Creating a Large-Scale Content-Based Airphoto Image Digital Library

Bin Zhu, Marshall Ramsey, and Hsinchun Chen

***Abstract*—This paper describes a content-based image retrieval digital library that supports geographical image retrieval over a testbed of 800 aerial photographs, each 25 megabytes in size. In addition, this paper also introduces a methodology to evaluate the performance of the algorithms in the prototype system. The major contributions of this paper are two. 1) We suggest an approach that incorporates various image processing techniques including Gabor filters, image enhancement, and image compression, as well as information analysis technique such as self-organizing map (SOM) into an effective large-scale geographical image retrieval system. 2) We present two experiments that evaluate the performance of the Gabor-filter-extracted features along with the corresponding similarity measure against that of human perception, addressing the lack of studies in assessing the consistency between an image representation algorithm or an image categorization method and human mental model.**

***Index Terms*—Content-based image retrieval, digital library, Gabor wavelets, self-organizing map, system evaluation.**

I. Introduction

The emergence of digital libraries makes content-based image retrieval an issue of rapidly increasing importance. To make an image system accessible requires the integration of image processing and information retrieval technologies. Currently, these two research areas are completely separate and rarely interact. To counteract the nonscalability of the traditional approach of textual annotation, effective algorithms for image feature extraction and image representation have been developed in the image processing field. However, incorporation of those algorithms into an effective image retrieval system has not been completely explored. This work presents a texture-based geographical image retrieval system, called SOM-AIR, that supports geographical image retrieval over a testbed of 800 aerial photographs, each 25 Mbytes in size. The description of this system proposes an approach to incorporating various image processing techniques including Gabor filters, image enhancement, and image compression into creation of an effective image digital library. In addition, The system applies SOM, an information analysis technique that has been successfully applied in textual information retrieval in the Illinois/Arizona digital library project [1], [18] to categorize geographical images based on texture features extracted from them. To assist users in specifying their queries and to translate users' high level queries into low-level texture features, the SOM-AIR system takes advantage of the two–dimensional (2-D) display of the SOM output as a part of its user interface for convenient on-click image retrieval.

In a large-scale content-based image digital library, it is desirable that the features extracted and the corresponding similarity measure map to the human perception of image similarity. However, most image retrieval research compares potential algorithms by running them on data sets from the Brodatz album [20] and has rarely evaluated the performance of an algorithm against that of human subjects. To address this weakness, this paper presents a way to evaluate the performance of image representation and categorization techniques that the SOM-AIR system uses. Two evaluations were conducted in such a manner that the performance of the system was directly compared with that of human subjects.

The structure of this correspondence is as follows. Section II provides an overview of the SOM-AIR system. Section III reviews the procedure of feature extraction using Gabor filters and the evaluation. Section IV describes the SOM algorithm and its application in image retrieval, followed by an evaluation. The implementation and the functionality of the SOM-AIR system are presented in Section V. Section VI provides a discussion and conclusions.

Manuscript received December 15, 1998; revised July 27, 1999.This work was supported by the NSF/ARPA/NASA Digital Library Initiative (1996–1998), the NSF/ARPA/NASA Digital Library Initiative IRI-9 411 381 (1994–1998), and by DARPA through Grant N6601-97-C-8535 (1997-2000). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Thomas S. Huang.

The authors are with the Management Information Systems Department, University of Arizona, Tucson, AZ 85721 USA (e-mail: bzhu@bpa.arizona.edu; mramsey@bpa.arizona.edu; hchen@bpa.arizona.edu).

Publisher Item Identifier S 1057-7149(00)00186-X.

II. The Overview of the SOM-AIR System

A. The Testbed

The testbed of the SOM-AIR system is a collection of 800 aerial photos provided by the Map and Imagery Laboratory of Davidson Library at the University of California, Santa Barbara (UCSB). The areas that the data set covers include Santa Barbara County, Channel Islands, Anacapa Island, and Santa Barbara Island.

