V(H_j)\) for a query \(Q_k(H_j)\) are defined as follows:

\[
T_k(H_j) = \Omega_{s_k} = \{s_{ip} \text{ s.t. } s_{ip} \leq s_{ik}\},
\]

\[
V(H_j) = \{l \mid c(H_l) = c(H_j)\},
\]

where \(c(H) \in \{\text{Forest, Fields, Urban area, Mixed, Others}\}\) denotes the class of the image \(H\). The query scope \(k\) is class-normalized when it is defined as a multiple of the class size, i.e. \(k(H_j) = n|V(H_j)|\). Class normalized scope is recommended to compensate for class size [23,30].

The definitions of precision \(P_k(H_j)\) and recall \(R_k(H_j)\) for a query \(Q_k(H_j)\) are slightly different in this situation than those given in Eqs. (12) and (13) because the set of relevant images does not depend on the scope, they now read as follows:

\[
P_k(H_j) = \frac{|V(H_j) \cap T_k(H_j)|}{|T_k(H_j)|},
\]

\[
R_k(H_j) = \frac{|V(H_j) \cap T_k(H_j)|}{|V(H_j)|}.
\]

---

\(^3\) http://www.OpAiRS.aero
The average precision \( P_k \) and recall \( R_k \) are defined as in Eqs. (14) and (15). The Rank of \( H_a \) is defined as in Eq. (16). There is no order defined in the ground truth classes, therefore, the rank at which the \( i \)th relevant image is retrieved is \( \phi_i = i \cdot \delta(c(H_a) = c(H_{\text{true}})) \), where \( \delta(x) = 1 \) if \( x \) is true, 0 otherwise. This measure is 0 for perfect performance, and approaches 1 as the performance worsens. The average normalized rank (ANR) for the whole dataset is given by Eq. (17).

There are proposals of forcing an upper bound for the value of \( \phi_i \) in order to avoid bias in ANR against the smaller classes [30]. However, we find it a much more delicate issue setting of the appropriate upper bound, and its fine tuning to avoid the exact reverse effect, that of biasing the results too much against more populated classes. Therefore, we stick with the conventional ANR. Besides, the results comparing the effect of dissimilarity functions, endmember induction method, and endmember distances are computed marginalizing the effect of the class size.

Finally, we compute the generalization plots proposed by [23]. Generality, defined [23] as \( g = (N_i/N) \), measures the degree of embedding of the relevant class, it goes to zero as the size of the complete database grows. When the relative scope \( a = (N_i/k) \) is \( a=1 \) the recall and precision values are equal. The generality-precision-recall plots show the values of the precision at relative scope \( a=1 \) for growing \( g \) in a logarithmic scale. The aim is to assess the effect of the database growth on the system performance.

6.3. Results

Table 4 contains some statistics of the dataset: the number of image blocks per class, and the mean and standard deviation of the number of endmembers induced from the data. Note the high standard deviations of the numbers of endmembers induced within the classes.

We have explored the sensitivity of the robust LTS Hausdorff dissimilarity to its \( L \) parameter, computing the precision, recall and ranking for a regular sampling in the interval [0, 1]. Figs. 11 and 12 show the average precision–recall curves for the EIHA and N-FINDR endmember induction algorithms, respectively, using the complete database of image blocks extracted from the DLR HyMap image. The plots are the average of the precision and recall for the three main classes: Forest, Fields, and Urban. The Others and Mixed classes are considered nuisance classes. The plots correspond to the different dissimilarity measures considered: The Grana distance, the Hausdorff distance and the robust LTS Hausdorff dissimilarity for the specific optimal value of \( L \). The Grana distance performs better than the remaining dissimilarities, regardless of the endmember distance (Euclidean or SAM) and endmember induction algorithm (EIHA or N-FINDR), with the only exception of the robust LTS Hausdorff dissimilarity on N-FINDR endmembers using the SAM distance. The robust LTS Hausdorff dissimilarity performs better with the N-FINDR endmembers than with the EIHA, contrary to the Hausdorff distance. Notice that the optimal setting for the \( L \) parameter seems to be rather arbitrary, it changes completely depending on the endmember induction algorithm and the distance between endmembers.

