Further results on dissimilarity spaces for hyperspectral images RF-CBIR
Further results on dissimilarity spaces for hyperspectral images RF-CBIR

Miguel A. Veganzones\textsuperscript{a,b,*}, Mihai Datcu\textsuperscript{c,**}, Manuel Graña\textsuperscript{a,*}

\textsuperscript{a}Grupo de Inteligencia Computacional, Basque Country University (UPV/EHU), Spain
\textsuperscript{b}Gipsa-lab, Grenoble-INP, France
\textsuperscript{c}German Aerospace Center (DLR), Germany

Abstract

Content-Based Image Retrieval (CBIR) systems are powerful search tools in image databases that have been little applied to hyperspectral images. Relevance Feedback (RF) is an iterative process that uses machine learning techniques and user’s feedback to improve the CBIR systems performance. We pursued to expand previous research in hyperspectral CBIR systems built on dissimilarity functions defined either on spectral and spatial features extracted by spectral unmixing techniques, or on dictionaries extracted by dictionary-based compressors. These dissimilarity functions were not suitable for direct application in common machine learning techniques. We propose to use a RF general approach based on dissimilarity spaces which is more appropriate for the application of machine learning algorithms to the Hyperspectral RF-CBIR. We validate the proposed RF method for hyperspectral CBIR systems over a real hyperspectral dataset.

Keywords: Hyperspectral imaging, CBIR systems, dissimilarity spaces,
relevance feedback.

1. Introduction

The increasing interest in hyperspectral remote sensing [23] will yield to an exponential growth of hyperspectral data acquisition in a short time. Most spatial agencies have scheduled the launch of hyperspectral sensors on satellite payloads such as in EnMAP [6] or PRISMA [9] missions. That will involve the storage of a huge quantity of hyperspectral data. The problem of searching through these huge databases using Content-Based Image Retrieval (CBIR) techniques has not been properly addressed for the case of hyperspectral images until recently. Recent works on hyperspectral CBIR systems [8, 28] make use of spectral and spectral-spatial dissimilarity functions to compare hyperspectral images. The spectral and spatial features are extracted by means of spectral unmixing algorithms [10]. In [27], authors define dissimilarity functions built upon Kolmogorov complexity [15] and its approximation by compression and dictionary distances [29, 14]. Compression-based distances require a high computational cost that make it unaffordable for the definition of CBIR systems. Dictionary distances operate over dictionaries extracted from the hyperspectral images by the off-line application of a lossless dictionary-based compressor such as the Lempel-Ziv-Welch (LZW) compression algorithm [30]. In this work we pursued to extend these hyperspectral CBIR systems by using the feedback of the user.

Relevance Feedback (RF) is an iterative process that makes use of the feedback provided by the user to reduce the gap between the low-level feature representation of the images and the high-level semantics of the user’s
queries \cite{24}. Often, the user’s feedback comes on the form of a labelling of
the previously retrieved images as relevant or irrelevant for the query. The
set of labelled images is then used by the CBIR system to adapt the search
to the query semantics. If each image is represented by a point in a feature
space, the RF with both, positive and negative training examples, becomes
a two-class classification problem or an online learning problem in a batch
mode \cite{31}.

Dictionaries and spectral-spatial features extracted from hyperspectral
images cannot be directly represented as points in a feature space. Thus, they
do not fit easily in feature-based machine learning techniques employed for
the definition of RF processes. It is possible to treat dissimilarity functions as
kernel functions in order to use them in kernel-based method, for instance in
Support Vector Machine (SVM) \cite{21}. However, these dissimilarity functions
do not comply often with valid kernel conditions \cite{21}. Authors in \cite{20, 5}
propose the definition of dissimilarity spaces as an alternative to feature
spaces for machine learning. In dissimilarity spaces some data instances are
used as reference points named prototypes. The data samples are compared
to these prototype instances by some dissimilarity function. Then, for each
data sample, the dissimilarities to the prototypes define the data coordinates
in a so-called dissimilarity space. Thus, each prototype defines a dimension
in this dissimilarity space. The dissimilarity space is analogous to a feature
space so, once the data samples are represented as points in the dissimilarity
space, all the available potential of machine learning techniques can then be
used.

In this paper we propose the use of dissimilarity spaces to define a RF
methodology for hyperspectral CBIR making use of the already available spectral, spectral-spatial and dictionary dissimilarity functions. The use of dissimilarity spaces to define RF processes is scarce on the literature. In [19], authors propose the use of dissimilarities to prototypes selected by an offline clustering process as the entry to a RF process defined as an one-class classification problem. Authors in [7] perform an online prototypes selection instead, where the images retrieved to the user for evaluation are at the same time the prototypes and the training set. The RF process is defined as a new dissimilarity function based on the combination of the database images dissimilarities to the set of prototypes and the prototypes labeling. In [2], authors propose different strategies to characterize an image by a feature vector based on the combination of dissimilarities to a set of prototypes. We propose an hyperspectral RF process defined as a two-class classification problem based on dissimilarity spaces. The input to the classifier is a dissimilarity representation defined over the unmixing and dictionary-based hyperspectral dissimilarity functions respect to offline and online selected prototypes.

