

# Web image retrieval based on cloud computing model

ZHIFANG ZHANG<sup>1</sup>

**Abstract.** As Internet technology has been popularized in people's lives; network data grow at an explosive rate and there is growing demand from users, it is increasingly difficult to obtain demand information from massive data. Since computer and server architecture in distributed computing systems can hardly meet actual demand in the case of the surge in data volume, enterprises and individuals have to overcome many obstacles in the process of obtaining data by using computers. How to build high-performance distributed applications has become a key concern of the vast majority of scholars. For image retrieval, web image retrieval based on cloud computing models has been researched, which has markedly improved the efficiency of retrieval. IBM Almaden Research Center developed the QBIC system for image retrieval in the 1990s and used it for commercial search. It has diverse query features, including color, texture, and silhouette, and enables automatic segmentation of image objects [1]. The QBIC system has a relatively rich set of interfaces that can be interfaced with multiple applications. Virage, Inc. has developed a content-based image search engine that is similar to QBIC and supports retrieval of color, texture, and silhouette features. In addition, this engine can effectively combine image features, so that users can change the query mode by giving different weights in tune with actual demand.

**Key words.** Cloud computing, WEB, image retrieval.

## 1. Introduction

As Internet technology has been popularized in people's lives; network data grow at an explosive rate and there is growing demand from users, it is increasingly difficult to obtain demand information from massive data. Since computer and server architecture in distributed computing systems can hardly meet actual demand in the case of the surge in data volume, enterprises and individuals have to overcome many obstacles in the process of obtaining data by using computers. How to build high-performance distributed applications has become a key concern of the vast majority of scholars. For image retrieval, web image retrieval based on cloud computing models has been researched, which has markedly improved the efficiency of retrieval. IBM Almaden Research Center developed the QBIC system for image retrieval in

---

<sup>1</sup>Ordos Vocational College, Inner Mongolia Ordos, 017009, China

the 1990s and used it for commercial search. It has diverse query features, including color, texture, and silhouette, and enables automatic segmentation of image objects [1]. The QBIC system has a relatively rich set of interfaces that can be interfaced with multiple applications. Virage, Inc., has developed a content-based image search engine that is similar to QBIC and supports retrieval of color, texture, and silhouette features. In addition, this engine can effectively combine image features, so that users can change the query mode by giving different weights in tune with actual demand.

TinEye is a reverse image search engine. Users can find out where an image came from only by inputting the image in the software. Compared to traditional image search software, TinEye is more accurate. It can only retrieve images that have been indexed in the database, and cannot retrieve unindexed images, which means the search will return no results. ImageRover can use themes as a clue for browsing, and is equipped with the traditional keyword search function, which means to search relevant content by entering keywords in the search bar. In addition to precise search, the system is also capable of fuzzy searching, which is to look for contents similar to keywords in the webpage [2].

## 2. Materials and methods

### 2.1. Overview

Cloud computing is a new information resource processing model that relies on back-end computer capabilities. It connects many end-users to the system by interface means which are different from traditional ones, thereby achieving multiple resource retrieval functions [3]. Users can use this model via the Internet, which can reduce costs of background operations and maintenance. Moreover, suppliers only need to install a unified set of hardware and software systems and develop a network environment that is suitable for all users. In this case, many users can use this platform.

Hadoop is a distributed system foundation framework. When using Hadoop, users are freed from needing to understand all those underlying details, so it is practical and promotional [4]. At present, many companies use Hadoop as the foundation to support the building of their computer platforms. The core components include HDFS, Map Reduce and HBase. HDFS frame structure is shown in Fig. 1.

An HDFS uses a structural model of master/slave. In general, an HDFS contains a single Namenode and a number of Datanodes, in which Namenode acts as central server. It manages the file system's namespace and users' access through the client. Datanodes are generally used in conjunction with node significance, with the effect of managing its own node storage. The primary role of Datanodes is to serve read and write requests from the clients on the file system, and also perform block creation and deletion through unified scheduling. As per HDFS frame structure shown in Fig. 1, an HDFS contains a single Namenode and a number of Datanodes [5].