We plot in Fig. 13 the average ANR over the cover classes weighted by the number of samples of each class obtained using the robust LTS Hausdorff distance for varying values of its \( L \) parameter. The value of \( L \) has strong influence on the ANR with some trend of improvement towards the higher range of values in some instances of the plot. However, optimal values appear almost anywhere. There is a definite strong effect of the endmember induction algorithm and the endmember distance, the ANR improves with N-FINDR and SAM.

Tables 5 and 6 show the average normalized rank (ANR) values for the Grana and the Hausdorff distances, and the robust LTS Hausdorff dissimilarity (for its optimal value of \( L \) given inside brackets, for the EIHA and N-FINDR endmember induction algorithms, respectively. Overall, the Grana distance obtains the best average ANR. We specify the ANR for each image block class and combination of distances. There is no definite effect of the

### Table 4

Statistics of the dataset constituted by the image blocks extracted from the real hyperspectral image.

<table>
<thead>
<tr>
<th>Class</th>
<th>EIHA</th>
<th>N-FINDR</th>
<th># blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>Forests</td>
<td>6.84</td>
<td>2.10</td>
<td>7.30</td>
</tr>
<tr>
<td>Fields</td>
<td>7.75</td>
<td>2.67</td>
<td>6.16</td>
</tr>
<tr>
<td>Urban</td>
<td>7.58</td>
<td>2.14</td>
<td>4.33</td>
</tr>
<tr>
<td>Mixed</td>
<td>6.92</td>
<td>2.36</td>
<td>5.83</td>
</tr>
<tr>
<td>Others</td>
<td>5.82</td>
<td>1.74</td>
<td>5.31</td>
</tr>
</tbody>
</table>

Please cite this article as: M. Gran˜a, M.A. Veganzones, An endmember-based distance for content based hyperspectral image retrieval, Pattern Recognition (2012), http://dx.doi.org/10.1016/j.patcog.2012.03.015
endmember distance, though there are some instances where the Euclidean distance between endmembers show a dramatic increase in value (worsening). The effect of the feature distance on each class’ ANR has some big variations. For the Forest, Mixed and Others classes, the robust LTS Hausdorff semimetric improves the Hausdorff distance, while the contrary happens in the Fields, Urban classes. All distances considered, the worst results correspond to the Mixed and Others classes, which can be attributed to their heterogeneity and high confusion with the other classes in terms of their endmembers, because these classes contain heterogeneous regions. We do not appreciate a clear bias of the ANR measure against small sized classes as claimed in [30]. The smaller classes, Forest, Urban and Others, have different average ANR values. The worst is the Others class, but the others have values comparable to the most abundant classes. In fact, the Mixed class is very large and has worse results than Forest and Urban.

Figs. 14 and 15 show the average precision-recall plot for the real image database. Endmembers induced by EIHA.

<p>| Table 5 | ANR values for real image database (optimal column values in bold). Endmembers computed with EIHA. Dissimilarities: Grana (G), Hausdorff distance (HD), robust LTS Hausdorff dissimilarity (HD-LTS) for the optimal value of L. |</p>
<table>
<thead>
<tr>
<th>Distances</th>
<th>Forest</th>
<th>Fields</th>
<th>Urban</th>
<th>Mixed</th>
<th>Others</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>ED</td>
<td>0.16</td>
<td>0.16</td>
<td>0.40</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.20</td>
<td>0.17</td>
<td><strong>0.22</strong></td>
<td><strong>0.36</strong></td>
<td>0.36</td>
</tr>
<tr>
<td>HD</td>
<td>ED</td>
<td>0.22</td>
<td><strong>0.15</strong></td>
<td>0.29</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.25</td>
<td>0.19</td>
<td>0.25</td>
<td>0.37</td>
<td>0.40</td>
</tr>
<tr>
<td>HD-LTS</td>
<td>ED (L=0.9)</td>
<td><strong>0.15</strong></td>
<td>0.19</td>
<td>0.29</td>
<td><strong>0.36</strong></td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>SAM (L=0.35)</td>
<td>0.20</td>
<td>0.20</td>
<td><strong>0.22</strong></td>
<td><strong>0.36</strong></td>
<td><strong>0.32</strong></td>
</tr>
</tbody>
</table>