B. System Architecture

As a joint effort between the Illinois Digital Library project [3], [23] and the Alexandria Digital Library Project [24], our prototype aerial photo image digital library attempts to provide a user-friendly, scalable approach to air photo image retrieval. Content-based image retrieval in a large map and imagery library such as that of the UCSB (which owns over 3.5M maps and aerial photographs, one of the largest collections in the world) requires significant research effort in both image analysis and digital library content management and interface development.

Fig. 1 presents the architecture of the SOM-AIR system. The system consists of a user interface, five subsystems, and three databases. The functionality of major components of the system and what technique is applied at which part of the system are described below.

The Image Enhancement Subsystem applies the normalized contrast stretching algorithm [22] on the digitized aerial photos to improve the visual quality of some low contrast images. An advantage of using this algorithm is that it gives both the convenience of image brightness control and adjustability of image contrast by simply assigning values of the desired mean and standard deviation to the processed image.

The feature extraction subsystem extracts salient texture features of input images by using Gabor filters. Each image is divided into tiles, each of which is represented by the features extracted. The low-level features created are stored in the feature database.

The image compression subsystem compresses an image to a set of images with various resolutions by applying the Joint Photographic Experts Group (JPEG) compression. Users often expect an image retrieval system to return a set of images that match their queries; maintaining a hierarchical resolution of images will enable the system to meet this user requirement without sacrificing performance, especially in web-based image retrieval. The lowest resolution images can be used as thumbnails for rapid browsing and higher resolution images can provide more detailed visual information. The subsystem stores image data generated into the image database.

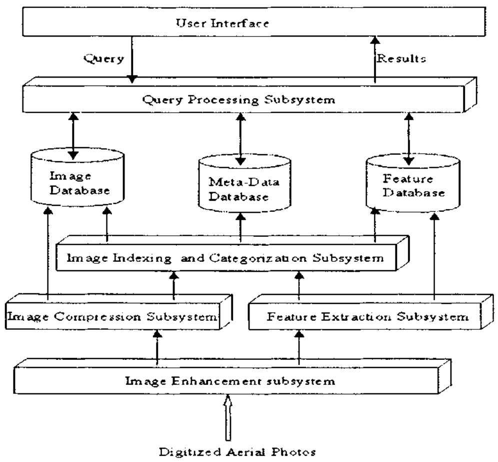

Fig. 1. Architecture of the SOM-AIR system.

The image indexing and categorization subsystem applies SOM to categorize images based on the texture features extracted. Each image consists of tiles and is associated with a set of image data with various resolutions. Every tile is represented by its texture features and is grouped into a category. All this information is stored in the meta-data database.

The user interface consists of four frames: the query frame, the texture-info frame, the results frame, and the image frame. The user interface uses the 2-D display of SOM output to support "retrieve by browsing" or "query by example." The query results with different resolutions are also displayed. A detailed description of the interface is presented in Section V of this paper.

III. Texture Feature Extraction

A. Algorithm Description

The traditional algorithm of textual annotation represents an image based on its author, date, and content. This approach, however, fails to capture the complete content of an image and requires manual effort to define and enter the necessary annotation. To counteract the nonscalability of this approach, effective algorithms for image feature extraction and image representation have been developed, as evidenced by several recent prototypes such as the Photobook system at MIT [19] and by commercial systems, such as IBM's QBIC system [7]. A variety of algorithms can be employed to extract low-level features in image retrieval systems. For instance, QBIC calculates the texture features of an image according to its coarseness, contrast, and directionality. Photobook consists of three parts: the Appearance Photobook, the Shape Photobook, and the Texture Photobook. In the texture Photobook, Wold-based representations are used to extract the texture features of an image. In the Los Alamos National Lab's CANDID Project, Laws' texture energy maps are applied to extract textural features from pulmonary CT images [13]. In [16], in the prototype system for the Alexandria Digital Library Project, Manjunath and Ma used Gabor filters to extract texture features of an aerial photo.

