Image Processing in the Alexandria Digital Library Project

B. S. Manjunath

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560
manj@ece.ucsb.edu

ABSTRACT

The management of images, video, and in general, multimedia data, is an important issue in the design of digital libraries. In particular, the following problems stand out: efficient storage, fast retrieval, and protection of intellectual property. We outline below some of the recent advances in image processing in the context of the UCSB Alexandria Digital Library (ADL) project whose goal is to create a database of spatially indexed data. Maps and satellite images are among the main data sets in this project. The focus of this overview is on image retrieval using texture and on digital watermarking. A texture thesaurus for browsing aerial photographs and a wavelet based digital watermarking scheme are presented.

1. INTRODUCTION

The ADL project's goal is to establish an electronic library of spatially indexed information, providing internet access to a wide collection of geographic information. A significant part of this collection includes maps, satellite images, and aerial photographs. A typical airphoto can take over 25MB of disk space and providing access to such data raises several important issues such as network bandwidth and new tools for browsing and selecting images based on content.

A quick visual access to the stored data is essential for efficient navigation through image collections. While textual and image content based queries help in narrowing down the search space, visual browsing is used to obtain an overview of the data retrieved. Low resolution image thumbnails is one approach to such browsing. If image browsing is performed at multiple resolutions, then the average bandwidth requirements can be further reduced. In the ADL project, considerable progress has been made in multiresolution browsing using wavelets and in lossless compression. These include a new average interpolation subdivision scheme for multiresolution data representation and reversible wavelet transforms for lossless image compression. A detailed discussion of these can be found in [1]. As mentioned before, the focus of this paper is on the image retrieval using texture and on digital watermarking.

Image retrieval by example is an interesting problem. While manual image annotations can be used to help an image search, the feasibility of such an approach to large databases of images is questionable. In some cases, such as faces or texture patterns, simple textual descriptions can be ambiguous and inadequate for search and retrieval. Several systems have been developed recently to search through image databases using color, texture, and shape attributes. QBIC [2] by IBM is perhaps the best known example. Others include the PhotoBook [3], VisualSEEK [4] and the Virage system [5].

Image texture provides a powerful low level image description in the context of satellite images. Texture could be used to select a large number of geographically salient features in airphotos, such as vegetation patterns, parking lots, and building developments. We have developed a *texture thesaurus* for aerial photographs that facilitates fast search and retrieval of image data and this is currently being integrated into the ADL project testbed. In addition to efficient access, protecting intellectual property rights is another important issue in digital libraries research. Our current work on digital watermarking is presented in Section 3.

2. TEXTURE BASED IMAGE RETRIEVAL

In recent years image texture has emerged as an important visual primitive to search and browse through large collections of similar looking patterns. An image can be considered as a mosaic of textures and texture features associated with the regions can be used to index the image data. For instance, a user browsing an aerial image database may want to identify all parking lots in the image collection. A parking lot with cars parked at regular intervals is an excellent example of a textured pattern when viewed from a distance, such as in an airphoto. Similarly, agricultural areas and vegetation patches are other examples of textures commonly found in aerial and satellite imagery. Examples of queries that could be supported in this context could include "Retrieve all Landsat images of Santa Barbara which have less than 20% cloud cover" or "Find a vegetation patch that looks like this region.'

In [6] we have investigated the role of textures in annotating image collections and report on the performance of several state-of-the-art texture analysis algorithms with performance in similarity retrieval being the objective. It is demonstrated that simple statistics computed from Gabor filtered images provide a good feature descriptor for content based search. These texture features

