7th International Conference on Communication, Computing and Virtualization 2016

Architecture for Software as a Service (SaaS) Model of CBIR on Hybrid Cloud of Microsoft Azure

Mamta Meena, Anamika Ram Singh, Vinayak A. Bharadi

PhD Scholar, Shri Jagdishprasad Jharmal Tibrewala University, jhunjhunu -333001
ME Student, Thakur College of Engineering & Technology, Kandivali (E),Mumbai-400101
Associate Professor, Thakur College of Engineering & Technology, Kandivali (E),Mumbai-400101

Abstract

In various areas of government, academia, hospitals and commerce large collections of digital images are produced. Many of these collections are the merchandise of digitizing existing collections of analogue drawings, photographs, paintings and prints. Generally the only way of searching these collections was by basically browsing or keyword indexing. Digital images databases however, open the way to content-based searching. Content Based Image Retrieval (CBIR) is concerned with the retrieval of images similar to a specified image, from an image repository. Content Based Image Retrieval (CBIR) is an efficient retrieval of relevant images from large databases based on features extracted from the image.

This paper proposes a system that can be used for retrieving images related to a query image from a large set of distinct images. It follows an image segmentation based approach to extract the different features present in an image. The above features which can be stored in vectors called feature vectors and therefore these are compared to the feature vectors of query image and the image information is sorted in decreasing order of similarity. The processing of the same is done on cloud. The CBIR system is an application built on Windows Azure platform. It is a parallel processing problem where a large set of images have to be operated upon to rank them based on a similarity to a provided query image by the user. Numerous instances of the algorithm run on the virtual machines provided in the Microsoft data centers, which run Windows Azure. Windows Azure Stack is the operating system for the cloud by Microsoft Incorporation. Windows azure Stack is responsible for creating ideal hybrid architecture.

Keywords: Content Based Image Retrieval (CBIR), Web Role, Blob Storage, Azure Stack Introduction

1. Introduction

Modern technology has leading to an accelerated growth of digital media collections, often containing both still images and videos. Storage devices are filled with terabytes of digital images, making it increasingly harder to
retrieve images of interest from such collections. It is clear that search capabilities are needed for finding what we are looking for in such large collections, but how can we make such searches useful? Manual annotation of images with keywords describing the image content can make it easier to find images of interest, but this takes a lot of time, making this approach very costly. It is also only helpful to a certain extent, since we do not always know in advance what kind of searches that will be made in the future. Furthermore, different persons are likely to annotate the same image using different keywords, making it difficult to create a suitable taxonomy and annotate images with the “correct” keywords. For all of the above reasons, the use of content-based image retrieval (CBIR) has been suggested. Content Based Image Retrieval (CBIR) is any technology that in principle helps to organize digital image archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. The most common form of CBIR is an image search based on visual. The increasing amount of digitally produced images requires new methods to archive and access this data. Conventional databases allow for textual searches on metadata only. Content Based Image Retrieval (CBIR) is a technique which uses visual contents, normally called as features, to search images from large scale image databases according to users’ requests in the form of a query image. Apart from the usual features like color and texture. This model expects the input as Query by Example (QBE) and any combination of features can be selected for retrieval. Then the texture features such as contrast, energy, correlation, and homogeneity are retrieved. Human being gets images, sound and any other information by seeing, hearing and perception and analysis. Human judge similarity of images and sounds according to their semantic contents, for instance the searching for a star’s picture is based on his facial characters or other contents. So the 2 retrieval methods based on text or keywords for the digital multimedia apparently can’t meet the demand that human being get multimedia information exactly. With more and more multimedia information appear on the Internet and other digital multimedia as well as human beings’ thirst for exact and fast retrieval based on contents multimedia information retrieval becoming the focus of the academe research as well as images retrieval of contents is one of the important study aspect of multimedia information retrieval[1].

2. Review of Literature Survey

2.1 NIR: Content Based Image Retrieval on Cloud Computing

Zhuo YANG, Sei-ichiro KAMATA and Alireza AHRARY in 2009 proposed NIR - an open source cloud based content based image retrieval system. As the growth of cloud computing, researcher from different domain do search with the help of cloud computing. CBIR is one of the challenging and emerging technologies as high computation task because of the algorithm computation complexity and big amount of data. As based on cloud computing infrastructure, NIR is easy to extent and flexible for deployment. Due to this there has been significant increase in scalability and availability. As an open source project, NIR can be improved on demand and integrated to other existing systems [2].

