MIT International Journal of Computer Science & Information Technology, *Vol. 2, No. 1, Jan. 2012, pp. (32-39)*
ISSN 2230-7621 © MIT Publications

Information Retrieval through Image Based Search Engine

Shivanshu Rastogi
Assistant Professor
MIT, Moradabad
e-mail: rshivanshu1145@gmail.com

Prachi Gupta
Assistant Professor
MIT, Moradabad
e-mail: prachi.g19@gmail.com

INTRODUCTION

Recent years have seen a rapid increase in the size of digital image collections. Everyday, both military and civilian equipment generates gigabytes of images. A huge amount of information is out there. However, we cannot access 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 1970s, 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 1970s. 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. Many advances, such as data modelling, multidimensional indexing, and query evaluation, 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 labour 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 later retrieval processes.

In the early 1990s, 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 key words, images would be indexed by their own visual content, such as colour and texture. 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.

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.

There are three fundamental bases for content-based image retrieval, i.e. visual feature extraction, multidimensional 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 multidimensional 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.

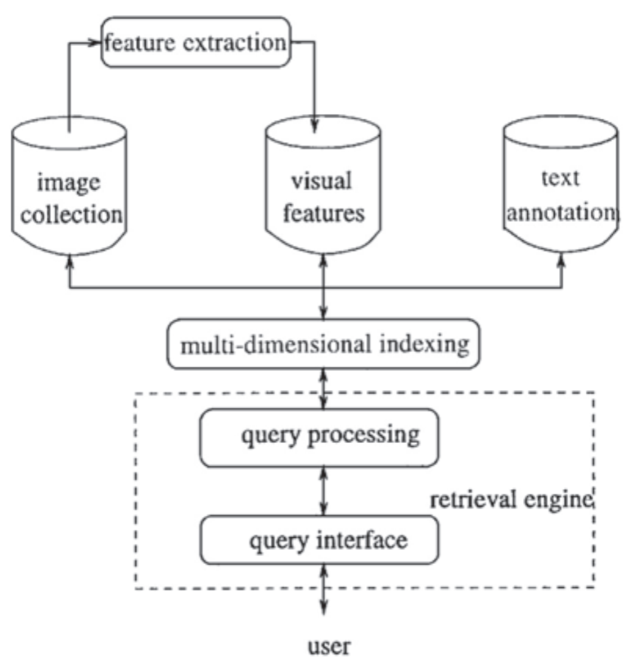

Figure 1: Image based Image Retrieval

ISSN 2230-7621 © MIT Publications

Image Based Search Engine is basically have three Components which are as follows:

- Query Dependent Part
- Query Independent Part
- Image Crawling

With reference to Figure 2 description is as follow:

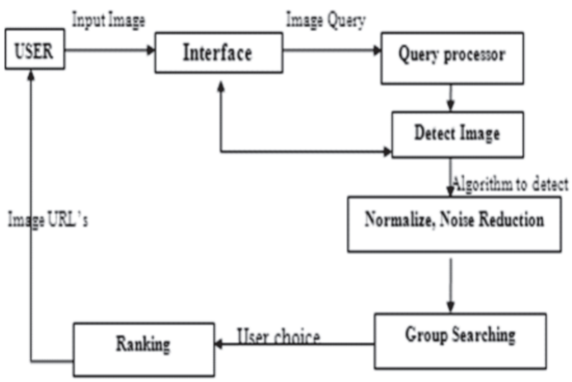

Figure 2: Query Dependent Part

DESCRIPTION OF IMAGE DEPENDENT QUERY CHART

STEP 1: User

This step include the initiation by the user i.e. the user will input the image of the person/thing about whom he/she wants to convoke the information. A user can upload the image directly from upload button or he/she can drag the particular image from the particular location and drop that image to drop textbox. User can input the input in following form i.e. Text based, Voice Based and Image. We are working with Imaged Based, We are going to input the information or input in the form of Images.

STEP 2: Interface

After inputting the image, the image will pass through the interface, this interface will generate the query, for processing the image further.