The selection of an algorithm for image representation varies with the image type. For instance, an algorithm may work well with medical images but may not be appropriate for geographical images. In our prototype system, since we used aerial photos as the input of the prototype system, we employed Gabor filters as our image representation algorithm and Euclidean distance as the similarity measure in the feature space. Gabor filters were first proposed by Gabor [8] in 1946 to analyze one-dimensional (1-D) signals such as audio signals, and were extended into 2-D by Daugman [6] in 1980. As indicated in [16], Gabor filters perform well in representing aerial photos. Our experiment, presented in the next subsection, also indicated that Gabor-filter-extracted features and associated similarity measure could map the human perception of aerial photo similarity.

B. Evaluation

To evaluate how well the human perception of geographical image similarity is mapped by Gabor-filter- extracted texture features along with corresponding similarity measures, we designed an experiment that involved ten human subjects and ten different sample images. Every image consisted of 192 tiles, each of which was made up of 128×128 pixels. Both the system and human subjects performed the same tasks. An expert with three years of experience in analyzing remote sensing pictures evaluated both sets of results. For every image, we randomly selected six tiles as reference tiles. We assigned every subject one image and the corresponding reference tile. For each reference tile, we asked the subject to evaluate every tile on the image and assign a score, from 0–10, based on its similarity to the reference tile. The quality of the performances of subjects and the system was measured by the values of *precision* and *recall*. *Precision* represented the relevance of the retrieved information, while *recall* indicated how much of the relevant information in the database was retrieved [21].

Our geographical image system determines the similarity between image tiles based on Euclidean distances in their feature space. We decided on a threshold of Euclidean distance and considered an image tile to be similar to a reference tile when the Euclidean distance between them was shorter than that threshold value. This threshold was determined during the pilot study according to our visual judgment. We also set a threshold value for the scores assigned by human subjects in the first part of the experiment. We believed other images were either similar or not similar to a reference tile. Janes [12] conducted an experiment to study humans' perception of relevance. During his experiment, human subjects were asked to give a "break point" on a continuum of relevance from zero relevance to complete relevance, he found that the break points assigned by human subjects exhibited a wide range. However, the mean value of the break point remained at around 50 on a scale of 100. Therefore, we considered an image tile to be similar to the reference tile if human subjects assigned it a score higher than or equal to six.

The results of this part of the experiment is summarized as follows.

1) The system exhibited no statistically significant difference from human subjects in the precision comparison (subject precision = 42.55%, system precision = 47.64%, $P = 0.428$), while it performed less well than human subjects in the comparison of recall(subject recall = 78.00%, system recall = 66.38%, $P = 0.034$).
2) The top five similar tiles suggested by the system and by human subjects were often different. The rationale behind this appears to be that the measure of Euclidean distance only partially resembles human perception of similarity. This finding is consistent with the results of [15], which indicated that nearest-neighbor search failed to retrieve some other more relevant patterns.
3) Both the system and the human subjects did well in retrieving tiles with distinguishable texture but had difficulty in retrieving image tiles without apparent texture. For instance, both the system and human subjects did well in retrieving image tiles of orchard, which has apparent texture, while both the human subjects and the system had difficulty in retrieving the pure water tiles because water tiles had the same homogeneous