The paper is divided as follows. In section 2 we outline the dissimilarity functions used in the definition of hyperspectral CBIR systems and in section 3 we outline the dissimilarity spaces approach. In section 4 we introduce the proposed hyperspectral RF process. In section 5 we define the experimental methodology and in section 6 we comment on the results. Finally, we contribute with some conclusions in section 7.
2. Hyperspectral dissimilarity functions

Here, we outline the dissimilarity functions used on the literature to compare hyperspectral images. Firstly, we describe the spectral and spectral-spatial dissimilarity functions defined over the results of a spectral unmixing process. Secondly, we describe the dictionary distance defined over dictionaries extracted from the hyperspectral images by means of lossless dictionary-based compressors.

2.1. Unmixing-based dissimilarity functions

Spectral unmixing pursues the decomposition of an hyperspectral image into the spectral signatures of its main constituents and their corresponding spatial fractional abundances. Most of the unmixing methods are based on the Linear Mixing Model (LMM) \[11, 1\]. The LMM states that an hyperspectral sample is formed by a linear combination of the spectral signatures of pure materials present in the sample (endmembers), plus some additive noise. Often, the spectral signatures of the materials are unknown, and the set of endmembers must be built by either manually selecting spectral signatures from a spectral library, or by automatically inducing them from the image itself. The latter involves the use of some endmember induction algorithm (EIA). The hyperspectral literature features plenty of such algorithms. Some reviews on the topic can be found in \[22, 26, 1\]. Once the set of endmembers has been induced, their corresponding per-pixel abundances can be estimated by a Least Squares method \[12\].

The dissimilarity functions based on the spectral unmixing make use of the spectral and spectral-spatial characterization of the hyperspectral im-
Given an hyperspectral image, $H_{\alpha}$, whose pixels are vectors in a $q$-dimensional space, its spectral characterization is defined by the set of endmembers, $E_{\alpha} = \{e_{\alpha}^1, e_{\alpha}^2, \ldots, e_{m_{\alpha}}^\alpha\}$, where $m_{\alpha}$ denotes the number of induced endmembers from the $\alpha$-th image. The spectral-spatial characterization is defined as the tuple $(E_{\alpha}, \Phi_{\alpha})$, where $\Phi_{\alpha} = \{\phi_{\alpha}^1, \phi_{\alpha}^2, \ldots, \phi_{m_{\alpha}}^\alpha\}$ is the set of fractional abundance maps resulting from the unmixing process. To implement this approach, an EIA is first used to induce the endmembers from the image and then, their respective fractional abundances are estimated by a Least Squares Unmixing algorithm.

In order to compute the unmixing-based dissimilarities, the Spectral Distance Matrix (SDM), $D_{\alpha,\beta}$, between two given hyperspectral images, $H_{\alpha}$ and $H_{\beta}$, has first to be computed. The SDM is the matrix $D_{\alpha,\beta} = [d_{ij}]$, $i = 1, \ldots, m_{\alpha}$, $j = 1, \ldots, m_{\beta}$, whose elements $d_{ij}$ are the pairwise distances between the endmembers $e_{\alpha}^i, e_{\beta}^j \in \mathbb{R}^q$ of each image. The spectral distance function $d : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}^+$ is often the angular pseudo-distance:

$$d(e_i, e_j) = \cos^{-1}\frac{e_i e_j}{\|e_i\| \|e_j\|}. \tag{1}$$

The Spectral dissimilarity [8] is then given by:

$$s_E(H_{\alpha}, H_{\beta}) = \|m_r\| + \|m_c\|, \tag{2}$$

where $\|m_r\|$ and $\|m_c\|$ are the Euclidean norms of the vectors of row and column minimal values of the SMD, respectively. The Spectral-Spatial dissimilarity [28] is given by:

$$s_{E,\Phi}(H_{\alpha}, H_{\beta}) = \sum_{i=1}^{m_{\alpha}} \sum_{j=1}^{m_{\beta}} r_{ij}d_{ij}, \tag{3}$$
where \(d_{ij}\) is the aforementioned spectral distance and \(r_{ij}\) is the significance associated to \(d_{ij}\). The significance matrix \(R_{\alpha,\beta} = [r_{ij}], i = 1, \ldots, m_\alpha, j = 1, \ldots, m_\beta\) is calculated on base to the normalized average abundances \(\bar{\Phi}_\alpha\) and \(\bar{\Phi}_\beta\) by the most similar highest priority (MSHP) principle [13].