Map Reduce. Research has found that most abstracted operations can be transformed into Map Reduce. Map Reduce can be divided into two parts: Map and Reduce, in which Map is responsible for decomposing Input into Key/Value, and

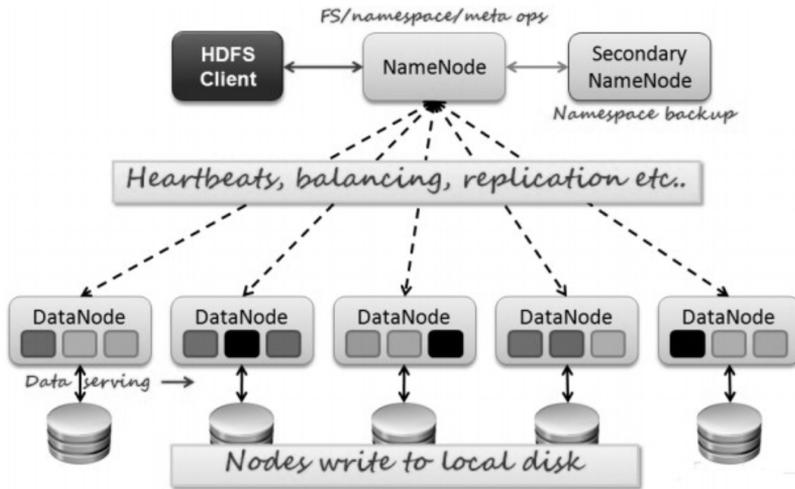


Fig. 1. HDFS frame structure

then synthesizing and outputting the Key/Value obtained through Reduce. Map and Reduce can be run in the cluster, and the results will be stored in the distributed file system. The system algorithm is mainly edited by the programmer who assigns corresponding functions according to actual needs. The Map Reduce framework is shown in Fig. 2.

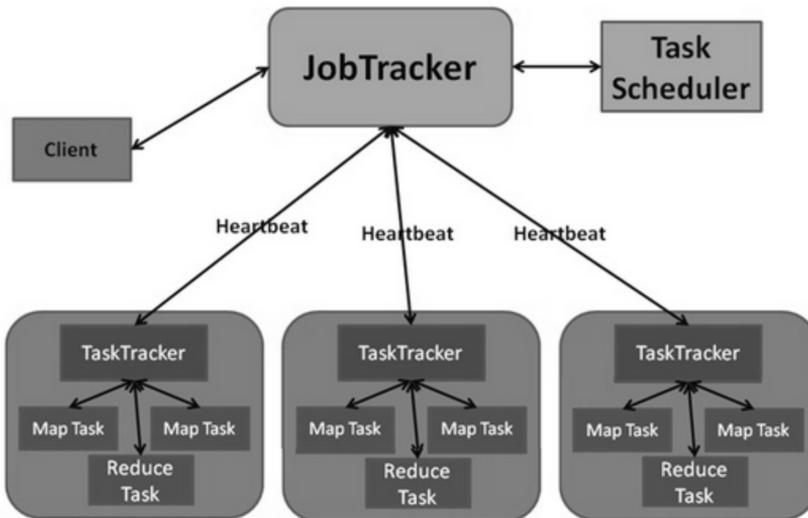


Fig. 2. Diagram of the Map/Reduce framework

Fig. 2 shows that Map Reduce often breaks data down into number of independent modules in the operation. First of all, Map processes these independent modules,

sorts the output of the processed data, and feeds back the results to Reduce. In the Map Reduce operation, input and output are stored in the file system, and task scheduling and monitoring are carried out by the whole frame. The task system that fails in system running is automatically reexecuted.

HBase is an open-source, distributed database that differs from traditional databases [6]. HBase uses a data model very similar to that of Bigtable. In the data storage process, each data row has a sortable key and an arbitrary number of columns. The table is stored sparsely, and users can make different definitions of it. In the visit process, HBase is random and can store large user data in coordination with the system in real time.

## *2.2. Content-based image retrieval*

The image retrieval process is to find image features quickly, timely and effectively, and to find objects with similar characteristics. Retrieval characteristics include image characteristics, shape and texture, among others [7].

Color is an important feature for image retrieval, and is currently the most common bases to retrieve images. Color is the most widely used underlying visual feature. Compared to visual features of other attributes, color has the least reliance on image rotation, translation and size changes, and makes image retrieval relatively simple in the system retrieval process, so it is a retrieval way that is easy to implement [8].

Shape is another important feature for image retrieval. The shape feature provides a high-level visual retrieval way, which is an advanced retrieval method developed on the traditional retrieval methods, and can further find the semantics of images during the retrieval process. Firstly, we need to combine the bottom-level features with high-level features, and express the features by using advanced algorithms. It is hard to obtain the shape of an object. Therefore, it is difficult to retrieve images based on the shape feature in applications and this retrieval method is suitable for images that are easily recognizable.

Texture feature. Texture is one of the basic characteristics of an image, and an important basis for visual presentation. It can effectively reflect the surface features of an object. Texture is independent of color, brightness and other characteristics of an image. The essence is the distribution of pixels in the image space. Pixels are distributed either densely or loosely. The basic components that make up texture are called primitives. Texture features mainly include directions, linearity, roughness, etc. These can be used as an important basis for texture retrieval [9].