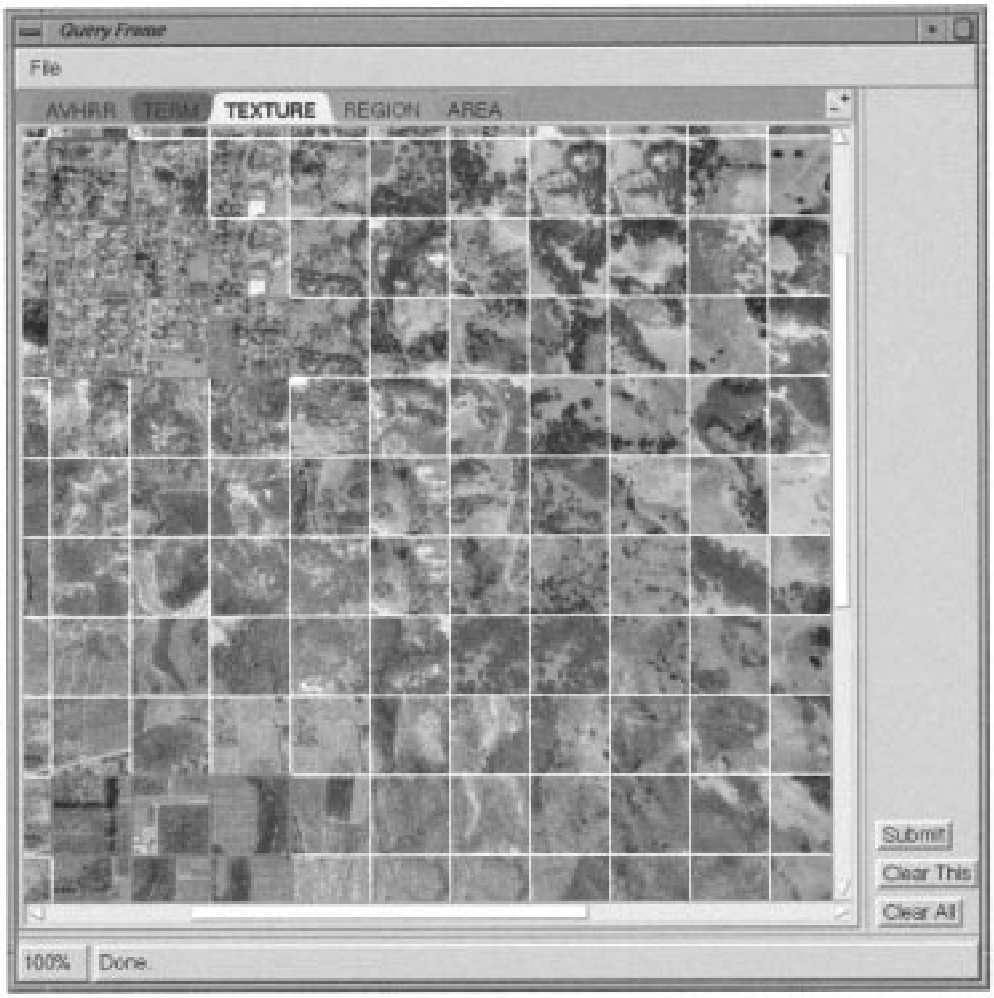

Fig. 2. Query frame of the SOM-AIR interface. The other tabs such as "AVHRR" and "TERM" are used to integrate image retrieval with textual document retrieval and numerical satellite data retrieval. In this work, we focus on image retrieval. The query frame is a 2-D display of SOM output. User-selected tiles are highlighted.

texture as soil or forest tiles. This suggests that under some circumstances, texture alone is insufficient to represent image.

IV. Apply SOM in Image Categorization

A. Algorithm Description

This section describe how we adopted an unsupervised classification technique, self-organizing map (SOM) for image categorization. SOM was first proposed by Kohonen, who based his neural network on the associative neural properties of the brain [14]. The network consists of an input layer and an output layer. The number of the input nodes equates to the number of attributes associated with the input. After all of the input is processed, the result is a spatial representation of the input data, organized into clusters of similar regions. SOM is defined as a mapping from a high-dimensional input space into a 2–D array of output nodes, and the output nodes that are topographically close are considered to be similar to each other.

Several recent studies adopted the SOM approach to textual analysis. Examples are the Distributed Script Processing and Episodic Memory Network (DISCERN) developed by Mikkulainen [17] as a natural language processing system, the WEBSOM system developed by Kohonen's group for newsgroup classification [11], and the multilayered SOM ˝system developed by the Arizona Artificial Intelligence Laboratory for Internet home page categorization [4]. Their work suggests a high applicability of the SOM approach to large-scale classification.

To apply SOM to image categorization, the SOM-AIR system divides each image into small tiles, each of which has the size of 128 × 128 pixels. Tiles are the smallest units in the SOM system and are represented by the Gabor-filter-extracted features. The adaptive SOM algorithm applied in the SOM-AIR image retrieval system is summarized as follows..