2.2 Content-based Image Retrieval (CBIR) using Hybrid Technique

This method of CBIR was researched by Zainab Ibrahim Abood, Israa Jameel Muhsin, and Nabeel Jameel Tawfiq in year 2013. The research delivered Content Based Image Retrieval (CBIR) using four feature extraction techniques. The four techniques used are colored histogram features technique, properties features technique, gray level co-occurrence matrix (GLCM) statistical features technique and hybrid technique. The features are extracted from the database images and query images in order to find the similarity measure between them. The similarity-based matching is post extraction stage in CBIR. The three types of similarity measure used are, normalized Mahalanobis distance, Euclidean distance and Manhattan distance. The research concluded that CBIR using hybrid technique have higher match performance in all kind of similarity measures used [3].

2.3 Content Based Image Retrieval Using Fusion of Gabor Magnitude and Modified Block Truncation Coding

Dr. H B Kekre, V A. Bharadi have introduced Content Based Image Retrieval using Fusion of Gabor Magnitude and Modified Block Truncation Coding. Gabor filters comprises of wavelets, where each wavelet captures energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features 5 in the form of energy of the signal. Texture features can be extracted from the group of energy distributions. And modified block truncation is used to retrieve color feature from image.
The proposed system was giving higher Precision and Recall as compared to only Gabor and only MBTC based CBIR. Gabor feature gives good response to texture of the image and Modified BTC give good response to color content of image [4].

2.4 Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns

Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu and Vincent S. Tseng in 2011 proposed a method, Navigation-Pattern-based Relevance Feedback (NPRF), to achieve the high efficiency and effectiveness of CBIR in coping with the large-scale image data. In terms of efficiency, the iterations of feedback are reduced substantially by using the navigation patterns discovered from the user query log. In terms of effectiveness, the proposed search algorithm NPRF Search makes use of the discovered navigation patterns and three kinds of query refinement strategies, Query Point Movement (QPM), Query Reweighting (QR), and Query Expansion (QEX), to converge the search space toward the user’s intention effectively. By using NPRF method, high quality of image retrieval on RF can be achieved in a small number of feedbacks. The experimental results reveal that NPRF outperforms other existing methods significantly in terms of precision, coverage, and number of feedbacks [5].

2.5 Content Based Image Retrieval Using Independent Component Analysis

ArtiKhaparde, B L Deekshatulu, M.Madhavilatha, ZakiraFarheen, Sandhya Kumari presented a new approach for global feature extraction using an emerging technique known as Independent Component Analysis (ICA). A comparative study was delivered between ICA feature vectors and Gabor feature vectors for 180 different texture and natural images in a databank. Result analysis show that extracting color and texture information by ICA provides significantly improved results in terms of retrieval accuracy, computational complexity and storage space of feature vectors as compared to Gabor approaches [6].

2.6 Content Based Image Retrieval Algorithm Using Colour Models

Ms. K. Arthi and Mr. J. Vijayaraghavan proposed CBIR using Colour Models in 2013. Digital Image Processing mainly deals with changing the nature of the image as required. An image can be viewed as group of pixels in terms of Image processing. Images are processed mainly to improve the Pictorial information. Content Based Image Retrieval (CBIR) is an emerging and developing trend in Digital Image Processing. CBIR is used to search and retrieve the query image from wide range of databases. Many Features and algorithms can be used for efficient image retrieval. In this paper an efficient image retrieval algorithm based on CCM (Colour Co-occurrence Matrix) is proposed. The CCM for each pixel of an image is found using the Hue Saturation Value (HSV) of the pixel and then compared with CCM of the images in the database and the images are retrieved [7].

