# IMAGE RETRIEVAL: PAST, PRESENT, AND FUTURE

*Yong Rui and Thomas S. Huang*

Dept. of ECE & Beckman Institute
University of Illinois at Urbana-Champaign
Urbana, IL 61801
{yrui,huang}@ifp.uiuc.edu

*Shih-Fu Chang*

Dept. of EE & New Media Technology Center
Columbia University
New York, NY 10027
sfchang@ee.columbia.edu

## ABSTRACT

This paper provides a comprehensive survey of the technical achievements in the research area of Image Retrieval, especially Content-Based Image Retrieval, an area so active and prosperous in the past few years. The survey includes 100+ papers covering the research aspects of image feature representation and extraction, multi-dimensional indexing, and system design, three of the fundamental bases of Content-Based Image Retrieval. Furthermore, based on the state-of-the-art technology available now and the demand from real-world applications, open research issues are identified, and future promising research directions are suggested.

## 1. INTRODUCTION

Recent years have seen a rapid increase of the size of digital image collections. Everyday, both military and civilian equipment generates giga-bytes of images. Huge amount of information is out there. However, we can not access to or make use of the information unless it is organized so as to allow efficient browsing, searching and retrieval. Image Retrieval has been a very active research area since the 1970's, with the thrust from two major research communities, Database Management and Computer Vision. These two research communities study Image Retrieval from different angles, one being text-based and the other visual-based.

The text-based Image Retrieval can be traced back to the late 1970's. A very popular framework of Image Retrieval then was to first annotate the images by text and then use text-based Database Management Systems (DBMS) to perform Image Retrieval. Representatives of this approach are [20, 21, 24, 26]. Two comprehensive surveys on this topic are [157, 25]. Many advances, such as Data Modeling, Multi-Dimensional Indexing, Query Evaluation, etc., have been made along this research direction. However, there exist two major difficulties, especially when the size of image collections is large (tens or hundreds of thousands). One is the vast amount of labor required in manual image annotation. The other difficulty, which is more essential, results from the rich content in the images and the subjectivity of human perception. That is, for the same image content, different people may perceive it differently. The perception subjectivity and annotation impreciseness may cause unrecoverable mismatches in the later retrieval processes.

In the early 90's, because of the emergence of large-scale image collections, the two difficulties faced by the manual annotation approach became more and more acute. To overcome these difficulties, content-based Image Retrieval was proposed. That is, instead of being manually annotated by text-based keywords, images would be indexed by their own visual content, such as color, texture, etc. Since then, many techniques in this research direction have been developed and many Image Retrieval systems, both research and commercial, have been built. The advances in this research direction are mainly contributed by the Computer Vision community. Many special issues of leading journals have been dedicated to this topic [55, 105, 97, 4, 130].

This approach has established a general framework of Image Retrieval from a new perspective. However, there are still many open research issues to be solved before such retrieval systems can be put into practice. Regarding Content-Based Image Retrieval, we feel there is a need to survey what has been achieved in the past few years and what are the potential research directions which can lead to compelling applications.

Since excellent surveys for text-based Image Retrieval paradigm already exist [157, 25], in this paper

Work of the first author was supported in part by a CSE Fellowship, College of Eng., UIUC; work of the second author was supported in part by ARL Cooperative Agreement No. DAAL01-96-2-0003; and work of the third author was supported in part by the National Science Foundation under a CAREER award (IRI-9501266), a STIMULATE award (IRI-96-19124), and industrial sponsors of Columbia's ADVENT project.

we will devote our effort primarily to the Content-based Image Retrieval paradigm. There are three fundamental bases for Content-Based Image Retrieval, i.e. Visual Feature Extraction, Multi-Dimensional Indexing, and Retrieval System Design. The remainder of this paper is organized as follows. In Section 2, we review various visual features and their corresponding representation and matching techniques. To facilitate fast search in large-scale image collections, effective indexing techniques need to be explored. Section 3 evaluates various such techniques, including dimension reduction and multi-dimensional indexing. State-of-the-art commercial and research systems and their distinct characteristics are described in Section 4. Based on the current situation and what is demanded from real-world applications, promising future research directions and suggested approaches are presented in Section 5. Section 6 gives some concluding remarks.

## 2. FEATURE EXTRACTION

Feature (content) extraction is the basis of content-based Image Retrieval. In broad sense, features may include both text-based features (keywords, annotations, etc.) and visual features (color, texture, shape, faces, etc.). However, since there already exist rich literatures on text-based feature extraction in the DBMS and Information Retrieval research communities, we will confine ourselves in the techniques of visual feature extraction. Within the visual feature scope, the features can be further classified as general features and domain-specific features. The former include color, texture and shape features while the latter is application dependent and may include, for example, human faces and finger prints. The domain-specific features are better covered in Pattern Recognition literatures and may involve in much domain knowledge which we will not have enough space to cover in this paper. Therefore, the remaining of the section will concentrate on the general features which can be used in most applications.

Because of perception subjectivity, there does not exist a single best presentation for a given feature. As we will see soon, for any given feature, there exist multiple representations which characterize the feature from different perspectives.

### 2.1. Color

Color feature is one of the most widely used visual features in Image Retrieval. It is relatively robust to background complication and independent of image size and orientation. Some representative studies of color perception and color spaces can be found in [90, 95, 163].

In Image Retrieval, Color Histogram is the most commonly used color feature representation. Statistically, it denotes the joint probability of the intensities of the three color channels. Swain and Ballard proposed Histogram Intersection, a $L_1$ metric, as the similarity measure for the Color Histogram [152]. To take into account the similarities between similar but not identical colors, Ioka [68] and Niblack et al. [99] introduced a $L_2$-related metric in comparing the histograms. Furthermore, considering that most Color Histograms are very sparse and thus sensitive to noise, Stricker and Orengo proposed to use the cumulated Color Histogram. Their research results demonstrated the advantages of the proposed approach over the conventional Color Histogram approach [151].

Besides Color Histogram, several other color feature representations have been applied in Image Retrieval, including Color Moments and Color Sets. To overcome the quantization effects as in Color Histogram, Stricker and Orengo proposed to use Color Moments approach [151]. The mathematical foundation of this approach is that any color distribution can be characterized by the its moments. Furthermore, since most of the information is concentrated on the low-order moments, only the first moment (mean), and the second and third central moments (variance and skewness) were extracted as the color feature representation. Weighted Euclidean distance was used to calculate the color similarity.

To facilitate fast search over large-scale image collections, Smith and Chang proposed Color Sets as an approximation to Color Histogram [139, 140]. They first transformed the (R,G,B) color space into a perceptually uniform space, such as HSV, and then quantized the transformed color space into $M$ bins. A Color Set is defined as a selection of the colors from the quantized color space. Because Color Set feature vectors were binary, a binary search tree was constructed to allow fast search. The relationship between the proposed Color Sets and the conventional Color Histogram was further discussed [139, 140].

### 2.2. Texture

Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity [141]. It is an innate property of virtually all surfaces, including clouds, trees, bricks, hair, fabric, etc. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment [59]. Because of its importance and usefulness in Pattern Recognition and Computer Vision, there existed rich research results in the past three decades. Now, it further finds its way in Image Retrieval. More and more research achievements are being added to it.

In the early 70's, Haralick et al. proposed the co-occurrence matrix representation of texture feature [59]. This approach explored the gray level spatial dependence of texture. It first constructed a co-occurrence matrix based on the orientation and distance between image pixels and then extracted meaningful statistics from the matrix as the texture representation. Many other researchers followed the same line and further proposed enhanced versions. For example, Gotlieb and Kreyszig studied the statistics originally proposed in [59] and experimentally found out that *contrast*, *inverse deference moment* and *entropy* had the biggest discriminatory power [52].

Motivated by the psychological studies in human visual perception of texture, Tamura et al. explored the texture representation from a different angle [156]. They developed computational approximations to the visual texture properties found to be important in psychology studies. The six visual texture properties were *coarseness, contrast, directionality, linelikeness, regularity*, and *roughness*. One major distinction between the Tamura texture representation and the co-occurrence matrix representation is that all the texture properties in Tamura representation are visually meaningful whereas some of the texture properties used in co-occurrence matrix representation may not (for example, entropy). This characteristic makes the Tamura texture representation very attractive in Image Retrieval, as it can provide a more friendly user interface. The QBIC system [44] and MARS system [64, 102] further improved this texture representation.