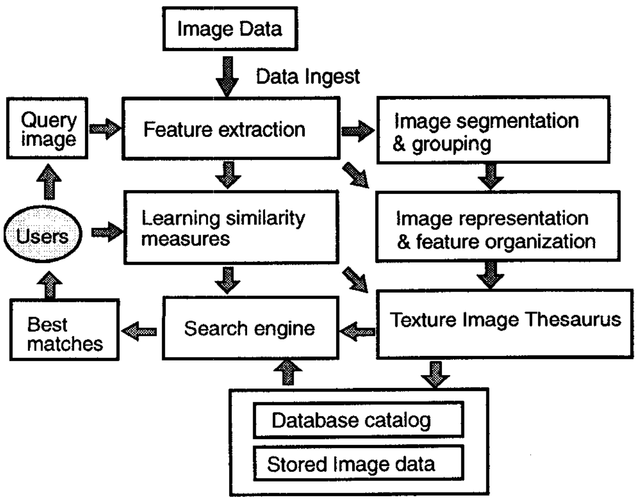
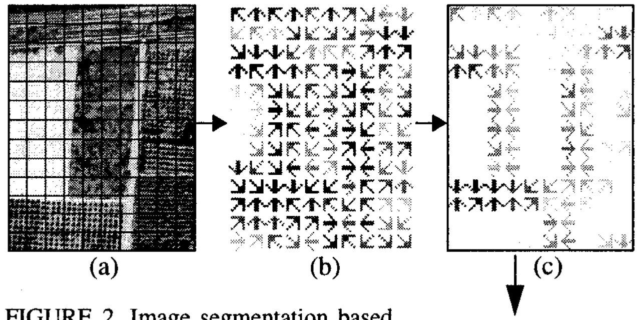


FIGURE 1. A schematic of the image search system using the texture thesaurus.



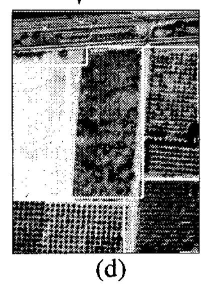


FIGURE 2. Image segmentation based on local texture flow. The texture flow is indicated by the arrows in (b) whose directions point to the texture boundaries, with darker arrows representing stronger texture gradient. After a few iterations the propagation reaches a stable state (c). The final image segmentation is shown in (d).

are used in developing a *texture thesaurus* for fast search and retrieval [7],[8]. A schematic diagram of our thesaurus based image browsing system is shown in Figure 1.

Salient aspects of this work are:

- A new image segmentation scheme for region based search [9].
- Learning similarity in the texture feature space [10].
- Multidimensional indexing and dimensionality reduction [11].

In the following we provide a brief description of each of these components. The objective is provide the reader with a very general overview of recent research and technical details can be found in the cited references.

2.1 Image Segmentation

Automated image segmentation is clearly a significant bottleneck in enhancing the retrieval performance. Although some of the existing systems have demonstrated a certain capability in extracting regions and providing a region-based search, their performances on large and diverse image collections have not been demonstrated. We believe that it is important to localize the image feature information. Region or object based search is more natural and intuitive than search using the whole image information. With automated segmentation as the primary objective, we have developed a robust segmentation scheme, called EdgeFlow, that has yielded very promising results on a diverse collection of a few thousand images [9].

The EdgeFlow scheme utilizes a simple predictive coding model to identify and integrate the change in visual cues such as color and texture at each pixel location. As a result of this computation, a flow vector which points in the direction of the closest image boundary is constructed. This edgeflow is iteratively propagated to its neighbor if the edgeflow of the corresponding neighbor points in a similar direction. The flow stops propagating if the corresponding neighbor has an opposite flow direction. In this case the two image locations have flow vectors pointing at each other indicating the presence of a discontinuity between them.

This EdgeFlow framework results in a *dynamic* boundary detection scheme. The flow direction gives the direction with the most information change in the image feature space. Since any of the image attributes such as color, texture, or their combination can be used to define the edgeflow, this provides a simple framework for integrating diverse visual cues for boundary detection. Figure 2 shows different stages of the segmentation algorithm on a small region in an aerial photograph. For computational reasons, the image texture is computed in blocks of 64 x 64 pixels. The initial texture flow vectors so computed are shown in Figure 2(b) and after convergence, in Figure 2(c). The final detected boundaries are shown in Figure 2(d).

2.2 Learning Similarity

In a typical database search, several top matching patterns are retrieved for a given query. These matches are rank-ordered based on their similarity to the query pattern. Ideally, a distance metric in the texture feature space should preserve the visual pattern similarity. Computing such a similarity is an important problem in content based image retrieval.

In order to improve the retrieval performance of the texture image features, we have proposed the use of a learning algorithm [10]. For this, a hybrid neural network algorithm is used to cluster texture patterns in the feature space. This algorithm contains two stages of training. The first stage performs an unsupervised learning using the Kohonen feature map to capture the underlying feature distribution. In the second stage, clusters are labelled using a winner-takes-all representation, and class boundaries are fine tuned using a learning vector quantization scheme. This results in a partitioning of the original feature space into clusters of visually similar patterns based on the class label information provided by human observers.