In computing, a graphical user interface (GUI, sometimes pronounced gooey) is a type of user interface that allows users to interact with electronic devices with images rather than text commands. GUIs can be used in computers, hand-held devices such as MP3 players, portable media players or gaming devices, household appliances and office equipment. A GUI represents the information and actions available to a user through graphical icons and visual indicators such as secondary notation, as opposed to text-based interfaces, typed command labels or text navigation. The actions are usually performed through direct manipulation of the graphical elements.

The term GUI is historically restricted to the scope of two-dimensional display screens with display resolutions able to describe generic information, in the tradition of the computer science research at the PARC (Palo Alto Research Center). The term GUI earlier might have been applicable to other high-resolution types of interfaces that are non-generic, such as video games, or not restricted to flat screens, like volumetric displays.

STEP 3: Query Processor

The generated query will be taken to the query processor. This query processor will processes the query and grab information that can be used to detect that particular Image. A image which passed through the interface with the help of drag and drop or directly upload through the channel is than inputted to query processor and than it will go through detect images.

STEP 4: Detect Images

It is the very step in which the face will be detected on the basis of algorithm, that are used to detect the face. The step it self include various steps of the algorithm that are used to detect the face. After detecting the faces they are available for the user to select the images they want.

STEP 5: Normalization, Noise Reduction

1. Image is rotated to align the eyes (eye coordinates must be known).
2. The image is scaled to make the distance between the eyes constant. The image is also cropped to a smaller size that is nearly just the face.
3. A mask is applied that zeros out pixels not in an oval that contains the typical face. The oval is generated analytically.
4. Histogram equalization is used to smooth the distribution of gray values for the non-masked pixels.
5. The image is normalized so the non-masked pixels have mean zero and standard deviation one.

STEP 6: Group Searching

This step include searching process of the images according to the inputted image (by the user). After applying the face detection algorithm on the image, the matched image will be found and then these images will be grouped into different categories based on the similarities of the images. Then the images will be distinguished into groups, (in case if there are thousands of images found and the user wants only 10 images then the rest images will be discarded automatically, without user interaction).

STEP 7: Ranking

In this step the recognition algorithm is applied to recognize different faces found, (based on detection algorithm) after the recognition of the faces, they will be ranked according to the features matched. The ranking of the faces include prioritizing the images so as to help the user in easy access of the images. These ranked images will be then send to the users to select the image according to their need.

ISSN 2230-7621 © MIT Publications

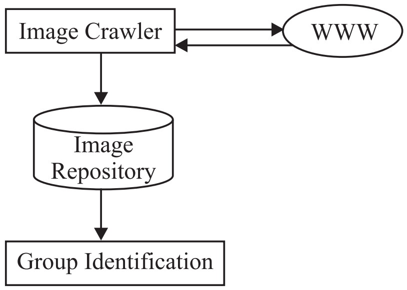

Figure 3: Query Independent Part

With refrence to Figure 3 description is as follows.

WEB CRAWLER

A Web crawler is a computer program that browses the World Wide Web in a methodical, automated manner or in an orderly fashion. Other terms for Web crawlers are ants, automatic indexers, bots, Web spiders, Web robots, or especially in the FOAF community Web scutters. This processiscalled Web crawling or spidering. Many sites, in particular search engines, use spidering as a means of providing up-to-date data. Web crawlers are mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches. Crawlers can also be used for automating maintenance tasks on a Web site, such as checking links or validating HTML code. Also, crawlers can be used to gather specific types of information from Web pages, such as harvesting e-mail addresses (usually for sending spam).

A Web crawler is one type of bot, or software agent. In general, it starts with a list of URLs to visit, called the seeds. As the crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the list of URLs to visit, called the crawl frontier. URLs from the frontier are recursively visited according to a set of policies. The large volume implies that the crawler can only download a fraction of the Web pages within a given time, so it needs to prioritize its downloads. The high rate of change implies that the pages might have already been updated or even deleted.

The number of possible crawlable URLs being generated by server-side software has also made it difficult for web crawlers to avoid retrieving duplicate content. Endless combinations of HTTP GET(URL-based) parameters exist, of which only a small selection will actually return unique content. For example, a simple online photo gallery may offer three options to users, as specified through HTTP GET parameters in the URL. If there exist four ways to sort images, three choices of thumbnail size, two file formats, and an option to disable user-provided content, then the same set of content can be accessed with 48 different URLs, all of which may be linked on the site. This mathematical combination creates a problem for crawlers, as they must sort through endless combinations of relatively minor scripted changes in order to retrieve unique content. As Edwards et al. noted, "Given that the bandwidth for conducting crawls is neither infinite nor free, it is becoming essential to crawl the Web in not only a scalable, but efficient way, if some reasonable measure of quality or freshness is to be maintained.