In early 90's, after Wavelet transform was introduced and its theoretical framework established, many researchers began to study the use of Wavelet transform in texture representation [138, 31, 73, 54, 72, 159]. In [138, 141], Smith and Chang used the statistics (mean and variance) extracted from the Wavelet subbands as the texture representation. This approach achieved over 90% accuracy on the 112 Brodatz texture images. To explore the middle-band characteristics, tree-structured Wavelet transform was used by Chang and Kuo in [31] to further improve the classification accuracy. Wavelet transform was also combined with other techniques to achieve better performance. Gross et al. used Wavelet transform, together with KL expansion and Kohonen maps, to perform texture analysis in [54]. Thyagarajan et al. [159] and Kundu et al. [72] combined Wavelet transform with co-occurrence matrix to take advantage of both the statistics based and transform based texture analysis.

There also existed quite a few review papers in this area. An early review paper, by Weszka et al., compared the texture classification performance of Fourier power spectrum, second-order gray level statistics (co-occurrence matrix), and first-order statistics of gray level differences [165]. They tested the three methods on two sets of terrain samples and concluded that Fourier method performed poorly while the other two were comparable. In [100], Ohanian and Dubes compared and evaluated four types of texture representations, namely Markov Random Field representation [40], multi-channel filtering representation, fractal based representation [107], and co-occurrence representation. They tested the four texture representations on four test sets, with two being synthetic (fractal and Gaussian Markov Random Field) and two being natural (leather and painted surfaces). They found out that co-occurrence matrix representation performed best in their test sets. In a more recent paper [82], Ma and Manjunath evaluated the texture image annotation by various Wavelet transform representations, including orthogonal and bi-orthogonal Wavelet transforms, tree-structured Wavelet transform, and Gabor wavelet transform. They found out that Gabor transform was the best among the tested candidates, which matched the Human vision study results [141].

### 2.3. Shape

In Image Retrieval, depending on the applications, some require the shape representation to be invariant to translation, rotation, and scaling; while others do not. Obviously, if a representation satisfies the former requirement, it will satisfy the latter as well. Therefore, in the following we will focus on shape representations that are transformation invariant.

In general, the shape representations can be divided into two categories, boundary-based and region-based. The former uses only the outer boundary of the shape while the latter uses the entire shape region [125]. The most successful representatives for these two categories are Fourier Descriptor and Moment Invariants.

The main idea of Fourier Descriptor is to use the Fourier transformed boundary as the shape feature. Some early work can be found in [170, 108]. To take into account the digitization noise in the image domain, Rui et al. proposed a modified Fourier Descriptor which is both robust to noise and invariant to geometric transformations [125].

The main idea of Moment Invariants is to use region-based moments, which are invariant to transformations, as the shape feature. In [62], Hu identified seven such moments. Based on his work, many improved versions emerged. In [169], based on the discrete version of Green's theorem, Yang and Albregtsen proposed a fast method of computing moments in binary images. Motivated by the fact that most useful invariants were found by extensive experience and trial-and-error, Kapur et al. developed algorithms to systematically generate and search for a given geometry's invariants [71]. Realizing that most researchers did not consider what happened to the invariants after image digitization, Gross and Latecki developed an approach which preserved the qualitative differential geometry of the object boundary, even after an image was digitized [71]. In [36, 75], a framework of algebraic curves and invariants is proposed to represent complex objects in cluttered scene by parts or patches. Polynomial fitting is done to represent local geometric information, from which geometric invariants are used in object matching and recognition.

Some recent work in shape representation and matching include Finite Element Method (FEM) [106], Turning Function [9], and Wavelet Descriptor [35]. FEM defines a stiffness matrix, which describes how each point on the object is connected to other points. The eigenvectors of the stiffness matrix are called modes and span a feature space. All the shapes are first mapped into this space and similarity is then computed based on the eigenvalues. Along a similar line of Fourier Descriptor, Arkin et al. developed a Turning Function based method for comparing both convex and concave polygons [9]. In [35], Chuang and Kuo used Wavelet transform to describe object shape. It embraced the desirable properties such as multiresolution representation, invariance, uniqueness, stability, and spatial localization. For shape matching, chamfer matching attracted much research attention. Barrow et al. first proposed the Chamfer matching technique, which compared two collections of shape fragments at a cost proportional to linear dimension, rather than area [12]. In [15], to speed up the Chamfer matching process, Borgerfos proposed a hierarchical Chamfer matching algorithm. The matching was done at different resolutions, from coarse to fine.

Some recent review papers in shape representations are [77, 93]. In [77], Li and Ma showed that the Geometric Moments method (region-based) and the Fourier Descriptor (boundary-based) were related by a simple linear transformation. In [93], Babu et al. compared the performance of boundary based representations (chain code, Fourier Descriptor, UNL Fourier Descriptor), region based representations (Moment Invariants, Zernike moments, pseudo-Zernike moments), and combined representations (Moment Invariants & Fourier Descriptor, Moment Invariants & UNL Fourier Descriptor). Their experiments showed that the combined representations outperformed the simple representations.

In addition to 2D shape representations, there were many methods developed for 3D shape representations. In [161], Wallace and Wintz presented a technique for normalizing Fourier Descriptors which retained all shape information, and was computationally efficient. They also took advantage of an interpolation property of Fourier Descriptor which resulted in efficient representation of 3D shapes. In [160], Wallace and Mitchell proposed to use a hybrid structural/statistical local shape analysis algorithm for 3D shape representation. Further, Taubin proposed to use a set of Algebraic Moment Invariants to represent both 2D and 3D shapes [158], which greatly reduced the computation required for shape matching.

## 2.4. Color Layout

Although the global color feature is simple to calculate and can provide reasonable discriminating power in Image Retrieval, it tends to give too many false positives when the image collection is large. Many research results suggested that using color layout (both color feature and spatial relations) is a better solution to Image Retrieval. To extend the global color feature to a local one, a natural approach is to divide the whole image into sub-blocks and extract color features from each of the sub-blocks [46, 34]. A variation of this approach is the quad-tree based color layout approach [80], where the entire image was split into quad-tree structure and each tree branch had it own histogram to describe its color content. Although conceptually simple, this regular-subblock based approach can not provide accurate local color information and is computation- and storage- expensive. A more sophisticated approach is to segment the image into regions with salient color features by Color Set Back-projection, and then store the position and Color Set feature of each region to support later queries [139]. The advantage of this approach is its accuracy while the disadvantage is the general difficult problem of reliable image segmentation.

To achieve a good trade-off between the above two approaches, several other color layout representations were proposed. In [116], Rickman and Stonham proposed a color tuple histogram approach. They first constructed a code book which described every possible combination of coarsely quantized color hues that might be encountered within local regions in an image. Then a histogram based on quantized hues was constructed as the local color feature. In [150], Stricker and Dimai extracted the first three color moments from five predefined partially overlapping fuzzy regions. The usage of the overlapping region made their approach relatively insensitive to small regions transformations.

In [104], Pass et al. classified each pixel of a particular color as either coherent or incoherent, based on whether or not it is part of large similarly-colored region. By using this approach, widely scattered pixels were distinguished from clustered pixels and thus improving the representation of local color features. In [63], Huang et al. proposed a color correlogram based color layout representation. They first constructed a color cooccurrence matrix and then used the auto-correlogram and correlogram as the similarity measures. Their experimental results showed that this approach was more robust than the conventional Color Histogram approach in terms of retrieval accuracy [63].

Along the same line of Color Layout feature, the layout of texture and other visual features can also be constructed to facilitate more advanced Image Retrieval.

### 2.5. Segmentation

Segmentation is very important to Image Retrieval. Both the shape feature and the layout feature depend on good segmentation. In this subsection we will describe some existing segmentation techniques used in both Computer Vision and Image Retrieval.

In [81], Lybanon et al. researched morphological operation (opening and closing) based approach in image segmentation. They tested their approach in various types of images, including optical astronomical images, infrared ocean images, and magnetograms. While this approach was effective in dealing with the above scientific image types, its performance needs to be further evaluated for more complex natural scene images. In [58], Hansen and Higgins exploited the individual strengths of watershed analysis and relaxation labeling. Since fast algorithm exists for watershed, they first used watershed to subdivide an image into catchmen basins. They then used relaxation labeling to refine and update the classification of catchmen basins initially obtained from watershed to take advantages of relaxation labeling's robustness to noise. In [78], Li et al. proposed a fuzzy entropy based segmentation approach. This approach is based on the fact that local entropy maxima correspond to the uncertainties among various regions in the image. This approach was very effective for images whose histogram do not have clear peaks and valleys. Other segmentation techniques based on Delaunay triangulation, Fractals, and Edge Flow can be found in [50, 134, 85].

All the above mentioned algorithms are automatic. A major advantage of this type of segmentation algorithms is that it can extract boundaries from large number of images without occupying the user's time and effort. However, in an unconstrained domain, for non-preconditioned images, the automatic segmentation is not always reliable. What an algorithm can segment in this case is only regions, but not objects. To obtain high-level objects, which is desirable in Image Retrieval, human assistance is needed.