Once the network is trained, the search and retrieval process is performed in the following way:

- When a query pattern is presented to the system, the network first identifies a subspace of the original feature space which is more likely to contain visually similar patterns.
- The final retrievals are then computed using a simple Euclidean distance measure with the patterns which belong to the corresponding sub-space.

In addition to retrieving perceptually similar patterns, an additional advantage of this clustering approach is that it provides an efficient indexing tree to narrow down the search space. The cluster centers are then used to construct a visual texture image thesaurus, as explained below.

2.3 Texture Thesaurus

A texture thesaurus can be visualized as an image counterpart of the traditional thesaurus for text search. It contains a collection of codewords which represent visually similar clusters in the feature space. A subset of the airphotos was used as training data for the hybrid neural network algorithm described earlier to create the first level of indexing tree [8]. Within each subspace, a hierarchical vector quantization technique was used to further partition the space into many smaller clusters. The centroids of these clusters were used to form the codewords in the texture thesaurus, and the training image patterns of these centroids were used as icons to visualize the corresponding codewords. In the current implementation, the texture thesaurus is organized as a two-level indexing tree which contains 60 similarity classes and about 900 codewords. Figure 3 shows some examples of the visual codewords in the texture thesaurus designed for airphotos. Associations of the code words can be made to semantic concepts as well. This is being investigated in a related project [12].

When an airphoto is ingested into the database, texture features are extracted using 64x64 subpatterns. These are then grouped to form regions. These features are used to compare with the codewords in the thesaurus. Once the best match is identified, a two-way link between the image region and the corresponding codeword is created and stored as the image meta-data. During query time, the feature vector of the selected pattern is used to search for the best matching codeword, and by tracing back the links to it, all similar patterns in the database can be retrieved (Figure 4). Some examples of retrievals are shown in Figure 5 and Figure 6. An on-line web demo of this texture based retrieval can be found at http://vivaldi.ece.ucsb.edu/AirPhoto.

2.4 Dimensionality Reduction

Image retrieval using the texture thesaurus reduces the search complexity by providing a tree-structured indexing while preserving the similarity between patterns.

However, even greater improvements in retrieval efficiency are desirable. Because the database to be searched is large and the feature vectors are of high dimension, search complexity is still high. Promising new results suggest that non-linear PCA may be useful in reducing the dimension of the feature vectors without destroying too much of the information they contain. Preliminary experiments show that non-linear PCA can project 60 dimensional feature vectors to just 6 dimensions, while maintaining very high retrieval rates [11]. Thus an efficient retrieval system architecture might use neural nets to initially direct query vectors to subclasses as described above, and then apply a non-linear projection to a lower dimension before searching the subclass for the best matches.

The basic idea is to compute a projection map that maps the high dimensional feature vectors to a lower dimensional space subject to certain constraints. For example, the distances in the new space should approximate user provided perceptual distances between pairs of patterns.

In our experiments with this approach to dimensionality reduction, we chose a mapping that reduced the 60 dimensional features to a 6 dimensional vector. A training feature set is used to compute the parameters of the transformation. To evaluate the quality of the projection map, each of the database images that was not used in training was used as a query image. For each query, the 10 closest vectors from the database were retrieved. Ideally, the retrieved vectors should belong to the same image class as the query vector. We found that the average correct retrieval percentage considering only those images not used in training was 87%. For comparison, we also evaluated retrieval performance using the full 60 dimensional feature vectors; we found 90% correct retrieval. Thus a factor of 10 reduction in the dimension did not greatly reduce retrieval performance. For further comparison, a traditional linear PCA was used to project the vectors to 6 dimensions. For this linear projection we found only 28% retrieval. The results are summarized in Figure 7. Details can be found in [11].