A crawler must carefully choose at each step which pages to visit next. The behaviour of a Web crawler is the outcome of a combination of policies:

- a selection policy that states which pages to download.
- a re-visit policy that states when to check for changes to the pages.
- a politeness policy that states how to avoid overloading Web sites, and
- a parallelization policy that states how to coordinate distributed Web crawlers.

WORLD WIDE WEB

The World Wide Web (or the proper World Wide Web; abbreviated as WWW or W3, and commonly known as the Web) is a system of interlinked hypertext documents accessed via the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via hyperlinks. Using concepts from earlier hypertext systems, British engineer and computer scientist Sir Tim Berners-Lee, now Director of the World Wide Web Consortium (W3C), wrote a proposal in March 1989 for what would eventually become the World Wide Web. At CERN in Geneva, Switzerland, Berners-Lee and Belgian computer scientist Robert Cailliau proposed in 1990 to use hypertext "to link and access information of various kinds as a web of nodes in which the user can browse at will", and they publicly introduced the project in December. "The World-Wide Web was developed to be a pool of human knowledge, and human culture, which would allow collaborators in remote sites to share their ideas and all aspects of a common project." The term WWW refers to the World Wide Web or simply the Web. The World Wide Web consists of all the public Web sites connected to the Internet worldwide, including the client devices (such as computers and cell phones) that access Web content. The WWW is just one of many applications of the Internet and computer networks.

The World Web is based on these technologies:

- HTML—Hypertext Mark-up Language
- HTTP—Hypertext Transfer Protocol
- Web servers and Web browser

Researcher Tim Berners-Lee led the development of the original World Wide Web in the late 1980s and early 1990s. He helped build prototypes of the above Web technologies and coined the term WWW. Web sites and Web browsing exploded in popularity during the mid-1990s." The World Wide Web Consortium exists to realize the full potential of the Web. The W3C is an industry consortium which seeks to promote standards for the evolution of the Web and interoperability

ISSN 2230-7621 © MIT Publications

between WWW products by producing specifications and reference software. Although W3C is funded by industrial members, it is vendor-neutral, and its products are freely available to all. The Consortium is international; jointly hosted by the MIT Laboratory for Computer Science in the United States and in Europe by INRIA who provide both local support and performing core development. The W3C was initially established in collaboration with CERN, where the Web originated, and with support from DARPA and the European Commission.

"Organizations may apply for membership to the Consortium; individual membership isn't offered. The W3C has taken over what was formerly called the CERN Hypertext Transfer Protocol daemon or Web server.

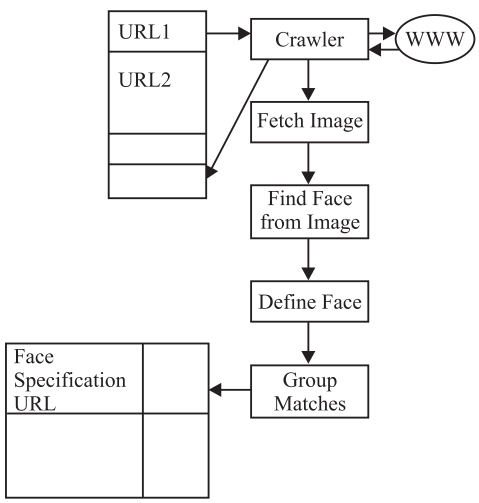

Figure 4: Image Crawling

With reference to Figure 4 description is as follows:

DESCRIPTION OF IMAGE CRAWLING CHART

STEP 1: Crawler

Crawler is a computer program that browses the World Wide Web automated manner or in an orderly fashion. It starts with a list of URLs to visit, called the seeds. As the crawler visits these URLs, it identifies all the hyperlinks in the page and adds them to the list of URLs to visit, called the crawl frontier.

STEP 2: Fetch Image

At this step the images will be fetched from the database, as the database contains a huge amount of images, so all the images will be automatically fetched during this step. This step is also done without the interaction of the user.