<p>| Table 6 | ANR values for real image database (optimal column values in bold) endmembers computed with N-FINDR. Dissimilarities: Grana (G), Hausdorff distance (HD), robust LTS Hausdorff dissimilarity (HD-LTS) for the optimal value of L. |</p>
<table>
<thead>
<tr>
<th>Distances</th>
<th>Forest</th>
<th>Fields</th>
<th>Urban</th>
<th>Mixed</th>
<th>Others</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>ED</td>
<td>0.11</td>
<td><strong>0.20</strong></td>
<td>0.14</td>
<td>0.36</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.10</td>
<td><strong>0.20</strong></td>
<td><strong>0.12</strong></td>
<td>0.35</td>
<td>0.39</td>
</tr>
<tr>
<td>HD</td>
<td>ED</td>
<td>0.21</td>
<td>0.21</td>
<td>0.22</td>
<td>0.38</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>SAM</td>
<td>0.19</td>
<td><strong>0.20</strong></td>
<td>0.20</td>
<td>0.38</td>
<td>0.43</td>
</tr>
<tr>
<td>HD-LTS</td>
<td>ED (L=0.4)</td>
<td><strong>0.09</strong></td>
<td>0.22</td>
<td>0.16</td>
<td><strong>0.34</strong></td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>SAM (L=0.55)</td>
<td><strong>0.09</strong></td>
<td>0.21</td>
<td>0.14</td>
<td><strong>0.34</strong></td>
<td><strong>0.34</strong></td>
</tr>
</tbody>
</table>

Fig. 13. Weighted average over the cover classes of the ANR computed for the robust LTS Hausdorff dissimilarity as a function of its parameter L.
Fields, Forests, and Urban classes, providing three points of the generality plot per tested distance. The results show that the Grana distance is the most robust dissimilarity function. In general, the performance decrease with the size of the database is less for NFINDR than for EIHA induced endmembers. We find that there are strong handicaps for the use of the robust LTS Hausdorff dissimilarity in practical applications: (1) the need to tune the L parameter and (2) the uncertainty of its optimality when minor changes in the system are introduced, or even depending on the performance measure considered.

Fig. 16 shows the response to a Fields image block query on the basis of the EIHA induced endmembers, given by the first ten ranked image blocks. The query image contains a small road crossing the fields with diverse degrees of growth covering. The best collection of responses is Fig. 16(d) provided by the SAM endmember distance and the Grana distance. Overall, responses to this query are rather natural, even images containing some building or big road features, have big regions corresponding to fields.

Fig. 17 shows the response to a Urban image block query on the basis of the NFINDR induced endmembers, given by the first ten ranked image blocks. The Urban areas contain some trees and green regions, therefore, there is some confusion between this class and the vegetation classes. All responses contain several images corresponding to Fields class. The Hausdorff dissimilarities contain more such confusions than the Grana distance. In fact, the best response is Fig. 17(d) corresponding to the Grana distance.

7. Conclusions

In this paper we introduce a feature extraction process and a distance between hyperspectral images which can be used for CBIR system on databases of hyperspectral images. Image features are the endmembers induced from the image data by some endmember induction algorithm (EIA). The distance between images is computed over these induced endmembers. We have tested the sensitivity of the CBIR system defined on this distance to the EIA used, the individual endmember distance, the image size, and the diversity of the underlying ground-truth on a large database of synthetic hyperspectral images. We find the CBIR system to be sensitive to the ground-truth diversity and the image size. We find that the proposed distance improves the Hausdorff distance for the same CBIR task. Besides, we have tested the CBIR system on a dataset built from a large real hyperspectral image, specifically we find that the proposed distance improves over the Hausdorff distance, and a robust LTS Hausdorff semimetric.

Among the avenues of future research, the study of appropriate image normalization procedures is of paramount importance, because data variations due to the image capture conditions and parameters may have a big impact on system performance. The definition of user interaction interfaces and
The authors very much acknowledge the support of Drs. Martin Bachmann and Mihai Datcu from DLR. Miguel A. Veganzones was supported by predoctoral grant BFI_07.225 from the Basque Government.

Appendix A. Miscellaneous results

Lemma 11. The dissimilarity function $s_t(H_s,H_B)$ of Eq. (8) is a semimetric.

Proof. The dissimilarity function $s_t(H_s,H_B)$ of Eq. (8) complies with the axioms required for a semimetric for any hyperspectral images $H_s$, $H_B$.