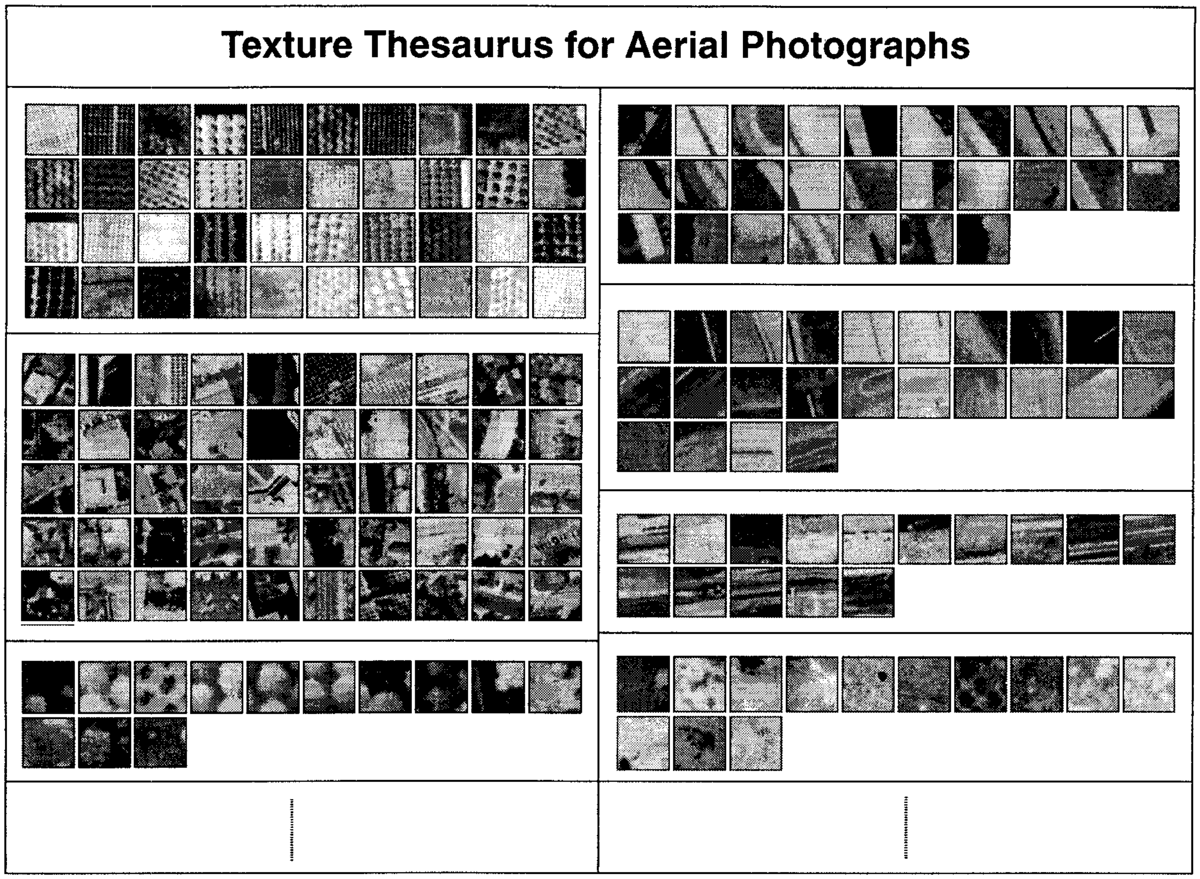


FIGURE 3. Examples of the codewords obtained for the aerial photographs. The patterns inside each block belong to the same class.

