Validation of a Hyperspectral Content-Based Information Retrieval (RS-CBIR) System upon scarce data

Miguel Angel Veganzones, Manuel Graña

Abstract Validation of Remote Sensing Content-Based Information Retrieval (RS-CBIR) systems requires innovative strategies due to the scarcity of labelled data. CBIR systems validation by means of precision/recall measures based on either user feedback or a-priori known categories, are hard to apply to RS-CBIR systems. We propose to apply a data-driven (unsupervised) quality assessment strategy analogous to the DAMA strategy applied for the validation of classification methods used in thematic mapping. The strategy is intended for quality assessment when little or no ground truth is available. The proposed strategy deals with the RS-CBIR validation problem by giving a quantitative and qualitative evidence of the relative (subjective) quality of RS-CBIR systems without need of *a-priori* knowledge. We apply the proposed strategy to validate a Hyperspectral CBIR system.

1 Introduction

Modern imaging sensors continuously deliver enormous amounts of Earth Observation data, which couldn't be systematically exploited for a lack of appropriate methodology and analytical techniques. For instance, the German Aerospace Center (DLR) expects to launch along year 2014 a hyperspectral satellite mission, the Environmental Mapping and Analysis Program (EnMAP) [11], wich will generate a huge amount of hyperspectral data. Content Base Image Retrieval (CBIR) systems are relevant to the geosciences because they provide automated tools to explore and understand the contents of large and highly complex images [17, 9, 7]. There have been several efforts along this decade to develop CBIR tools for remote sensing images. The main focus has been multispectral and synthetic aperture radar (SAR)

images [8, 15, 16, 6, 5]. Exploitation of the spectral information provided by hyperspectral sensors by CBIR systems has not been deeply pursued although there are some instances in the literature[10, 14, 19].

In previous works [18] we dealt with the validation of the Hyperspectral CBIR system proposed in [19] using synthetic hyperspectral images. In this paper we consider the case of scarce ground truth knowledge about the data. We overcome the lack of available benchmark datasets whose data samples are divided into well defined categories. Such reliable benchmarks do not exist yet for RS-CBIR systems validation. We propose a methodology similar to [1] to we asses the problem of CBIR systems validation in a Remote Sensing (RS) context. The work in [1] deals with the quality of thematic maps produced by competing unsupervised classification algorithms, that must be applied because of the lack of ground truth data. Their data-driven quality map assesment (DAMA) technique is an alternative to the supervised classification building techniques that are useless when little or no ground truth are available. Similar to DAMA, our methodology creates a reference truth by the application of clustering algorithms on the image data. This reference truth validates the performance of the hyperspectral CBIR system. We apply the proposed quality assessment strategy to test the Hyperspectral CBIR system introduced in [19].

Section 2 and 3 give a brief overview of the Hyperspectral CBIR system and the DAMA strategy respectively. In section 4 we introduce the common quality measures used for CBIR systems validation, and we explain our proposed DAMA extension strategy for RS-CBIR validation. Section 5 provides a sample experiment of applicability over the Hyperspectral CBIR system. Finally we provide some conclusions in section 6.

2 Hyperspectral CBIR system

We describes here the Hyperspectral CBIR system introduced in [19]. A dissimilarity function between two hyperspectral images, $s\left(H_{\alpha}, H_{\beta}\right)$ is defined on the basis of the distances between their corresponding set of endmembers E_{α} and E_{β}. Let it be $E_{\alpha} = \left\{\mathbf{e}_1^{\alpha}, \mathbf{e}_2^{\alpha}, \ldots, \mathbf{e}_{p_{\alpha}}^{\alpha}\right\}$ the set of endmembers induced from the hyperspectral image H_{α} in the database, where p_{α} is the number of induced endmembers from the α-th image. Given two images, H_{α}, H_{β}, we compute the following matrix whose elements are the distances between the endmembers of each image:

$$D_{\alpha,\beta} = [d_{i,,j}; i = 1, \ldots, p_{\alpha}; j = 1, \ldots, p_{\beta}], \tag{1}$$