In [128], Samadani and Han proposed a computer-assisted boundary extraction approach, which combined manual inputs from the user with the image edges generated by the computer. In [41], Daneels et al. developed an improved method of active contours. Based on the user's input, the algorithm first used a greedy procedure to provide fast initial convergence. Secondly, the outline was refined by using dynamic programming. In [124], Rui et al. proposed a segmentation algorithm based on clustering and grouping in spatial-color-texture space. The user defines where the attractor (object of interest) is, and the algorithm groups regions into meaningful objects.

Last comment worth mentioning in segmentation is that the requirements of segmentation accuracy are quite different for shape features and layout features. For the former, accurate segmentation is highly desirable while for the latter, a coarse segmentation may suffice.

### 2.6. Summary

As we can see from the above descriptions, many visual features have been explored, both previously in Computer Vision applications and currently in Image Retrieval applications. For each visual feature, there exist multiple representations, which model the human perception of that feature from different perspectives.

What features and representations should be used in Image Retrieval is application dependent. There is a need of developing an image content description (model) to organize the features. The features should not only be just associated with the images, but also should be invoked at the right place and right time, whenever they are needed to assist retrieval. Such a research effort has taken place. MPEG has started a new work item called MPEG-7, whose formal name is Multimedia Content Description Interface [2, 1, 5, 120]. It will specify a standard set of descriptors (features) that can be used to describe various types of multimedia information. The descriptions shall be associated with the content itself, to allow fast and efficient searching for information of a user's need [2].

## 3. HIGH DIMENSIONAL INDEXING

To make the content-based Image Retrieval truly scalable to large size image collections, efficient multi-dimensional indexing techniques need to be explored. There are two main challenges in such an exploration for Image Retrieval:

- High dimensionality
  The dimensionality of the feature vectors is normally of the order of $10^2$.

- Non-Euclidean similarity measure
  Since Euclidean measure may not effectively simulate human perception of a certain visual content, various other similarity measures, such as Histogram Intersection, Cosine, Correlation, etc., need to be supported.

Towards solving these problems, one promising approach is to first perform dimension reduction and then use appropriate multi-dimensional indexing techniques, which are capable of supporting Non-Euclidean similarity measures.

### 3.1. Dimension Reduction

Even though the dimension of the feature vectors in Image Retrieval is normally very high, the *embedded dimension* is much lower [166, 167]. Before we utilize any indexing technique, it is beneficial to first perform dimension reduction. There are at least two approaches appeared in the literature, i.e. Karhunen-Loeve Transform (KLT) and column-wise clustering.

KLT and its variation in face recognition, eigenimage, and its variation in information analysis, Principal Component Analysis (PCA), have been studied by researchers in performing dimension reduction. In [98], Ng and Sedighian followed the eigenimage approach to carry out the dimension reduction, and in [47] Faloutsos and Lin proposed a fast approximation to KLT to perform the dimension reduction. Experimental results from their research showed that most real data set (visual feature vectors) can be considerably reduced in dimension without significant degradation in retrieval quality [98, 47, 167]. Recently, Chandrasekaran et al. developed a low-rank Singular Value Decomposition (SVD) update algorithm which was efficient and numerically stable in performing KLT [19]. Considering that the Image Retrieval system is a dynamic system and new images are continuously added to the image collection, a dynamic update of indexing structure is indispensably needed. This algorithm provides such a tool.

In addition to KLT, clustering is another powerful tool in performing dimension reduction. Clustering technique is used in various disciplines such as Pattern Recognition [43], Speech Analysis [115], and Information Retrieval [127], etc. Normally it is used to cluster similar objects (patterns, signals, and documents) together to perform recognition or grouping. This type of clustering is called row-wise clustering. However, clustering can also be used in column-wise to reduce the dimensionality of the feature space [127]. Experiments show that this is a simple and effective approach.

One thing worth pointing is that blind dimension reduction can be dangerous, since information can be lost if the reduction is below the embedded dimension. To avoid blind dimension reduction, a post-verification stage is needed. Among different approaches, Fisher's discriminant analysis can provide useful guidance [138].

### 3.2. Multi-dimensional Indexing Techniques

After we identify the *embedded dimension* of the feature vectors, we need to select appropriate multi-dimensional indexing algorithms to index the reduced but still high dimensional feature vectors. There are three major research communities contributing in this area, i.e. Computational Geometry, Database Management, and Pattern Recognition. The existing popular multi-dimensional indexing techniques include Bucketing algorithm, k-d tree, priority k-d tree [167], quad-tree, K-D-B tree, hB-tree, R-tree and its variants $R^+$-tree and $R^*$-tree [57, 132, 53, 13, 118]. In addition to the above approaches, clustering and Neural Nets, widely used in Pattern Recognition, are also promising indexing techniques [43, 171].

The history of multi-dimensional indexing techniques can be traced back to middle 1970's, when Cell methods, quad-tree and k-d tree were first introduced. However, their performance were far from satisfactory. Pushed by then urgent demand of spatial indexing from GIS and CAD systems, Guttman proposed the R-tree indexing structure in 1984 [57]. Based on his work, many other variants of R-tree were developed. Sellis et al. proposed $R^+$ tree in [132]. Greene proposed her variant of R-tree in [53]. In 1990, Beckman and Kriegel proposed the best dynamic R-tree variant, $R^*$ tree [13]. However, even for $R^*$-tree, it was not scalable to dimensions higher than 20 [46].

Very good review and comparison of various indexing techniques in Image Retrieval can be found in [167, 98]. The research goal of White and Jain in [167] was to provide general purpose and domain independent indexing algorithms. Motivated by k-d tree and R-tree, they proposed VAM k-d tree and VAMSplit R-tree. Experimentally they found that the VAMSplit R-tree provided the best performance, but the trade-off is the loss of dynamic nature of R-tree. In [98], Ng and Sedighian proposed a three-step strategy towards Image Retrieval indexing, i.e. dimension reduction, evaluation of existing indexing approaches, and customization of the selected indexing approach. After dimension reduction by using eigenimage approach, the following three characteristics of the dimension-reduced data can be used to select good existing indexing algorithms:

- the new dimension components are ranked by decreasing variance
- the dynamic ranges of the dimensions are known
- the dimensionality is still fairly high

On their test data sets, they found that BA-KD-tree gave the best performance.

Considering that most of the tree indexing techniques were designed for traditional database queries (point queries and range queries) but not for the similarity queries used in Image Retrieval, there was a need to explore the new characteristics and requirements for indexing structures in Image Retrieval. Such a technique was explored in [155], where Tagare developed a tree adaptation approach which refined the tree structure by eliminating inefficient tree nodes for similarity queries.

So far, the above approaches only concentrated on how to identify and improve indexing techniques which are scalable to high dimensional feature vectors in Image Retrieval. The other nature of feature vectors in Image Retrieval, i.e. non-Euclidean similarity measures, has not been deeply explored. The similarity measures used in Image Retrieval may be non-Euclidean and may even be non-metric (e.g. Histogram Intersection). There are two promising techniques towards solving this problem, i.e. clustering and Neural Nets. In [32], Charikar et al. proposed an incremental clustering technique for dynamic information retrieval. This technique had three advantages, i.e. dynamic structure, capable of handling high dimensional data, and the potential to deal with Non-Euclidean similarity measures. In [118], Rui and Chakrabarti et al. further extended this technique in the directions of supporting Non-Euclidean similarity measure and faster and more accurate search strategies.

In [171], Zhang and Zhong proposed to use Self-Organization Map (SOM) Neural Nets as the tool for constructing the tree indexing structure in Image Retrieval. The advantages of using SOM were its unsupervised learning ability, dynamic clustering nature and the potential of supporting arbitrary similarity measures. Their experimental results over the Brodatz texture collection demonstrated that SOM was a promising indexing technique.

## 4. IMAGE RETRIEVAL SYSTEMS

Since the early 90's, content-based Image Retrieval has become a very active research area. Many Image Retrieval systems, both commercial and research, have been built. Most Image Retrieval systems support one or more of the following options [22]:

- random browsing
- search by example
- search by sketch
- search by text (including keyword or speech)
- navigation with customized image categories

We have seen the provision of a rich set of search options today, but systematic studies involving actual users in practical applications still need to be done to explore the tradeoffs among different options mentioned above. Here, we will select a few representative systems and highlight their distinct characteristics.

### 4.1. QBIC

QBIC [99, 48, 46, 44, 129, 74, 41], standing for Query By Image Content, is the first commercial content-based Image Retrieval system. Its system framework and techniques have profound effects on later Image Retrieval systems.

