# ISSUES FOR IMAGE/VIDEO DIGITAL LIBRARIES

*B. S. Manjunath and Yining Deng*

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

## ABSTRACT

Creation of digital image and video libraries poses several interesting and challenging problems. New tools are needed for managing such multimedia content. These include methods to search, retrieve, and manipulate digital media by using the media content information and mechanisms to protect intellectual property rights. This paper outlines some of the recent advances in image processing as related to digital libraries in the context of the UCSB Alexandria Digital Library project.

## 1. INTRODUCTION

With the emergence of large multimedia databases, there is a growing need for new technologies for search, retrieval, and manipulation of the digital media. Digital media now pervades many aspects of day-to-day life, including television and broadcasting, medicine, engineering, and life sciences. There has been significant advances in the technology related to generating, storage, and transmission of large scale images, video, and audio data. Examples of potential application domains include geographic information systems, entertainment industry, and medical databases. The importance of these and other applications can be seen by the fact that there is now an organized effort to standardize some of the content description. The recently started MPEG-7 activity attempts to standardize description schemes and descriptors for audio and video data together with a way to specify these description schemes and descriptors.

However, in order to realize some of the benefits of digital media, a key bottleneck remains - that of data overload. There is a need to develop "intelligent agents" to filter out relevant information from large amounts of data so as to facilitate browsing and search based on content information.

The problem of data overload exists in many imaging domains such as satellite images (where it is expected that the earth observing systems will generate trillions of bytes of image data every day), aerial photographs, astronomical pictures, stock photo galleries, human face images, trademark symbols, etc. The proliferation of the world wide web adds a new dimension to this problem, making internet based search tools another potential application.

In addition to efficient access and manipulation, protecting the intellectual property rights is another important issue in digital libraries research. Our current work on digital watermarking is presented in Section 3.

## 2. IMAGE/VIDEO RETRIEVAL

In the past few years, several prototype systems have been developed for content based access to image/video databases. The QBIC system from IBM is perhaps the best known example [1]. Since then, several impressive demonstrations have been made in utilizing low level image features such as color, texture, and shape information to search through collections of images.

### 2.1 IMPORTANT ISSUES

**Segmentation/Region based Search.** While many of the existing work focus on extracting the global image features, we believe that it is important to localize the image feature information. Region or object based search is more natural and intuitive than whole image information. This requires segmentation schemes that identify salient regions in an image. Our main contribution to this is the development of a robust segmentation scheme, called EgdeFlow, that has yielded very promising results on a diverse collection of a few thousand images [2].

**Image features.** A compact and robust characterization of the local image features is necessary for search and retrieval. As mentioned above, image features such as texture, color, and shape are the obvious choice. In the past few years, we have developed algorithms for texture feature computations that have shown to be competitive for image retrieval applications [3]. However, much work is needed in building a semantic level description from these image features. Similarity metrics for comparing distances in the feature domain remains as an active research problem. Typical image features are in a high dimensional space and this is of concern when the database contains a large number (more than few hundred thousand) of images. Our current work on developing a visual thesaurus addresses some of these issues.

### 2.2 A VISUAL THESAURUS

Conceptually, the image thesaurus model can be visualized as an image counterpart of the traditional one for text search. At the time of data ingest image features of interest are computed and are then clustered to give a feature code book. Each codeword in the dictionary will have a corresponding iconic representation to help users visualize the code words. Associations can be made to semantic concepts as well. This scheme reduces the search complexity by providing a tree-structured indexing while preserving the similarity between patterns. Note that traditional indexing structures (such as B-tree or R-trees) do not generalize well to large dimensions such as a 60 component texture feature. At the query time, an example image or a sequence will be provided by the user. Image features computed from the query pattern will then be used to look-up from the table the closest matching codeword or a set of codewords.

Our long term goal is to construct a visual thesaurus for images/video where the thesaurus code-words are created at vari-