where $d_{i,j}$ is any defined distance between the endmembers $\mathbf{e}_i^{\alpha}, \mathbf{e}_j^{\beta} \in \mathbb{R}^q$; i.e. the Euclidean distance, d_{euc}, or the Angular distance, also know as Spectral Angle Mapper (SAM) distance in remote sensing applications, d_{sam}:

$$d_{euc}\left(\mathbf{e}_1, \mathbf{e}_2\right) = \sqrt{\sum_{k=1}^{q}\left(e_{1,k} - e_{2,k}\right)^2} \tag{2}$$

$$d_{sam}\left(\mathbf{e}_1, \mathbf{e}_2\right) = \cos^{-1}\left(\frac{\sum_{k=1}^{q}\left(e_{1,k} \cdot e_{2,k}\right)}{\sqrt{\sum_{k=1}^{q}\left(e_{1,k}\right)^2}\sqrt{\sum_{k=1}^{q}\left(e_{2,k}\right)^2}}\right) \tag{3}$$

Then, the dissimilarity between two hyperspectral images, H_α, H_β, is given as a function of the distance matrix (1) by the following equation:

$$s\left(H_\alpha, H_\beta\right) = \left(m_r + m_c\right)\left(\left|p_\alpha - p_\beta\right| + 1\right) \tag{4}$$

where m_r and m_c are the mean of the vectors of minimal values, $\min_{\mathbf{r}}$ and $\min_{\mathbf{c}}$, of the distance matrix, $D_{k,l}$, computed by rows and columns, respectively. The value of m_r represents the degree of inclusion of the set E_α in E_β, that is a meassure of how the materials presented in the image H_β are presented in the image H_β. In the same way, the value of m_c represents the degree of inclusion of the set E_β in E_α. The factor $\left(\left|p_\alpha - p_\beta\right| + 1\right)$ of 4 penalizes the difference on the number of materials found in each image p_α, p_β. Note that the endmember induction algorithm can give different number of endmembers for each image. The proposed dissimilarity function can cope with this asymmetry avoiding the combinatorial problem of trying to decide which endmembers can be matched and what to do in case that the number of endmembers is different from one image to the oher.

3 The DAMA strategy

DAMA is a data-driven thematic map quality assessment strategy suitable for comparative purposes when competing discrete mapping products are provided with little or no ground truth knowledge. It exploits a large number of implicit reference samples extracted from multiple reference cluster maps generated from unlabelled blocks of the input RS image, that are clustered separately to detect genuine, but small, image details at the cost of little human supervision. Thus, the output consists of unsupervised relative quantitative indexes (unsupervised *map quality* measures, in contrast to traditional supervised *map accuracy* measures) of labelling and segmentation consistency between every competing map and the set of multiple reference cluster maps.

The goal is to compute labelling and segmentation indexes of the consistency between a map x generated from a digital input image z, and multiple cluster maps generated from z without employing any prior knowledge. The procedure consists of three steps:

1. Locate across raw image z several blocks of unlabelled data, $\{s_{z_i} \subseteq z, i = 1, \ldots, Q\}$, using no prior knowledge and with a minimum of human intervention. These *unlabelled candidate representative raw areas*, s_{z_i}, have to satisfy some heuristic constraints: (a) be sufficiently small so that it is easy to analyze it by clustering algorithms, and (b) contain at least two of the cover types of interest according to photo-interpretation criteria. Each land cover type must appear in one or more blocks, and the set of blocks should be sufficiently large to provide a statistically valid dataset of independent samples and to be representative of all possible variations in each land cover.

2. Each block s_{z_i} is subject to clustering separately, generating Q independent so-called multiple reference cluster maps, $\{x_i^*, i = 1, \ldots, Q\}$.

3. Estimate the labelling (class) and segmentation (spatial) agreement between each reference cluster map x_i^* and the portion of the test map, x_i corresponding to the block.

4. Combine independently the spatial and agreement fidelity results collected by submaps according to empirical (subjective) image quality criteria.