## *2.3. Feature matching*

Objects identical or similar to the matching image are found during image retrieval. Currently there are two types of feature matching: exact match and similar match. Exact matching mainly means that the image is exactly the same as the reference object. In this case, the two images are basically a copy of the same object. In addition, we can also set a threshold. If the image falls within the specified

range, it indicates a similar match. In other words, the search image is similar to the reference image. Through using image content as a retrieval basis, image matching can be done by a corresponding measurement strategy after retrieving the image feature, which is to determine the similarity between the retrieval image and the feature vector distances in the database. The image feature can be converted to any number of eigenvectors, e.g.,  $n$ -dimensional feature vectors in images  $A$ ,  $B$ , and  $C$ ,  $d$  refers to the distance on the basis of which the following axiom constraints are satisfied [10]:

$$d(A, A) = d(B, B) = 0, \quad (1)$$

$$d(A, B) \geq d(A, A) = 0, \quad (2)$$

$$d(A, B) \geq d(B, A), \quad (3)$$

$$d(A, C) \leq d(A, B) + d(B, C), \quad (4)$$

The algorithm used in the actual image retrieval process does not completely meet the above four axioms, but generally meets one or several of them. In view of this, image feature matching can be conducted. At present, common image feature matching algorithms include the Minkowski distance, the Manhattan distance, the Euclidean distance, and the Mahalanobis distance, respectively given by [11]

$$d(A, C) \leq d(A, B) + d(B, C), \quad (5)$$

$$d(X, Y) = \sum_{i=1}^n |x_i - y_i|, \quad (6)$$

$$d(X, Y) = \left( \sum_{i=1}^n w(x_i - y_i)^2 \right)^{\frac{1}{2}}, \quad (7)$$

$$d(X, Y) = \sum_{i=0}^n \sum_{j=0}^n (x_i - y_i) \cdot a_{ij} (x_j - y_j), \quad a_{ij} \in A, \quad (8)$$

where  $A$  in (8) denotes the covariance matrix. Symbol  $w$  refers to the weight and  $\alpha$  refers to the attribute value.

### 3. Results

As previously mentioned, cloud computing platform and images serve as the basis. In this paper, a set of image retrieval systems are designed on the basis of cloud computing platforms. The system supports retrieval of image input information, and retrieves similar or same images in webpage.

### ***3.1. Needs analysis of image retrieval system***

The traditional image retrieval system based on cloud computing platforms fails to use images to retrieve input information [12]. The current image retrieval system is built upon the traditional database, which has a limited image storage capacity and thus can hardly adapt to needs of retrieving massive image information. The traditional retrieval method can retrieve some simple images intelligently, while many images are intricate at present, so the traditional image retrieval cannot meet the actual demand.

Focusing on the above problems, the system needs to meet the following demands in the design: the system is built upon efficient databases and can store images efficiently. It has large storage capacity and can meet the future needs. Moreover, it is scalable to some extent. When the capacity or function cannot meet the demand, users can increase its functionality and capacity; users can quickly retrieve the required image through the image retrieval system, which is faster and more accurate than traditional retrieval systems. The system can provide users with a more convenient, user-friendly retrieval mode, and users can master the retrieval function through basic learning. The system can provide users with a good interactive interface, through which users can retrieve images efficiently.

### ***3.2. Overall design***

In combination with the needs analysis, subsystems like image retrieval, feature extraction, cloud computing and user interaction are set up when designing the web image retrieval system based on cloud computing models. The structure of web image retrieval system based on cloud computing model is shown in Fig. 3.

As shown in Fig. 3, image retrieval is the central module of the whole system and has an important effect on the system image retrieval results [13]. The image retrieval module can realize image analysis and processing, and extract features from the image. The control input feature extraction module mainly provides platform support for image retrieval and feature extraction and enhances the system's operation efficiency. The user interaction module is responsible for providing users with a scientific interactive interface, through which users can achieve their own needs and view the test results on the terminal display interface. The IQL engine module is responsible for interpreting the statement.