1) Initialize the weights connecting the N input nodes and the M output nodes at a small random value and initialize the neighborhood size r_e. We use w_{ij} to denote the weight between the input node i and the output node j.
2) Present the feature vector of one tile as the input vector, which is represented as $\{x_0, x_1, \cdots, x_{N-2}, x_{N-1}\}$. Calculate the Euclidean distance (d_j) between each output node and the input vector.
3) Find the "winning" output node j^*, such that $d_{j*} = \min(d_j)$, $0 \leq j \leq M-1$.

4) Update weights to Nnode j^* and its neighbors. The neighbors at time t of the node j^* are defined by $N_c(t)$.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t)) \qquad j \in N_c(t),\quad 0 \le i \le N-1$$

where $n(t)$ is a learning rate that is decreased with time.
5) Repeat steps 2, 3, and 4, until all feature vectors have been used.
6) Repeat steps 2, 3, 4, and 5 with a decreasing learning rate and neighborhood size.
7) Assign every tile to an output node that has minimum Euclidean distance to the feature vector of that tile.
8) Label each output node with a tile whose feature vector has minimum Euclidean distance to this node.
9) Merge adjacent output nodes that have the same label to form a category. The tile that labels the category is called the representative tile of that category.

Thus the SOM map created can be considered to be a graphical categorization of the images.

B. Evaluation

We also designed an experiment to evaluate the performance of the SOM. We continued to use the same ten images in this part of the experiment, but a different group of ten human subjects participated. Every human subject worked with one image that was categorized by using the SOM algorithm. We used the system-selected representative tiles as suggested categories. Human subjects categorized the image by selecting all the tiles in the image into one of the suggested categories.

Again, both the expert and the system completed the same task the human subjects performed in this experiment. We continued to use the measures of *precision* and *recall* to evaluate the performance of the system.

1) The results of this part of the experiment are listed below. The system did at least as well as human subjects in image categorization. The precision and recall comparisons indicated that there were no statistically significant differences between performances of the human subjects and of the system (subject recall = 40.00%, system recall = 41.60%, subject precision = 35.08%, system precision = 33.88%).
2) As in the previous part of the experiment, both human subjects and the system did well with tiles of distinguishable texture and had difficulty in dealing with tiles without an apparent texture pattern.
3) Most of the subjects complained that there were too many suggested categories and that some of the representative tiles were similar to each other. This probably was due to the small size of the input data. A set of 192 tiles was too small for the adaptive SOM algorithm. However, the pilot studies indicated that 192 tiles were as many as human subjects could handle with the assistance of the interface.

V. The Implementation of the SOM-AIR System

An eight-node SGI Origin2000 machine was used to process the images in the SOM-AIR system. As the first step of system implementation, we applied the contrast enhancement algorithm to most of the images with preselected values of mean and standard deviation. This algorithm was implemented by a C program, which took 1 h of wall-clock time on a single node to process the whole image set.

We then divided each image into tiles, each of which contained 128 × 128 pixels, and consequently we derived 6.1 million tiles. We applied 30 Gabor filters to each tile and then represented that tile with its texture features extracted. Thus, each tile was represented by a feature vector with 60 elements. The whole operation was implemented in a C program that used 451 wall-clock hours summing over all eight nodes to extract texture features from 6.1 million image tiles. Each image was compressed to JPEG format by using a standard JPEG C library.