### 2.2. Dictionary-based dissimilarity functions

Given a signal \(x\), a dictionary-based compression algorithm looks for patterns in the input sequence from signal \(x\). These patterns, called *words*, are subsequences of the incoming sequence. The compression algorithm result is a set of unique words called *dictionary*. The dictionary extracted from a signal \(x\) is hereafter denoted as \(D(x)\), with \(D(\lambda) = \emptyset\) only if \(\lambda\) is the empty signal. The Normalized Dictionary Distance (NDD) [16] is given by:

\[
s_{\text{NDD}}(x, y) = \frac{D(x \cup y) - \min \{D(x), D(y)\}}{\max \{D(x), D(y)\}} ,
\]

where \(D(x \cup y)\) and \(D(x \cap y)\) respectively denote the union and intersection of the dictionaries extracted from signals \(x\) and \(y\). The NDD is a normalized admissible distance satisfying the metric inequalities. Thus, it results in a non-negative number in the interval \([0, 1]\), being zero when the compared signals are equal and increasing up to one as the signals are more dissimilar.

### 3. Dissimilarity spaces

The dissimilarity space is a vector space in which the dimensions are defined by dissimilarity vectors measuring pairwise dissimilarities between individual objects and reference objects (prototypes) [5]. Given a set of prototypes \(P = \{p_1, \ldots, p_r\}\), where \(r\) denotes the number of prototype objects
on \( P \), and a set of objects \( X = \{x_1, \ldots, x_N\} \), where \( N \) denotes the number of individual objects on \( X \), the dissimilarity representation \( D(X, P) \) is a data-dependent mapping \( D(\cdot, P) : X \rightarrow \mathbb{R}^r \) from a set of objects \( X \) to the dissimilarity space specified by the prototypes set \( P \). Each dimension in the dissimilarity space corresponds to a dissimilarity to a prototype object, \( D(X, p_i) \). The dissimilarity representation \( D(X, P) \) is thus defined as a \( N \times r \) dissimilarity matrix, where each object \( x \in X \) is described by a vector of dissimilarities \( s_x = D(x, P) = [s(x, p_1), \ldots, s(x, p_r)] \). The pairwise dissimilarity function \( s(x, p_i) \) is not required to be metric and can be defined \textit{ad-hoc} for the given prototype. The dissimilarity space is a vector space equipped with an inner product and an Euclidean metric. Thus, the vector of dissimilarities to the set of prototypes, \( s_x \), can be interpreted as a feature, allowing the use of machine learning techniques commonly defined over feature spaces.

4. Relevance feedback by dissimilarity spaces

The use of dissimilarity spaces allows one to use the previously mentioned hyperspectral dissimilarity functions to define a RF process based on conventional machine learning techniques. The proposed hyperspectral RF process follows the general approach in [7, 2, 19] and it is depicted in Fig. [1]. First, the user defines a \textit{zero-query} by feeding the system with some positive sample. Next, an initial ranking is obtained comparing the database images to the query sample by some hyperspectral dissimilarity function and some images are retrieved for user’s evaluation. Then, the user labels the images retrieved by the system, a set of prototype images is selected and the RF process starts.
We follow by describing the zero-query and the relevance feedback processes in detail, and then we discuss on the prototypes selection and the selection of the images retrieved by the system for evaluation.

4.1. Zero query

First, a query \( Q_l (H_\alpha) \) is defined following the query-by-image approach. \( H_\alpha \) denotes the hyperspectral image selected as the query and \( l \in \mathbb{Z}^+ \), named the scope of the query, denotes the number of images that should be retrieved by the system. Every image \( H_\beta \) in the dataset is compared to the query image by some hyperspectral dissimilarity function, \( s(H_\alpha, H_\beta) \). The dissimilarities
to the query image are represented as a vector $s_\alpha = [s_{\alpha,1}, \ldots, s_{\alpha,N}]$, where $N$ is the number of images in the dataset and $s_{\alpha,\beta}$ is the dissimilarity between the query image $H_\alpha$ and the dataset image $H_\beta$, with $\beta = 1, \ldots, N$. Then, we sort the components of $s_\alpha$ in increasing order, and the resulting shuffled image indexes constitute the zero ranking $\Omega_\alpha^0 = [\omega_q \in \{1, \ldots, N\}]$, $q = 1, \ldots, N$, so that $s_{\alpha,\omega_q} \leq s_{\alpha,\omega_{q+1}}$. Then, some selection criterion is followed to select $l$ images from the zero ranking and retrieve them for user’s evaluation. The user labels these images as relevant or non-relevant for the query. The set of relevant images, denoted as $R$, and the set of non-relevant images, denoted as $NR$, form the training set, $T = \{R \cup NR\}$, with which the relevance feedback process starts.

### 4.2. Relevance feedback

We propose a RF process defined as a two-class problem where the classes are the set of relevant (positive class) and the set of irrelevant (negative class) images respect to the query. The input to the two-class classifier is a feature vector composed of the dissimilarity values computed from a given image respect to each of the images of the prototypes set. The output of the classifier should be an scalar representing any measure of an image identification with the positive class respect to the negative class, for instance a class probability. The classifier outputs are ordered to define a ranking of the database images respect to the user’s query. Finally, the ranking is used to select some database images that will be retrieved for the user’s evaluation and so, proceed with a new RF iteration. Thus, the RF process is divided in two steps, a training phase and a testing phase.
4.2.1. Training phase