### ***3.3. Module functions***

Feature extraction subsystem module. Feature extraction needs to be based on the database generation subsystem, which is a core of the whole system. It is responsible for preprocessing the image information collected to get a clearer image, extract the feature of the processed image and store the feature values acquired in a specific database. On this basis the feature extraction process is realized. Image preprocessing is to preprocess an image. The image processing process consists of rotation, translation, zooming and mirroring. It also provides a basis for later image

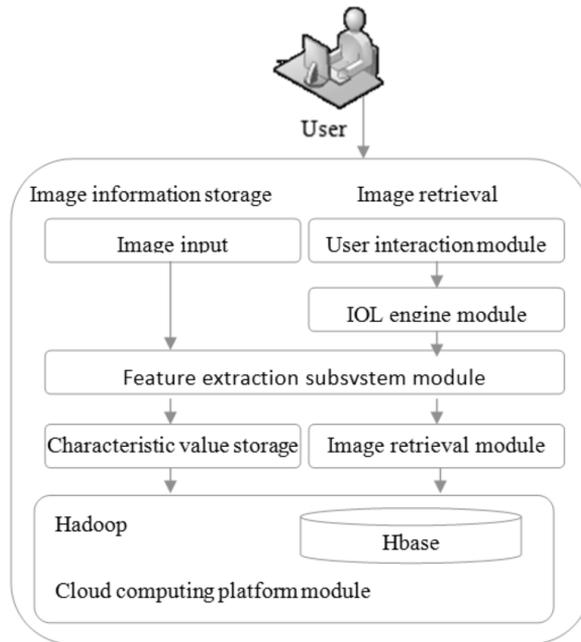


Fig. 3. Structure diagram of web image retrieval system based on cloud computing models

segmentation, such as sharpening and blurring the image, removing the image noise. A diagram of image preprocessing functions is shown in Fig. 4.

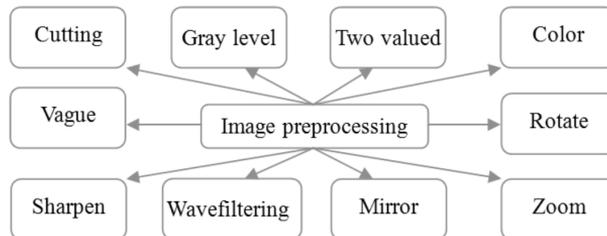


Fig. 4. Image preprocessing process

Image extraction is to extract a key factor from the image, and use the factor as an index of image retrieval. Key factors encompass color, texture, shape and other characteristics, which are the focus of this paper. The image retrieval submodule is a core module of the entire system. It is responsible for retrieving images that resemble the target image from the database. In the retrieval process, the feature is extracted from the target image according to the extraction function, and then it judges whether single or multiple query is used to retrieve images. The appropriate interface is selected in tune with the actual situation, and the image feature is matched with the feature database to object a distance value close to the

target image feature value and sort in ascending order to display the image to the user through the display terminal [14].

### 3.4. Module functions

Cloud computing platform module. Cloud computing platform module is the basis of establishing this system. In this paper, Hadoop is taken as a basis to quickly build the cloud computing platform through HDFS. The Hadoop system has high stability and builds the system operating platform together with Ubuntu. The cloud computing platform framework for image retrieval is shown in Fig. 5.

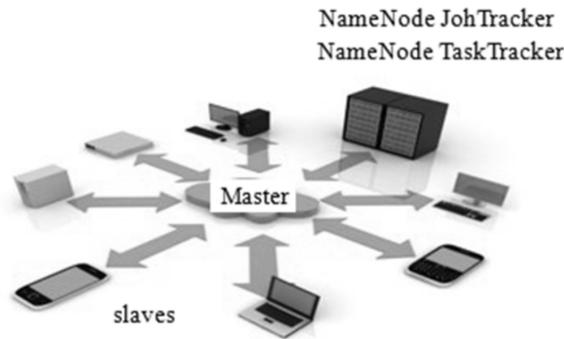


Fig. 5. The cloud computing platform framework for image retrieval

It can be seen from Fig. 5 that the cloud computing model assigns tasks to server nodes via a master node. On the basis of building the framework, the image retrieval process can be realized by combining subsystems such as user interaction and IQL engine module. In combination with the needs analysis, subsystems like image retrieval, feature extraction, cloud computing and user interaction are set up when designing the web image retrieval system based on cloud computing models [15]. Image retrieval is the central module of the whole system and has an important effect on the system image retrieval results. The image retrieval module can realize image analysis and processing, and extract features from the image. The control input feature extraction module mainly provides platform support for image retrieval and feature extraction and enhances the system's operation efficiency.

