

Content Based Image Retrieval “Public and Private Search Engines”

Dr T. Santha

Principal, Dr GRD College of Science.

M .Abhayadev

PhD Research Scholar, Dr GRD College of Science
abhayadevmalayil@gmail.com.

Abstract

This paper presents a brief overview of content-based image retrieval search engines. Search engines are commonly classified into private and public. Both are used for commercial purpose. This paper explains the Image Retrieval and Content Based Image Retrieval (CBIR) systems, and a set of CBIR working levels. It also discusses the various challenges of CBIR, which need to be taken into account when designing any kind of image retrieving system.

Keywords: CBIR, image retrieval, TBIR, QBIR, Semantic Content.

1. INTRODUCTION

Nowadays the use of multimedia data is massive, because of the difference in the type of electronic devices used. Humans cannot have grip on things like capturing, storing, indexing, retrieving, analyzing, and summarizing the data used in these devices. The images have an irreplaceable role for all multimedia applications. It is highly imperative for the user to have a perfect image mining system. Earlier, this image retrieval system was termed as text based image retrieval. Unfortunately, this system has many drawbacks. Textural description is not capable of capturing visual content. The Image content can be expressed in different ways and images are beyond the description of words. In order to overcome these issues, a new retrieval technique called CBIR (Content Base Image Retrieval) for image mining is being widely used and practiced.

The image retrieval feather is primitive and semantic but CBIR technique is totally automatic. There are two different ways in which the CBIR process can be explained. A single picture compares with an individual database record to verify the identity. In this process only two images are taken for comparison. The whole database may be searched to find the closest matching image this process is a genuine example of CBIR process.

In 1990 a new technology was introduced called as content based image retrieval. Content based image retrieval helps the image data management system. Image mining is rapidly gaining attention among researchers in the field of data mining, information retrieval, multimedia data base. It is because of its potential in discovering useful pattern in different image patterns. (Hsu, Wynne, et al 2002). CBIR contains many components such as feature extraction and representation, similarity measurement, databases of pre-analyzed image collections, and relevance feedback.

The images are retrieving from a large collection of image data base. The retrieving images having some features such as colour, texture and shape. The feathers can be automatically extracted from the images data base. Different types of image retrieval systems are used for image retrieval (Remco C. et al 2001). This paper explains the importances of CBIR search engines for Private and Public use and the challenges facing CBIR technology in image mining area.

2 REVIEW OF LITERATURE

Usage of digital cameras would increase the volume of the digital images. It was hard for computer engineers to retrieve useful information from them. Content Based Search Engines are tools that utilize the rich image content and apply pattern recognition methods in order to retrieve similar images. (Brilakis, et al 2005). Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were annotated with text and then searched using a text-based approach from traditional database management systems (Singh, B., & Singh, 2010).

From the mid 90's to early 2000, the process of Image Retrieval using content was introduced. Before its introduction, content based image retrieval system was an old system called text based image retrieval. It had many drawbacks like an image feathers cannot be explain properly using a text. TBIR having two type of gap problem, Sensory and Semantic Gap, The Sensory Gap was the gap between the object in the world and the information in a description derived from the recording of a scène. The Semantic gap was the lack of coincidence between the information that one can extract from the visual data (Ritendra Datta, et al 2008). At that time three typses of image search were introduced Search by association, Aimed search and category search, all these type of image search

were divided in to two parts namely image processing and feature construction but the problem related to the gap was not solved.

3 IMAGE RETRIEVAL

It is a process of searching and receiving image from a huge image data base, the image retrieval process contain information retrieval and image processing, a common image content having both visual content and semantic content , visual content is very general and domain specific it include color, texture, relationship. Some time visual content may be domain specific, the semantic content uses some complex interface procedure for image content analysis.

3.1 CONTENT BASED IMAGE RETRIEVAL

Content-Based Image Retrieval (CBIR), also known as Query By Image Content (QBIC) and Content-Based Visual Information Retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based image retrieval is opposed to traditional concept based approaches.