Let \( P = \{ H_{p_i} \}_{i=1}^{r} \) be the set of prototypes where \( p_i \) is an index pointing to a database image and \( r \) is the number of prototype instances. Let \( T = \{ H_{q_j} \}_{j=1}^{t} \) be the set of training samples where \( q_j \) is an index pointing to a database image, \( t \) denotes the number of training samples and each image \( H_{q_j} \) has been labelled as belonging to the positive class, \( C^+ \), or to the negative class, \( C^- \). Then, the system calculates the \( t \times r \) dissimilarity matrix \( D(T, P) = [s(H_{q_j}, H_{p_i})]_{j=1}^{t}, i=1,\ldots,r; \) using some given hyperspectral dissimilarity function \( s(\cdot, \cdot) \). The rows of \( D(T, P) \) are the geometrical coordinates of the training samples in the dissimilarity space defined by the set of prototypes, and would be used as feature vectors to train the two-class classifier.

4.2.2. Testing phase

For each image \( H_{\beta} \) in the dataset we calculate the dissimilarity vector \( s_{\beta} = D(H_{\beta}, P) = [s(H_{\beta}, H_{p_i})]_{i=1}^{r} \), given the hyperspectral dissimilarity function \( s(\cdot, \cdot) \). The dissimilarity vector, \( s_{\beta} \), represents a point in the dissimilarity space and is used as the input to the trained classifier. The classifier will return an scalar, \( c_{\beta} \), measuring the probability or the degree of inclusion of the image \( H_{\beta} \) respect to the query class \( C^+ \). An image \( H_{k} \) having a classification value higher than an image \( H_{l} \), that is \( c_{k} \geq c_{l} \), should be ranked in a better position. The values obtained by the classifier for all the images in the dataset are then represented as a vector \( c_{\alpha} = [c_1, \ldots, c_N] \), where \( N \) is the number of images in the dataset. The vector of classification values \( c_{\alpha} \) is sorted in decreasing order and the resulting shuffled image indexes constitute the ranking \( \Omega^*_\alpha = [\omega^*_q \in \{1, \ldots, N\}], q = 1, \ldots, N \), so that \( c_{\omega^*_q} \geq c_{\omega^*_{q+1}} \).
superscript $t$ in $\Omega^t_{\alpha}$ denotes the iteration in turn on the RF process, being $t$ a positive integer, $t > 0$. The ranking serves to select some images that are retrieved to the user for evaluation, and then included in the training set. The RF process ends when the user is satisfied, a maximum number of iterations, $t_{\text{max}}$, is achieved, or no new images are being incorporated to the training set.

4.3. Prototypes selection

The general RF process depicted in Fig.1 requires of a set of prototypes. We distinguish between two criteria to build the prototypes set, an offline selection and an online selection. In the former, the prototypes are a priori representative subset of the images in the database. A common procedure is to perform a clustering and keep the centres of the clusters as the prototypes. This criterion could lead to a dramatical reduction in the computational costs of the CBIR system, but on the other hand it defines a fixed set of prototypes for all the possible queries, limiting the adaptability of the CBIR system. The later builds the set of prototypes during the RF process. In each iteration some images are retrieved to the user for evaluation and then included on the training set. These same images or a subset of them are also used as prototypes. This allows to adapt the set of prototypes to the query. However, it increases the computational burden.

4.4. Image retrieval

A key aspect of RF-CBIR systems is the criterion to select from a given ranking those images that will be retrieved to the user for evaluation. Let $l$ denote the scope of the query, that is, the number of images that should be
retrieved to the user. If the criterion is to return the best $l$ images given by the $l$ best ranked images on the database, is likely that the training set is biased towards the positive class. So, a better criterion seems to retrieve the $l/2$ best images and the $l/2$ worst images, hereafter denoted as the Best-Worst (BW) criterion. However, the best and worst images are not necessarily the most informative ones. The active learning paradigm \[3\] states that the most ambiguous images, those that are close to the class boundaries, are the most informative. Thus, the Active Learning (AL) criterion will return the $l/2$ most ambiguous images labelled as belonging to the positive class, and the $l/2$ most ambiguous images labelled as belonging to the negative class.

5. Experimental methodology

5.1. 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\[1\]. The scene corresponds to a flight line over the facilities of the DLR center in Oberpfaffenhofen (Germany) and its surroundings, mostly fields, forests and small towns. The data cube has 2878 lines, 512 samples and 125 spectral bands. We have removed non-informative bands due to atmospheric absorption and 113 spectral bands remained.

We cut the scene in patches of $64 \times 64$ pixels size for a total of 360 patches forming the hyperspectral database used in the experiments. We grouped the patches by visual inspection in five rough categories. The three main cat-

\[1\]http://www.OpAiRS.aero
Figure 2: Examples of the five categories patches: (a) Forests, (b) Fields, (c) Urban Areas, (d) Mixed, (e) Others.