0-7803-4455-3/98/$10.00 © 1998 IEEE

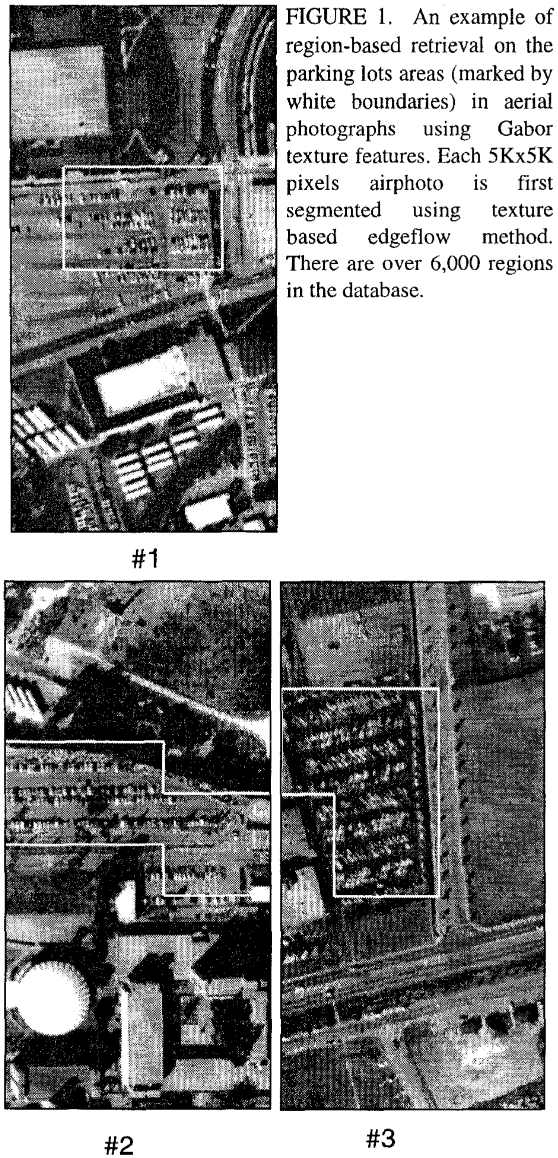


FIGURE 1. An example of region-based retrieval on the parking lots areas (marked by white boundaries) in aerial photographs using Gabor texture features. Each 5Kx5K pixels airphoto is first segmented using texture based edgeflow method. There are over 6,000 regions in the database.

ous 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 more complex set of image attributes.

We have developed preliminary versions of a visual thesaurus for two distinct databases: one containing aerial photographs and the other containing natural images from a Corel stock photo library.

**Aerial Photographs [4].** A preliminary version of this visual thesaurus has been implemented for an airphoto database using image texture features. Texture features can provide powerful low-level cues for object groupings and delineating functional areas in geographic databases. Using texture alone, one can search for a surprisingly diverse set of salient image areas which include parking lots, airport turmacs, building developments, vegetation patterns, and highways. Texture features are computed by filtering the image with a bank of Gabor filters [3]. A hybrid neural network algorithm is used to learn the visual similarity by clustering patterns in the feature space. A texture image thesaurus is created by combining similarity learning with a hierarchical vector quantization scheme. The texture thesaurus facilitates the indexing process while maintaining a good retrieval performance. An example of retrieving parking lots in airphotos is shown in Figure 1. This texture based search of airphotos is part of the ADL testbed that is being developed at UCSB (*http://www.alexandria.ucsb.edu*).

**NeTra: Region based search using color, texture, shape, and location [5,6].** As mentioned before, it is important that local image attributes be used for search and retrieval. Lack of robust image segmentation algorithms is one of the reasons why shape and other local image features have not been extensively used in image queries. We have recently developed a prototype system called NeTra (which means *eyes*, in Sanskrit) which offers region based search functionality. NeTra uses a new segmentation scheme that appears quite promising in the context of large image/video databases. This technique, which we call "edgeflow", utilizes a predictive coding model to identify and integrate the direction of change in color and texture at each image pixel location. Details of this scheme are presented in [2].

In NeTra, images are segmented into homogeneous regions and image attributes that represent each of these regions are then computed. We use color, texture, shape of the boundary, and location of the region as features for search and indexing. Texture features are similar to the ones used for the aerial photographs. In NeTra, each image region color is represented by a subset of colors from a color codebook. The codebook is constructed using a vector quantization scheme using a training dataset. A color table is constructed for efficient indexing of the region colors.

A query interface is provided that allows users to form visual queries of the type "retrieve all images that contain regions that have the color of object A, texture of object B, shape of object C, and lie in the upper one-third of the image", where the individual objects could be regions belonging to different images. A Java

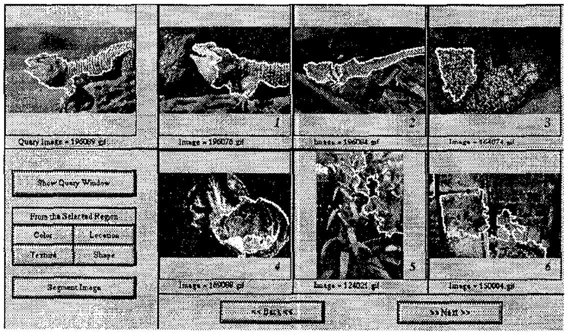


FIGURE 2. An example of region based image retrieval using color and texture.