## 4. Conclusion

Cloud computing is a new information resource processing model that relies on back-end computer capabilities. It connects many end-users to the system by interface means which are different from traditional ones, thereby achieving multiple resource retrieval functions. The image retrieval process is to find image features quickly, timely and effectively, and to find objects with similar characteristics. Retrieval characteristics include image characteristics, shape and texture, among others. Color is an important feature for image retrieval, and is currently the most

common bases to retrieve images. Shape is another important feature for image retrieval. The shape feature provides a high-level visual retrieval way, which is an advanced retrieval method developed on the traditional retrieval methods, and can further find the semantics of images during the retrieval process. Texture is one of the basic characteristics of an image, and an important basis for visual presentation. It can effectively reflect the surface features of an object. Texture is independent of color, brightness and other characteristics of an image. Objects identical or similar to the matching image are found during image retrieval. Currently there are two types of feature matching: exact match and similar match. In combination with the needs analysis, subsystems like image retrieval, feature extraction, cloud computing and user interaction are set up when designing the web image retrieval system based on cloud computing models.

## References

- [1] A. Q. H. AL-NEAMI: *Measurement of trace elements association with diabetes mellitus based on atomic absorption spectrophotometers*. *Advances in Life Science and Technology* 24 (2014), No. 2, 21–27.
- [2] M. Z. JORA, J. A. NÓBREGA, J. J. R. ROHWEDDER, C. PASQUINI: *Double tungsten coil atomic absorption spectrometer based on an acousto-optic tunable filter*. *Spectrochimica Acta Part B: Atomic Spectroscopy* 103–104 (2015), 1–8.
- [3] H. TAVALLALI, H. NOROZIKHAN: *Design of cold vapor system and assembled on atomic absorption spectrometer for mercury determination in several waste water samples*. *International Journal of ChemTech Research* 1 (2009), No. 2, 390–393.
- [4] X. T. YANG, Y. X. ZHANG, Y. WANG: *Tow novel equipments of zeeman background correction for atomic absorption spectrophotometer*. *Chinese Journal of Analytical Chemistry* 39 (2011), No. 10, 1517–1520.
- [5] M. XIA, T. J. YING: *Indirect determination of lactoferrin in milk powder by atomic absorption spectrophotometer*. *Chinese Journal of Pharmaceutical Analysis* 31 (2011), No. 1, 173–175.
- [6] Y. G. ZHANG, M. XIAO, Y. H. DONG, Y. JIANG: *Determination of soil exchangeable base cations by using atomic absorption spectrophotometer and extraction with ammonium acetate*. *Spectroscopy and Spectral Analysis* 32 (2012), No. 8, 2242–2245.
- [7] X. Y. BIAN, R. X. ZHU, Z. T. WANG: *Software design for atomic absorption spectrometer based on LabVIEW and database*. *Microcomputer Information* 24 (2008), No. 16, 117–119.
- [8] S. K. LU, D. X. HE, X. P. LI, S. TIAN: *Adaptability of atomic absorption spectrophotometer based on semi-floating optical base for vehicle application*. *Chinese Journal of Analysis Laboratory* 30 (2011), No. 4, 26–29.
- [9] J. WU, J. H. ZHANG, X. S. ZHU: *A self-made circulating water cooling system of atomic absorption spectrometer*. *Chinese Journal of Spectroscopy Laboratory* 28 (2011), No. 6, 3268–3270.
- [10] T. G. KAZI, N. JALBANI, J. A. BAIG, G. A. KANDHRO, H. I. AFRIDI, M. B. ARAIN, M. K. JAMALI: *Assessment of toxic metals in raw and processed milk samples using electrothermal atomic absorption spectrophotometer*. *Food and Chemical Toxicology* 47, (2009), No. 9, 2163–2169.
- [11] I. A. M. SHAIKAL, F. KHANOM, M. A. RAHMAN, A. M. SHAFIQU ALAM: *Separation and determination of metals in mixed solvent system on anion exchanger using atomic absorption spectrophotometer*. *Pakistan Journal of Analytical & Environmental Chemistry* 6 (2005), No. 2, 35–41.

- [12] Z. M. MEMON, A. M. SHAH, T. G. KAZI: *Study of calcium in active pulmonary tuberculosis patients by flame atomic absorption spectrophotometer*. Indo-American Journal of Pharmaceutical Research 4, (2014), No. 11, 45–48.
- [13] Z. F. DU, Y. QIU: *Discussion of improving energy of 4600 atomic absorption spectrophotometer with a tandem atomizer*. Optical Instruments (2011), No. 4, 34–39.
- [14] S. NIMMAGADDA, A. JYOTHI: *Detection and identification of chemical agent using atomic absorption spectrophotometer*. International Journal of Research in Engineering and Technology 2 (2013), No. 5, 764–768.
- [15] X. ZHENG: *Application of automatic sampler mixed mouth function of shimadzu AA-7000 series atomic absorption spectrometer*. Chemical Analysis and Meterage (2014), No. 6, 23–29.

Received May 22, 2017