5.2. Methodology

We test the use of the proposed hyperspectral RF-CBIR using the unmixing and dictionary-based hyperspectral dissimilarity functions. For the unmixing-based dissimilarities, the spectral [2] and the spectral-spatial [3] dissimilarity functions, we conduct for each image in the database an unmixing process in order to obtain the set of induced endmembers and their corresponding fractional abundances. In order to do that we use the Vertex Component Analysis (VCA) [18] endmember induction algorithm and a partially constrained least squares unmixing (PCLSU) [12] algorithm. As VCA
is an stochastic algorithm we perform 20 independent runs for each image and we keep the one with the lowest averaged root squared mean reconstruction error:

$$\epsilon(H, \hat{H}) = \frac{1}{M} \sum_{i=1}^{M} \sqrt{\frac{1}{q} \sum_{j=1}^{q} \left( H_i^{(j)} - \hat{H}_i^{(j)} \right)^2}$$

(5)

where $H_i^{(j)}$ denotes the $j$-th band value of the $i$-th pixel in the hyperspectral image $H$ and $\hat{H} = \Phi E$ is the hyperspectral image reconstructed by the set of induced endmembers $E$ and their corresponding fractional abundances $\Phi$.

For the Normalized Dictionary Distance (4), we first convert each hyperspectral image $H$ and $\hat{H} = \Phi E$ is the hyperspectral image reconstructed by the set of induced endmembers $E$ and their corresponding fractional abundances $\Phi$.

For the Normalized Dictionary Distance (4), we first convert each hyperspectral image to a text string in two ways, using the average of the spectral bands and band-by-band. For the former, we calculate the mean of each hyperspectral pixel along the spectral bands. For the later we transform each spectral band independently. In both cases we traverse the image in a zig-zag way.

The averaged band transformation incurs in a big lost of spectral information compared to the band by band transformation, but by contrast it yields to a more compact dictionary and so, to speed up the NDD computation.

Thus, we compare the use of the four hyperspectral dissimilarities, the Spectral, the Spectral-Spatial, the Averaged Band NDD and the Band-by-Band NDD, in the RF process respect to their use in the zero-query. In order to do that, we run independent retrieval experiments over the HyMAP dataset. Each of the 360 patches was a priori labelled as belonging to one of the five categories defined above. The query is a categorical search, where the images belonging to the same category than the query image form the positive class and the remaining ones form the negative class. We perform an independent search for each of the 360 patches. Thus, user’s evaluation
was not required and the experiment was fully automatized. The maximum number of iterations on the retrieval feedback process was set to \( t_{\text{max}} = 5 \).

For the RF process we compare the use of a \( k \)-NN classifier and a two-class SVM classifier with a radial basis kernel. The \( k \)-NN classifier does not require of a training phase and returns the fraction of the \( k \) most similar images in the training set respect to the query image belonging to the positive class, that is, 
\[
c = \frac{1}{k} \sum_{i=1}^{k} I(H_i, H_\alpha),
\]
where \( I \) denotes an indicator function returning 1 if the two images belong to the same class, and 0 otherwise. The SVM classifier outputs the probability that the tested image belongs to the positive class. The parameters of the SVM classifier where selected using a 5-fold cross validation. For the \( k \)-NN the \texttt{knnclassify} MATLAB function was used. For the SVM, we used the \( C \)-SVM classifier of the LIBSVM library.

We also compare the use of online and offline prototypes selection processes. For the offline prototypes selection process we performed a hierarchical segmentation using each of the four hyperspectral dissimilarity functions and we keep 10 clusters. Then, for each cluster \( \zeta \) we selected the image \( H^\circ_\zeta \) minimizing the averaged distance to the rest of images grouped into the same cluster:
\[
H^\circ_\zeta = \arg \min_i \frac{1}{|\zeta|} \sum_{H_j \in \zeta} s(H_i, H_j)
\]
where \( |\zeta| \) denotes the cardinality of the cluster \( \zeta \).

Finally, we compare the results obtained using three different criteria to select the images to be retrieved to the user for evaluation: the BW criterion, the AL criterion and a combination of both, BW+AL. For the BW criterion the system retrieves the 5 best and worst ranked images in the database. For the AL criterion the system retrieves the 5 most ambiguous positive and
negative instances, that is, the ones closed to the class boundary on each side. For both, BW and AL criteria, the scope is then $l = 10$. For the BW+AL criterion the system returns the 3 best and worst ranked images, and the 3 most ambiguous positive and negative instances, for a total scope of $l = 12$.

5.3. Performance measures

Evaluation metrics from information retrieval field have been adopted to evaluate CBIR systems quality. The two most used evaluation measures are precision and recall \[25, 4\]. Precision, $p$, is the fraction of the returned images that are relevant to the query. Recall, $q$, is the fraction of retrieved relevant images respect to the total number of relevant images in the database according to a priori knowledge. If we denote $L$ the set of returned images and $R$ the set of all the images relevant to the query, then $p = \frac{|L \cap R|}{|L|}$ and $r = \frac{|L \cap R|}{|R|}$. Precision and recall follow inverse trends when considered as functions of the scope of the query. Precision falls while recall increases as the scope increases. Thus, precision and recall measures are usually given as precision-recall curves for a fixed scope. To evaluate the overall performance of a CBIR system, the Average Precision and Average Recall are calculated over all the query images in the database. For a query of scope $l$, these are defined as:

$$\bar{p}_l = \frac{1}{N} \sum_{\alpha=1}^{N} p_l(H_{\alpha})$$ \hfill (7)$$

and

$$\bar{r}_l = \frac{1}{N} \sum_{\alpha=1}^{N} r_l(H_{\alpha})$$ \hfill (8)