4 RS-CBIR validation

4.1 CBIR quality assessment

Evaluation metrics from information retrieval field have been adopted to evaluate CBIR systems quality. The two most used evaluation measures are *precision* and *recall*. Precision, p, is the fraction of the retrieved 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 T the set of returned images and R the set of all the images relevant to the query, then

$$p = \frac{|T \cap R|}{|T|} \tag{5}$$

$$r = \frac{|T \cap R|}{|R|} \tag{6}$$

Precision and recall follow an inverse relationship as function of the number of retrieved images, named *scope*. Precision falls while recall increases as the scope increases. Results are usually summarized as precision-recall or precision-scope curves.

The main handicaps for the evaluation of RS-CBIR systems is the lack of ground truth knowledge (categories) and the users difficulties to evaluate the

retrieved images giving a positive/negative feedback. The former is due to the expensive, tedious and error prone groundtruth gathering process, and it is a well known problem in RS classification [2]. The later is an specific problem of CBIR systems in a Remote Sensing context. This kind of images are not easily interpreted by visual inspection, what implies that RS-CBIR feedback retrieval requires domain-specific skills and new interaction methodologies yet to be developed.

Our proposed RS-CBIR validation strategy inspired on DAMA overcomes these problems by giving a quantitative and qualitative measure of RS-CBIR performance using only the RS data inherent structures.

4.2 Proposed validation strategy

A CBIR system model is a tupla $\mathcal{M} = \langle D, \phi, d, \psi \rangle$, where $\mathcal{D}$ is a dataset with n images, $\mathcal{D} = \{x_i\}_{i=1}^{n}$, $\phi(x)$ is a *feature extraction* process which maps any image x onto a feature space Φ; a *disimilarity function*, $d : \Phi \times \Phi \rightarrow \mathbb{R}^{+}$, which is a distance function measuring the disimilarity between two images defined on their features; and, an optional *retrieval feedback* process, ψ, which allows the user to provide a feedback to the CBIR system to improve the data search process.

The input to a CBIR system is an user's *query*, q. This is usually done by providing one or more sample images. Then, the response of the CBIR system model $\mathcal{M}$ to the query q is a ranked list, $f_{\mathcal{M}}(q) = \{x_{i_1}, \ldots, x_{i_n}\}$, of the images in $\mathcal{D}$, where $I = \{i_1, \ldots, i_n\}$ is a permutation of the set of image indices, $i = 1, \ldots, n$, such that the returned images are ordered by increasing disimilarity relative to the query, $d(q, x_{i_1}) < d(q, x_{i_2}) < \cdots < d(q, x_{i_n})$. The number of images returned to the user is limited by the *scope s*, $0 < s \leq n$, of the query, so only the first s images, $\{x_{i_1}, \ldots, x_{i_s}\}$, on the ranked list $f_{\mathcal{M}}(q)$ are returned.

We propose to build the groundtruth of a potential query by a clustering process. Thus, the groundtruth modelled by a clustering process is a ranked list given by

$$f_{\mathcal{M}}(q_i) = \{x_{i_1}, \ldots, x_{i_k}, x_{i_{k+1}}, \ldots, x_{i_n}\} \tag{7}$$

where $q_i = \{x_i\}$ is the query, images $\{x_{i_1}, \ldots, x_{i_k}\}$ belong to the same cluster $\mathcal{C}_q$ as the query image q_i, and images $\{x_{i_{k+1}}, \ldots, x_{i_n}\}$ belong to any of the remaining clusters $\mathcal{C}_j$. The set of all the queries, $Q = \{q_i\}_{i=1}^{n}$, represents a simulated family of queries whose groundtruth is given by a clustering process.

We can take advantage of this to provide precision and recall measures in absence of ground truth data or a-priori categorization, which is the com-