STEP 3: Define Images

At this step the fetched images will be defined by giving a tag that helps in distinction of the images, or more precisely all the fetched (similar) images will be grouped together under one tag/category.

STEP 4: Group Matches

In group matches, the group which was already manipulated, we have to check in which group the defined image match the most.

FEATURE EXTRACTION

Feature (content) extraction is the basis of content-based image retrieval. In a broad sense, features may include both text-based features (key words, annotations) and visual features (color, texture, shape, faces). However, since there already exists rich literature on text-based feature extraction in the DBMS and information retrieval research communities, we will confine ourselves to 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 literature and may involve much domain knowledge which we will not have enough space to cover in this paper. Therefore, the remainder of the section will concentrate on those 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.

Color

The 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.

In image retrieval, the 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, an L_1 metric, as the similarity measure for the color histogram. To take into account the similarities between similar but not identical colors, Ioka and Niblack et al introduced an 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 using the cumulated color histogram. Their research results demonstrated the advantages of the proposed approach over the conventional color histogram approach.

Besides the 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 the color histogram, Stricker and Orengo proposed using the color moments approach. The mathematical foundation of this approach is that any color distribution can be characterized by 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 the color histogram. They first transformed the (RGB) 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 colors from the quantized color space. Because color set feature vectors were binary, a binary search tree was constructed to allow a fast search. The relationship between the proposed color sets and the conventional color histogram was further discussed.

Texture

In the early 1970s, Haralick et al. proposed the co-occurrence matrix representation of texture features. 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 and experimentally found out that contrast, inverse deference moment, and entropy had the biggest discriminatory power.

There also were 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. They tested the three methods on two sets of terrain samples and concluded that the Fourier method performed poorly while the other two were comparable. Ohanian and Dubes compared and evaluated four types of texture representations, namely Markov random field representation, multichannel filtering representation, fractal-based representation 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 that co-occurrence matrix representation performed best in their test sets. In a more recent paper, Ma and Manjunath evaluated the texture image annotation by various wavelet transform representations, including orthogonal and bi-orthogonal wavelet transforms, the tree-structured wavelet transform, and the Gabor wavelet transform. They found that the Gabor transform was the best among the tested candidates which matched human vision study results.

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. The most successful representatives for these two categories are Fourier descriptor and moment invariants.

The main idea of a Fourier descriptor is to use the Fourier transformed boundary as the shape feature. 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. The main idea of moment invariants is to use region-based moments which are invariant to transformations, as the shape feature. Hu identified seven such moments. Based on his work, many improved versions emerged 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. 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 a framework of algebraic curves and invariants is proposed to represent complex objects in a 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 includes the Finite Element Method (FEM), the turning function and the wavelet descriptor. The FEM defines a stiffness matrix which describes how each point on the object is connected to the 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 the Fourier descriptor, Arkin et al. developed a turning function-based method for comparing both convex and concave polygons Chuang and Kuo used the 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 the linear dimension, rather than area to further 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 Li and Ma showed that the geometric moments method (region-based) and the Fourier descriptor (boundary-based) were related by a simple linear transformation. 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 and Fourier descriptor, moment invariants and 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 Wallace and Mitchell proposed using a hybrid structural/statistical local shape analysis algorithm for 3D shape representation. Further, Taubin proposed using a set of algebraic moment invariants to represent both 2D and 3D shapes, which greatly reduced the computation required for shape matching.

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 subblocks and extract color features from each of the subblocks. A variation of this approach is the quadtree-based color layout approach, where the entire image was split into a quadtree structure and each tree branch had its own histogram to describe its color content. Although conceptually simple, this regular subblock-based approach cannot 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 to store the position and color set feature of each region to support later queries. 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 lay-out representations were proposed. 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. 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 region transformations. Pass et al. classified each pixel of a particular color as either coherent or incoherent, based on whether or not it is part of a large similarly colored region. By using this approach, widely scattered pixels were distinguished from clustered pixels, thus improving the representation of local color features. Huang et al. proposed a color correlogram based color layout representation. They first constructed a color co-occurrence 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.

Along the same line of the color layout feature, the layout of texture and other visual features can also be constructed to facilitate more advanced image retrieval.

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.