The Normalized Rank \[17\] was used to summarize the system performance into an scalar value. The normalized rank for a given image query, denoted
as Rank \( (H_\alpha) \), is defined as:

\[
\text{Rank} \left( H_\alpha \right) = \frac{1}{NN_\alpha} \left( \sum_{i=1}^{N_\alpha} \Omega^i_\alpha - \frac{N_\alpha(N_\alpha - 1)}{2} \right),
\]

where \( N \) is the number of images in the dataset, \( N_\alpha \) is the number of relevant images for the query \( H_\alpha \), and \( \Omega^i_\alpha \) is the rank at which the \( i \)-th image is retrieved. This measure is 0 for perfect performance, and approaches 1 as performance worsens, being 0.5 equivalent to a random retrieval. We calculated the Rank \( (H_\alpha) \) for each of the images in the dataset and then we calculated the average normalized rank (ANR):

\[
\text{ANR} = \frac{1}{N} \sum_{\alpha=1}^{N} \text{Rank} \left( H_\alpha \right).
\]

6. Results

Tables 2-3 show the ANR \([10]\) values of the comparing hyperspectral dissimilarities, using the proposed RF-CBIR respect to the zero-query, for the Forest, Fields and Urban areas categorical queries respectively. We run the experiments using different values of \( k \) for the \( k \)-NN classifier, but we only show the results using \( k = 7 \) as in general it outperforms the other \( k \) values. The ANR results correspond to the ranking obtained in the fifth RF iteration. In general, the hyperspectral RF process yields to better ANR results than the zero query for the four compared hyperspectral dissimilarity functions. The online prototype selection leads to better results than the offline selection, and so it does the 7-NN classifier compared to the SVM classifier. The use of AL for the image retrieval selection outperforms the BW criterion, and often the combination of both, BW+AL. As it was expected, the results
Table 1: ANR values of the hyperspectral RF-CBIR for the Forests categorical search.

| Zero Query | Avg.Band NDD | By-Band NDD | Spectral | Spectral-Spatial |
|------------|--------------|-------------|----------|-----------------|
| Zero Query | 0.0809       | 0.0613      | 0.1360   | 0.0552          |
| Online Prot. |              |             |          |                 |
| 7NN BW     | 0.0343       | 0.0426      | 0.1394   | 0.0630          |
| 7NN AL     | **0.0280**   | **0.0258**  | 0.0869   | 0.0337          |
| 7NN BW+AL  | 0.0287       | 0.0281      | **0.0770** | **0.0330**     |
| SVM BW     | 0.0383       | 0.1392      | 0.2600   | 0.0852          |
| SVM AL     | 0.0596       | 0.1155      | 0.3947   | 0.2371          |
| SVM BW+AL  | 0.0462       | 0.0358      | 0.2143   | 0.2430          |
| Offline Prot. |              |             |          |                 |
| 7NN BW     | 0.0662       | 0.0723      | 0.1922   | 0.0543          |
| 7NN AL     | 0.0329       | 0.0631      | 0.1735   | 0.0494          |
| 7NN BW+AL  | 0.0448       | 0.0633      | 0.1848   | 0.0473          |
| SVM BW     | 0.0758       | 0.0478      | 0.2502   | 0.1063          |
| SVM AL     | 0.0542       | 0.0409      | 0.3116   | 0.1678          |
| SVM BW+AL  | 0.0642       | 0.0538      | 0.3180   | 0.1055          |

using the Band-by-Band NDD and the Spectral-Spatial dissimilarity functions outperform the Averaged Bands NDD and the Spectral dissimilarity functions.

There are however some discrepancies depending on the categorical query. This effect is specially relevant for the Urban areas category and it is related to the asymmetry in the number of images present in the database for each class. The low number of images belonging to the Urban areas category makes the training set very unbalanced yielding to poor classification results, and so, to a low performance in the CBIR ranking. Figures [45] show
Table 2: ANR values of the hyperspectral RF-CBIR for the Fields categorical search.