CBIR image retrieval process will help for overcoming the image mining problems. Google, Yandex and other search engines are used for image retrieval purpose. The CBIR technology is much better, than old technologies.

Chabot(V. E. Ogle and M. Stonebraker. Chabot. 1995) integrates image content retrieving based on color information with text-based queries. Its interface allows user to search and update the image database. This system does not include texture and shape descriptors.

In recent years CBIR model has been a major topic of research. It has been explored from many different points of view from early heuristic-based feature weighting schemes are introduced (Y.Ishikawa, et al 1998).There are many proposed optimal learning algorithms, probabilistic Bayesian learning algorithms, boosting techniques, discriminate E M algorithm (Y.Rui, et al 1998), biased discriminate algorithms (X.S. Zhou, et al 2001).

The CBIR consists of four main steps Query Image, Feather Extraction, Initial Feedback, and Retrieved Similar Image. Besides investigating suitable frameworks for image mining.Early image miners have attempted to use existing techniques to mine for image information. The techniques frequently used include object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network (Ji Zhang, et al).

3.2 LEVELS OF CBIR AND WORK FLOW

DIAGRAM

Level 1:

Based on Color, Texture, Shape Features

In level one the image having basic feathers like color texture and shape features, color having so much of

patterns (RGB), visual patterns in the image is called as textures brightness of an image. The shape is not directly involved in to the image but it is comes on a particular regain of that image.

Level 2:

Bring Semantic Meanings into the Search

Identify the domain of the mining image example (identifying human beings, horses, trees, beaches) then requires retrieval techniques of level 1

Level 3:

Retrieval with Abstract and Subjective Attributes

This stage requires retrieval techniques and getting relevant image from the image data base.

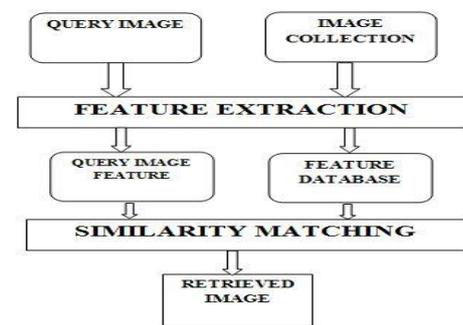


Fig. 1 Architecture of Content based image retrieval system.

The above fig.1 represents the CBIR search engine architecture. CBIR search engine retrieving images based on above work flow diagram.

4 APPLICATIONS OF CBIR

The CBIR technology was used in several applications such as Automatic Face Recognition Systems, Medical Image Databases and Diagnosis, Biodiversity Information Systems, Manage Image Archives and Historical Research.

4.1 AUTOMATIC FACE RECOGNITION SYSTEMS

Face retrieval problem is concerned with retrieving facial images that are relevant to users' requests from a collection of images. The retrieving is based on the visual contents or by information associated with this facial image. The basic image retrieval systems mostly use low level visual features such as color, texture and shape features [Datta, R., et al.2008) [Lew, M.S., et al. 2006).

4.2 MEDICAL IMAGE DATABASES AND DIAGNOSIS

There is large number of images generated by hospitals and clinics every day. These images play very important role in diagnosis of diseases, medical research and education. Major hospitals store their patients records as image database namely CT, MRA, Ultrasound scan images, they are used for disease identification with the help of

CBIR system. The scanned images are analyzed and finding the human decease.

4.3 BIODIVERSITY INFORMATION SYSTEMS

Many images of living beings are analyzed by Biologists and identify the species. The biodiversity information system (BIS) help the researchers to fullfill the knowledge and understanding different species by textual image and geographical queries.

4.4 MANAGE IMAGE ARCHIVES AND DIGITAL LIBRARY

Digital museum of butterflies [J. S. Hong, et al 2000), is an example of CBIR system supporting Digital library. The library contains digital collection of Taiwanese butterflies. The library or image archive works the feathers like color, texture, and patterns. Many important historical books are stored in government library as image archive.