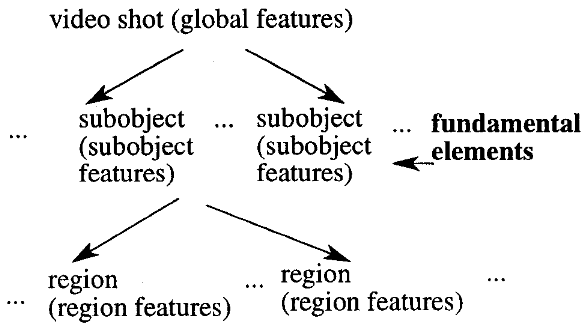

video shot (global features)
...
subobject (subobject features)
...
subobject (subobject features)
...
fundamental elements
...
region (region features)
...
region (region features)
...

FIGURE 3. Structure of low-level content description. Subobjects are the fundamental elements of this representation level.

based web implementation of NeTra is available at *http://vivaldi.ece.ucsb.edu/NeTra*. Figure 2 shows an example of region based retrieval using color and texture.

## 2.3 VIDEO SEARCH [7]

In extending our approach to video data, we are working on a new spatio-temporal segmentation and object-tracking scheme, and a hierarchical object-based video representation model. The spatio-temporal segmentation scheme combines the spatial color/texture image segmentation (using the edgeflow method) and affine motion estimation techniques. The computed features are used to form the low-level content description of our video representation model (Figure 3).

The basic idea is to develop techniques that can segment and track meaningful physical objects so that these objects can be indexed into a video object database. However, in reality, this is not possible using just the low level image features. As a compromise, we define as "sub-objects" those collections of coherent regions which could be tracked for a certain minimum number of frames in a video sequence, all of which have similar motion vectors.

Similar to MPEG coding, we form a group of frames to begin the analysis. One of the frames in the group is chosen for spatial segmentation. The regions obtained as a result of this segmentation are the ones that are tracked over the entire group. An example of region tracking over several consecutive groups of frames is illustrated in Figure 4.

Each video shot is thus composed of a set of subobjects. A video shot now can be characterized by its subobject information, and the spatial and temporal relations between these subobjects. Such a representation allows the user to track regions in a video sequence and search for regions with similar color, texture, shape, motion pattern, location, or size in the database.

# 3. DIGITAL WATERMARKING

As multimedia data becomes wide spread, such as on the internet, there is a need to address issues related to the security and protection of such data. While access restriction can be provided using electronic keys, they do not offer protection against further (illegal) distribution of such data. Digital watermarking is one approach to managing this problem by encoding user or other copyright information directly in the data while not restricting access. Watermarking of image data could be visible, for example, a background transparent signature, or could be perceptually invisible. A visible watermark acts like a deterrent but may not be acceptable to users in some contexts. In order to be effective, an invisible watermark should be secure, reliable, and resistant to common signal processing operations and intentional attacks. Recovering the signature from the watermarked media could be used to identify the rightful owners and the intended recipients as well as to authenticate the data.