- Non-negativity: $s_t(H_s,H_B) \geq 0$ because all $m_{x_i}^0 \geq 0$ and there is not any operation in the computation of $s_t(H_s,H_B)$ that may produce a negative outcome.

- Identity of indiscernibles: it is restricted to the equivalence of images whose endmembers obtained by the EIA are the same.
  - If $E_s = E_B$ then $m_{x_i}^0 = m_{x_j}^0 = 0$, and consequently $s_t(H_s,H_B) = 0$. This is true even if the EIA finds a permutation of the endmembers when applied to the same or different images.
  - On the other hand, if $s_t(H_s,H_B) = 0$ then all $m_{x_i}^0 = m_{x_j}^0 = 0$. Therefore, for each endmember in $E_s$ there is another identical to it in $E_B$ and, conversely, for each endmember in $E_B$ there is another identical to it in $E_s$. Thus, $E_s = E_B$.

- Symmetry: is immediate if we note that $D_{x,y} = (D_{y,x})^T$: $s_t(H_s,H_B) = \max(h_t(m_n), h_t(m_p)) = \max(h_t(m_p), h_t(m_n)) = s_t(H_B,H_s)$.

Lemma 12. The dissimilarity function $s_t(H_s,H_B)$ of Eq. (8) is a semimetric.

Proof. The proof is identical to that of Lemma 11.

Lemma 13. The triangle inequality does not hold everywhere for dissimilarity function $s_t(H_s,H_B)$ of Eq. (8).

Proof. The triangle inequality on this dissimilarity $s_t(H_s,H_B) \leq s_t(H_s,H_f) + s_t(H_f,H_B)$, is rewritten

$$\max(h_t(m_n),h_t(m_p)) \leq \max(h_t(m_n),h_t(m_p)) + \max(h_t(m_p),h_t(m_n)).$$

From Lemma 3, for each $e_j \in E_s$ there is a $e_j \in E_B$ such that $d(e_j,e_j') = \min_{i=1,...,p} d(e_j,e_j') = m_{x_i}^0$. In the same way, there is a $e_j \in E_S$ such that $d(e_j,e_j') = \min_{k=1,...,q} d(e_j,e_j') = m_{x_k}^0$. By the triangle inequality of the distance between endmembers we have that $d(e_j,e_j') = d(e_j,e_j') + d(e_j',e_j')$. We proved in Lemma 3 that $d(e_j,e_j') = \min_{i=1,...,p} d(e_j,e_j') = m_{x_i}^0$. Therefore, the triangle inequality $m_{x_i}^0 \leq m_{x_i}^0 + m_{x_i}^0$ holds. Taking into account that $\rho_i$ is monotonically increasing, we have $\rho_i(m_{x_i}^0) \leq \rho_i(m_{x_i}^0) + \rho_i(m_{x_i}^0)$.

Adding all the inequalities for the endmembers in $E_s$

$$\sum_{i=1}^{p_s} \rho_i(m_{x_i}^0) \leq \sum_{i=1}^{p_s} \rho_i(m_{x_i}^0) + \sum_{j=1}^{p_f} \rho_j(m_{x_j}^0),$$

where $J = \{d(e_j,e_j') = m_{x_i}^0 | i = 1,\ldots,p_s\}$. We allow $J$ to be a superset, i.e. allowing repeating elements corresponding to the
situation of several endmembers in \( E_s \) mapped to the same endmember in \( E_p \). Hence, we cannot prove

\[
h_t(m^g_j) \leq h_t(m^g_j) + h_t(m^g_j),
\]

unless \( h_t(m^g_j) \geq \sum_{j \notin F} \rho_t(m^g_j) \). We have three situations

1. \( \sum_{j \notin F} \rho_t(m^g_j) = h_t(m^g_j) \). This condition is true only if \( |J| = \rho_t \) and there are no repeated elements in \( J \).
2. \( \sum_{j \notin F} \rho_t(m^g_j) > h_t(m^g_j) \). This condition is true when the repeated elements in \( J \) are large enough to add up above \( h_t(m^g_j) \), i.e., repetitions of the maximum value, whatever the relation between \( J \) and \( F \).

Following a similar reasoning we find that the triangle inequality does not hold everywhere in data domain.