QBIC supports queries based on example images, user-constructed sketches and drawings and selected color and texture patterns, etc. The color feature used in QBIC are the average (R,G,B), (Y,i,q),(L,a,b) and MTM (Mathematical Transform to Munsell) coordinates, and a $k$ element Color Histogram [46]. Its texture feature is an improved version of the Tamura texture representation [156], i.e. combinations of coarseness, contrast and directionality [44]. Its shape feature consists of shape area, circularity, eccentricity, major axis orientation and a set of algebraic moments invariants [129, 46]. QBIC is one of the few systems which take into account the high dimensional feature indexing. In its indexing subsystem, KLT is first used to perform dimension reduction and then $R^*$-tree is used as the multi-dimensional indexing structure [74, 46]. In its new system, text-based keyword search can be combined with content-based similarity search. The on-line QBIC demo is at http://wwwqbic.almaden.ibm.com/.

### 4.2. Virage

Virage is a content-based image search engine developed at Virage Inc. Similar to QBIC, Virage [11, 56] supports visual queries based on color, composition (color layout), texture, and structure (object boundary information). But Virage goes one step further than QBIC. It also supports arbitrary combinations of the above four atomic queries. The users can adjust the weights associated with the atomic features according to their own emphasis. In [11], Jeffrey et al. further proposed an open framework for image management. They classified the visual features ("primitive") as general (such as color, shape, or texture) and domain specific (face recognition, cancer cell detection, etc.). Various usefully "primitives" can be added to the open structure depending on the domain requirements. To go beyond the query-by-example mode, Gupta and Jain proposed a nine-component *query language* framework in [56]. The corresponding demos of Virage are at http://www.virage.com/cgi-bin/query-e.

### 4.3. RetrievalWare

RetrievalWare is a content-based Image Retrieval engine developed by Excalibur Technologies Corp. [42, 3]. From one of its early publications, we can see that its emphasis was in Neural Nets to Image Retrieval [42]. Its more recent search engine uses color, shape, texture, brightness, color layout, and aspect ratio of the image, as the query features [3]. It also supports the combinations of these features and allows the users to adjust the weights associated with each feature. Its demo page is at http://vrw.excalib.com/cgi-bin/sdk/cst/cst2.bat.

### 4.4. Photobook

Photobook [106] is a set of interactive tools for browsing and searching images developed at MIT Media Lab. Photobook consists of three sub-books, from which shape, texture, and face features are extracted respectively. Users can then query based on corresponding features in each of the three sub-books.

In its more recent version of Photobook, FourEyes, Picard et al. proposed to include human in the image annotation and retrieval loop [112, 94, 110, 113, 109, 111, 114, 79]. The motivation of this was based on the observation that there was no single feature which can best model images from each and every domain. Furthermore, human's perception is subjective. They proposed a "society of model" approach to incorporate the human factor. Experimental results show that this approach is effective in interactive image annotation [94, 114].

### 4.5. VisualSEEk and WebSEEk

VisualSEEk [135, 144] is a visual feature search engine and WebSEEk [137] is a World-Wide Web oriented text/image search engine, both of which are developed at Columbia University. Main research features are spatial relationship query of image regions and visual feature extraction from compressed domain [162, 27, 28, 29].

The visual features used in their systems are Color Set and Wavelet Transform based texture feature [139, 140, 138, 141]. To speed up the retrieval process, they also developed binary tree based indexing algorithms [136, 30, 142, 143].

VisualSEEk supports queries based on both visual features and their spatial relationships. This enables a user to submit a "sunset" query as red-orange color region on top and blue or green region at the bottom as its "sketch". WebSEEk is a web oriented search engine. It consists of three main modules, i.e. image/video collecting module, subject classification and indexing module, and search, brows and retrieval module. It supports queries based on both keywords and visual content. The on-line demos are at http://www.ee.columbia.edu/~sfchang/demos.html.

### 4.6. Netra

Netra is a prototype image retrieval system developed in the UCSB Alexandria Digital Library (ADL) project [86]. Netra uses color, texture, shape and spatial location information in the segmented image regions to search and retrieve similar regions from the database. Main research features of the Netra system are its Gabor filter based texture analysis [7, 82, 89, 88], Neural Nets based image thesaurus construction [87, 84, 83] and edge flow based region segmentation [85]. The on-line demo is at http://vivaldi.ece.ucsb.edu /Netra/.

### 4.7. MARS

MARS (Multimedia Analysis and Retrieval System) is developed at University of Illinois at Urbana- Champaign [64, 91, 92, 102, 124, 125, 121, 119, 122, 123]. MARS differs from other systems in both the research scope and the techniques used. It is an interdisciplinary research effort involving multiple research communities: Computer Vision, Database Management System (DBMS), and Information Retrieval (IR). The research features of MARS are the integration of DBMS and IR (exact match with ranked retrieval) [64, 102], integration of indexing and retrieval (how the retrieval algorithm can take advantage of the underline indexing structure) [118], and integration of computer and human. The main focus of MARS is not on finding a single "best" feature representation, but rather on how to organize various visual features into a meaningful retrieval architecture which can dynamically adapt to different applications and different users. MARS formally proposes a relevance feedback architecture in Image Retrieval [123] and integrates such technique at various levels during retrieval, including query vector refinement [119], automatic matching tool selection [121], and and automatic feature adaption [122, 123]. The on-line demo is at http://jadzia.ifp.uiuc.edu:8000.

### 4.8. Other Systems

ART MUSEUM [61], developed in 1992, is one of the earliest content-based Image Retrieval systems. It uses edge feature as the visual feature for retrieval. Blobworld [18] developed at UC-Berkeley provides a transformation from the raw pixel data to a small set of localized coherent regions in color and texture space. This system allows the user to view the internal representation of the submitted image and the query results; and therein enables the user to know why some "non-similar" images are returned and can therefore modify his or her query accordingly. The distinct feature of CAETIIML (http://www.videolib.princeton.edu/test/retrieve), built at Princeton University, is its combination of the on-line similarity searching and off-line subject searching [60]. More Image Retrieval systems can be found in [49, 10, 131, 154, 51, 168, 33, 153, 101].

## 5. FUTURE RESEARCH DIRECTIONS

From the above review, we can see that many advances have been made in various research aspects, including visual feature extraction, multi-dimensional indexing, system design, etc. [148, 6, 66, 65, 67, 69, 70]. However, there are still many open research issues need to be solved before current Image Retrieval can be of practical use.

### 5.1. Human in the Loop

A fundamental difference between a Computer Vision pattern recognition system and an Image Retrieval system is that human is an indispensable part in the latter system. We need to explore the synergy of human and computer [69, 70]. This research trend has already been reflected in the evolution of content-based Image Retrieval. Early literatures emphasize "fully automated system" and try to find a "single best feature". But such an approach does not lead to a success, as the Computer Vision technique is not there yet. More recent research emphasis is given to "interactive systems" and "human in the loop". For example, QBIC team uses interactive region segmentation [41]. Based on the observation that each of the different texture representations, MRSAR, EV, and Wold-decomposition [94], has its own advantages in different domains, MIT team moves from the "automated" Photobook to "interactive" FourEyes [112, 94]. WebSEEk system allows for dynamic feature vector recomputation based on user's feedback [146]. UCSB team incorporates supervised learning in texture analysis [87, 84]. MARS team formally proposes a *Relevance Feedback* architecture in Image Retrieval, where human and computer can interact with each other to improve the retrieval performance [119, 121, 123]. Other relevance feedback based approaches include PicHunter [38] and [63].

### 5.2. High-level Concepts and Low-level Visual Features

Human tend to use high-level concepts in everyday life. However, what current Computer Vision techniques can automatically extract from images are mostly low-level features. In constrained applications, such as human face and finger print, it is possible to link the low level features to high level concepts (faces or finger prints). In general setting, however, the low-level features do not have a direct link to the high-level concepts. To narrow down this semantic gap, some off-line and on-line processing is needed. The off-line processing can be achieved by using either supervised learning, unsupervised learning, or the combination of the two. Neural Nets, Genetic Algorithm, and clustering are such learning tools [87, 84, 112, 94]. For on-line processing, a powerful and user-friendly intelligent query interface is needed to perform this task. It should allow the user to easily provide his or her evaluation of current retrieval result to the computer. The relevance feedback technique proposed in MARS is one possible tool [119, 123].

### 5.3. Web Oriented

The expansion of the World Wide Web is astonishing. Each day thousands of documents, among which many are images, are added to the web. To better organize and retrieve the almost unlimited information, Web-based search engines are highly desired. Such solution exists for text-based information. The fact that Alta Vista, Inforseek, etc. are among the most frequently visited web sites indicates the need of Web-based search engine [2]. For images on the web, even though some good work has been taken place [137, 49, 10, 131, 76], technical breakthroughs are needed to make the image search engines comparable to their text-based counterpart.

