Metadata Generation and Retrieval of Geographic Imagery*†

Aidong Zhang and Lei Zhu
Department of Computer Science and Engineering
State University of New York at Buffalo
Buffalo, NY 14260
{azhang, lzhu}@cse.buffalo.edu

Abstract

Novel approaches are needed to support content-based retrieval on large-scale geographic image databases. In this paper, we present a new approach termed *Keyblock* for content-based geographic image retrieval, which is a generalization of the text-based information retrieval technology in the image domain. In this approach, methods for extracting comprehensive geographic image features are provided, which are based on the frequency and correlation of representative blocks, termed keyblocks, of the geographic image database. Keyblocks, which are analogous to index terms in text document retrieval, can be constructed by exploiting various clustering algorithms. By comparing the performance of our approach with conventional techniques using color feature and wavelet texture feature, the experimental results demonstrate the effectiveness of our approach for geographic image retrieval.

1 Introduction

Geographic images are being gathered from civil, defense, and intelligence satellites at an explosive rate. For example, NASA has terabytes of space data for exploration of space and atmospheric sciences. The United States Geological Survey (USGS) provides the archiving of the Nation's largest repository of remotely sensed data along with other USGS natural science data. To effectively and efficiently access such enormous data resources, novel approaches are needed to support content-based retrieval on large-scale geographic images. Despite the advance in content-based image retrieval [2, 12, 8, 1], systems that are designed particularly for geographic images are scarce. Effective representation of geographic content has not been adequately addressed.

In general, content-based image retrieval (CBIR) using low-level features such as color, texture, shape and others extracted from the images has been well studied. Various image querying systems including QBIC [2], VisualSeek [12], PhotoBook [8] and Virage [1] have been built based on the low-level features for general or specific image retrieval tasks. However, effective and precise image retrieval still remains to be an open problem because of the extreme difficulty in image understanding.

In contrast, many techniques in text-based information retrieval (IR) have been developed, and some keyword-based text information retrieval systems such as Yahoo, Lycos, and Google have achieved great success for indexing and querying web sites. The success has also shed light on the area of content-based image retrieval because the relatively mature theories and techniques of text-based information retrieval may be applied to the image domain. One of the representatives is to annotate images with metadata such as text keywords, captions and then to index the image database using text-based techniques on the metadata.

*This research is supported by National Science Foundation Digital Government Program.
†URL of demo: http://vangogh.cse.buffalo.edu:8080/.

Since both text and image retrieval are involved in information retrieval, the success of the text-based information retrieval motivates us to develop a theory of content-based image retrieval which is analogous to the techniques developed in text-based information retrieval.

The generalization of information retrieval from text domain to image domain is however non-trivial. The greatest obstacle is due to the intrinsic difference between text and image as different media in representing and expressing information. With respect to representation, syntactically, a text document is 1-dimensional while an image is 2-dimensional. With respect to expression, semantically, the units (words) of a text document, especially those keywords, carry direct semantics which are related to the semantics of text documents. In contrast, the units of an image, either in the pixel level or in the segment level after segmentation, provide generally no clue at all about the semantics of the image in the first case, or give unreliable object description in the second case.

In recognizing the existing problems in the CBIR field, this research investigates the theories and techniques of information retrieval for geographic images which are analogous to keyword-based text retrieval. Our work will result in practical solutions for querying and browsing of geographic images over the web. A critical issue to be addressed is how to construct metadata (or feature segments) in images which are similar to keywords. We term such metadata or feature segments as *"keyblocks"*. The codebook consists of the set of keyblocks that are selected. If the codebook of the keyblocks can be constructed, then an image can be encoded as the indices of the keyblocks in the codebook. Based on this image representation, information retrieval and database analysis techniques developed in text domain can be generalized for image retrieval. We present a new approach for content-based geographic image retrieval, which is a generalization of the text-based information retrieval technology in the image domain.

This paper is organized as follows. Section 2 introduces the framework of the keyblock-based image retrieval. Section 3 describes the approaches for generating keyblocks of images and for encoding images. Section 4 presents the codebook based image feature extraction and retrieval models. And finally, the conclusion is provided in Section 5.