|                | Avg.Band NDD | By-Band NDD | Spectral | Spectral-Spatial |
|----------------|--------------|-------------|----------|------------------|
| **Zero Query** |              |             |          |                  |
|                | 0.2171       | 0.1641      | 0.1594   | 0.1599           |
| **Online Prot.** |              |             |          |                  |
| 7NN            |              |             |          |                  |
| BW             | 0.1552       | 0.0634      | 0.1776   | 0.1494           |
| AL             | **0.1388**   | **0.0495**  | 0.1573   | 0.1514           |
| BW+AL          | 0.1433       | 0.0587      | 0.1862   | 0.1883           |
| SVM            |              |             |          |                  |
| BW             | 0.1898       | 0.2462      | 0.1511   | 0.1983           |
| AL             | 0.1808       | 0.0914      | 0.1526   | **0.0924**       |
| BW+AL          | 0.1567       | 0.0812      | 0.1477   | 0.1184           |
| **Offline Prot.** |              |             |          |                  |
| 7NN            |              |             |          |                  |
| BW             | 0.1847       | 0.0756      | 0.2607   | 0.1779           |
| AL             | 0.1802       | 0.0533      | 0.2660   | 0.2158           |
| BW+AL          | 0.1694       | 0.0569      | 0.2957   | 0.1994           |
| SVM            |              |             |          |                  |
| BW             | 0.2033       | 0.0724      | 0.2136   | 0.1936           |
| AL             | 0.1831       | 0.0660      | 0.2112   | 0.1700           |
| BW+AL          | 0.2008       | 0.0497      | 0.2171   | 0.1442           |
Table 3: ANR values of the hyperspectral RF-CBIR for the Urban areas categorical search.

|                  | Zero Query | Avg.Band NDD | By-Band NDD | Spectral | Spectral-Spatial |
|------------------|------------|--------------|-------------|----------|------------------|
| Online Prot.     |            |              |             |          |                  |
| 7NN              | BW         | 0.1920       | **0.0082**  | **0.0509** | 0.0416           |
|                  | AL         | 0.1900       | 0.0096      | 0.0626   | **0.0392**       |
|                  | BW+AL      | 0.2702       | 0.0282      | 0.1230   | 0.0654           |
| SVM              | BW         | 0.2675       | 0.0437      | 0.1120   | 0.2126           |
|                  | AL         | 0.5870       | 0.0416      | 0.2501   | 0.1603           |
|                  | BW+AL      | 0.3825       | 0.0415      | 0.1459   | 0.1712           |
| Offline Prot.    |            |              |             |          |                  |
| 7NN              | BW         | 0.2578       | 0.0545      | 0.0799   | 0.0762           |
|                  | AL         | 0.2713       | 0.0276      | 0.0698   | 0.0570           |
|                  | BW+AL      | 0.3425       | 0.1061      | 0.1509   | 0.1224           |
| SVM              | BW         | 0.1562       | 0.0103      | 0.0833   | 0.1240           |
|                  | AL         | 0.2276       | 0.0273      | 0.2164   | 0.2651           |
|                  | BW+AL      | 0.1763       | 0.0246      | 0.0561   | 0.2032           |
the average number of relevant (R) and non-relevant (NR) images in the training set for each RF iteration using the BW and the AL image retrieval selection criteria for the Forests, Fields and Urban areas categorical queries respectively. It is clear that the Urban areas category presents the most asymmetrical distribution of the training set into relevant and non-relevant images, what it can explain the poor results on the RF process for this category. In general, the asymmetry in the R/NR ratio is not so important as soon as there are some critical number of each on the training set. It is also possible to observe that the AL selection criterion yields to better training sets compared to the BW selection criterion, expressed as bigger and more equally distributed training sets. This issue seems to be a major drawback for the SVM classifier while the impact on the 7-NN classifier is less severe as soon as there are enough positive samples present on the training set. This issue should be further addressed in future research in order to develop an operative hyperspectral RF-CBIR system.

Finally, Figures 6 and 7 show the P-R curves (7) (8) for the zero-query and the best RF results respectively, using the four comparing dissimilarity functions. The improve on the P-R curves by the RF process is clear except for the Urban areas categorical search, due to the pernicious effect of the lack of positive samples and the consequent asymmetrical distribution of R/NR samples on the training sets.
Figure 3: Average number of relevant (R) and non-relevant (NR) images in the training set for each RF iteration and comparing hyperspectral dissimilarity functions, using the BW and the AL image retrieval selection criteria for the Forests categorical search.
Online Prototypes - 7NN

Online Prototypes - SVM

Figure 4: Average number of relevant (R) and non-relevant (NR) images in the training set for each RF iteration and comparing hyperspectral dissimilarity functions, using the BW and the AL image retrieval selection criteria for the Fields categorical search.
Figure 5: Average number of relevant (R) and non-relevant (NR) images in the training set for each RF iteration and comparing hyperspectral dissimilarity functions, using the BW and the AL image retrieval selection criteria for the Urban areas categorical search.
Figure 6: Precision-Recall curves for the zero query.

Figure 7: Precision-Recall curves for the best RF results.
7. Conclusions

We have extended the hyperspectral CBIR systems present on the literature by a RF process based on dissimilarity spaces. To define a relevance feedback process for hyperspectral CBIR systems is not easy as most of the available hyperspectral CBIR systems rely on feature representations and dissimilarity functions that do not fulfil the conditions to be used in common machine learning RF processes. The proposed approach expands the available dissimilarity-based hyperspectral CBIR systems on the literature in a simple way by using dissimilarity space instead of the usual feature space. The proposed approach proved to improve the performance of the hyperspectral CBIR systems in the preliminary experiments presented on this paper. Also, the selection of a proper training set for the RF process was pointed as a major issue affecting the performance of the proposed hyperspectral RF-CBIR system. Further research will focus on this aspect and on the validation of the proposed system in a real scenario with a big database of hyperspectral images and real users.