One major technical barrier lies in linking the low-level visual feature indexes used in most systems today to more desired semantic-level meanings. Based on preliminary on-line experiments, we have observed that subject browsing and text-based matching are still more popular operations than feature-based search options [145]. That's partly the reason that commercial Image Retrieval systems on the Web typically use customized subject categories to organize their image collection. Usually, different image retrieval systems focus on different sections of users and content. As a result, the indexing features and the subject taxonomies are also different, causing the concern of interoperability. Several recent efforts in standards have started to address this issue [2, 164]. Several research systems on image metaservers [14, 23, 96] have also investigated frameworks for integrated access to distributed image libraries.

### 5.4. High Dimensional Indexing

A by-product of the web expansion is the huge collection of images. Most currently existing research prototype systems only handle hundreds or at most a few thousand of images, therefore a sequential scan of the all the images will not degrade the system's performance seriously. Because of this very reason, only a few existing systems explored the multi-dimensional indexing aspect of Image Retrieval [99, 48, 64, 137]. However, as the image collections are getting larger and larger, the retrieval speed is becoming a bottle neck. Although some progresses have been made in this area, as described in Section 3, effective high dimensional indexing techniques are still in urgent need to be explored.

### 5.5. Performance Evaluation Criterion and Standard Testbed

Any technique is pushed forward by its domain's evaluation criterion. SNR is used in data compression, and precision and recall are used in text-based information retrieval. Good metric will lead the technique to the correct direction while bad ones may mislead the research effort. Currently, some Image Retrieval systems measure performance based on "cost/time" to find the right images [144]. Others evaluate performance using precision and recall, terms borrowed from text-based retrieval.

Although these criteria measure the system's performance to some extend, they are far from satisfactory. One major reason causing the difficulty of defining a good evaluation criterion is the perception subjectivity of image content. That is, the subjectivity of image perception prevents us from defining an objective evaluation criteria. But still, we need to find a way of evaluating the system performance to guide the research effort in the correct direction [69, 70].

An equally important task is to establish a well-balanced large scale testbed. For image compression, we have the Lena image, which has a good balance in various textures. For video compression, MPEG community developed well balanced test video sequences. For text-based Information Retrieval, standard large-scale testbed also exists. For image retrieval testbed, MPEG-7 community has recently started to collect test data. For a testbed to be successful, it has to be large in scale to test the scalability (for multi-dimensional indexing); balanced in image content to test image feature's effectiveness and system's overall performance.

### 5.6. Human Perception of Image Content

The ultimate end user of an Image Retrieval system is human, therefore the study of human perception of image content from a psychophysical level is crucial. This topic is closely related to the topics in sections 5.1 and 5.2. Because of its importance, however, we will re-emphasize it here.

This topic is gaining increasing attention in the recent years, aiming at exploring how human perceive image content and how can we integrate such a "human model" into the Image Retrieval systems. The early research was conducted independently by MIT team [112, 94], NEC team [39, 38, 37, 103], and UIUC team [119, 123]. Interesting enough, these teams are also the teams who initiated the study of relevance feedback in Image Retrieval. This is because, after realizing the difficulty in interpreting human perception subjectivity of image content, they naturally resorted relevance feedback to "decode" the human perception.

More recent study of human perception focuses on the psychophysical aspects of human perception [103, 117]. In [103], Papathomas et al. conducted experiments studying the importance of using (a) semantic information, (b) memory of previous input, and (c) relative vs. absolute judgment of image similarities, using PicHunter [38] as the underlying Image Retrieval system. Results show that the best performance is achieved when uses only semantic cues, with memory and relative similarity judgment. The combination of semantic and visual cues only achieves second best result. We feel that one of the reasons that visual cues did not help in this case may due to the limited test dataset size – if the dataset is not big enough, the system may not be able to utilize the additional information from the visual cues. Our conjecture matches the experimental results from another group. In [117], Rogowitz et al. conducted a series experiments on analyzing human psychophysical perception of image content. According to their results, even though visual features do not capture the whole semantic meaning of the images, they do correlate a lot with the semantics. This result encourages us to develop perceptually-based image features and metrics to achieve semantically meaningful retrievals.

### 5.7. Integration of Disciplines and Media

Both the Database community literature and the Computer Vision community literature have used "Image Database" as the title of many articles [33]. However, in reality, most Database community systems are Non-image (text-based keywords or graphics-based icons) Databases, while most Computer Vision systems are Image Non-databases (just a large file containing thousands of images is not a database, since most fundamental Database units such as data model, indexing, etc. are not addressed at all). To the authors' knowledge, even though there are research effort ongoing to build true Image Databases [102, 45], the systems are not at the complete stage yet.

A successful Image Database System requires an interdisciplinary research effort. Besides the integration of Database Management and Computer Vision, the research from the traditional Information Retrieval area [17, 127, 133, 126, 16, 8] is also an indispensable part. Although the traditional Information Retrieval area's research focus was in text-based document retrieval, many useful retrieval models and techniques can be adapted to Image Retrieval. Some successful examples of such research effort include the adaption of Boolean retrieval model in Image Retrieval [92, 102], and the utilization of relevance feedback in Image Retrieval [119, 123].

Another observation is that integration of multimedia, multi-modalities provides great potential for improved indexing and classification of images in general domains. Research in [149, 147, 145] has shown promising results in using both textual and visual features in automatic indexing of images. More sophisticated techniques for cross-mapping image classification between the high level using textual cues and the low level using the visual cues will bear fruits.

## 6. CONCLUDING REMARKS

Past and current technical achievements in visual feature extraction, multi-dimensional indexing, and system design are reviewed. Open research issues are identified and future research directions suggested. From the previous section, we can see that a successful Image Retrieval system requires the seamless integration of multiple research communities' effort. Both progress in each individual research community and an overall system architecture are equally important. We propose one possible integrated system architecture, as shown in Figure 1.

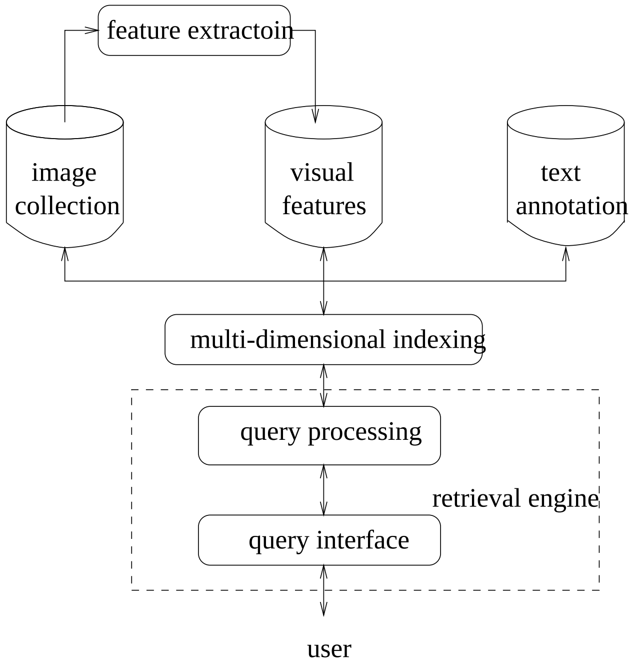


Figure 1: An Image Retrieval System Architecture

There are three databases in this system architecture. The Image Collection database contains the raw images for visual display purpose. During different stages of Image Retrieval, different image resolutions may be needed. In that case, wavelet-compressed image is a good choice [64]. Image Processing and Compression research communities contribute to this database.

The Visual Feature database stores the visual features extracted from the images using techniques described in Section 2. This is the information needed to support content-based Image Retrieval. Computer Vision and Image Understanding are the research communities contributing to this database.

The Text Annotation database contains the keywords and free-text descriptions of the images. It is becoming clear in the Image Retrieval community that content-based Image Retrieval is not a replacement of, but rather a complementary component to, the text-based Image Retrieval. Only the integration of the two can result in satisfactory retrieval performance. The research progress in IR and DBMS is the main thrust to this database.

As discussed earlier, to achieve fast retrieval speed and make the retrieval system truly scalable to large-size image collections, effective multi-dimensional indexing module is an indispensable part of the whole system. This module will be pushed forward by Computational Geometry, Databases Management, and Pattern Recognition research communities.