2 Keyblock-based Image Retrieval Model

To generalize techniques of information retrieval from the text domain to the GIS image domain, we propose a practical framework called keyblock-based image retrieval, which includes the following main stages:

(1). *Codebook generation*: generate codebooks which contain keyblocks of different resolutions. Although objects are good candidates to be considered as visual keywords in the images, object recognition for natural images is still an unsolved problem and may remain to be an open problem in the long term. With a limited degree of sacrificing the accuracy, one practical approach is to partition/segment the images into smaller blocks, and then select a subset of representative blocks using clustering algorithms. These representative blocks can be used as the keyblocks to represent the image contents.

(2). *Image encoding*: for each image in the database as well as in the query, decompose it into blocks. Then, for each of the blocks, find the closest entry in the codebook and store the index correspondingly. Each image is then a matrix of indices, which can be treated as 1-dimensional codes of the keyblocks in the codebook. This property is similar to a text document which is considered as a linear list of keywords in text-based information retrieval. The image can also be re-constructed by using the codebook.

(3). *Image feature representation and retrieval*: extract comprehensive image features, which are based on the frequency of the keyblocks within the image, and provide retrieval techniques to support content-based image retrieval. There are four main components in this stage: (a). Database $D = \{I_1, ..., I_j, ..., I_M\}$: a list of encoded images; (b). Codebook $C = \{c_1, ..., c_i, ..., c_N\}$: a list of keyblocks; (c). CBIR model $\varphi = (f, s)$: f is a feature extraction mapping which generates the feature vector for each image, and s is a similarity measure between the feature vectors of two images; and (d). Q: a set of visual queries, where

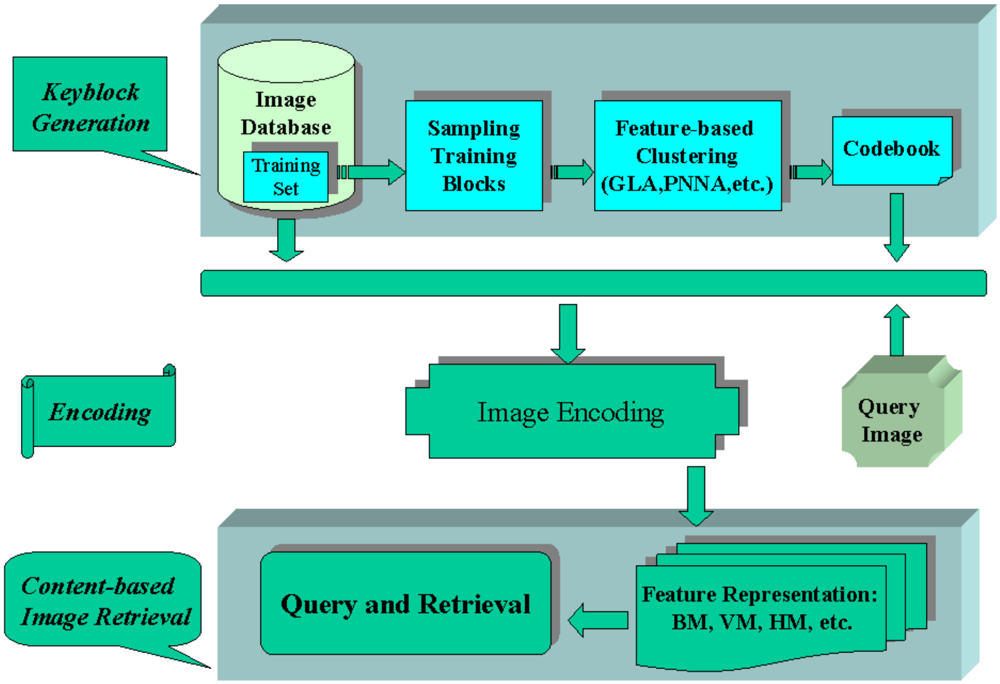

Figure 1: Flowchart of the keyblock-based image retrieval.

each query q has a feature vector which is similar to the feature vector of an image.