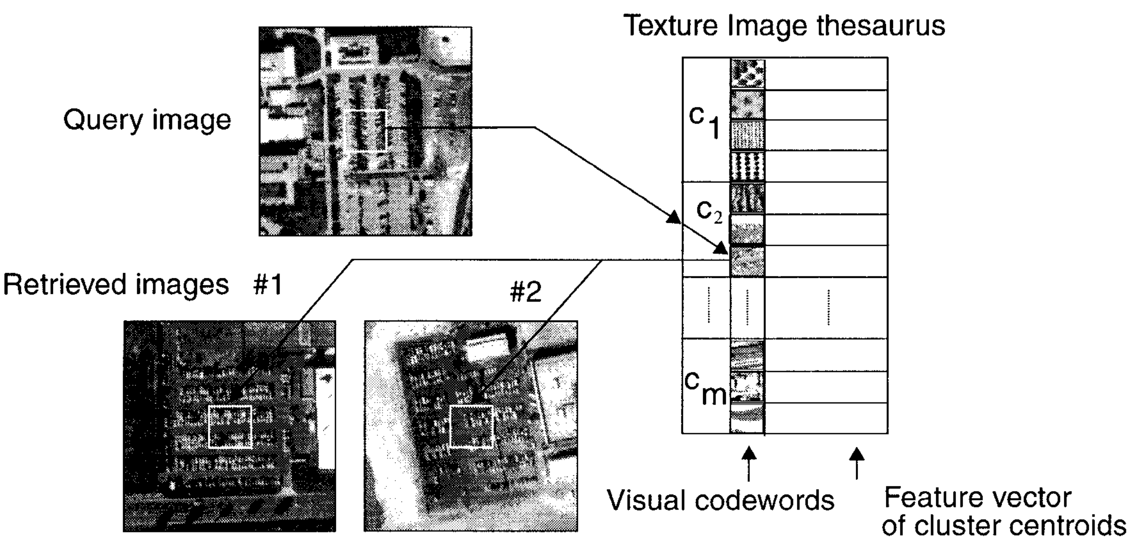


FIGURE 4. A texture image thesaurus for content-based image indexing. The image tiles shown here contain parking lots.

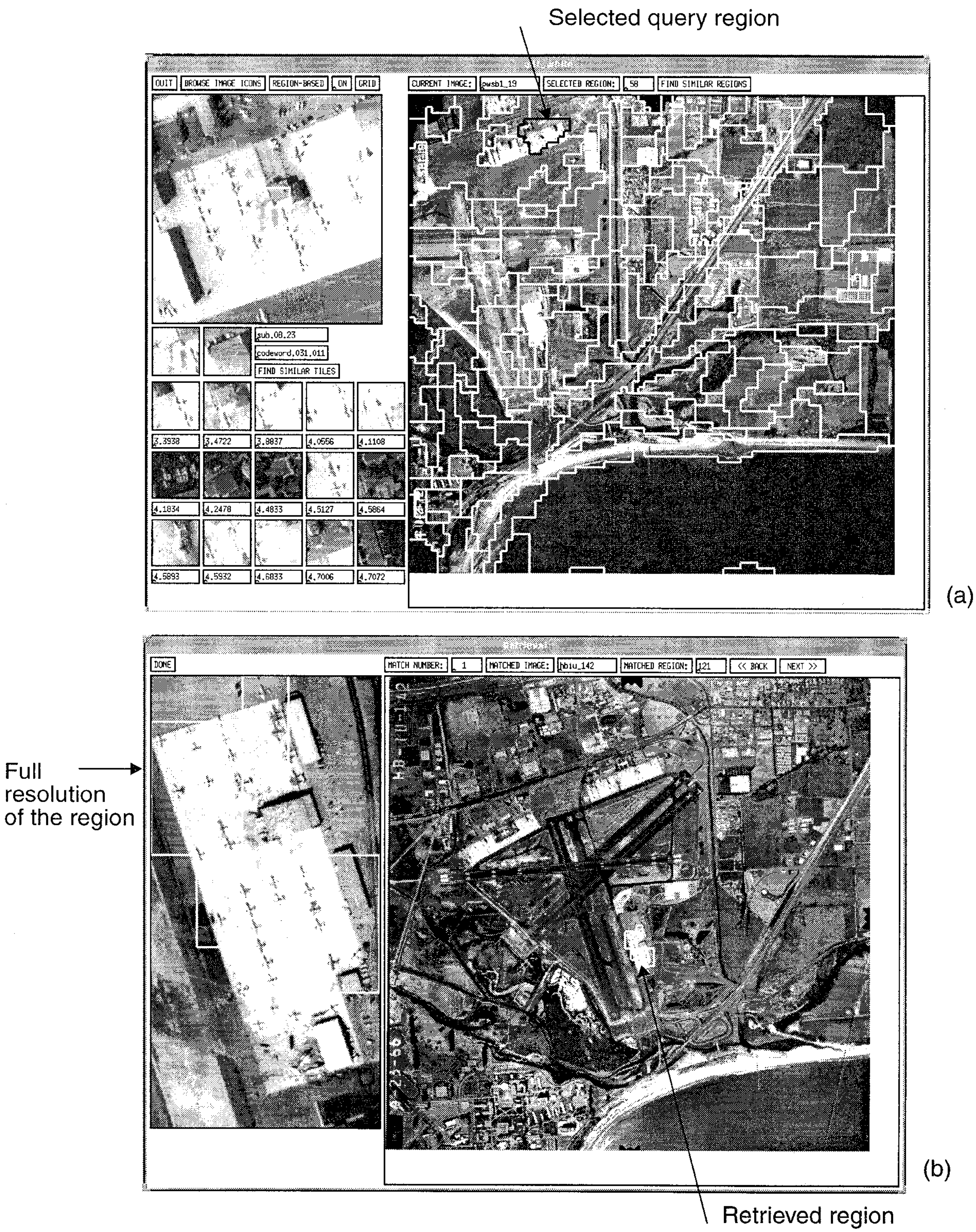


FIGURE 5. An example of the region-based search. (a) shows the down-sampled version of a segmented large airphoto (on the right), and higher resolution picture of the selected airport region (on the left). Some of the retrievals based on the 64x64 tiles are shown as well (left bottom). (b) shows a region based retrieval result using the airport region shown in (a). Both the query and retrieval patterns are from the airport areas.