5 CBIR SEARCH ENGINES

Almost 25 search engines used for CBIR image referral purpose in this only few important search engines are listed below for their growing importance in modern world. Many of the search engines are controlled by private companies. But public search engines are also available. The entire search engine’s images come under different domain and they use different platforms.

Table.1.Commercial CBIR Search Engines Private Usage

Name	Description	External Image Query	Meta Data Query	Index Size Or Image Size	Organization Type	License Open/Closed
Picalike	CBIR engine for Mobile and eCommerce	No	No (additional filters can be added)	No limit	Private Company	Closed
Id my pill	Automatic prescription pill identification (CBIR)	Yes	No	No limit	Private Company	Open (via API)
Imense image search portal	CBIR search engine, by Imense.	No	Yes	3M	Private Company	Closed
Id reader	CBIR search engine, by Imense.	No	Yes	3M	Private Company	Closed
Imprezzeo image search	CBIR search engine, by Imprezzeo.	No	Yes	No limit	Private Company	Closed

Table 1 explains some important fetchers of five Private Search Engines.

5.1 PICALIKE

Picalike is the search engine for mobile and e-commerce products it search based on visual and similarity search, based on the customer request the Picalike search engine detect product similar to the reference object in real time. All process is carried out according to the image we give as inputs. The final result based on some criteria such

as brand, price, name etc .The system analysis over 600 product characteristics such as color, shape, color gradient, texture etc for searching similar images.

5.2 ID MY PILL

This is an iphone application to select right pill. It identifies each person’s prescription of pills and provides correct drug for their needs also gave drug name and detailed note for that drug usage

5.3 IMENSE IMAGE SEARCH PORTAL

It works on the basses of “SEARCH INSIDE THE IMAGE “use computer vision technology to extract visual feathers from image as mathematical model, this model compared against defined parameters for particular concepts and where the feathers occur spatially. And the image is assigned to the visual ontology. Image fetchers are extracted in the classification process. The feathers are stored in visual index along with Meta Data and User query Meta Data related image can be searched out.

5.4 ID READER

Id reader captures and automatically reads different format of ID cards all with a single click. The application run on android smart phone, the systems use Optical Character Reorganization Technology (OCR) to enable near instantiates id reading in a single click. The captured image is send to the data base for validation and within seconds the valid id holder picture to the mobile device is determine whether the valid id belong to the holder or if it is stolen.

5.5 IMPREZZEO IMAGE SEARCH

This search engine provides the user for a second solution to quickly find image even if the data base is large. It use two base technologies content base image retrieval CBIR and face recognitions FR, the similarity measure is done analyzing actual contend and attributes of the sample set of image color, shape, layout, texture and face recognition information. Imprezzeo image site is an enterprise solution that can scale to work across image collections ranging from thousand to ten thousand.

Table.2.Commercial CBIR Search Engines Public Usage

Name	Description	External Image Query	Meta Data Query	Index Size Or Image Size	Organization Type	License Open/Closed
Google Image Search	Google’s CBIR system, note: does not work on all images	Yes	Yes	No limit	Public Company	Closed
Yandex Image Search	Yandex CBIR system	Yes	Yes	10000M	Public Company	Closed
Baidu Image Search	Baidu’s CBIR system.	Yes	Yes	10000M	Public Company	Closed
eBay Image Search	Image Search for eBay Fashion.	No	Yes	20M	Public Company	Closed

Table 2 explains some important fetchers of four Public Search Engines.

5.6 GOOGLE IMAGE SEARCH

Google image is an image search engine owned by Google Corporation. It is introduced in 2001. The Google image search allow user to search web content image. The search engine works based on CBIR technology.