Figure 1 illustrates a flowchart of our approach. This idea was initially explored in [5, 4]. This paper expands the idea and apply it to the geographic image domain.

3 GIS Codebook Generation

Keyblock generation is critical to our approach. Two general domains are considered in our approach:

- *Original space*: with a limited degree of sacrificing the accuracy, one practical approach is to partition/segment the images into smaller blocks, and then select a subset of representative blocks, which can be used as the keyblocks to represent the images.
- *Feature space*: another practical approach is to extract low-level feature vectors, such as color, texture, and shape, from image segments/blocks, and then select a subset of representative feature vectors, which can be used as the keyblocks to represent the images.

We design various clustering approaches to generate keyblocks from either the original image space or the feature space. The keyblocks are selected from the centroids of the clusters of either the original space or feature space. Fundamentally, let $C = \{c_1, \ldots, c_i, \ldots, c_N\}$ be the "codebook" of the keyblocks representing the images, where N is the codebook size and c_i, $1 \leq i \leq N$, are the keyblocks. Let F be a mapping:

$$F : R^k \longrightarrow C = \{c_1, \ldots, c_i, \ldots, c_N \mid c_i \in R^k\},$$

where R^k is the Euclidean space of dimension k. Given a sequence $T = \{t_1, \ldots, t_j, \ldots, t_l \mid t_j \in R^k\}$, the mapping F gives rise to a partition of T which consists of N cells $P = \{p_1, \ldots, p_i, \ldots, p_N\}$, where $p_i = \{t \mid t \in T, F(t) =$

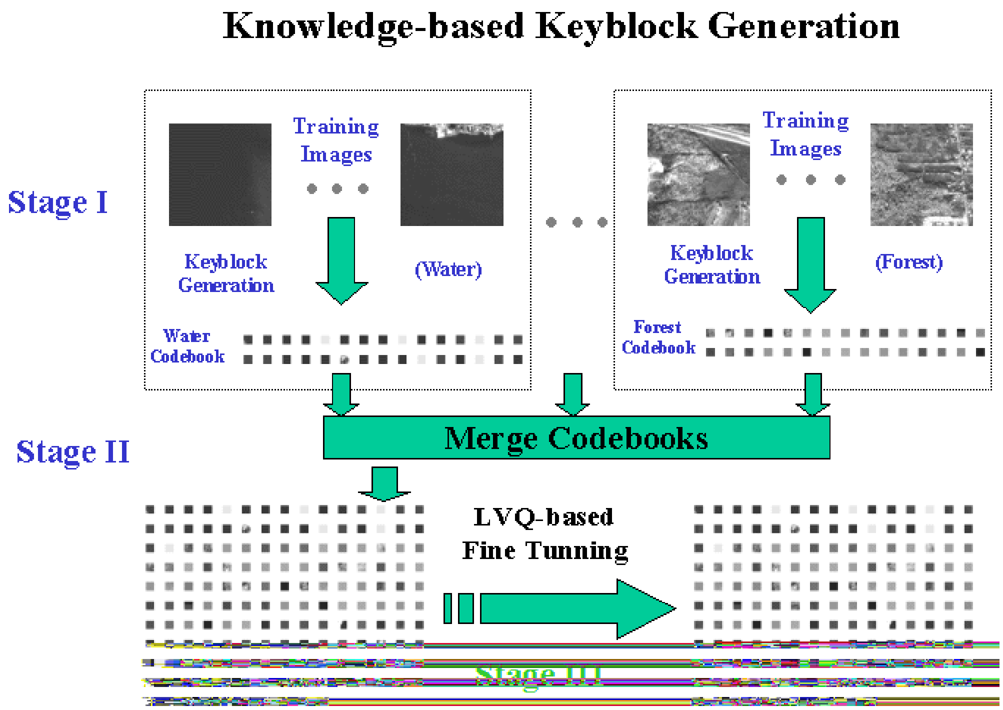

Figure 2: 3-Stages of the knowledge-based keyblock generation.