Lybanon et al. researched a morphological operation (opening and closing) 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. Hansen and Higgins exploited the individual strengths of watershed analysis and relaxation labeling. Since fast algorithm exists for the watershed method, they first used the watershed to subdivide an image into catchmen basins. They then used relaxation labeling to refine and update the classification of catchment basins initially obtained from the watershed to take advantage of the relaxation labeling's robustness to noise. 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 histograms do not have clear peaks and valleys. Other segmentation techniques based on Delaunay triangulation, fractals.

All the above-mentioned algorithms are automatic. A major advantage of this type of segmentation algorithms is that it can extract boundaries from a large number of images without occupying human 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.

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. 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. Second, the outline was refined by using dynamic programming. Rui et al. pro-posed 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.

One 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.

IMAGE RETRIEVAL SYSTEMS

QBIC

QBIC 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. Its texture feature is an improved version of the Tamura texture representation; i.e. combinations of coarseness, contrast, and directionality. Its shape feature consists of shape area, circularity, eccentricity, major axis orientation, and a set of algebraic moment invariants QBIC is one of the few systems which takes 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 multidimensional indexing structure. In its new system, text-based key word search can be combined with content based similarity search.

Virage

Virage is a content-based image search engine developed at Virage Inc. Similar to QBIC, Virag 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. 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 useful "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.

RetrievalWare

RetrievalWare is a content-based image retrieval engine developed by Excalibur Technologies Corporation. From one of its early publications, we can see that its emphasis was in neural nets to image retrieval. Its more recent search engine uses color, shape, texture, brightness, color layout, and aspect ratio of the image, as the query feature. It also supports the combinations of these features and allows the users to adjust the weights associated with each feature.

Photobook

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

In its more recent version of Photobook, FourEyes, Picard et al. proposed including human in the image annotation and retrieval loop. 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, a human's perception is subjective. They proposed a "society of model" approach to incorporate the human factor.

VisualSEEk and WebSEEk

VisualSEEk is a visual feature search engine and WebSEEk 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.

The visual features used in their systems are color set and wavelet transform based texture feature. To speed up the retrieval process, they also developed binary tree based indexing algorithms.

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, browse, and retrieval module. It supports queries based on both keywords and visual content.

Netra

Netra is a prototype image retrieval system developed in the UCSB Alexandria Digital Library (ADL) project. 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, neural net-based image thesaurus construction and edge flow-based region segmentation.

MARS

MARS (multimedia analysis and retrieval system) was developed at University of Illinois at Urbana-Champaign. 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), integration of indexing and retrieval (how the retrieval algorithm can take advantage of the underline indexing structure), 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 and integrates such a technique at various levels during retrieval, including query vector refinement, automatic match-ing tool selection, and automatic feature adaption.

CONCLUDING REMARKS

In this paper, past and current technical achievements in visual feature extraction, multidimensional 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' efforts. 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.

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, a wavelet-compressed image is a good choice. 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 key words 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

REFERENCES

[1] A. Andoni. http://web.mit.edu/andoni/www/LSH/i ndex.html.

[2] J.R. Bach, C. Fuller, and et al. The virage image search engine: An open framework for image management. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, volume 2670, pp. 76-87,1996.

[3] N. Beckmann, H.P. Kriegel, R. Schneider and B. Seeger. The r*-tree: An efficient and robust access method for points and rectangles. In *Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, Atlantic City, NJ, May 23-25, 1990*, pp. 322-331. ACM Press, 1990.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. PAMI, p. 509, 2002.

[5] S.A. Berrani, L. Amsaleg and P. Gros. Robust content-based image searches for copyright protection. In *MMDB '03: Proceedings of the1st ACM International Workshop on Multimedia Databases*, pp. 70-77, New York, USA, 2003. ACM Press.

[6] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Proc. 20th Annual Symposium on Computational Geometry*, pp. 253-262, New York, NY, USA, 2004.

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

[8] DP Challenge. http://www.dpchallenge.com/.

[9] R. Egas, D.P. Huijsmans, M.S. Lew and N. Sebe. Adapting k-d trees to visual retrieval. In *Visual Information and Information Systems*, pp. 533-540, 1999.

[10] M.A. Fischler and R.C. Bolles. Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM*, 24(6):381-395, 1981.