2.7 Content Based Image Retrieval Using Color and Shape Features

ReshmaChaudhari and A. M. Patil have introduced Content Based Image Retrieval Using Color and Shape Features in year 2012. This research proposed an algorithm which incorporates the advantages of various other algorithms to improve the accuracy and performance of retrieval. The accuracy of color histogram based matching can be increased by using Color Coherence Vector (CCV) for successive refinement. The speed of shape based retrieval can be enhanced by considering approximate shape rather than the exact shape. In addition to this a combination of color and shape based retrieval is also included to improve the accuracy of the result [8].

2.8 Content based Image Retrieval using Sectorisation of Self Mutated Hybrid Wavelet Transforms

Yogita D. Shinde and Sudeep D. Thepade have proposed CBIR using Sectorisation of Self Mutated Hybrid Wavelet Transforms in year 2015. The research delivered Self Mutated Hybrid Wavelet transform (SMHWT) is used which is formed by using same component transform. In Proposed algorithm, feature extraction is done by applying sectorisation on Self Mutated Hybrid Wavelet transformed images. To test the performance of the proposed method, total 1000 queries were fired on the image database containing 1000 images of 10 categories. Manhattan Distance is used for similarity measurement. Performances proposed algorithm is evaluated using average precision. 7 Results show that the proposed Self Mutated Hybrid Wavelet Transform containing Sine transform as a component gives better performance improvement across all tried variations of SMHWTs [9].
2.9 CCBIR: A Cloud Based Implementation of Content Based Image Retrieval

CCBIR has introduced by R. MADANA MOHANA and Dr. A. RAMA MOHAN REDDY in year 2015. This research proposes a system that can be used for retrieving images related to a query image from a large set of distinct images. It follows an image segmentation based approach to extract the different features present in an image. The above features which can be stored in vectors called feature vectors and therefore these are compared to the feature vectors of query image and the image information is sorted in decreasing order of similarity. The processing of the same is done on cloud. The CBBIR system is an application built on Windows Azure platform. It is a parallel processing problem where a large set of images have to be operated upon to rank them based on a similarity to a provided query image by the user. Numerous instances of the algorithm run on the virtual machines provided in the Microsoft data centers, which run Windows Azure. Windows Azure is the operating system for the cloud by Microsoft Incorporation [10].

2.10 Content Based Image Retrieval Using Color Histogram

A.Ramesh Kumar and D.Saravanan have introduced CBIR method using Color Histogram in year 2013. This research proposed Content-based image retrieval (CBIR) scheme searches the most similar images of a query image that involves in comparing the feature vectors of all the images in the database with that of the query image using some pre-selected similarity measure, and then sorting of the results. On querying an image, a reduced set of candidate images which have the same Grid Code as that of the query image is obtained. The color histogram for an image is constructed by quantizing the colors within the image and counting the number of pixels of each color. The feature vector of an image can be derived from the histograms of its color components and finally can set the number of bins in the color histogram to obtain the feature vector of desired size. Thus the grid code of an image is obtained through the quantization of the feature vector derived from the histogram of the desired color component of the image. In order to have similar features of the images the grid code must be same for all Images in the grid. Experimental result show confirms that the proposed method is most effective [11].

2.11 Comparison of Content Based Image Retrieval Systems Using Wavelet and Curvelet Transform

Suchismita Das, Shruti Garg and G. Sahoo have introduced Comparison of Content Based Image Retrieval Systems Using Wavelet and Curvelet Transform in year 2012. This research implements a CBIR system using different feature of images through four different methods, two were based on analysis of color feature and other two were based on analysis of combined color and texture feature using wavelet coefficients of an image. To extract color feature from an image, one of the standard ways i.e. color histogram was used in YCbCr color space and HSV color space. Daubechies’ wavelet transformation and Symtel’s wavelet transform were performed to extract the texture feature of an image. In this paper a color image retrieval system is illustrated, in which the novelty lies in the use of a fuzzy partition of the HSV color space and wavelet transformation of the fuzzified new image. To increase efficiency of the system finally an image retrieval method was proposed using curvelet transform of an image, which provides an opportunity to extract more accurate texture feature for image retrieval. After obtaining all experimental results, a comparative study was done. From the result it was inferred that curvelet based method gave a better performance as compared to other methods [12].