$c_i\}$. For a given distortion function $d(t_j, c_i)$, which is the distance between the input t_j and output code c_i (For example, the Euclidean distance, which is also called the square error), an optimal mapping must satisfy the following conditions:

- *Nearest Neighbor Condition:* For each p_i, if $t \in p_i$, then $d(t, c_i) \leq d(t, c_j)$, for all $j \neq i$.
- *Centroid Condition:* For a given partition P, the optimal code vectors satisfy

$$c_i = \frac{\sum_{t \in p_i} t}{k_i}, 1 \leq i \leq N, k_i \ is\ the\ cardinality\ of\ p_i.$$

There are a variety of clustering algorithms available which can be applied to different types of data sets [6, 13, 14, 7, 9, 10]. We use a practical and efficient algorithm. In this algorithm, clustering is applied to the set of data obtained from a training set of the images (either from the original space or the feature space) and then the centroid of each cluster is used as a codebook entry. Two popularly used algorithms are *Generalized Lloyd Algorithm* (GLA) [3] and *Pairwise Nearest Neighbor Algorithm* (PNNA). We use an integrated approach of the two algorithms to efficiently generate the codebook of keyblocks. Furthermore, we also incorporate the GIS domain knowledge into the keyblock generation process. To generate codebooks which reflects the semantic content of the geographic features, we adopt a 3-stage keyblock generation strategy which is illustrated in Figure 2.

At stage I, for each geographic feature (For example, for geographic image database GEO used in our testing, there are five geographic features: water, agriculture areas, forest, grass lands, and residential areas.), a corresponding codebook will first be generated. For each geographic feature, domain experts only need to provide some training images, then the clustering algorithm mentioned above is used to select the keyblocks. At stage II, codebooks generated in the stage I will be merged to a big codebook. This codebook

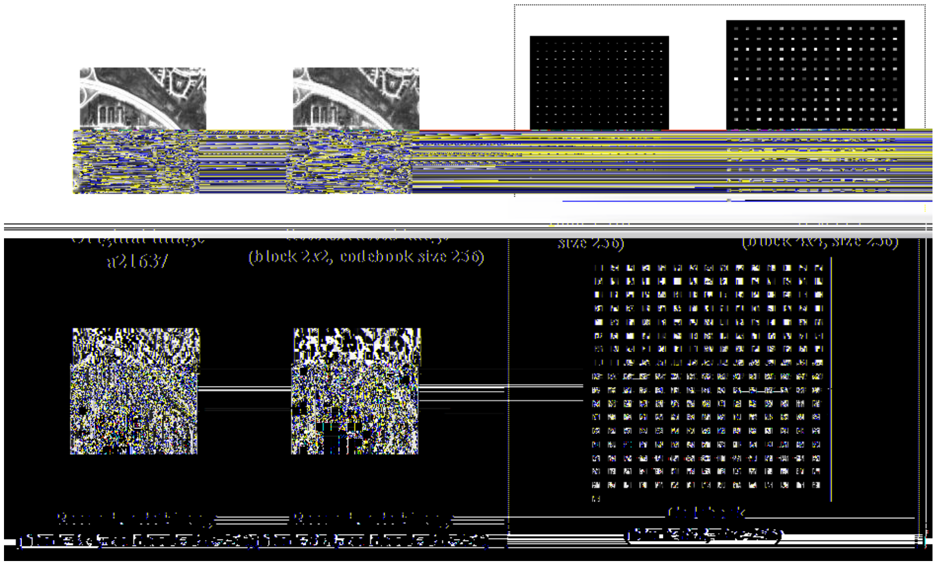

Figure 3: A raw image and the encoded images after re-construction with different codebooks. Each codebook is obtained with the same training set.

has keyblocks of different semantic meaning and can be used in image coding and decoding. At stage III, a fine tuning process, termed learning vector quantization (LVQ)-based approach, will be used.

In our experiments, to generate keyblocks, 405 remote-sensing images are randomly selected as the training set from our GIS testbed. Three block sizes, 2×2, 4×4, and 8×8, are used. Intuitively, blocks with different sizes may capture information with different granularity. Usually smaller blocks exploit local information of the image content, such as edges and regions with high spatial frequency. Larger blocks may provide correlation among neighboring sub-blocks as well as an overview of the global variation.

