Relevance Feedback for Content-Based Visual Information Retrieval using Active Learning Method

Vikrant Gunjal *Institute of Management & Computer Studies, University of Mumbai , India*

and

Virendra Baranwal *Institute of Management & Computer Studies, University of Mumbai , India*

Abstract-- In content-based Visual Information retrieval, relevance feedback has been introduced to reduce the gap between low-level image properties and high-level properties. Furthermore, to speed up the converging to the query, several active learning methods have been proposed including random filtering to select images for labeling by the user. relevance feedback (RF) schemes improves the performance of content-based Visual Information retrieval (CBVIR) requiring the user to describe a large number of images. To reduce the labeling effort of the user, this paper presents a novel active learning (AL) method to drive RF for retrieving remote sensing images or sattelite images from large archives in the framework of the support vector machine classifier. The proposed AL method is specifically designed for CBVIR and defines an effective and as small as possible set of related and irrelated images with regard to a general query image by jointly evaluating three criteria: unreliabillity ; diversity; and density of images in the archive. The unreliabillity and diversity mainly aims at selecting the most informative images in the collection of images, whereas the density criterion goal is to choose the images that are representative of the underlying distribution of data in the collection. In first Step the most uncertain images are selected from the collection on the basis of margin sampling strategy. In second Step the images that are both diverse to each other and associated to high-density regions of the image feature space in the archive are chosen from most uncertain images. This step is achieved by a novel clustering-based strategy which contributes to solve the problems of unbalanced and biased set of related and irrelated images.

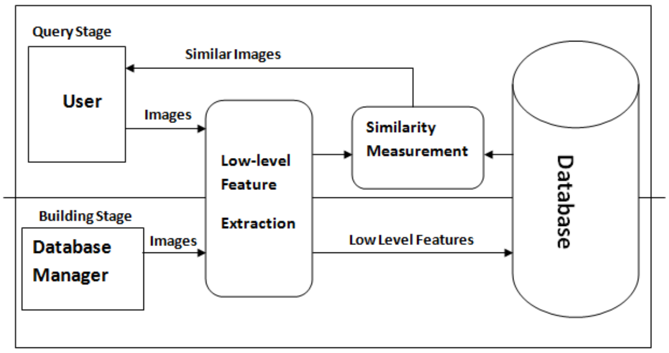
Keywords-- Active Learning (AL), Content–Based Visual Information Retrieval (CBVIR), Relevance feedback (RF).

I.INTRODUCTION

Visual information retrieval is the field of study concerned with searching and exploring digital images from database collection. This area of research is very active research since the 1970s. Due to more and more images have been generated in digital form around the world, image retrieval attracts interest among researchers in the fields of image processing, multimedia, digital libraries, remote sensing,sattelite imagery, astronomy, database applications and other related area.

In 1980s, Content-based visual information retrieval (CBVIR) then has been used as an alternative to text based image retrieval. CBVIR can be categorized based on the type of features used for retrieval which could be either low level or high level properties.

Low level properties include color (distribution of color intensity across image), texture (Homogeneity of visual patterns), shape (boundaries, or the interiors of objects depicted in the image), spatial relations (the relationship or arrangement of low level features in space) or combination of above features were used. General Framework of Content based Image Retrieval is shown in Fig.2. All images will undergo the low level properties extraction process before being added to the images database colletction. In propeties fetching stage, properties such as color, shape or texture are fetched from the image. [1]



The last few decades witnessed an explosion in the large amount of digital images, which necessitates an efficient scheme for exploring and indexing large image databases collection. To find this issue, people have proposed an integrated framework named content-based visual information retrieval (CBVIR).

The learning method in relevance feedback has been extensively studied. Traditional learning methods can be categorized into three major groups [9]: query reviewing , query point movement, and query rewriting. However, because these methods do not fully utilize the information embedded in feedback images, their performance is far from satisfactory. More recently, **statistical learning** methods have been applied to relevance feedback.