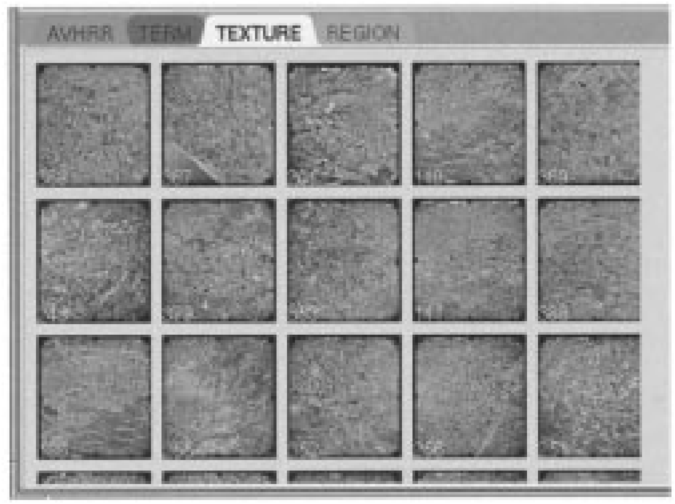

Fig. 3. Result frame of the SOM-AIR interface. The number at the left-bottom corner of each image is the image number. A set of images with low resolution is displayed in this frame. The images are ranked according to the numbers of tiles that match user queries.

Knowing that mountain is the dominant and visually uniform land surface type for the northern parts of both Santa Barbara county and Ventura counties, we used aerial photos taken over the southern parts of these two counties, which contain rich and diverse image patterns such as residential areas, farmland, and coast areas, to train the SOM according to the algorithm described in Section III in order to improve the precision and the richness of the categories created. A subset of 185 images was used to train the SOM categorization process. We then mapped all tiles, including mountain-rich photos, into categories based on the Euclidean distance between their feature vectors and the feature vector of representative tiles.

The SOM was implemented in a C++ program; it took 15 h to train the SOM, and 27 h to map each image tile to an appropriate category.

Fig. 2 is an example of the query frame of the system. This frame displays the representative tiles generated by SOM. Each tile is in low-resolution (64 × 64 pixels). A user could get a close look at a tile of interest by right clicking on it, upon which the system would highlight the tile clicked in blue and bring up another window to display the full resolution version of that tile and other tiles within the same group. In Fig. 2, the user selected some tiles with an urban pattern and some tiles with a farmland pattern. After the user clicked on the "submit" button on the query frame, the system displayed a list of thumbnails of retrieved images in the results frame (Fig. 3). The images in this list were sorted according to the numbers of related tiles they contained. The results frame displayed a set of the retrieved images in low resolution, from which the user could select any image in the list to activate the system to present a high-resolution version (700 × 700) of that image in the image frame, along with the related tiles being highlighted in green (does not show). In addition, the system could also display the place names associated with this image (Fig. 4) by cross-referencing with the Geographical Name Information System (GNIS) gazetteer. Place names provide rich and important contextual information for aerial photo browsing.

Fig. 4. Texture frame of the SOM-AIR interface. In this figure, a high resolution version (700 × 700) of image 368 is displayed and place names associated with the presented image are extracted from the Geographical Name Information System (GNIS) gazetteer and are displayed on the image according to the locations.

VI. Discussion and Conclusion

This work describes a prototype system that supports geographical image retrieval over a collection of 800 aerial photos. The system integrates various image processing techniques with information analysis techniques to support effective large-scale content-based image retrieval. By comparing the performance of the system with that of human subjects, we found that the combination of Gabor features and Euclidean distance could, to a certain extent, match the human perception of geographical image similarity and that the adaptive SOM algorithm is at least as good as human subjects in image categorization. Moreover, the display of representative tiles of each output node produced by SOM appears to be both an effective way to help a user specify queries and to translate high-level visual queries into low-level features.

Differing from the Alexandria Digital Library Project that also applies SOM to Gabor-filters-extracted features in its system, our system uses the output of SOM as its interface to help users to specify their queries. The prototype system suggests a promising way to integrate image representation techniques, such as Gabor filters, with information analysis techniques, such as SOM, to support large-scale content-based image retrieval in digital libraries. However, as indicated in the previous section, image feature extraction costs extensive implementation time. To support larger scale image retrieval will require improvement of the image representation algorithm used in our system.

Acknowledgment

The authors would like thank Prof. Manjunath of UCSB for providing the software for the Gabor filters, and L. Carver and M. Larsgaard of the Map and Imagery Laboratory, UCSB, for providing the aerial photo collections. Members of Arizona Artificial Intelligence Laboratory who have directly contributed to this paper are Y. Santoso, W. Wyzga, and D. Ng.