For each block size, experiments have been performed to generate codebooks of three different sizes 256, 512, and 1024. In the implementation, the distortion, which is the objective function for optimization when generating a codebook, is the square error commonly used for image compression. The square error is the Euclidean distance between the vectors of the intensity values of an original block and that of the corresponding keyblock. In short, the testing is conducted with 9 (3 block sizes $\times$ 3 codebook sizes) codebooks. After the codebooks are generated, all images in the database are then encoded correspondingly. As an example, Figure 3 shows an image and its reconstructed images with different codebooks.

To encode an image based on the codebook, the image is partitioned or segmented into blocks and then each block (or its feature vector) is replaced by the index of the nearest entry in the codebook. Now each image is an matrix of indices of the keyblocks. We can re-construct the image using the codebook to measure if the codebook is properly selected. To reconstruct the image, each index is replaced by the code vector in the codebook which is actually a lookup table. Obviously, the reconstructed image is only an approximation of the original one. Figure 4 illustrates the general procedure for image encoding and decoding.

4 Keyblock-Based Image Feature Representation and Retrieval

Image feature representation models similar to text-based retrieval models can be designed in this context. We have designed various models suitable for geographic images:

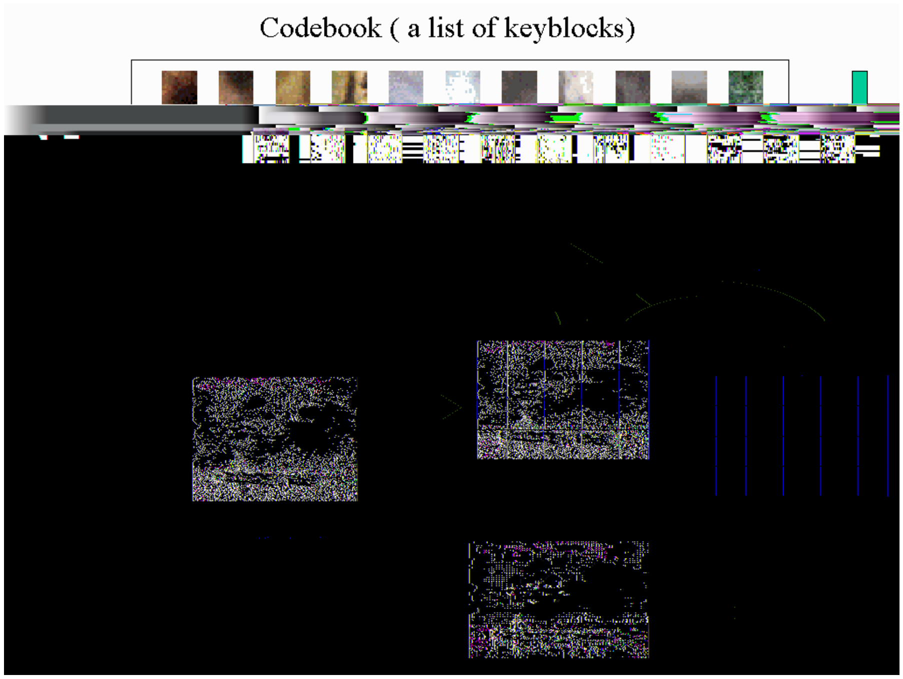

Figure 4: The general procedure of image encoding and decoding.

- *Models based on single keyblocks*: the features are calculated based on the appearance of individual keyblocks (termed uni-block models). In particular, we have designed Boolean Model (BM), Vector Model (VM), and Histogram Model (HM).

- *Models based on multiple keyblocks*: the features are calculated based on the correlations between keyblocks in images. Our purpose is to extract feature vectors which not only include the occurrence information of blocks but also carry context information of the neighboring blocks. We call such models n-block model, where n is the number of blocks considered. Particularly interested models are bi-block and tri-block models.

- *Models based on combined features*: The above models capture different image content under various contexts. For example, the uni-block model only considers single keyblock's occurrence, while the bi-block and tri-block models consider multiple keyblocks' co-occurrence. It is reasonable to combine them to improve the retrieval performance.