Acknowledgements

The authors very much acknowledge the support of Dr. Martin Bachmann from DLR.

References

[1] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Qian Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. IEEE
Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(2):354–379, 2012.

[2] E. Bruno, N. Moenne-Loccoz, and S. Marchand-Maillet. Learning user queries in multimodal dissimilarity spaces. In M. Detyniecki, J.M. Jose, A. Nurnberger, and C.J. Rijsbergen, editors, Adaptive Multimedia Retrieval: User, Context, and Feedback, volume 3877 of Lecture Notes in Computer Science, pages 168–179. Springer Berlin Heidelberg, 2006.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.

[4] H. Daschiel and M. Datcu. Information mining in remote sensing image archives: system evaluation. Geoscience and Remote Sensing, IEEE Transactions on, 43(1):188–199, 2005.

[5] R.P.W. Duin and E. Pekalska. The dissimilarity space: Bridging structural and statistical pattern recognition. Pattern Recognition Letters, 33(7):826 – 832, 2012.

[6] German Spatial Agency (DLR). Environmental mapping and analysis program (enmap), 2011.

[7] G. Giacinto and F. Roli. Dissimilarity representation of images for relevance feedback in content-based image retrieval. In P. Perner and A. Rosenfeld, editors, Machine Learning and Data Mining in Pattern Recognition, volume 2734 of Lecture Notes in Computer Science, pages 202–214. 2003.
[8] Manuel Graña and Miguel A. Veganzones. An endmember-based distance for content based hyperspectral image retrieval. *Pattern Recognition*, 45(9):3472 – 3489, 2012.

[9] Italian Spatial Agency (ASI). Precursore iperspettrale of the application mission (prisma), 2011.

[10] N. Keshava and J. F Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1):44–57, January 2002.

[11] N. Keshava and J.F. Mustard. Spectral unmixing. *Signal Processing Magazine, IEEE*, 19(1):44–57, 2002.

[12] Charles L. Lawson. *Solving Least Squares Problems*. Prentice Hall, 1974.

[13] J. Li, J.Z. Wang, and G. Wiederhold. IRM: integrated region matching for image retrieval. In *Proceedings of the eighth ACM international conference on Multimedia*, MULTIMEDIA ’00, pages 147–156, Marina del Rey, California, United States, 2000. ACM.

[14] Ming Li, Xin Chen, Xin Li, Bin Ma, and P.M.B. Vitanyi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, 2004.

[15] Ming Li and Paul Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 2nd edition, February 1997.

[16] A. Macedonasa, D. Besiris, G. Economou, and S. Fotopoulos. Dictionary based color image retrieval. *Journal of Visual Communication and Image Representation*, 19(7):464–470, 2008.
[17] H. Muller, W. Muller, D.McG. Squire, S. Marchand-Maillet, and T. Pun. Performance evaluation in content-based image retrieval: overview and proposals. Pattern Recognition Letters, 22(5):593–601, 2001.

[18] J.M.P. Nascimento and J.M. Bioucas Dias. Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 43(4):898–910, 2005.

[19] G.P. Nguyen, M. Worring, and A.W.M. Smeulders. Similarity learning via dissimilarity space in cbir. In ACM International Multimedia Conference and Exhibition, pages 107–116, 2006.

[20] E. Pekalska and Robert P.W. Duin. The Dissimilarity Representation for Pattern Recognition: Foundations And Applications. World Scientific Pub Co Inc, 2005.

[21] E. Pekalska and B. Haasdonk. Kernel discriminant analysis for positive definite and indefinite kernels. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(6):1017–1032, 2009.

[22] A. Plaza, P. Martinez, R. Perez, and J. Plaza. A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing, 42(3):650–663, 2004.

[23] Antonio Plaza, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, and Giovanna Trianni. Recent advances in techniques
for hyperspectral image processing. *Remote Sensing of Environment, 113*, Supplement 1(0):S110–S122, September 2009.

[24] J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis.* Cambridge University Press, 2004.

[25] A.W.M. Smeulders, M. Worrin, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(12):1349–1380, 2000.

[26] M. A. Veganzones and M. Graña. Endmember extraction methods: A short review. In *Knowledge-Based Intelligent Information and Engineering Systems, 12th International Conference, KES 2008, Zagreb, Croatia, September 3-5, 2008, Proceedings, Part III*, volume 5179 of *Lecture Notes in Computer Science*, pages 400–407. Springer, 2008.

[27] M.A. Veganzones, M. Datcu, and M. Graña. Dictionary based hyperspectral image retrieval. In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 426–432. SciTePress, 2012.

[28] M.A. Veganzones and M. Graña. A spectral/spatial cbir system for hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5*(2):488–500, april 2012.

[29] T. Watanabe, K. Sugawara, and H. Sugihara. A new pattern representation scheme using data compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5):579–590, 2002.
[30] T.A. Welch. A technique for high-performance data compression. *Computer*, 17(6):8–19, 1984.

[31] Xiang Sean Zhou and Thomas S. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, 2003.