Among others, some researchers apply inductive methods to the learning task, aiming to create a selector that generalizes well on unseen examples. For example, the authors of [15] first calculate a large number of highly selective properties, and then use pump off to learn a classification function in this property space; similarly, the learning method proposed in trains a support vector machine (SVM) from labeled examples, hoping to obtain a small generalization error by maximizing the margin between the two types of images. On the other hand, some researchers consider visual information retrieval as a transductive learning problem, aiming to accurately identify the relevance of unlabeled images attainable during the training stage. For example, the authors of propose a discriminant-EM algorithm. It makes use of unlabeled data to construct a generative model, which will be used to measure diffrence between the query and database images.

Recent studies have shown that the content of the RS data is more relevant than manual tags. Accordingly, content-based visual information retrieval (CBVIR) has attracted increasing attentions in the RS community particularly for its potential practical applications to RS image management. This will become particularly important in the next years when the number of acquired images will dramatically increase. Any CBVIR system essentially consists of (at least) two modules [1], [2]: 1) a property extraction module that derives a set of properties for diffrencing and describing images and 2) a retrieval module that finds and retrieves images similar to the query image. Querying image contents from large RS data collection depends on the capability and effectiveness of the property fetching techniques in describing and representing the images. In the RS literature, several primitive (i.e., low level) properties have been presented for retrieval purposes, such as the following: intensity property [5], color property [6], [7], shape property [8]–[10], property features [10]–[16], and local invariant property [17]. However, the low level property from an image have a very limited capability in representing and analyzing the high-level concept conveyed by RS images This issue is known as the semantic gap that occurred between the low level property and the high-level semantic content and leads to poor CBVIR performance. Consequently, the semantic gap is the crucial challenge in CBVIR applications.

An effective approach to reduce the annotation effort in RF is active learning (AL) that aims at finding the most explainatory images in the collection that, when annotated and included in the set of related and irrelated images (i.e., the training set), can significantly improve the retrieval performance [10]. Moreover, selecting the most explainatory images results in the following: 1) a smaller number of RF iterations to optimize the CBVIR and 2) a reduced annotation time due to the optimization of the training set with a minimum number of highly explainatory images. In the RS community, most of the previous studies in AL have been developed in the context of classification problems. In particular, the unlabeled samples that are highly unpredictable and diverse to each other are usually selected as explainatory samples to be labeled and included in the training set for the classification of RS images [18]. The uncertainty of a sample is related to the confidence of the supervised algorithm in correctly classifying it, whereas the diversity among samples is associated to their correlation in the property space.

From the AL perspective, the CBVIR problem is more complex than the standard classification problem due to the following facts: 1) In general, the class of irrelated images (which is dynamically driven on the basis of the specific query image given as the input to the classifier) is much larger than the class of relevant images because the irrelated class consists of the huge number of images that, in a real archive, are irrelevant to the query image; 2) the classifier is trained with a largely incomplete number of annotated images (training set) due to the absence of many irrelevant image categories (those that exist in the archive) within the training set; and 3) in real large-scale RS archives, the total number of images is usually very large. All of the aforementioned reasons result in strongly imbalanced and biased training sets. As a result, the boundary between two classes is initially unstable and inaccurate, and thus, it does not allow a reliable modeling of the problem. Accordingly, AL methods defined for classification problems that only assess uncertainty and diversity of samples are not efficient for CBVIR problems.

To overcome the aforementioned critical issues, in this paper, we propose a CBVIR approach that includes a novel triple criteria AL (TC) method to drive RF in CBVIR. For the selection of the most informative as well as representative unlabeled images of images to be annotated, the proposed TC method jointly evaluates three criteria: 1) uncertainty; 2) diversity; and 3) density of images in the archive. In order to assess the aforementioned three criteria, the proposed TC method exploits a two-step procedure defined in the framework of the SVM classifier. In the first step, the most uncertain (i.e., ambiguous) images are selected by the well-known MS strategy [10], whereas in the second step, the diverse images among the most uncertain ones are selected from the highest density regions of the image feature space. The latter step is achieved by a novel clustering-based strategy that evaluates the density and diversity of unlabeled images in the image property space to drive the selection of images to be annotated.