Algorithm 1 Proposed RS-CBIR validation strategy

1. Perform a clustering over the dataset $\mathcal{D}$ of n images to obtain a clustering map $\mathbf{c} = \{c_1, \ldots, c_n\}$ where c_i indicates the cluster $C_j \in [1, p]$, patch i belongs to.
2. For each image, $x_i \in \mathcal{D}$, select the image as the query q, and calculate the ranked list $f_{\mathcal{M}}(q) = \{x_{i_1}, \ldots, x_{i_n}\}$ returned by the CBIR system.
3. Being s the query scope, the set of returned images T and the set of all the images relevant to the query R are given by:

$$T = \{x_{i_1}, \ldots, x_{i_s}\} \tag{8}$$

$$R = \left\{ \cup_{c_i = j} x_i; q \in C_j \right\} \tag{9}$$

4. Now the precision and recall can be calculated by substituting (8),(9) in equations (5),(6). The average of the precision and recall measures estimated by all the queries q_i, $i = 1, \ldots, n$, is a quality assesment of the RS-CBIR system $\mathcal{M}$ response respect to the family of queries $Q = \{q_i\}_{i=1}^{n}$ on $\mathcal{D}$.

mon scenario in RS-CBIR. Algorithm (1) describes the proposed validation strategy.

5 Experiment and results

Here we apply the proposed RS-CBIR validation strategy to assess the performance of the Hyperspectral CBIR system introduced in [10]. For doing that, we applied the Hyperspectral CBIR over a scene taken by the HyMap sensor, kindly provided by the German Aerospace Center (DLR). The scene is a big image of 2878×512 pixels and 125 spectral bands. Twelve bands corresponding to water absorption bands have been removed, remaining 113 bands. The image has been captured over the DLR facilities in Oberpfaffenhofen (Germany), and consist mainly of vegetation and fields, in addition to the DLR facilities and some small towns buildings.

We built six datasets by cutting the scene in patches of increasing sizes, from 8×8 pixels (23040 patches) to 256×256 pixels (24 patches). For each dataset we performed several clusterings on the average radiance of each patch sample by means of the ELGB clustering algorithm [12, 13], setting different values of the number of clusters, $k = 2, \ldots, 7$. The ELGB is an enhanced k-means clustering algorithm which has a strong robustness against initial condition variations. Such robustness is needed for the generalization of the validation results. For each dataset and cluster, the mean and standard deviation were calculated in order to purge those patches away from two times the standard deviation (eliminating ambiguous samples and making the dataset more simple). Complexity grows with the number of clusters k, as the number of image samples belonging to the relevant cluster decreases.

Each clustering is assumed to be the ground truth of the expected response to a simulated family of queries, against which the RS-CBIR must compete.

Figures 1-6 show the precision-recall curves estimated by our proposed validation strategy. Each figure corresponds to an experiment over each dataset, and shows a precision-recall curves for each of the different complexities, given by the number of clusters k. We can see a decrease on the performance of the Hyperspectral CBIR system as the complexity of the simulated queries, given by the number of clusters k, increases. It also can be observed that the precision-recall curves are similar for the different datasets, for which we can conclude that the size of the images does not affect significantly the performance of the Hyperspectral CBIR system.
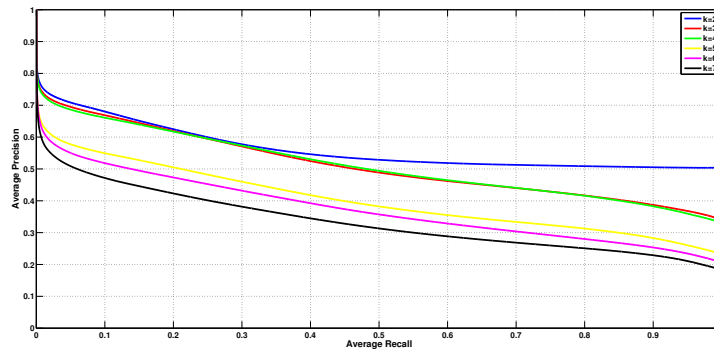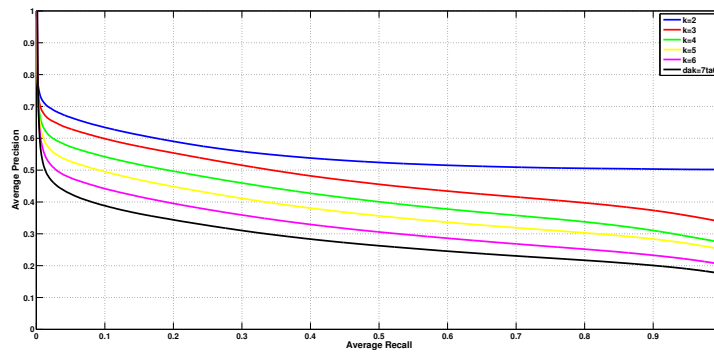