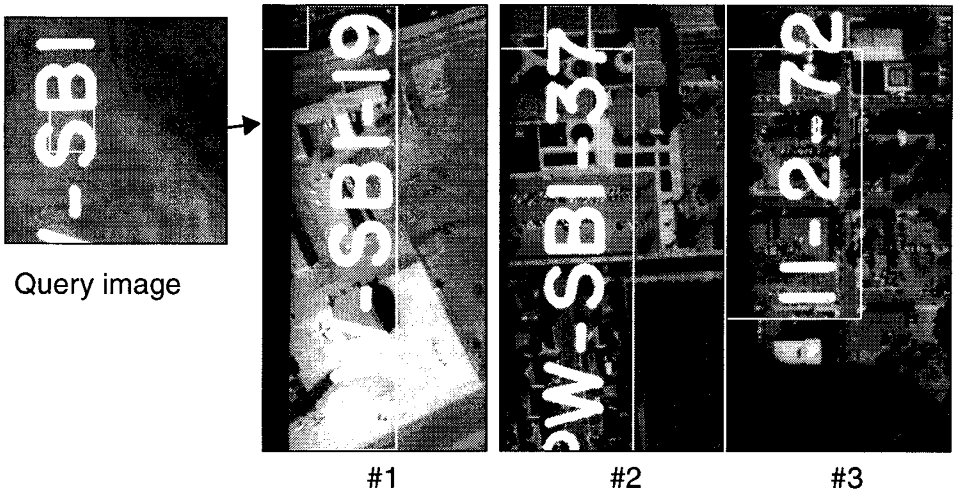


FIGURE 6. Shows another retrieval example. The query is from a region containing an image identification number.

Our long term goal is to construct a visual thesaurus for images/video where the thesaurus code-words are created at various levels of visual hierarchy by grouping primitives such as texture, color, shape, and motion. For complex structured patterns, these codewords take the form of a labelled graph, with the nodes in the graph representing primitive image attributes and the links the part relationships. The whole approach is hierarchical and can be extended to a more complex set of image attributes.

3. DIGITAL WATERMARKING

Intellectual property protection is another important issue for digital media content providers and in digital libraries. One approach to this problem is the use of digital watermarking. In digital watermarking, a signature is embedded into the original host data and data authentication can be done by checking for the presence of such signatures. For images and video, these signatures could be either visible (as in a transparent background) or invisible. The use of invisible signatures is of interest as one can distribute the data in its original form with little, if any, perceptual distortion.

In order to be effective, an invisible watermark should be secure, reliable, and resistant to common signal processing operations and intentional attacks. In our work on data embedding using signal processing techniques, we have focussed on hiding significantly larger amounts of signature data. This is in contrast to much of the related work in digital watermarking where the signatures are typically binary pseudo-random sequences. For example, we can embed signature images which are as much as 25% of the host image data. Embedding such large amounts of image and video data opens up other interesting applications in security and intelligence gathering, and offers an alternative to traditional encryption methods.

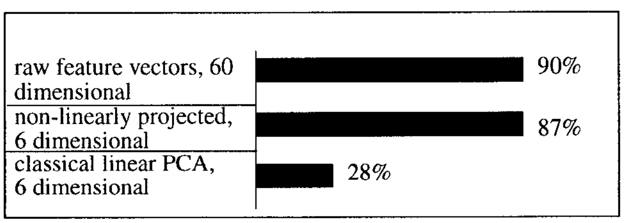


FIGURE 7. Comparison of retrieval rates.