The retrieval engine module includes a query interface submodule and a query processing submodule. To communicate with the user in a friendly manner, the query interface is graphics-based. The interface collects the information need from the users and displays back the retrieval results to the users in a meaningful way. Research progress in User Psychology and User Interface helps to improve the interface design. Furthermore, the same query from a user can be processed in different ways. The query processing submodule manipulates the user query into best processing procedures. This techniques is advanced by Database Management community.

There are two major characteristics of this system architecture. One is its multi-discipline and inter-discipline nature, as demonstrated in the above discussion. The other is its interactivity nature between human and computer. (Note that from the user to the three databases, the arrows are bi-directional). In all, integration of various disciplines, of multiple information sources, and of human and computer will lead to a successful Image Retrieval system.

## 7. REFERENCES


[1] MPEG-7 applications document. *ISO/IEC JTC1/SC29/WG11 N1922, MPEG97*, Oct 1997.

[2] MPEG-7: Context and objectives (v.5). *ISO/IEC JTC1/SC29/WG11 N1920, MPEG97*, Oct 1997.

[3] Retrievalware demo page. *http://vrw.excalib.com/cgi-bin/sdk/cst/cst2.bat*, 1997.

[4] Special issue on visual information management. *Communications of ACM*, Dec 1997. Guest Editor:Ramesh Jain.

[5] Third draft of MPEG-7 requirements. *ISO/IEC JTC1/SC29/WG11 N1921, MPEG97*, Oct 1997.

[6] ACM. Proc. int. conf. on multimedia. 1993-1997.

[7] A. D. Alexandrov, W.Y.Ma, A. El Abbadi, and B.S.Manjunath. Adaptive filtering and indexing for image databases. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.

[8] James Allan. Relevance feedback with too much data. In *Proc of SIGIR'95*, 1995.

[9] Esther M. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 13(3), March 1991.

[10] Vassilis Athitsos, Michael J. Swain, and Charles Frankel. Distinguishing photographs and graphics on the world wide web. In *Proc IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[11] Jeffrey R. Bach, Charles Fuller, Amarnath Gupta, Arun Hampapur, Bradley Horowitz, Rich Humphrey, Ramesh Jain, and Chiao fe Shu. The Virage image search engine: An open framework for image management. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*.

[12] H. G. Barrow. Parametric correspondence and chamfer matching: Two new techniques for image matching. In *Proc 5th Int. Joint Conf. Artificial Intelligence*, 1977.

[13] N. Beckmann, H.-P. Kriegel, R. Schneider, and B.Seeger. The R*-tree: an efficient and robust access method for points and rectangles. In *Proc ACM SIGMOD*, 1990.

[14] Mandis Beigi, Ana Benitez, and S.-F. Chang. Metaseek: A content-based meta search engine for images. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, San Jose, CA, 1998. demo and document: http://www.ctr.columbia.edu/metaseek.

[15] Gunilla Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 1988.

[16] Chris Buckley and Gerard Salton. Optimization of relevance feedback weights. In *Proc. of SIGIR'95*, 1995.

[17] James P. Callan, W. Bruce Croft, and Stephen M. Harding. The inquery retrieval system. In *Proc of 3rd Int Conf on Database and Expert System Application*, Sept 1992.

[18] Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik. Region-based image querying. In *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97*, 1997.

[19] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *CVGIP: Graphical models and image processing journal*, 1997.

[20] N. S. Chang and K. S. Fu. A relational database system for images. Technical Report TR-EE 79-28, Purdue University, May 1979.

[21] N. S. Chang and K. S. Fu. Query-by pictorial-example. *IEEE Trans. on Software Engineering*, SE-6(6), Nov. 1980.

[22] S.-F. Chang, A. Eleftheriadis, and Robert McClintock. Next-generation content representation, creation and searching for new media applications in education. *(to appear in) IEEE Proceedings*, 1998.

[23] S.-F. Chang, J. R. Smith, M. Beigi, and A. Benitez. Visual information retrieval from large distributed online repositories. *Communications of ACM, Special Issue on Visual Information Retrieval*, pages 12–20, Dec 1997.

[24] Shi-Kuo Chang. Pictorial data-base systems. *IEEE Computer*, Nov. 1981.

[25] Shi-Kuo Chang and Arding Hsu. Image information systems: Where do we go from here? *IEEE Trans. on Knowledge and Data Engineering*, 4(5), Oct. 1992.

[26] Shi-Kuo Chang, C. W. Yan, Donald C. Dimitroff, and Timothy Arndt. An intelligent image database system. *IEEE Trans. on Software Engineering*, 14(5), May 1988.

[27] Shih-Fu Chang. Compressed-domain content-based image and video retrieval. In *Proc Symposium on Multimedia Communications and Video Coding*, 1995.

[28] Shih-Fu Chang. Compressed-domain techniques for image/video indexing and manipulation. In *Proc ICIP95 Special Session on Digital Library and Video on Demand*, 1995.

[29] Shih-Fu Chang and john R. Smith. Finding images/video in large archives. *D-Lib Magazine*, 1997.

[30] Shih-Fu Chang and John Smith. Extracting multi-dimensional signal features for content-based visual query. In *Proc SPIE Symposium on Visual Communications and Signal Processing*, 1995.

[31] Tianhorng Chang and C.-C. Jay Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Proc.*, 2(4):429–441, October 1993.

[32] Moses Charikar, Chandra Chekur, Tomas Feder, and Rajeev Motwani. Incremental clustering and dynamic information retrieval. In *Proc. of the 29th Annual ACM Symposium on Theory of Computing*, pages 626–635, 1997.

[33] Biao Cheng. Approaches to image retrieval based on compressed data for multimedia database systems. *PhD Thesis, University of New York at Buffalo*, 1996.

[34] Tat Seng Chua, Kian-Lee Tan, and Beng Chin Ooi. Fast signiture-based color-spatial image retrieval. In *Proc. IEEE Conf. on Multimedia Computing and Systems*, 1997.

[35] Gene C.-H. Chuang and C.-C. Jay Kuo. Wavelet descriptor of planar curves: Theory and applications. *IEEE Trans. Image Proc.*, 5(1):56–70, January 1996.

[36] David Copper and Zhibin Lei. On representation and invariant recognition of complex objects based on patches and parts. *Spinger Lecture Notes in Computer Science series, 3D Object Representation for Computer Vision*, pages 139–153, 1995. M. Hebert, J. Ponce, T. Boult, A. Gross, editors.

[37] I. J. Cox, M. L. Miller, Thomas P. Minka, and P. N. Yianilos. An optimized interaction strategy for bayesian relevance feedback. In *IEEE Conf. CVPR*, 1998.

[38] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos. Pichunter: Bayesian relevance feedback for image retrieval. In *Intl. Conf. on Pattern Recognition.*

[39] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos. Target testing and the pichunter bayesian multimedia retrieval system. In *Advanced Digital Libraries Forum*, Washington D.C., May.

[40] G. C. Cross and A. K. Jain. Markov random field texture models. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 5:25–39, 1983.

[41] Dirk Daneels, D. Campenhout, Wayne Niblack, Will Equitz, Ron Barber, Erwin Bellon, and Freddy Fierens. Interactive outlining: An improved approach using active contours. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

[42] James Dowe. Content-based retrieval in multimedia imaging. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

[43] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*, chapter 6. John Wiley and Sons, Inc.

[44] Will Equitz and Wayne Niblack. Retrieving images from a database using texture – algorithms from the QBIC system. Technical Report RJ 9805, Computer Science, IBM Research Report, May 1994.

[45] Ronald Fagin and Edward L. Wimmers. Incorporating user preferences in multimedia queries. In *Proc of Int. Conf. on Database Theory*, 1997.

[46] C. Faloutsos, M. Flickner, W. Niblack, D. Petkovic, W. Equitz, and R. Barber. Efficient and effective querying by image content. Technical report, IBM Research Report, 1993.

[47] Christos Faloutsos and King-Ip (David) Lin. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proc. of SIGMOD*, pages 163–174, 1995.

[48] M. Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Q. Huang, Byron Dom, Monika Gorkani, Jim Hafine, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 1995.

[49] Charles Frankel, Michael J. Swain, and Vassilis Athitsos. Webseer: An image search engine for the world wide web. Technical Report TR-96-14, Computer Science Department, University of Chicago, 1996.

[50] T. Gevers and V. K. Kajcovski. Image segmentation by directed region subdivision. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[51] Yihong Gong, HongJiang Zhang, H.C. Chuan, and M.Sakauchi. An image database system with content capturing and fast image indexing abilities. In *Proc IEEE 1994*, 1994.

[52] Calvin C. Gotlieb and Herbert E. Kreyszig. Texture descriptors based on co-occurrence matrices. *Computer Vision, Graphics, and Image Processing*, 51, 1990.

[53] Diane Greene. An implementation and performance analysis of spatial data access. In *Proc ACM SIGMOD*, 1989.