Following a similar reasoning we find that the triangle inequality

\[
h_t(m^g_j) \leq h_t(m^g_j) + h_t(m^g_j),
\]

for the dissimilarity \( s_t(H_s, H_p) \) of Eq. (9).

Proof. The triangle inequality on this dissimilarity

\[
s_t(H_s, H_p) \leq s_t(H_s, H_p) + s_t(H_p, H_p),
\]

is rewritten

\[
\delta_{s,\beta}(|p_x - p_y| + 1) \leq \delta_{s,\beta}(|p_x - p_y| + 1) + \delta_{s,\beta}(|p_y - p_z| + 1),
\]

where \( \delta_{s,\beta}(\cdot) = (\|m^g_\beta\| + \|m^g_\beta\|) \). We can rewrite it as follows:

\[
\delta_{s,\beta} \leq \delta_{s,\beta} + \delta_{s,\beta} \quad \text{for } \beta = 1, \ldots, m^g_\beta.
\]

Because \( (|p_x - p_y| + 1) \geq 1 \). Though the triangle inequality holds for \( \delta_{s,\beta} \) as proved in Theorem 4, there may be hyperspectral image triples where

\[
(\|p_x - p_y\| + 1) \leq 1 \quad \text{or} \quad (\|p_y - p_z\| + 1) \leq 1,
\]

breaking the inequality of Eq. (A.1). Therefore, the triangle inequality does not hold everywhere for this dissimilarity.

Appendix B. Endmember induction algorithms

For the sake of completeness, in this Appendix we present a brief review of the EIAs employed in the experiments, the geometric approach represented by N-FINDER, and the lattice computing approach by the EIHA. We will denote \( \{i(i) \in \mathbb{R}^p : i = 1, \ldots, N \} \) the collection of spectra corresponding to the pixels of the image, and \( E = \{e_1, e_2, \ldots, e_p \} \) the collection of endmembers induced from the image. The vector \( \overline{r} \) denotes the vector of component-wise means over the image pixels. The vector \( \sigma \) denotes the vector of component-wise standard deviations over the image pixels.

B.1. N-FINDER

Algorithm 1 presents the N-FINDER [50], which works by growing a simplex inside the data, beginning with a random set of pixels. The vertices of the simplex with higher volume are assumed to identify the endmembers. Previously, data dimensionality has been reduced to \( p - 1 \) dimensions, with \( p \) the number of endmembers searched for.

Let \( E \) be the matrix of endmembers augmented with a row of ones

\[
E = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
e_1 & e_2 & \cdots & e_p
\end{bmatrix},
\]

(B.1)

where \( e_i \) is a column vector containing the spectra of the \( i \)th endmember. The volume of the simplex defined by the endmembers is proportional to the determinant of \( E \)

\[
V(E) = \frac{|\text{det}(E)|}{(p - 1)!},
\]

(B.2)
The N-FINDER starts by selecting an initial random set of pixels as endmembers. Then, for each pixel and each stored endmember, the endmember is replaced with the spectrum of the pixel and the volume recalculated by Eq. (B.2). If the volume of the new simplex increases, the endmember is replaced by the spectrum of the pixel. The procedure ends when no more replacements are done. The N-FINDER is a greedy algorithm, prone to fall in local maxima of the volume function defined by Eq. (B.2).

**Algorithm 1.** N-FINDER algorithm.

1. Compute Principal Component Analysis (PCA) retaining the first principal components accounting for 99% of the sum of the eigenvalues. Let \( p-1 \) be the number of eigenvectors retained.
2. Randomly select \( p \) vectors from the data to initialize the set of induced endmembers \( E \).
3. Calculate the volume of the simplex \( v = V(E) \) of Eq. (B.2).
4. For each endmember \( e_k, k = 1, \ldots, p \):
   a. For each input vector \( f(i) \) \( i = 1, \ldots, N \):
      i. Form a new matrix \( E' \) by substituting the endmember \( e_k \) by the data vector \( f(i) \).
      ii. Calculate the volume of the simplex \( v' = V(E') \).
      iii. If \( v' > v_{\text{actual}} \) then \( E' \) becomes \( E \). \( v_{\text{actual}} = v' \).
5. Go to step 4.