The approach we are currently investigating is based upon well established techniques from channel coding using lattice structures. In this, both the host and signature image data are first wavelet transformed. The signature image coefficients are then quantized to a given number of levels. The host image coefficients are then grouped to form multidimensional vectors which are then perturbed by the channel codes corresponding to signature coefficients. After this embedding in the wavelet domain, the inverse transformation gives the watermarked image. Signature recovery follows by inverting the embedding procedure assuming that the host image is available. Details of this scheme is available in [13] (this proceedings).

Our preliminary experiments demonstrate that this type of embedding is robust to lossy image compression. Figure 8 shows two examples. The Alexandria project symbol is embedded in a aerial photograph. The two examples correspond to signature recovery under lossy wavelet compression and lossy JPEG compression. In general, good quality signature recovery is possible for up to 80% compression.

4. DISCUSSIONS

Large spatial databases such as satellite imagery and aerial photographs pose several challenging research problems. We have presented some promising results on region based retrieval using texture. Dimensionality reduction is going to be critical for large scale databases. Representation of spatial and spatio-temporal relationships is another

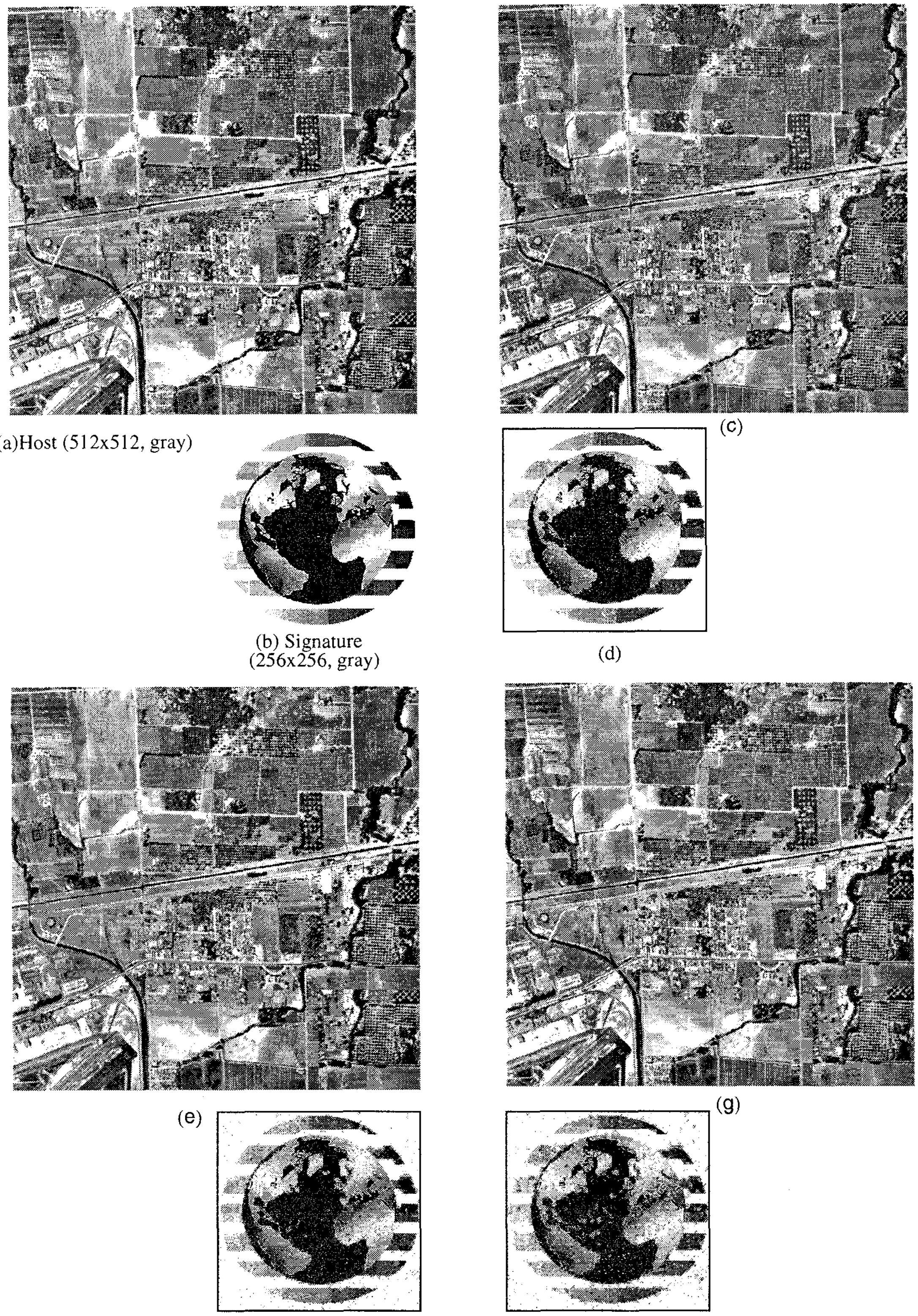


FIGURE 8: An example of image embedding. (a) an aerial photograph which is used as the host image, and (b) the Alexandria project symbol. (c) watermarked image after compressing by 75% using a wavelet transform based algorithm, and (d) is the recovered signature. (e), (g) watermarked image compressed by 75% and 85%, respectively, using JPEG. (f) recovered signature from (e), and (h) is recovered from (g).

important research problem. A visual thesaurus provides a conceptual framework for addressing many of these issues. A demonstration of an image thesaurus for airphoto browsing can be found on the web at *http://vivaldi.ece.ucsb.edu/AirPhoto*. Extensions of this work to include color and texture for natural photographs is also available on the web at *http://vivaldi.ece.ucsb.edu/Netra*.

ACKNOWLEDGMENTS. The research described here is the work of several people: Dr. Norbert Strobel contributed to the multiresolution representation schemes, Dr. Wei-Ying Ma developed the image texture thesaurus; Jong-Jin Chae and Debargha Mukherjee contributed to the digital watermarking research. Yining Deng helped in the preparation of this manuscript. This work is supported in part by NSF (award #IRI-9704785 and #IRI-9411330).