[54] M. H. Gross, R. Koch, L. Lippert, and A. Dreger. Multiscale image texture analysis in wavelet spaces. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[55] Venkat N. Gudivada and Jijay V. Raghavan. Special issue on content-based image retrieval systems. *IEEE Computer Magzine*, 28(9), 1995.

[56] Amarnath Gupta and Ramesh Jain. Visual information retrieval. *Communications of the ACM*, 40(5), 1997.

[57] A. Guttman. R-tree: a dynamic index structure for spatial searching. In *Proc ACM SIGMOD*, 1984.

[58] Michael Hansen and William Higgins. Watershed-driven relaxation labeling for image segmentation. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[59] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. Texture features for image classification. *IEEE Trans. on Sys, Man, and Cyb*, SMC-3(6), 1973.

[60] Hong heather Yu and Wayne Wolf. Hierarchical, multi-resolution algorithms for dictionary-driven content-based image retrieval. In *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[61] Kyoji Hirata and Toshikazu Kato. Query by visual example. In *Proc of 3rd Int Conf on Extending Database Technology*.

[62] M. K. Hu. Visual pattern recognition by moment invariants, computer methods in image analysis. *IRE Transactions on Information Theory*, 8, 1962.

[63] J. Huang, S. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih. Image indexing using color correlogram. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[64] Thomas S. Huang, Sharad Mehrotra, and Kannan Ramchandran. Multimedia analysis and retrieval system (MARS) project. In *Proc of 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.

[65] IEEE. Proc. ieee int. conf. on image proc. 1994-1997.

[66] IEEE. Proc. int. conf. on multimedia computing and systems. 1994-1997.

[67] IEEE. Proc. ieee conf. on computer vision and pattern recognition. 1995-1997.

[68] Mikihiro Ioka. A method of defining the similarity of images on the basis of color information. Technical Report RT-0030, IBM Research, Tokyo Research Laboratory, November 1989.

[69] R. Jain. Workshop report: NSF workshop on visual information management systems. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

[70] R. Jain, A. Pentland, and D. Petkovic. NSF-ARPA workshop on visual information management systems. Cambridge, MA, June 1995.

[71] Deepak Kapur, Y. N. Lakshman, and Tushar Saxena. Computing invariants using elimination methods. In *Proc. IEEE Int. Conf. on Image Proc.*, 1995.

[72] Amlan Kundu and Jia-Lin Chen. Texture classification using qmf bank-based subband decomposition. *CVGIP: Graphical Models and Image Processing*, 54(5):369–384, September 1992.

[73] Andrew Laine and Jian Fan. Texture classification by wavelet packet signatures. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 15(11):1186–1191, 1993.

[74] Denis Lee, Ron Barber, Wayne Niblack, Myron Flickner, Jim Hafner, and Dragutin Petkovic. Indexing for complex queries on a query-by-content image database. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[75] Z. Lei, D. Keren, and D. B. Cooper. Computationally fast bayesian recognition of complex objects based on mutual algebraic invariants. In *Proc. IEEE Int. Conf. on Image Proc.*

[76] Michael S. Lew, Kim Lempinen, and D. P. Huijsmans. Webcrawling using sketches. Technical report, Computer Science Department, Leiden University, The Netherlands, 1997.

[77] Bingcheng Li and Song De Ma. On the relation between region and contour representation. In *Proc. IEEE Int. Conf. on Image Proc.*, 1995.

[78] X. Q. Li, Z. W. Zhao, H. D. Cheng, C. M. Huang, and R. W. Harris. A fuzzy logic approach to image segmentation. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[79] Fang Liu and R. W. Picard. Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 18(7), July 1996.

[80] H Lu, B. Ooi, and K. Tan. Efficient image retrieval by color contents. In *Proc. of the 1994 Int. Conf. on Applications of Databases*, 1994.

[81] M. Lybanon, S. Lea, and S. Himes. Segmentation of diverse image types using opening and closing. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[82] W. Y. Ma and B. S. Manjunath. A comparison of wavelet transform features for texture image annotation. In *Proc. IEEE Int. Conf. on Image Proc.*, 1995.

[83] W. Y. Ma and B. S. Manjunath. A pattern thesaurus for browsing large aerial photographs. Technical Report 96-10, Univ. of California at Santa Barbara, 1996.

[84] W. Y. Ma and B. S. Manjunath. Texture features and learning similarity. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 425–430, 1996.

[85] W. Y. Ma and B. S. Manjunath. Edge flow: a framework of boundary detection and image segmentation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[86] W. Y. Ma and B. S. Manjunath. Netra: A toolbox for navigating large image databases. In *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[87] B. S. Manjunath and W. Y. Ma. Image indexing using a texture dictionary. In *Proceedings of SPIE conference on Image Storage and Archiving System*, volume 2606.

[88] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. Technical Report 95-06, Univ. of California at Santa Barbara, 1995.

[89] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE T-PAMI special issue on Digital Libraries*, Nov. 1996.

[90] C. S. McCamy, H. Marcus, and J. G. Davidson. A color-rendition chart. *Journal of Applied Photographic Engineering*, 2(3), Summer 1976.

[91] Sharad Mehrotra, Kaushik Chakrabarti, Mike Ortega, Yong Rui, and Thomas S. Huang. Multimedia analysis and retrieval system. In *Proc. of The 3rd Int. Workshop on Information Retrieval Systems*, 1997.

[92] Sharad Mehrotra, Yong Rui, Michael Ortega-B., and Thomas S. Huang. Supporting content-based queries over images in MARS. In *Proc. of IEEE Int. Conf. on Multimedia Computing and Systems*, 1997.

[93] Babu M. Mehtre, M. Kankanhalli, and Wing Foon Lee. Shape measures for content based image retrieval: A comparison. *Information Processing & Management*, 33(3), 1997.

[94] T. P. Minka and R. W. Picard. Interactive learning using a "society of models". In *Proc IEEE CVPR*, pages 447–452, 1996.

[95] Makoto Miyahara. Mathematical transform of (r,g,b) color data to munsell (h,s,v) color data. In *SPIE Visual Communications and Image Processing*, volume 1001, 1988.

[96] D. Murthy and A. Zhang. Webview: A multimedia database resource integration and search system over web. In *WebNet 97: World Conference of the WWW, Internet, and Intranet*, Oct. 1997.

[97] A. D. Narasimhalu. Special section on content-based retrieval. *Multimedia Systems*, 1995.

[98] Raymond Ng and Andishe Sedighian. Evaluating multi-dimensional indexing structures for images transformed by principal component analysis. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.

[99] W. Niblack, R. Barber, and et al. The QBIC project: Querying images by content using color, texture and shape. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, Feb 1994.

[100] Philippe P. Ohanian and Richard C. Dubes. Performance evaluation for four classes of texture features. *Pattern Recognition*, 25(8):819–833, 1992.

[101] Atsushi Ono, Masashi Amano, and Mituhiro Hakaridani. A flexible content-based image retrieval system with combined scene descrption keyword. In *Proc. IEEE Conf. on Multimedia Computing and Systems*, 1996.

[102] Michael Ortega, Yong Rui, Kaushik Chakrabarti, Sharad Mehrotra, and Thomas S. Huang. Supporting similarity queries in MARS. In *Proc. of ACM Conf. on Multimedia*, 1997.

[103] Thomas V. Papathomas, Tiffany E. Conway, I. J. Cox, Joumana Ghosn, M. L. Miller, Thomas P. Minka, and P. N. Yianilos. Psychophysical studies of the performance of an image database retrieval system. In *IS&T/SPIE Conf on Human Vision and Electronic Imageing III*, 1998.

[104] Greg Pass, Ramin Zabih, and Justin Miller. Comparing images using color coherence vectors. In *Proc. ACM Conf. on Multimedia*, 1996.

[105] A. Pentland and R. Picard. Special issue on digital libraries. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 1996.

[106] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 1996.

[107] A. P. Pentland. Fractal-based description of natural scenes. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 6(6):661–674, 1984.

[108] E. Persoon and K. S. Fu. Shape discrimination using fourier descriptors. *IEEE Trans. Sys. Man, Cyb.*, 1977.

[109] R. W. Picard. Computer learning of subjectivity. In *Proc ACM Computing Surveys*.

[110] R. W. Picard. Digital libraries: Meeting place for high-level and low-level vision. In *Proc Asian Conf. on Comp. Vis.*

[111] R. W. Picard. A society of models for video and image libraries. Technical report, MIT, 1996.

[112] R. W. Picard and T. P. Minka. Vision texture for annotation. *Multimedia Systems: Special Issue on Content-based retrieval*.