In order to restrict the semantic gap, relevance feedback (RF) schemes have been designed to iteratively improve the performance of CBVIR by taking user's (i.e., an oracle who knows the correct labeling of all images) feedback into account [3], [4]. At each iteration, the users feedback is used to provide related and irrelated images to the query image that are positive and negative feedback samples, respectively. RF can be considered as a binary-classification problem: One class includes related images, and the other one consists of the irrelated ones. Then, any supervised classification method can be used in the context of CBVIR by training the classifier with the already annotated images of two classes [3], [4]. Accordingly, during RF, the search strategy is refined iteration by iteration by improving the classification model with the recently annotated images. As mentioned previously, user involvement is required at each RF iteration for annotating images. However, labeling images as relevant or irrelevant is time-consuming and thus costly. Accordingly, despite the retrieval success of RF, the conventional RF schemes are not practical and efficient in real applications, especially when huge archives of RS images are considered.

From the AL perspective, the CBVIR problem is more complex than the standard classification problem due to the following facts: 1) In general, the class of irrelevant images (which is dynamically driven on the basis of the specific query image given as the input to the classifier) is much larger than the class of relevant images because the irrelevant class consists of the huge number of images that, in a real archive, are irrelevant to the query image; 2) the classifier is trained with a largely incomplete number of annotated images (training set) due to the absence of many irrelevant image categories (those that exist in the archive) within the training set; and 3) in real large-scale RS archives, the total number of images is usually very large. All of the aforementioned reasons result in strongly imbalanced and biased training sets. As a result, the boundary between two classes is initially unstable and inaccurate, and thus, it does not allow a reliable modeling of the problem. Accordingly, AL methods defined for classification problems that only assess uncertainty and diversity of samples are not efficient for CBVIR problems.

AL has been marginally considered in the framework of CBVIR problems in the RS community. To the best of our knowledge, only one AL method is presented [10], which is developed in the context of the support vector machine (SVM) classifier and inspired from AL methods used for classification problems. In this method, the uncertainty and diversity criteria have been applied in two consecutive steps. In the first step, the most uncertain images are selected from the archive. To this end, the unlabeled images closest to the current separating hyper-plane (those that are the most uncertain) are initially selected by margin sampling (MS). In the second step, the images that are diverse to each other among the uncertain ones are chosen on the basis of the distances calculated between them. An important shortcoming of the method presented in [10] is that it does not evaluate the representativeness of images in terms of their density in the archive. However, images that fall into the high-density regions of the image property (descriptor) space are crucial for CBVIR problems particularly when a small number of initially annotated images are available. This is due to the fact that they are visually very representative of the underlying image distribution in the collection. Therefore, the retrieval results on them affect much more the overall retrieval accuracy than the results obtained on images within the low-density regions. To overcome the critical issues, in this paper, we propose a CBVIR approach that includes a novel triple criteria AL (TC) method to drive RF in CBVIR. For the selection of the most informative as well as representative unlabeled images of images to be annotated, the proposed TC method jointly evaluates three criteria: 1) uncertainty; 2) diversity; and 3) density of images in the archive. In order to assess the aforementioned three criteria, the proposed TC method exploits a two-step procedure defined in the framework of the SVM classifier. In the first step, the most uncertain (i.e., ambiguous) images are selected by the well-known MS strategy [10], whereas in the second step, the diverse images among the most uncertain ones are selected from the highest density regions of the image property space. The next step is achieved by a novel clustering-based strategy that evaluates the density and diversity of unlabeled images in the image property space to drive the selection of images to be annotated. The novelties of the proposed AL method for RF in CBVIR consist in the following: 1) the design and development of a strategy to jointly evaluate the three criteria (i.e., uncertainty, diversity, and density) for the selection of the most explainatory and representative images in the context of CBVIR problems and 2) the use of the prior term of the distributions based on the density of unlabeled images in the image feature space to assess the representativeness of images and thus to identify the images to annotate. Owing to the joint use of the three criteria and to the use of the density of images in the image property space, the proposed TC method can effectively avoid various problems caused by the insufficient number of annotated samples in RF, and thus, it is appropriate and effective for RS image retrieval. Moreover, we introduce the use of the histogram intersection (HI) kernel in the RS community in the framework of the pro-posed CBVIR approach as a similarity measure of image features in the kernel space. Note that, in recent years, the HI kernel has gained an increasing interest for image retrieval problems in the computer-vision communities, whereas its use in RS has not been explored yet. Experiments carried out on an archive of aerial images demonstrate the effectiveness of the proposed method.