**B.2. EIHA**

The Endmember Induction Heuristic Algorithm (EIHA) proposed in [19–21] is based on the fact that sets of Strong Lattice Independent vectors are affine independent [39] and thus, they can be interpreted as a collection of endmembers for the analysis of hyperspectral data. The necessary conditions for Strong Lattice Independence are Lattice Independence and max/min dominance. Lattice Independence is decided based on results on fixed points for Lattice Autoassociative Memories (LAMs) [34,38,44,39]. Max/min dominance can be tested directly [45], but in this algorithm it is a byproduct of the selection of the endmember candidate. Given a set of vector patterns \( X = \{x_1, \ldots, x_k\} \) and mimicking the constructive procedure for linear autoassociative memories [37,36] propose the constructions of dual LAMs as follows: \( W_{XX} = \sum_{i=1}^{k} [x_i \times (-x_i')] \) and \( M_{XX} = \sqrt[k]{\sum_{i=1}^{k} [x_i \times (-x_i')]^2} \), where \( \times \) is any of the \( \oplus \) or \( \ominus \) operators. Here \( \oplus \) and \( \ominus \) denote the max and min matrix products defined in [37,36]. A lattice dependent vector will be a fixed point of anyone of the dual LAMs constructed with the current \( E \).

The EIHA specified by Algorithm 2 is a fast heuristic endmember induction method that passes only once over each pixel of the hyperspectral image. It starts with a randomly picked input vector as the single initial endmember, tests if each pixel spectrum in the input hyperspectral image data is lattice independent relative to the already discovered endmembers. If so, the pixel spectrum is added to the set of endmembers. The following notation is used in Algorithm 2. The expression \( x > 0 \) denotes a vector of 0’s and 1’s, where the \( i \)-th component is 0 if \( x_i \leq 0 \), and 1 otherwise. Therefore, \( X \) is a collection of binary vectors corresponding to the sign of the endmembers’ components. Also, \( \Gamma(i) \) and \( f(i) \) are binary vectors, respectively, corresponding to the sign of the components of the dilated and eroded pixel spectrum \( f(i) \) by the vector of standard deviations \( \sigma \) scaled by a gain factor \( z \), whose conventional value is \( z = 2 \). The lattice independence is tested by computing the vectors \( y(i) \) and \( y(i) \) recalled from the erosive and dilative LAM, respectively, built from \( X \). If neither recall returns a vector already in \( X \) then we declare that the corresponding pixel spectrum is a new endmember. If we do not detect a new endmember but the pixel spectrum can be considered a dilation or an erosion of an already selected endmember, then it becomes the new endmember substituting the old one, ensuring the max/min dominance in the set of endmembers.

**Algorithm 2.** Endmember Induction Heuristic Algorithm (EIHA) for the induction of endmembers.

1. Shift the data sample to a global zero mean \( f'(i) = f(i) - \bar{f}, i = 1, \ldots, n \).
2. Initialize the set of endmembers \( E = \{e_1 = f'(i^0)\} \) where \( i^0 \) is a randomly picked sample index. The initial set of endmember sample indices is \( I = \{i^0\} \).
3. Initialize the set of lattice independent binary signatures \( X = \{x_1\} = \{e_1 > 0\} \).
4. Construct the LAM’s based on \( X \): \( M_{XX} \) and \( W_{XX} \).
5. For each input image feature vector \( f(i) \):
   a. Compute the variance induced dilations and erosions sign vectors \( f'(i) = f(i) + \sigma > 0 \) and \( f(i) = f(i) - \sigma > 0 \).
   b. Check Lattice Independence by computing \( y(i) = M_{XX} \ominus f(i) \) and \( y(i) = W_{XX} \ominus f(i) \).
   i. If \( y(i) X \) and \( y(i) X \) then \( f(i) \) is a new vertex to be added to \( E \), and \( f(i) = 0 \) is added to \( X \). Execute step 4 with the new \( X \) and resume the exploration of the data samples.
   ii. If \( y(i) X \) and \( f(i) > e_y \), then the pixel spectral signature is a dilation of the endmember \( e_y \), we substitute it by \( f(i) \).
   iii. If \( y(i) X \) and \( f(i) < e_y \), then the pixel spectral signature is an erosion of the endmember \( e_y \), we substitute it by \( f(i) \).
6. The output set of endmembers is the set of original data vectors \( f(i) : i \in I \) corresponding to the data vectors selected as members of \( E \).

**References**