We have conducted comprehensive experiments to demonstrate the effectiveness of these models. In the GIS domain, users are more interested in "subimage match": given a small image pattern which represents kinds of geographic features such as forest and water, find images having similar geographic features and mark out those areas. It is relatively similar to object recognition. We compared these models with the popularly used wavelet-based models.

For wavelet transforms, first, we use Nona Tree [11] to decompose each image to subimages recursively until a certain subimage size (in this experiment, it is 32×32.) is reached. Then, based on these subimages as well as those original images, different types of wavelet transforms such as gabor, haar and daubechies are applied to extract texture features of images at different scales/resolutions from fine to coarse. Similarly, for codebook-based feature extraction and retrieval, we first use Nona Tree to decompose each encoded image, then conduct models such as the uni-block model, bi-block model, tri-block model and feature combination model respectively on the subimage level.

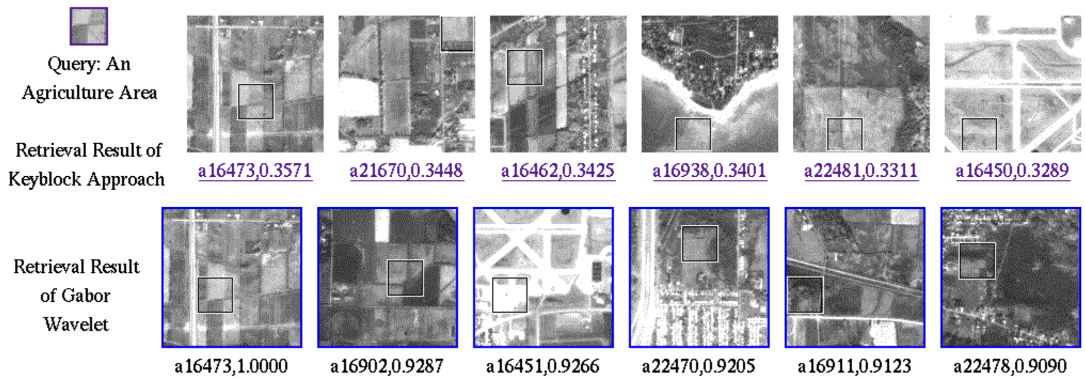

Figure 5: Example of retrieval results on an agriculture image in GEO.

There are 33 query images which are subimages of 32×32 chosen from the images in the database by GIS experts from NCGIA at Buffalo. These query images are divided into 5 categories: agriculture, grass, forest, residential area, and water. Their feature vector are generated correspondingly. For each query, the average precision corresponding to the top 1, 2, ..., up to 40 retrieved images are calculated. Finally, the average precision is calculated over all queries.

Figure 5 shows the retrieval results of an agriculture image. The query image is at the upper-left corner. The returned images as well as the marked areas of our approach is apparently better than those of gabor wavelet. Figure 6 indicate that the performance of our approach outperform the wavelet transforms. For example, in the case of Feature Combination Model, the average precision is always 20% to 30% higher than wavelet transforms.

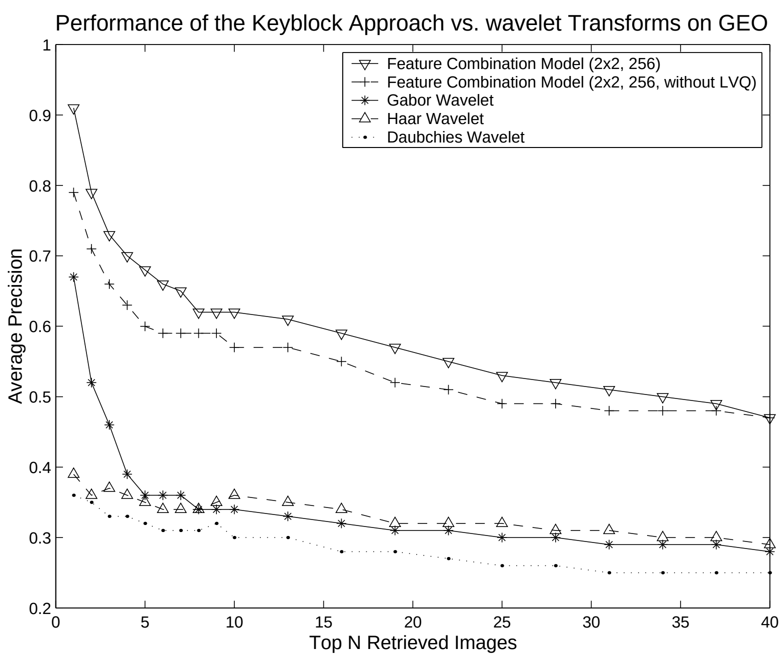

Figure 6: Average precision: comparison of codebook approach with Wavelet Transforms.