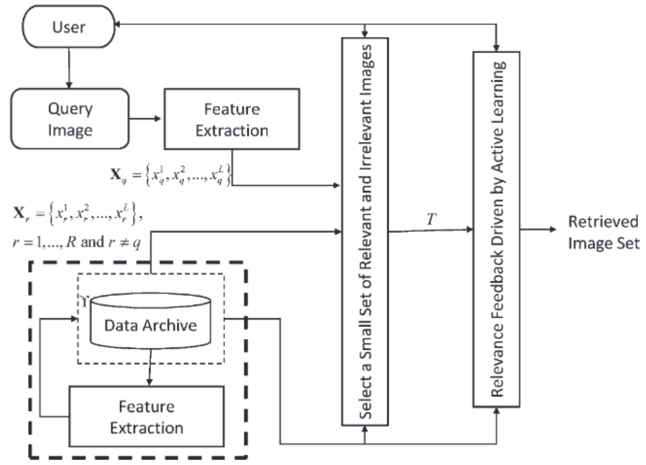

User
Query Image
Feature Extraction
$\mathbf{X}_q = \{x_q^1, x_q^2, \ldots, x_q^L\}$
$\mathbf{X}_r = \{x_r^1, x_r^2, \ldots, x_r^L\}$,
$r = 1, \ldots, R$ and $r \neq q$
Υ
Data Archive
Feature Extraction
Select a Small Set of Relevant and Irrelevant Images
T
Relevance Feedback Driven by Active Learning
Retrieved Image Set

Fig. 1. General architecture of a CBIR system with RF driven by AL.

A. Problem Formulation

Let us consider an archive Y made up a very large number of R RS images{X1,X2,...,XR}, where Xi is the Ith image defined as {x1i,x2i,...,x$_{Li}$}, i=1,...,R.x$_{li}$, l=1,...,L is the lth feature characterizing the content of the ith image in Y and L is the total number of features. Let X q = { x 1 q ,x 2 q,...,x$_{Lq}$} be a query image that can be selected by the user from the archive Y (i.e. X q ∈ Y) or outside the archive Y (i.e., X q / ∈ Y). A general CBVIR system with RF driven by AL consists of three modules: 1) the primitive (low level) feature extraction module that is applied to both query image and all images in the archive; 2) the initial training set definition module that builds an initial training set T with a small number of relevant and irrelevant images with respect to query; and 3) the RF driven by an AL module that enriches the training set T defined by the previous module and returns the set δ of images from the archive Y.

Fig. 1 shows the general block scheme of the CBVIR with RF driven by AL. In this paper, we mainly focus on the RF driven by the AL module (see Fig. 2) which is a crucial part for the success of the CBVIR system. Then, we briefly present the feature extraction module and the important choices adopted for assessing the similarities of image features in the proposed system. AL iteratively expands the size of an initial labeled training set T, selecting the most informative images from the archive Y for their annotation. At each RF iteration, the most informative unlabeled images for a given classifier are:

1) Selected based on an AL function;
2) Annotated by a supervisor (i.e., an oracle);
3) Added to the current training set T.

Finally, the supervised classifier is retrained with the images moved from Y to T. It is worth noting that the initial training set T requires few annotated images for the first training of the classifier and then is enriched iteratively by including the most informative images selected from Y. At each iteration, after the classifier is trained, the retrieval of the images under investigation is carried out. These processes are repeated until the user is satisfied

with retrieval results. The general flowchart of the AL-based RF approach is given in Fig. 2. The selection of the most informative samples from Y to be included in the training set T

on the basis of AL offers two main advantages:

1) The annotation cost is reduced due to the avoidance of redundant Images
2) an accurate retrieval accuracy can be obtained due to the improved class models estimated on a high-quality training set on the basis of the classification rule used from the considered classifier (the images to be annotated are selected from the classifier as the most informative for its classification rule). Of course, the success of the RF strongly depends on the capability of the specific AL method considered to select the most informative and representative images to be annotated in order to limit as much as possible the effort of the user for reaching the final relevant result.