References

[1] H. Chen, A. L. Houston, R. R. Sewell, and B. R. Schatz, “Internet browsing and searching: User evaluations of category map and concept space techniques,” *J. Amer. Soc. Inform. Sci.*, vol. 49, pp. 582–603, 1998.

[2] H. Chen *et al.*, “A geographic knowledge representation system (GKRS) for multimedia geospatial retrieval and analysis,” *Int. J. Digital Libraries*, vol. 1, pp. 132–135, 1998.

[3] H. Chen *et al.*, “A parallel computing approach to creating engineering concept spaces for semantic retrieval: The Illinois digital library initiative project,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 771–78 2, 1996.

[4] H. Chen, C Schuffels, and R. Orwig, “Internet categorization and search: A machine learning approach,” *J. Vis. Commun. Image Represent.*, vol. 7, pp. 88–102, 1996.

[5] P. C. Chen and T. Pavlidis, “Segmentation by texture using correlation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, pp. 64–69, Jan. 1983.

[6] J. G. Daugman, “Two-dimensional spectral analysis of cortical receptive field profiles,” *Vis. Res.*, vol. 20, pp. 847–856, 1980.

[7] M. Flickner, H. Sawhney, W. Niblack, and J. Ashley, “Query by image and video content: The QBIC system,” *IEEE Computer*, vol. 28, pp. 23–33, 1995.

[8] D. Gabor, “Theory of communication,” *J. Inst. Electr. Eng.*, vol. 93, pp. 429–459 , 1946.

[9] R. M. Haralick, K. Shanmugan, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 610–621, Nov. 1973.

[10] T. Huang, S. Methrotra, and K. Ramchandran, “Multimedia analysis and retrieval system (MARS) project,” *Proc. Clinic on Library Applications of Data Processing, Digital Image Access and Retrieval*, 1996.

[11] T. Honkela, S. Kaski, K. Lagus, and T. Kohonen, “Newsgroup exploration with WEBSOM method and browsing interface,” Helsinki Univ. Technol., Helsinki, Finland, Tech. Rep. A32, Jan. 1996.

[12] J. W. Janes, “Other people’s judgments: A comparison of users’ judgments of document relevance, topicality, and utility,” *J. Amer. Soc. Inform. Sci.*, vol. 45, pp. 160–171, 1994.

[13] P. M. Kelly and T. M. Cannon, “CANDID: Comparison algorithm for navigating digital image databases,” in *Proc.7th Int. Working Conf. Scientific and Statistical Database Management*, 1994, pp. 252–258.

[14] T. Kohonen , *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag, 1995, ch. 3.

[15] W. Y. Ma and B. S. Manjunath, “A texture thesaurus for browsing large aerial photographs,” *J. Amer. Soc. Inform. Sci.*, vol. 49, pp. 633–648, 1998.

[16] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Trans. Pattern Anal. Machine Intell., Special Issue Digital Libraries*, vol. 18, pp. 837–842, 1996.

[17] R. Mikkulainen, *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*. Cambridge, MA: MIT Press, 1993.

[18] R. E. Orwig, H. Chen, and J. F. Nunamaker, “A graphical, self-organizing approach to classifying electronic meeting output,” *J. Amer. Soc. Inform. Sci.*, vol. 48, pp. 157–170, 1997.

[19] A. Pentland, R. W. Picard, and S. Sclaroff, “Photobook: Tools for content based manipulation of image databases,” *Proc. SPIE*, vol. 2185, pp. 34–47, 1994.

[20] R. W. Picard and T. Kabir, “Finding similar patterns in large database,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Minneapolis, MN, 1993, pp. 161–164.

[21] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*, New York: McGraw Hill , 1993.

[22] R. A. Schowengerdt, *Remote Sensing Models and Methods for Image Processing*, New York: Academic , 1997.

[23] B. R. Schatz *et al.*, “Federating repositories of scientific literature: The Illinois digital library project,” *IEEE Computer*, vol. 29, pp. 28–37, 1996.

[24] T. R. Smith, “A digital library for geographically referenced material,” *IEEE Computer*, vol. 29, pp. 54–60, 1996.