5.7 YANDEX IMAGE SEARCH

Yandex image search helps the user to find particular format, specific color and various size images. It is a Russian internet company. This Russian internet company is the largest search engine in Russia. Lunched on September 1993, it comes forth place all around the world.

5.8 BAIDU IMAGE SEARCH

Baidu image search engine is the search engine of the people republic of china. It is a pioneering content-based image retrieval (CBIR) search engine. Different from traditional image search engines matching user query key words with surrounding text of images, stu.baidu.com enables users to directly use image as the query. By Simply uploading a local image or pasting the URL of a web image into the search box, users can easily get access to many other web images same or similar to the query. In addition, to this image send also gives stories, news, knowledge and other web information.

5.9 EBAY IMAGE SEARCH

eBay is a search engine based on CBIR technology. It is an image search engine as well as an online shopping cart for purchasing product.

6 CHALLENGES IN CBIR

The whole image retrieval research area does not have a universally acceptable algorithm for characterizing digital images. Hence, it is not surprise to see much research focus in this area. By the nature of its task, the CBIR technology collapse down to two intrinsic problems. An image cannot be mathematically described based on the abstracted description. And it is impossible be predict the similarity of pair of images.

- The semantic gap in the image and their textual description need not fully obey with the available technique. So new tool need to be developed for image retrieval.
- The Image contends description and its related service has to be followed by some rules, it can help the implementation of new application based on image.
- Taking advantage of domain ontology's better semantically enriched descriptions can be created for CBIR technology.
- Need for tools that automatically extract semantic features from images, extract high level concepts contained in multimedia data.
- To develop new data fusion algorithm to support text based and content based image retrieval.
- Finding new connection and mining pattern, Text mining techniques might be combined with visual based description.

- Researchers in Indian institutions lag so much of information resources.

7 COMPARATIVE STUDY OF PRIVATE AND PUBLIC SEARCH ENGINES

Score and worth of each search engines are calculated and represented in the following tableIII. The score of each search engines are also plotted in a bar diagram <http://www.worthofweb.com/website-value/picalike>. The name of the score generator site is "worthofweb". This sight using public ranking systems which are more accurate for score generation. Based on the public ranking system each search engines score and worth listed in the table.

Table.3.Score and worth of each search engines

Search Engine Name	Public/Private	Score	Worth
Picalike	Private	58/100	\$2496
Idmy pill	Private	52/100	\$1155
Idreader	Private	46/100	\$503
Imprezzeo	Private	36/100	\$130
Google image	Public	49/100	\$ 10,255
Yandex	Public	88/100	\$544,000,000
Baidu	Public	98/100	\$10,000,000,000
Ebay	Public	96/100	\$6,070,000,000

Table.3 report generates score and worth of search engines. The score generator site totally evaluated 9,708,708 sites at a time.

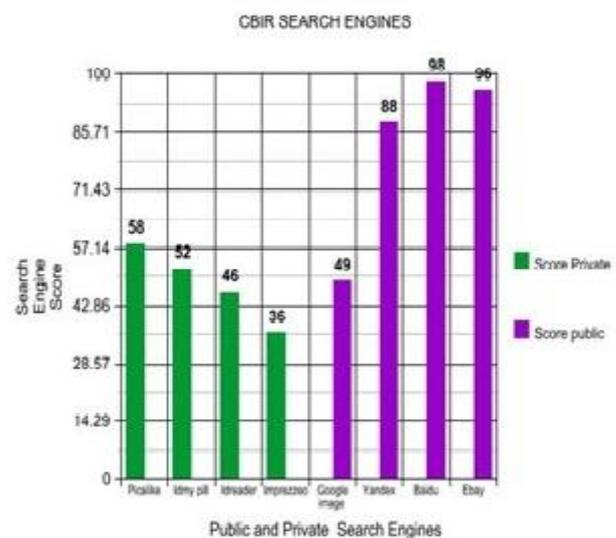


Fig.2. Bar diagram for public and private search engine comparison

The fig.2 represents the graphical comparison of public and private search engines. The graph is plotted based on the Table 3 score values. It is clear that all public search engines score are more than private search engines. Because many users access public search engine than private and all public search engines are platform

independents. All public search engines are open access search engines.