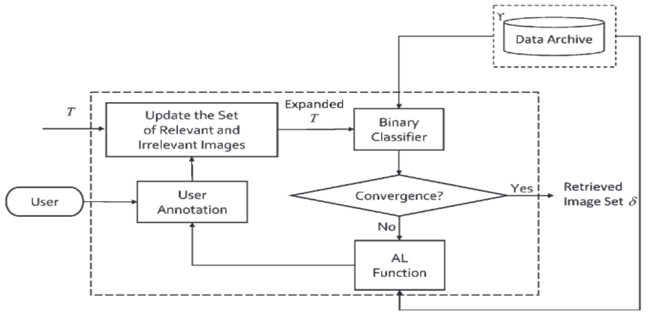
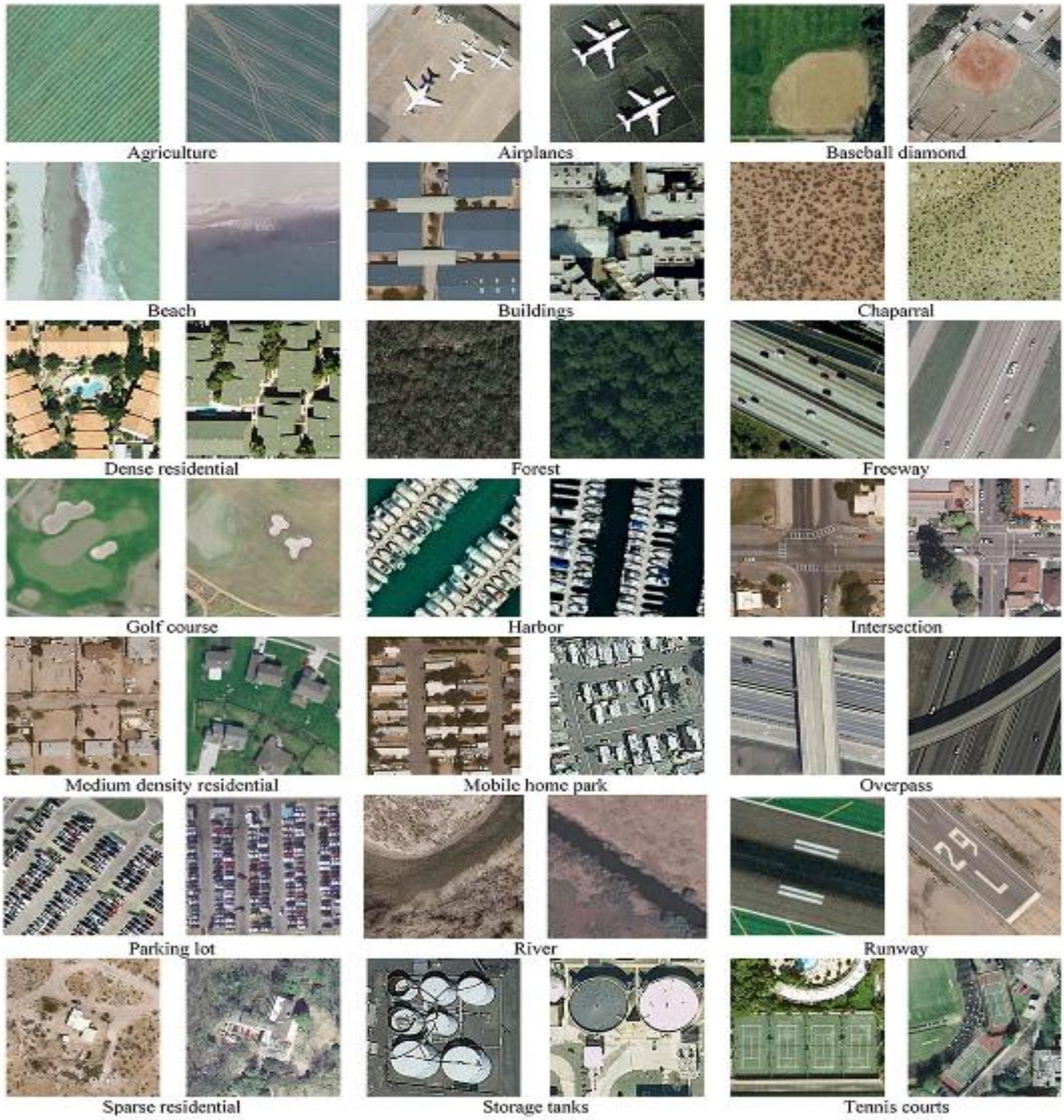