Fig. 1 Precision-recall results for 8×8 pixels dataset

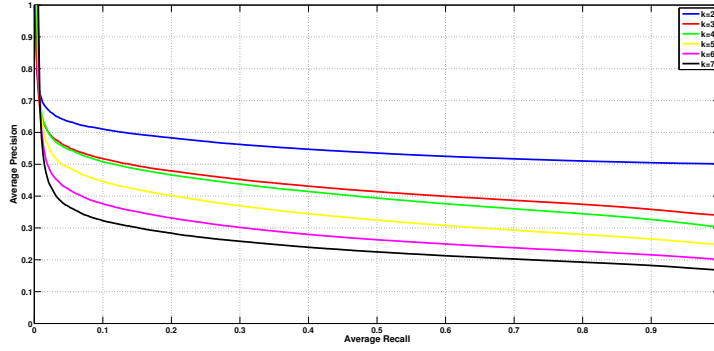**Fig. 2** Precision-recall results for 16×16 pixels dataset

Fig. 3 Precision-recall results for 32×32 pixels dataset
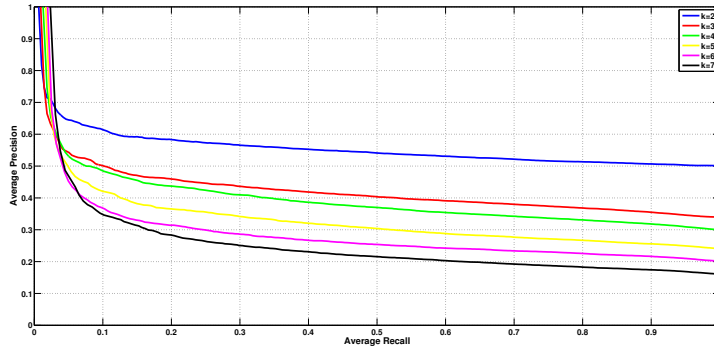
Fig. 4 Precision-recall results for 64×64 pixels dataset
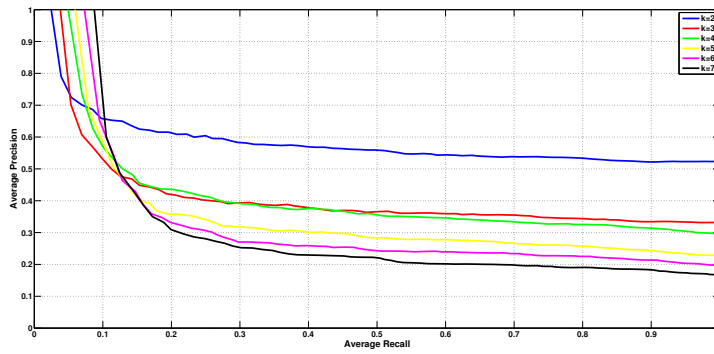
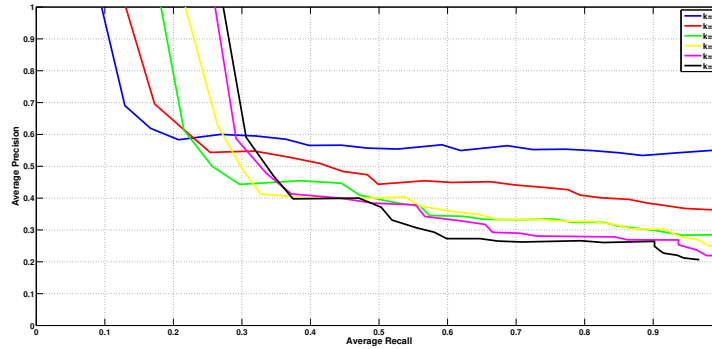**Fig. 5** Precision-recall results for 128×128 pixels dataset

Fig. 6 Precision-recall results for 256×256 pixels dataset

6 Conclusions

There is a big need of new strategies to validate RS-CBIR systems that could successfully overcome the lack of ground truth data. We have developed a methodology for RS-CBIR systems quality assessment, inspired in the DAMA strategy for unsupervised segmentation quality assessment in remote sensing images. Our methodology works when little or no ground truth data are available. We show an example of its applicability to test a Hyperspectral CBIR system.