2.12 A Novel approach for CBIR using Cloud Computing

Vinutha.B.S and Mrs. Savitha.T have introduced a novel approach for CBIR using cloud computing in year 2015. The research delivered technique in the content-based image retrieval (CBIR) systems is to search images in database that are —close to the query image using some similarity measure. Images from large database can be retrieved efficiently from the image content like color, shape and texture. Number of images retrieved depends on similarity matching measure of image features with the query image by utilizing machine learning algorithm. The current work deploys cloud computing enabled CBIR system with the aid of image shape features to retrieve similar images efficiently by reducing computation complexity [13].
3. Description of Windows Azure Cloud And Storage Services

3.1 Windows Azure Cloud Services

Azure Cloud Services provides a place to run highly scalable custom code on a Platform as a Service (PaaS) environment. Cloud service consist of two roles i.e. web role and worker role to choose from when creates an instance, both based on Windows Server. The main difference between the two is that an instance of a web role runs IIS, while an instance of a worker role does not. Both are managed in the same way, however, and it's common for an application to use both. A web role instance might accept requests from users, then pass them to a worker role instance for processing. Web Roles is that it doesn’t need to maintain the operating systems or virtual machines. Windows Azure Storage Services (figure 1).

![Figure 1: Windows Azure Storage Services](image)

3.2 Windows Azure Storage Service

Windows Azure provides multiple storage services that are highly durable, scalable as well as constantly available. Azure storage provides users with following capabilities to persist both structure and unstructured data: Anywhere and anytime access, Store data for any length of time, Scale to store any amount of data, pay for only what is used/stored. Windows Azure Storage provides blobs, tables, and queues (Figure 1).

1. **Blobs** - A blob contains binary information, and there is an easy hierarchy. A storage account can have one or several containers, every one of that holds one or plenty of blobs. There are two types of blob storage. Block blob-it can store up to 200 GB of data and is optimized for streaming workloads. Page blob-it can store up to 1 TB of data and meant for random access.
2. **Tables** - It permits operating with applications data in a more fine-grained way. The data every table holds is stored in a group of entities that contain properties. There is no limit on how many entities can be stored in one table but can typically store billions of entities in one table.
3. **Queues** - A primary function of queues is to afford some way for Web role instances to converse asynchronously with Worker role instances.

3.3 Microsoft Azure Stack

Microsoft Azure Stack brings proven innovation from Microsoft’s hyper-scale public cloud into your datacenter, providing agility and productivity for application owners, flexibility and control for IT, and assurance that corporate assets are protected.

Azure Stack delivers an Azure unified application development model and self-service IaaS and PaaS experiences supported by cloud-inspired infrastructure. The Azure-based services deliver a “consistent core” of Azure capabilities that can be deployed to any datacenter. IaaS and PaaS will be delivered as Resource Providers under common Azure Resource Manager, Portal, IT and Dev tools. Azure Stack shares infrastructure and service implementations with Azure cloud. The Microsoft Azure Stack will be available as part of the 2016 wave of products. Those interested in our vision can start the journey by trying the new Azure Portal and Azure Resource Manager, or with Windows Server 2012 R2, System Center 2012 R2, and the Azure Pack.
4. Implementation

Content-based image retrieval (CBIR) scheme searches the most-similar images of a query image that involves in comparing the feature vectors of all the images in the database with that of the query image using some pre-selected similarity measure, and then sorting of the results. In this application, user send the image in form of query image. Extraction process takes place after sending query image and calculates the feature vectors of query image on the basis of color, texture and shape etc. using extraction technique. Hybrid Wavelet is the feature extraction technique (figure 2) which is implemented by web services deployed on web role on the public cloud. These feature vector is compared with extracted features of the images residing in image database with help of similarity measures.

Fig 2. Feature Extraction Technique

Similarity measures is nothing but precision and recall both are used for comparison process. 
Precision=Number of relevant images retrieved/Total number of images retrieved 
Recall=Number of relevant images retrieved/Total Number of images in Database.