[113] Rosalind W. Picard. Toward a visual thesaurus. In *Springer Verlag Workshops in Computing, MIRO 95*, 1995.

[114] Rosalind W. Picard, Thomas P. Minka, and Martin Szummer. Modeling user subjectivity in image libraries. In *Proc. IEEE Int. Conf. on Image Proc.*, Lausanne, Sept. 1996.

[115] Lawrence Rabiner and Biing-Hwang Juang. *Fundamentals of Speech Recognition*, chapter 5. Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[116] Rick Rickman and John Stonham. Content-based image retrieval using colour tuple histograms. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.

[117] Bernice E. Rogowitz, Thomas Frese, John Smith, Charles A. Bouman, and Edward Kalin. Perceptual image similarity experiments. In *IS&T/SPIE Conf on Human Vision and Electronic Imageing III*, 1998.

[118] Yong Rui, Kaushik Chakrabarti, Sharad Mehrotra, Yunxin Zhao, and Thomas S. Huang. Dynamic clustering for optimal retrieval in high dimensional multimedia databases. In *TR-MARS-10-97*, 1997.

[119] Yong Rui, Thomas S. Huang, and Sharad Mehrotra. Content-based image retrieval with relevance feedback in MARS. In *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[120] Yong Rui, Thomas S. Huang, and Sharad Mehrotra. Mars and its applications to MPEG-7. *ISO/IEC JTC1/SC29/WG11 M2290, MPEG97*, July 1997.

[121] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega. Automatic matching tool selection using relevance feedback in MARS. In *Proc. of 2nd Int. Conf. on Visual Information Systems*, 1997.

[122] Yong Rui, Thomas S. Huang, Sharad Mehrotra, and Michael Ortega. A relevance feedback architecture in content-based multimedia information retrieval systems. In *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97*, 1997.

[123] Yong Rui, Thomas S. Huang, Michael Ortega, and Sharad Mehrotra. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Tran on Circuits and Systems for Video Technology, Special Issue on Interactive Multimedia Systems for the Internet*, Sept 1998.

[124] Yong Rui, Alfred C. She, and Thomas S. Huang. Automated shape segmentation using attraction-based grouping in spatial-color-texture space. In *Proc. IEEE Int. Conf. on Image Proc.*, 1996.

[125] Yong Rui, Alfred C. She, and Thomas S. Huang. Modified fourier descriptors for shape representation – a practical approach. In *Proc of First International Workshop on Image Databases and Multi Media Search*, 1996.

[126] Gerard Salton and Chris Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 1988.

[127] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, 1983.

[128] Ramin Samadani and Cecilia Han. Computer-assisted extraction of boundaries from images. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.

[129] B. Scassellati, S. Alexopoulos, and M. Flickner. Retrieving images by 2d shape: A comparison of computation methods with human perceptual judgments. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1994.

[130] Bruce Schatz and Hsinchun Chen. Building large-scale digital libraries. *Computer*, 1996.

[131] Stan Sclaroff, Leonid Taycher, and Marco La Cascia. Imagerover: A content-based image browser for the world wide web. In *Proc IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[132] T. Sellis, N. Roussopoulos, and C. Faloutsos. The R+ tree: A dynamic index for multi-dimensional objects. In *Proc 12th VLDB*, 1987.

[133] W. M. Shaw. Term-relevance computations and perfect retrieval performance. *Information Processing and Management.*

[134] Alfred C. She and Thomas S. Huang. Segmentation of road scenes using color and fractal-based texture classification. In *Proc ICIP*, Austin, Nov. 1994.

[135] J. R. Smith and S.-F. Chang. *Intelligent Multimedia Information Retrieval, edited by Mark T. Maybury*, chapter Querying by Color Regions Using the VisualSEEk Content-Based Visual Query System. 1996.

[136] J. R. Smith and S.-F. Chang. Local color and texture extraction and spatial query. In *Proc. IEEE Int. Conf. on Image Proc.*, 1996.

[137] J. R. Smith and S.-F. Chang. Visually searching the web for content. *IEEE Multimedia Magazine*, 4(3):12–20, Summer 1997. also Columbia U. CU/CTR Technical Report 459-96-25.

[138] John R. Smith and Shih-Fu Chang. Transform features for texture classification and discrimination in large image databases. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[139] John R. Smith and Shih-Fu Chang. Single color extraction and image query. In *Proc. IEEE Int. Conf. on Image Proc.*, 1995.

[140] John R. Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *IS & T/SPIE proceedings Vol.2670, Storage & Retrieval for Image and Video Databases IV*, 1995.

[141] John R. Smith and Shih-Fu Chang. Automated binary texture feature sets for image retrieval. In *Proc ICASSP-96*, Atlanta, GA, 1996.

[142] John R. Smith and Shih-Fu Chang. Automated binary texture feature sets for image retrieval. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, May 1996.

[143] John R. Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.

[144] John R. Smith and Shih-Fu Chang. Visualseek: A fully automated content-based image query system. In *Proc ACM Multimedia 96*, 1996.

[145] John R. Smith and Shih-Fu Chang. Enhancing image search engines in visual information environments. In *IEEE 1st Multimedia Signal Processing Workshop*, June 1997.

[146] John R. Smith and Shih-Fu Chang. An image and video search engine for the world-wide web. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1997.

[147] John R. Smith and Shih-Fu Chang. Multi-stage classification of images from features and related text. In *4th Europe EDLOS Workshop*, San Miniato, Italy, Aug 1997.

[148] SPIE. Proc. spie storage and retrieval for image and video databases. 1993-1997.

[149] R. K. Srihari. Automatic indexing and content-based retrieval of captioned images. *IEEE Computer Magzine*, 28(9), 1995.

[150] Markus Stricker and Alexander Dimai. Color indexing with weak spatial constraints. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.

[151] Markus Stricker and Markus Orengo. Similarity of color images. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.

[152] Michael Swain and Dana Ballard. Color indexing. *International Journal of Computer Vision*, 7(1), 1991.

[153] Michael J. Swain. interactive indexing into image databases. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, volume 1908, 1993.

[154] Daniel L. Swets and John J. Weng. Efficient content-based image retrieval using automatic feature selection. In *Proc IEEE 1995*, 1995.

[155] Hemant Tagare. Increasing retrieval efficiency by index tree adaption. In *Proc of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97*, 1997.

[156] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki. Texture features corresponding to visual perception. *IEEE Trans. on Sys, Man, and Cyb*, SMC-8(6), 1978.

[157] Hideyuki Tamura and Naokazu Yokoya. Image database systems: A survey. *Pattern Recognition*, 17(1), 1984.

[158] Gabriel Taubin. Recognition and positioning of rigid objects using algebraic moment invariants. In *SPIE Vol. 1570 Geometric Methods in Computer Vision*, 1991.

[159] K. S. Thyagarajan, Tom Nguyen, and Charles Persons. A maximum likelihood approach to texture classification using wavelet transform. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[160] Imothy Wallace and Owen Mitchell. Three-dimensional shape analysis using local shape descriptors. *IEEE Trans. Patt. Recog. and Mach. Intell.*, PAMI-3(3), May 1981.

[161] Imothy Wallace and Paul Wintz. An efficient three-dimensional aircraft recognition algorithm using normalized fourier descriptors. *Computer Graphics and Image Processing*, 13, 1980.

[162] Hualu Wang and S.-F. Chang. Compressed-domain image search and applications. Technical report, Columbia Univ., 1995.

[163] Jia Wang, Wen-Jann Yang, and Raj Acharya. Color clustering techniques for color-content-based image retrieval from image databases. In *Proc. IEEE Conf. on Multimedia Computing and Systems*, 1997.

[164] S. Weibel and E. Miller. Image description on the internet: A summary of the cni/oclc image metadata on the internet workshop, september 24 - 25, 1996, dublin, ohio. *D-Lib Magazine*, 1997.

[165] Joan Weszka, Charles Dyer, and Azeril Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Trans. on Sys, Man, and Cyb*, SMC-6(4), 1976.

[166] D. White and R. Jain. Algorithms and strategies for similarity retrieval. In *TR VCL-96-101*. University of California, San Diego, 1996.

[167] D. White and R. Jain. Similarity indexing: Algorithms and performance. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1996.

[168] J. K. Wu, A. Desai Narasimhalu, B. M. Mehtre, C. P. Lam, and Y. J. Gao. Core: a content-based retrieval engine for multimedia information systems. *Multimedia Systems*, 1995.

[169] Luren Yang and Fritz Algregtsen. Fast computation of invariant geometric moments: A new method giving correct results. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.

[170] C. T. Zahn and R. Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. on Computers*, 1972.

[171] HongJiang Zhang and Di Zhong. A scheme for visual feature based image retrieval. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.