5. REFERENCES

[1] N. Strobel, *Multiresolution based storage, browsing and retrieval for digital image libraries*, Ph.D. Thesis, ECE Department, University of California, Santa Barbara, March 1998.

[2] W. Niblack et al., "The QBIC project: querying images by content using color, texture, and shape," *Proc. SPIE*, vol. 1908, pp. 173-187, February 1993.

[3] Pentland, A., Picard, R. W. & Sclaroff, S. (1994). Photobook: tools for content based manipulation of image databases. *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases* - II, vol. 2185 (pp. 34-47). San Jose, CA.

[4] J. R. Smith and S. F. Chang, "VisualSEEK: a fully automated content-based image query system," ACM Multimedia'96, boston, MA, Nov. 20, 1996.

[5] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. C. Jain, and C. Shu, "Virage image search engine: an open framework for image management," in Proc. of SPIE, Storage and Retrieval for Image and Video Databases IV, San Jose, CA, pp. 76-87, February 1996.

[6] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. PAMI*, vol. 18(8), pp. 837-842, August 1996.

[7] B. S. Manjunath and W. Y. Ma, "Browsing large satellite and aerial photographs," *Proc. third IEEE international conference on image processing*, ICIP 96, Vol. II, pp. 765-768, Laussnne, Switzerland, September 1996 (invited paper).

[8] W. Y. Ma and B. S. Manjunath, "A texture thesaurus for browsing large aerial photographs," to appear in the Journal of the American Society for Information Science, 1998.

[9] W.Y. Ma and B.S. Manjunath, "Edge flow: a framework of boundary detection and image segmentation", *Proc. of IEEE CVPR'97*, pp 744-749, 1997.

[10] W. Y. Ma and B. S. Manjunath, "Texture features and learning similarity," *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 425-430, June, 1996.

[11] M. Beatty and B. S. Manjunath, "Dimensionality reduction using multidimensional scaling for image search," *Proc. IEEE International Conference on Image Processing*, Santa Barbara, California, Vol. II, pp. 835-838, October 1997.

[12] H. Chen, Semantic Interoperability for geographic information systems, http://ai.bpa.arizona.edu/gis.

[13] J. J. Chae, D. Mukherjee, and B. S. Manjunath, "Robust data hiding using the D_4 lattice," *Proc. of the ADL'98 Conference*, May 1998, Santa Barbara, CA.

[14] W.Y. Ma, Y. Deng and B.S. Manjunath, "Tools for texture/color based search of images", *Proc. of SPIE, Human Vision and Electronic Imaging II,* vol. 3106, pp 496-507, 1997.

[15] W.Y. Ma and B.S. Manjunath, "NeTra: A toolbox for navigating large image databases", *Proc. of IEEE ICIP'97*, vol. 1, pp 568-571, 1997.