5 Conclusion

A new approach for content-based geographic image retrieval is proposed by exploiting analogous text-based IR techniques. The proposed approach provides methods for extracting image features which is not in favor of any particular low-level feature such as color or texture. Instead, the features extracted for an image is a comprehensive description of the content of the image which is more semantics-related than the existing lower-level features. We have conducted substantial experiments to demonstrate the effectiveness of the proposed approach. Results have demonstrated that our approach is superior not only to color histogram and color coherent vector approaches which are in favor of color features, but also to Haar and Daubechies wavelet texture approaches which have been commonly used for texture-based geographic image retrieval.

References

[1] J.R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Jain, and C.F. Shu. The virage image search engine: An open framework for image management. In *Proceedings of SPIE, Storage and Retrieval for Still Image and Video Databases IV*, pages 76–87, San Jose, CA, USA, February 1996.

[2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, and B. Dom et al. Query by Image and Video Content: The QBIC System. *IEEE Computer*, 28(9):23–32, September 1995.

[3] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[4] Lei Zhu, Aibing Rao and Aidong Zhang. Advanced feature extraction for keyblock-based image retrieval. In *Proceedings of International Workshop on Multimedia Information Retrieval (MIR2000)*, pages 179–183, Los Angeles, California, USA, November 4 2000.

[5] Lei Zhu, Aidong Zhang, Aibing Rao and Rohini Srihari. Keyblock: An approach for content-based image retrieval. In *Proceedings of ACM Multimedia 2000*, pages 157–166, Los Angeles, California, USA, Oct 30 - Nov 3 2000.

[6] R. T. Ng and J. Han. Efficient and Effective Clustering Methods for Spatial Data Mining. In *Proceedings of the 20th VLDB Conference*, pages 144–155, Santiago, Chile, 1994.

[7] E.J. Pauwels, P. Fiddelaers, and L. Van Gool. DOG-based unsupervized clustering for CBIR. In *Proceedings of the 2nd International Conference on Visual Information Systems*, pages 13–20, San Diego, California, December 1997.

[8] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for Content-based Manipulation of Image Databases. In *Proceedings of the SPIE Conference on Storage and Retrieval of Image and Video Databases II*, pages 34–47, 1994.

[9] G. Sheikholeslami, S. Chatterjee, and A. Zhang. *WaveCluster*: A Multi-Resolution Clustering Approach for Very Large Spatial Databases. In *Proceedings of the 24th VLDB conference*, pages 428–439, August 1998.

[10] G. Sheikholeslami, S. Chatterjee, and A. Zhang. *WaveCluster*: A Wavelet-Based Clustering Approach for Multidimensional Data in Very Large Databases. *The VLDB Journal*, 8(4):289–304, February 2000.

[11] G. Sheikholeslami, E. Remias, J. Guo, and A. Zhang. Image decomposition and representation in large image database systems. *The Journal of Visual Communication and Image Representation*, 8(2):167–181, June 1997.

[12] John R. Smith and Shih-Fu Chang. VisualSeek: a fully automated content-based image query system. In *Proceedings of ACM Multimedia 96*, pages 87–98, Boston MA USA, 1996.

[13] Wei Wang, Jiong Yang, and Richard Muntz. STING: A Statistical Information Grid Approach to Spatial Data Mining. In *Proceedings of the 23rd VLDB Conference*, pages 186–195, Athens, Greece, 1997.

[14] Tian Zhang, Raghu Ramakrishnan, and Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, pages 103–114, Montreal, Canada, 1996.