Fig. 2. General flowchart of RF driven by AL.

B. RS Image property Extraction and Classification in the Context of CBVIR

We model RS images by exploiting a bag-of-visual-words (BOVW) representation of the local invariant features extracted by the scale invariant feature transform (SIFT). The SIFT is a translation, rotation, and scale invariant image feature ex- traction technique and has recently been found very effective and robust in the context of RS image retrieval [17]. The SIFT results in various local interest points within an image and their descriptors (i.e., SIFT descriptors) that characterize portions of images around the interest points. In order to summarize the SIFT descriptors by the BOVW representation (that is generally considered for the local image descriptors), we apply kernel k -means clustering to a subset of randomly selected SIFT descriptors. This process results in a codebook. Then, the descriptors extracted from each image are quantized by assigning the label of the closest cluster [17]. Accordingly, the final representation of an image is the histogram (i.e., frequency) of the codebook entries (known as code-words) in the image [17]. Note that the histogram-based image representation is very popular for the BOVW approaches that result to be the state-of-the-art in many image retrieval problems outside RS. In order to assess the similarities of the BOVW representations (histogram-based features) of the images in the kernel space, we introduce in RS the use of HI kernel. Note that the similarity is used in both the SVM classification and the pro-posed AL method. To measure the similarities between the images Xi = {x1i,x2i,...,x_{Li}} and Xj={x1j,x2j,...,x_{Lj}}, the HI kernel is defined as K(X_i, X_j) = Ll =1 min(x_{li}, x_{lj}) where x_{li}∈X_i, l=1,2 ,...,L and xlj ∈ Xj, l=1,2,...,L denote histogram features. Note that the HI kernel is a positive definite parameter-free kernel for nonnegative features, and it has been recently found very effective in various computer-vision tasks (where histograms are popular representations of images) such as CBVIR.



II.CONCLUSION

The unpredictable and diversity criteria aim to select the most informative images, whereas the density criterion aims to select the most representative images in terms of prior distribution. In the proposed AL method, the joint assessment of the three criteria is accomplished based on a two-step technique. In this paper, we have introduced a novel AL method to drive RF in CBVIR for the identification of effective images to annotate and to include in the training set. The proposed AL method selects both informative and representative explaintory images to be included in the training set at each RF round by the joint evaluation of the unpredictabillity, diversity, and density criteria. In the first step, the most uncertain (i.e., informative/ambiguous) images are selected by using the well-known MS approach. In the second step, the most diverse images among the uncertain ones are selected from the high-density regions in the image property space. In order to identify the highest density regions in the image property space, a novel clustering-based strategy has been introduced. The proposed AL method overcomes the limitations of previously presented AL methods in CBVIR problems, which are due to the following: 1) unbalanced training sets and 2) biased initial training sets. Note that the unlabeled images located in the high-density regions of the image property space are highly important for CBVIR problems particularly when an unbalanced and biased training set is available. This is due to the fact that they are statistically very representative of the underlying image distribution, and thus, the retrieval results on them affect much more the overall accuracy of the CBVIR than those obtained on images within the low-density regions.