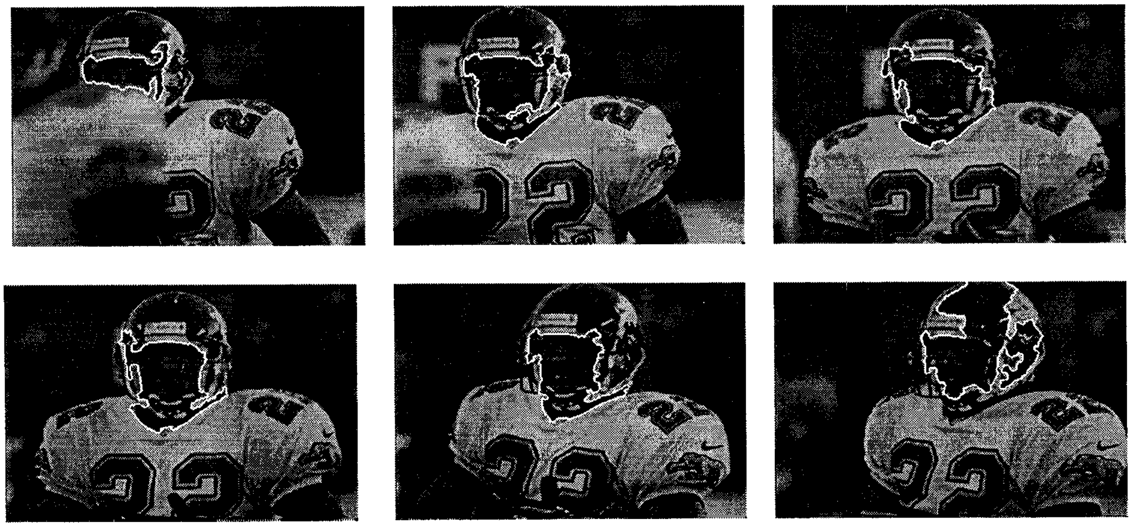
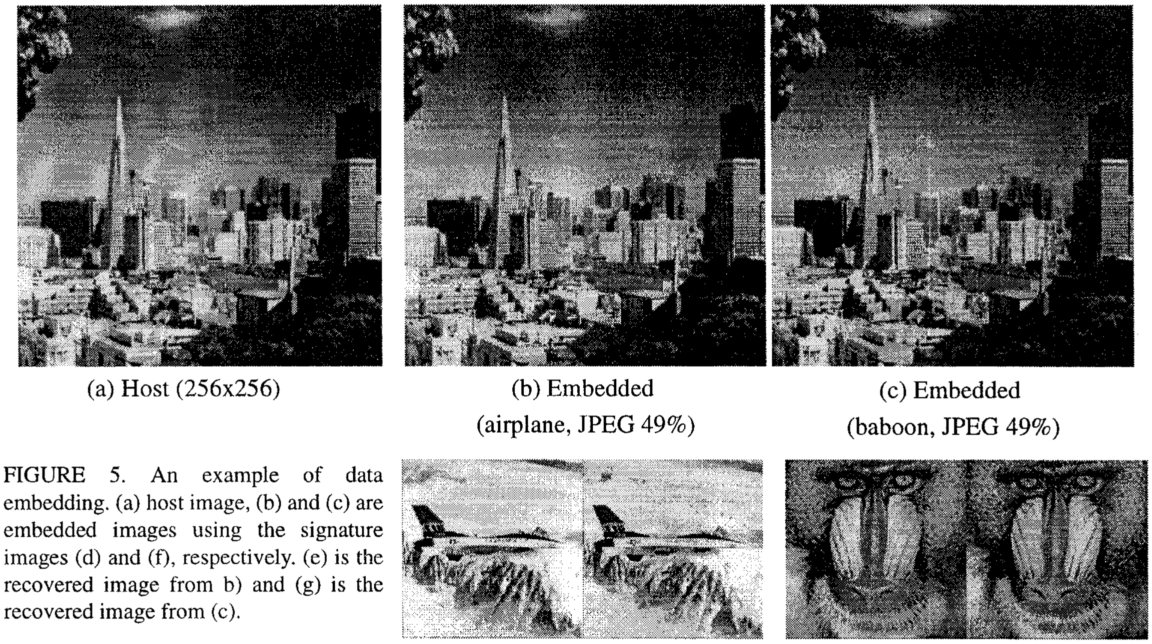


FIGURE 4. An identified subobject in a set of 6 consecutive groups of frames.



(a) Host (256x256)

(b) Embedded (airplane, JPEG 49%)

(c) Embedded (baboon, JPEG 49%)

(d)Signature (airplane)

(e) recovered

(f)Signature (baboon)

(g) recovered

FIGURE 5. An example of data embedding. (a) host image, (b) and (c) are embedded images using the signature images (d) and (f), respectively. (e) is the recovered image from b) and (g) is the recovered image from (c).

If the original host image is available, the operations of data injection and retrieval are, in fact, very similar to the channel coding and decoding operations in a typical digital communication system. In [8] we present an approach to signature embedding using noise-resilient channel codes based on the $D_4$ lattice. In watermarking in the transform domain, the original host data is transformed, and the transformed coefficients are perturbed by a small amount in one of several possible ways in order to represent the signature data. When the watermarked image is compressed or modified by other image processing operations, noise is added to the already perturbed coefficients. The retrieval operation subtracts the received coefficients from the original ones to obtain the noisy perturbations. The true perturbations that represent the injected data are then estimated from the noisy data as best as possible. Figure 5 shows an example of signature embedding in a wavelet transform domain using lattice structures. Our experimental results show that the watermarked image is transparent to embedding for large amounts of signature data, and the quality of the extracted signature is high even when the watermarked image is subjected to up to 85% JPEG lossy compression.

## 4. DISCUSSIONS

Multimedia databases pose several challenging research problems. Much progress has been made in the past few years in content based access using low level image/video features. However, several fundamental problems still remain: these include fully automated object segmentation, computing similarities in the feature space, encoding spatial and spatio-temporal relationships, multidimensional indexing and efficient multimodal retrieval that makes use of cues from several sources (such as audio, motion, color, texture, etc.). Success of object/semantic level representation and retrieval depends, to a large extent, on finding good solutions to these above problems.

**ACKNOWLEDGMENTS.** The authors would like to thank Jong-Jin Chae and Wei-Ying Ma for their help in preparing this paper. This work is supported in part by a grant from NSF (award #IRI-9704785).

## 5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] B. S. Manjunath and W. Y. Ma, "Browsing large satellite and aerial photographs," *Proc. IEEE ICIP'96*, vol. 2, pp. 765-768.

[5] 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.

[6] 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.

[7] Y. Deng and B.S. Manjunath, "NeTra-V: towards an object-based video representation", *Proc. of SPIE, Storage and Retrieval for Image and Video Databases VI*, vol. 3312, pp 202-213, 1998.

[8] 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.