Further work will address other soft-computing [4, 3] approaches to the problem.

Acknowledgements

References

1. A. Baraldi, L. Bruzzone, and P. Blonda. Quality assessment of classification and cluster maps without ground truth knowledge. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):857–873, 2005.

2. A. Baraldi, L. Bruzzone, P. Blonda, and L. Carlin. Badly posed classification of remotely sensed images-an experimental comparison of existing data labeling systems. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(1):214–235, 2006.

3. Emilio Corchado, Angel Arroyo, and Veronica Tricio. Soft computing models to identify typical meteorological days. *Logic Journal of IGPL*, 2010.

4. Emilio Corchado and Alvaro Herrero. Neural visualization of network traffic data for intrusion detection. *Applied Soft Computing*, In Press, Corrected Proof, 2010.

5. 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.

6. M. Datcu, H. Daschiel, A. Pelizzari, M. Quartulli, A. Galoppo, A. Colapicchioni, M. Pastori, K. Seidel, P.G. Marchetti, and S. D'Elia. Information mining in remote sensing image archives: system concepts. *Geoscience and Remote Sensing, IEEE Transactions on*, 41(12):2923–2936, 2003.

7. M. Datcu and K. Seidel. Human-centered concepts for exploration and understanding of earth observation images. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(3):601–609, 2005.

8. M. Datcu, K. Seidel, and M. Walessa. Spatial information retrieval from remote-sensing images. i. information theoretical perspective. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(5):1431–1445, 1998.

9. S. D'Elia. Image information mining for earth observation at ESA. volume 1, page 171, 2004.

10. J.O. Maldonado, D. Vicente, M.A. Veganzones, and M. Grana. Spectral indexing for hyperspectral image CBIR. Torrejon air base, Madrid (Spain), 2006.

11. R. Muller, M. Bachmann, C. Makasy, A. de Miguel, A. MÃ$\frac{1}{4}$ller, G. Palubinskas, R. Richter, M. Schneider, T. Storch, A. Neumann, T. Waizel, H. Kaufmann, and K. Segl. EnMAP - the future hyperspectral satellite mission: product generation. In *Proceedings of the ISPRS Hannover Workshop 2009*, Hannover, 2009.

12. Giuseppe Patane and Marco Russo. ELBG implementation. *International Journal of Knowledge based Intelligent Engineering Systems*, 2:2—4, 2000.

13. Giuseppe Patane and Marco Russo. The enhanced LBG algorithm. 2001.

14. A. Plaza, J. Plaza, A. Paz, and S. Blazquez. Parallel CBIR system for efficient hyperspectral image retrieval from heterogeneous networks of workstations. pages 285–291, 2007.

15. M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu. Spatial information retrieval from remote-sensing images. II. Gibbs-Markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(5):1446–1455, 1998.

16. M. Schroder, H. Rehrauer, K. Seidel, and M. Datcu. Interactive learning and probabilistic retrieval in remote sensing image archives. *Geoscience and Remote Sensing, IEEE Transactions on*, 38(5):2288–2298, 2000.

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

18. Miguel A. Veganzones and Carmen Hernandez. On the use of a hybrid approach to contrast endmember induction algorithms. In *Hybrid Artificial Intelligence Systems*, volume 6077 of *Lecture Notes in Computer Science*, pages 69–76. Springer Berlin / Heidelberg, 2010.

19. Miguel A. Veganzones, Jose O. Maldonado, and Manuel Grana. On Content-Based image retrieval systems for hyperspectral remote sensing images. In *Computational Intelligence for Remote Sensing*, volume 133 of *Studies in Computational Intelligence*, pages 125–144. Springer Berlin / Heidelberg, 2008.