VALIDATION OF REMOTE SENSING CONTENT-BASED INFORMATION RETRIEVAL (RS-CBIR) SYSTEMS UPON SCARCE DATA

Miguel A. Veganzones, Manuel Graña

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

ABSTRACT

Validation of Remote Sensing Content-Based Information Retrieval (RS-CBIR) systems requires innovative strategies to overcome the scarcity of labeled data. CBIR systems validation by means of precision/recall measures based on either, user's 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 *a-priori* knowledge.

Index Terms— CBIR, remote sensing, validation, DAMA strategy.

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. Content Base Image Retrieval (CBIR) systems are relevant to the geosciences because they provide automated tools to explore the contents of large and highly complex image databases [1, 2, 3]. The main efforts to develop CBIR tools for remote sensing images have been focused on multispectral and synthetic aperture radar (SAR) images [4, 5, 6, 7, 8]. 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 [9, 10, 11].

Validation is the process of assessing the performance of a system against the ground truth or some equivalent information. In previous works [12] we dealt with the validation of hyperspectral CBIR systems using synthetic hyperspectral images, where all the ground truth of the images is known. In this paper we consider the case of possessing scarce ground truth knowledge about the data. We propose a methodology similar to [13] to we asses the problem of CBIR systems validation in a Remote Sensing (RS) context. The work in [13] 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. The data-driven quality map assessment (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. In this paper we set the stage by giving a formalization of the data-driven quality CBIR assessment. We test the approach on a large hyperspectral image, working on image blocks or patches.

Section 2 gives a brief overview of CBIR systems and their quality assessment. In section 3 we introduce our DAMA-like strategy for RS-CBIR validation. Finally we provide some conclusions in section 4.

miguelangel.veganzones@ehu.es
manuel.grana@ehu.es

2. CBIR SYSTEMS

In this section we provide a formalization of CBIR systems and, then we introduce common measures for the quality assessment of such systems.

2.1. CBIR systems formulation

A CBIR system model is a tuple $\mathcal{M} = \langle D, \phi, d, \psi \rangle$, 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; d is a *dissimilarity function*, $d : \Phi \times \Phi \rightarrow \mathbb{R}^{+}$, which is a distance function measuring the dissimilarity between two images defined on their features; and, ψ is an optional *retrieval feedback* process, 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. 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 dissimilarity 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.

CBIR systems can be modeled as a two-class problem where images are classified as belonging to the relevant class, $\mathcal{C}_R(q)$, or to the irrelevant class, $\mathcal{C}_I(q)$, relative to a query q. If available, the retrieval feedback allows the user to provide a relevant/irrelevant labeling over the system's response images, so the system can use this new information to improve the classifier results. The ideal response $f_{\mathcal{M}}(q)$ of a CBIR system $\mathcal{M}$ to a query q is given by

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

where images $\{x_{i_1}, \ldots, x_{i_k}\}$ belong to the relevant class $\mathcal{C}_R(q)$, and images $\{x_{i_{k+1}}, \ldots, x_{i_n}\}$ belong to the irrelevant class $\mathcal{C}_I(q)$.

It is important to note that performance of a CBIR system is query-dependent, that is, its performance depends not only on the dataset $\mathcal{D}$, but on its ability to suit to the user goals, which are specified by the query q. Thus, the expected response of an ideal CBIR system, $f_{\mathcal{M}}(q)$, to a potential user's query, q, can be seen as a relevant/irrelevant mapping of the dataset $\mathcal{D}$ in an unknown feature space, Ω. We name this ideal mapping the *potential search map*. The performance of a CBIR system is related to its capacity to elaborate a mapping equivalent to the potential search map of any given query q. To do this, the CBIR system shall use the user's feedback to iteratively transform the original feature space Φ and/or modify the dissimilarity function d, so the response of the CBIR system to a query q will converge on a map equivalent to the potential search map corresponding to the query q.

2.2. CBIR quality assessment

Evaluation metrics from the information retrieval field have been adopted to evaluate CBIR systems quality. The two most frequently used evaluation measures are *precision* and *recall*. Precision, p, is the fraction of the returned images that are relevant to the query. Recall, q, is the fraction of returned 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

$$p = \frac{|T \cap R|}{|T|} \tag{2}$$

$$r = \frac{|T \cap R|}{|R|} \tag{3}$$

Precision and recall have opposite trends as functions of the scope of the query. Precision falls while recall increases as the scope increases. Results are usually summarized as precision-recall or precision-scope curves. Using the concepts defined above, for a given query q, the set R of images relevant to the query identifies with the relevant class $\mathcal{C}_R(q)$, specified by the potential search map of the query.

3. PROPOSED RS-CBIR VALIDATION STRATEGY

The main handicaps for the evaluation of RS-CBIR systems are the lack of ground truth knowledge (categories) and the user difficulties to evaluate the system's response images to provide 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 [14]. 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 strategy overcomes these problems by giving a quantitative and qualitative measure of RS-CBIR performance using only the RS data inherent structures. In the following, we briefly introduce the DAMA strategy before exposing our proposed strategy for RS-CBIR validation.