The experimental performances of the proposed system were evaluated on an archive of 2100 images describing 21 different categories. The results show that the proposed AL method provides efficient image retrieval performance requiring less RF iterations and thus with less annotation effort compared to previously presented AL methods based on CBVIR. We have emphasized that these are very important advantages because the main objective of AL in CBVIR is to optimize the search with a minimum number of annotated images and thus with a minimum cost in annotating images. It is worth emphasizing that, given the growing amount of RS image archives, CBVIR is becoming more and more important. One of the major challenges in CBVIR is the semantic gap which can be reduced by RF driven by AL. Accordingly the proposed method is very promising as it provides high retrieval accuracy with a small number of RF rounds. It is also worth noting that the proposed AL method is independent from the considered feature extraction method, and therefore, it can be used with any feature extraction technique presented in the literature. As a future development of this work, we plan to extend the validation of the proposed AL method to larger data sets and to use the proposed AL technique to drive the RF in image time series retrieval problem.

REFERENCES

[1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,”IEEE Trans. PatternAnal. Mach. Intell., vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[2] R. Datta, D. Joshi, J. Li, and J.-Z. Wang, “Image retrieval: Ideas, influ-ences, trends of the new age,”ACM Comput. Surveys, vol. 40, no. 2,pp. 1–60, Apr. 2008.

[3] P. Hong, Q. Tian, and T. S. Huang, “Incorporate support vector machines to content-based image retrieval with relevant feedback,” in Proc. IEEE
Int. Conf. Image Process., Vancouver, BC, Canada, 2000, pp. 750–753.

[4] X. S. Zhou and T. S. Huan, “Relevance feedback in image retrieval: A com-prehensive review,”MultimediaSyst., vol. 8, no. 6, pp. 536–544, Apr. 2003.

[5] Q. Bao and P. Guo, “Comparative studies on similarity measures for remote sensing image retrieval,” in Proc. IEEE Int. Conf. Syst., Man Cybern.
, Hague, The Netherlands, 2004, pp. 1112–1116.

[6] T. Bretschneider, R. Cavet, and O. Kao, “Retrieval of remotely sensed imagery using spectral information content,” inProc. IEEE Int. Geosci.Remote Sens. Symp., Toronto, ON, Canada, 2002, pp. 2253–2255.

[7] T. Bretschneider and O. Kao, “A retrieval system for remotely sensed imagery,” in Proc. Int. Conf. Imag. Sci., Syst., Technol. , Las Vegas, NV, USA, 2002, pp. 439–445.

[8] G. Scott, M. Klaric, C. Davis, and C.-R. Shyu, “Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases,” IEEE Trans. Geosci. Remote Sens. , vol. 49, no. 5, pp. 1603– 1616, May 2011

[9] Porkaew, K., and Chakrabarti, K. Query refinement for multimedia similarity retrieval in MARS.

[10] M. Ferecatu and N. Boujemaa, “Interactive remote sensing image re- trieval using active relevance feedback

[11] Y. Li and T. Bretschneider, “Semantics-based satellite image retrieval using low-level features

[12] Y. Hongyu, L. Bicheng, and C. Wen, “Remote sensing imagery retrieval based-on Gabor texture feature classification,” in Proc. Int. Conf. Signal Process., Troia, Turkey, 2004, pp. 733–736.

[13] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and re-trieval of image data,”IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 8, pp. 837–842, Aug. 1996.

[14] S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath, “Using texture to analyze and manage large collections of remote sensed image and video data,”
J. Appl. Opt., vol. 43, no. 2, pp. 210–217, Jan. 2004.

[15] A. Samal, S. Bhatia, P. Vadlamani, and D. Marx, “Searching satellite imagery with integrated measures,” Pattern Recognit. , vol. 42, no. 11, pp. 2502–2513, Nov. 2009.

[16] S. Newsam and C. Kamath, “Retrieval using texture features in high res-olution multi-spectral satellite imagery,” in Proc. SPIE Defense Security
Symp., Data Mining Knowl. Discov.—Theory, Tools, Technol. VI
, Orlando,FL, USA, 2004, pp. 21–32.

[17] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,”
IEEE Trans. Geosci. Remote Sens. , vol. 51, no. 2, pp. 818–832, Feb. 2013.

[18] L. Bruzzone, C. Persello, and B. Demir, “Active learning methods in classification of remote sensing images,” in
Signal and Image Processing for Remote Sensing
, 2nd ed, C. H. Chen, Ed. Boca Raton, FL, USA:
CRC Press, 2012, ch. 15, pp. 303–323