8 CONCLUSIONS

The Research work is a part of large project to express the usage of image Through Image Processing and CBIR technologies. In nutshell the aim is to explain some CBIR search engines (public, private) comparison and challenges. It is clear that the public search engines are open access and user friendly. It gives accurate image search results. It is impossible to manually retrieve image feathers directly from a digital image. There for many images cannot be captured and analyzed using human eyes. With the help of a digital camera and CBIR technology these types of images are captured and analyzed.

REFERENCES

- Brilakis, Ioannis, and Lucio Soibelman. "Content-based search engines for construction image databases." *Automation in Construction* 14.4 (2005): 537-550.
- Datta, R., et al., Image Rsetrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys (CSUR)*, 40 (2): (2008):1-60.
- Hsu, Wynne, Mong Li Lee, and Ji Zhang. "Image mining: Trends and developments." *Journal of Intelligent Information Systems* 19.1 (2002): 7-23.
- J. S. Hong, H.Y. Chen, and J. Hsiang. A Digital Museum of Taiwanese Butterflies. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, an Antonio, Texas, United States, ACM Press, (2000): 260–261.
- J.P. Eakins. Automatic image content retrieval are we getting anywhere? *De Montfort University*, May (1996):123–135.
- Ji, Zhang, , Wynne Hsu, and Mong Li Lee. "Image mining: Issues, frameworks and techniques." *Proceedings of the 2nd ACM SIGKDD International Workshop on Multimedia Data Mining (MDM/KDD'01)*. University of Alberta, (2001).
- Lew, M.S., et al., Content-Based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2(1) (2006): 1-19.
- Remco C. Veltkamp, Mirela Tanase, Danielle Sent ,State-of-the-Art in Content-Based Image and Video Retrieval , Features in Content-Based Image Retrieval Systems: A Survey, *Computational Imaging and Vision* Volume 22, (2001):97-124
- Ritendra Datta,, et al. "Image retrieval: Ideas, influences, and trends of the new age." *ACM Computing Surveys (CSUR)* 40.2 (2008): 5.

Singh, B., & Singh, H. K. Web data mining research: a survey. In *Computational Intelligence and Computing Research (ICCIC)*, IEEE International Conference (2010): 1-10.

V. E. Ogle and M. Stonebraker. Chabot: Retrieval from Relational Database of Images. *IEEE Computer*, 28(9), Sep(1995): 40–48

Xiang Sean Zhou, and Thomas S. Huang. "Relevance feedback in image retrieval: A comprehensive review." *Multimedia systems* 8.6 (2003): 536-544.

Y.Rui, T.S. Huang, S.Mehrotra."Relevance feedback technique s in inter active content-based image retrieval", *Storage and Retrieval for Image and Video Databases VI*, January(1998):25–36.

BIOGRAPHY

Dr.Santha T received her post graduation from P.S.G. College of Arts and Science, Coimbatore in 1987 and M.Phil. from Bharathiar University in 1991. She also completed her PhD in Computer Science from Bharathiar University during 2012. She is presently working as Principal of Dr.G.R.D. College of Science, Coimbatore. She has more than two decades of teaching and research experience. Her research centres on Computer Applications and Optimization Techniques, Multimodal Transportation and Computer Simulation.

Abhayadev M received MCA and M.Phil from Kongunadu Arts and Science College Coimbatore in 2011, 2014 respectively. He is presently a Computer Science PhD Research Scholar in Dr.G.R.D. College of Science, Coimbatore .He is doing his Research under the guidance of Dr.T Santha Principal of Dr.G.R.D. College of Science. His area of interest is Data Mining and Warehousing.