3.1. 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 unlabeled 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 labeling and segmentation consistency between every competing map and the set of multiple reference cluster maps.

The goal is to compute labeling 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 four steps:

1. Locate across raw image z several blocks of unlabeled data, $\{s_{z_i} \subseteq z, i = 1, \ldots, Q\}$, using no prior knowledge and with a minimum of human intervention. These *unlabeled 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 labeling (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.

3.2. Proposed strategy

We propose to perform clustering processes over a dataset $\mathcal{D}$ to discover data inherent equivalence relation structures, and to use this inherent structures to simulate potential user queries. Each clustering can be used to model potential search maps of a family of queries Q. Then, the simulated potential search maps are used to provide precision and recall measures, that show the RS-CBIR system capacity to solve the family of queries Q on a dataset $\mathcal{D}$.

A clustering models a family of queries Q by defining a set of potential search maps. Different clusterings can be obtained to represent different families of queries. It is supposed that potential users queries have some coherence, so the relevant classes are represented by some clustering mapping in an unknown feature space Ω. We do the inverse process, expecting a given clustering of the dataset $\mathcal{D}$ in some feature space, that could be different from the one used on the RS-CBIR system $\mathcal{M}$, represents a potential user's kind of queries, a query family.

The procedure is as following:

1. Perform a clustering over a dataset $\mathcal{D}$. The clustering process defines a mapping $X^* = \{x_1^*, \ldots, x_n^*\}$, where x_i^* indicates the identity of cluster c_k, $k = 1, \ldots, C$, which image i belongs to, where C is the number of clusters found in the clustering process. This clustering represents a family of queries $Q = \{q_1, \ldots, q_n\}$ where each query q_j, $j = 1, \ldots, n$, is given by the sample image x_j.

2. Given a RS-CBIR system $\mathcal{M}$, calculate its response to each of the queries $q_j \in Q$, $f_{\mathcal{M}}(q_j) = \{x_{i_1}, \ldots, x_{i_n}\}_{j=1}^n$, which can be represented as a matrix $M = \{m_{ij}\}$, $i, j = 1, \ldots, n$, so m_{ij} indicates the i-th most similar image to the query q_j.

3. Being s the query scope, the set of returned images T_j and the set of all the relevant images R_j to a query q_j are given by:

$$T_j = \{\cup_{i=1}^s m_{ij}\} \tag{4}$$

$$R_j = \left\{\cup_{x_i^*=k} x_i; q_j \in c_k\right\} \tag{5}$$

4. Now the precision and recall measures for the query q_j can be calculated by substituting (4),(5) in equations (2),(3). The average of the precision and recall measures estimated by all the queries $q_j \in Q$ is a quality assessment of the RS-CBIR system $\mathcal{M}$ response respect to the family of queries Q on $\mathcal{D}$.

4. CONCLUSIONS

There is a strong need of innovative strategies to validate RS-CBIR systems that could successfully overcome the lack of ground truth data. We were inspired by the DAMA strategy to work in quality assessment techniques for RS-CBIR systems validation. We have formalized an strategy to assess the quality of RS-CBIR systems when little or no ground truth data is available, modeling potential users queries by clustering processes, and providing a mechanism to define the set of relevant images for a given query, in order to provide precision and recall quality measures. Lack of space prevents us to present experimental results, which nevertheless can be found in the research report available through the Computational Intelligence group web site[1].

5. REFERENCES

[1] 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*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] S. D'Elia, "Image information mining for earth observation at ESA," 2004, vol. 1, p. 171.

[3] M. Datcu and K. Seidel, "Human-centered concepts for exploration and understanding of earth observation images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 601–609, 2005.

[4] 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*, vol. 36, no. 5, pp. 1431–1445, 1998.

[1] http://www.ehu.es/computationalintelligence

[5] 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*, vol. 36, no. 5, pp. 1446–1455, 1998.

[6] 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*, vol. 38, no. 5, pp. 2288–2298, 2000.

[7] 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*, vol. 41, no. 12, pp. 2923–2936, 2003.

[8] H. Daschiel and M. Datcu, "Information mining in remote sensing image archives: system evaluation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 1, pp. 188–199, 2005.

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

[10] A. Plaza, J. Plaza, A. Paz, and S. Blazquez, "Parallel CBIR system for efficient hyperspectral image retrieval from heterogeneous networks of workstations," 2007, pp. 285–291.

[11] 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*, vol. 133 of *Studies in Computational Intelligence*, pp. 125–144. Springer Berlin / Heidelberg, 2008.

[12] Miguel A. Veganzones and Carmen Hernandez, "On the use of a hybrid approach to contrast endmember induction algorithms," in *Hybrid Artificial Intelligence Systems*, vol. 6077 of *Lecture Notes in Computer Science*, pp. 69–76. Springer Berlin / Heidelberg, 2010.

[13] 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*, vol. 43, no. 4, pp. 857–873, 2005.

[14] 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*, vol. 44, no. 1, pp. 214–235, 2006.