The flow of model development is as follows (figure 3):
1. Efficient Selection of images depending upon our query set and upload it on Private Cloud.
2. Feature Extraction module to extract color or shape or texture features and build a single feature vector and store in the feature database i.e. in Blob Storage within Private Cloud.
3. Image Segmentation and Feature extraction process on Query image on Public Cloud.
4. Similarity comparison between the database image and query image through Euclidean Distance technique on Public Cloud.
5. Indexing and Retrieval based on semantically closer images are extracted from database and displayed as a thumbnail on user end.
3.1 Proposed technique for Uploading Operation

The extraction process done on public cloud through hybrid wavelet technique based on the content like color, shape and texture of image. This process is used to calculate feature vector of query image which is sent by user using web service which is running on Web role (figure 4).

1. User send query image through desktop application and upload image to Web Role.
2. Web Role Stores the query image in Blob storage.
3. Web role posts a work item to queue to have the feature vector calculated for query image which uploaded using Hybrid Wavelet technique.
4. Worker Role fetches work item from the queue.
5. Worker role connect to blob storage for retrieval images.
6. Worker Role retrieves the image from blob storage & process it to create a feature vector of image.
7. And store the calculated feature vector in blob storage.
3.2 Proposed technique for Classification Operation

The classification operation is used for verification and retrieval in database. This process done by comparison of feature vector of query image and feature vector of image of image database using similarities measures with high value than the threshold, then image is delivered from image database (figure 5).

1. User send query image through desktop application and upload image to Web Role.
2. Web Role Stores the query image in Blob storage.
3. Web role posts a work item to queue to have the feature vector calculated for query image using Hybrid Wavelet technique.
4. Worker Role fetches work item from the queue.
5. Worker role connect to blob storage for retrieval images.
6. Worker Role retrieves the image from blob storage & process it to create a feature vector of image.
7. Check to match between the calculated feature vector & feature vector stored in Blob storage with the help of Similarity measures i.e. Precision & Recall.
8. Send the relevant result to Worker Role.
9. Worker Role update the status of whether image is similar or not and send to the Web Role.
10. Web Role convert the result in a format understood by requesting and send it to appropriate user.
5. Results

The results include the proposed architecture of CBIR on hybrid cloud and the images uploaded in the blob storage on the cloud using Windows Azure storage services which are used to create cloud storage (figure 6 & 7).

6. Conclusion

Thus this paper describes Images have always been an inevitable part of human communication and its roots millennia ago. Images make the communication process more interesting, illustrative, elaborate, understandable and transparent. CBIR is the process of retrieving images from a database or library of digital images according to the visual content of the images. In other words, it is the retrieving of images that have similar content of colors, textures or shapes. The proposed system intends to satisfy the desired expectations by implementing CBIR software as a service on hybrid cloud where one can leverage all the possible cloud services like scalability, flexibility, availability with faster retrieval
of image. The proposed system CBIR SaaS model, a request (query image) is sent on cloud by desktop application. The image will be processed by web services which is running on Web Role. Feature extraction of query image will be done by hybrid wavelet. This extraction process will be done on public cloud. Feature vector will be obtained from feature extraction process and feature vector store in the Blob storage. After this process comparison process will occur on private cloud. In Comparison process the feature vector of query image will be compare with feature vector of stored images of database image (Database image-which stored large amount of images and present on private cloud).finally will get retrieval similar images on public cloud. Performance can be better due to comparison process and the two different process (Feature Extraction and Comparison) will be done on two architecture i.e. public and private cloud (Hybrid Cloud Architecture) which provides good speed for retrieving of an image.

References

1. Yogita Mistry and D.T. Ingole. Survey on Content Based Image Retrieval Systems. International Journal of Innovative Research in Computer and Communication Engineering (IJIRCE), Vol.1, October 2013.

2. Zhuo YANG, Sei-ichiro KAMATA and Alireza AHRARY. NIR: Content Based Image Retrieval on Cloud Computing. IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009.

3. Zainab Ibrahim Abood, Israa Jameel Muhsin, Nabeel Jameel Tawfiq. Content-based Image Retrieval (CBIR) using Hybrid Technique. International Journal of Computer Applications (O975 – 8887), December 2013.

4. H B Kekre, V A Bharadi, S D Thepade, B K Mishra, S E Ghosalkar, S M Sawant. Content Based Image Retrieval Using Fusion of Gabor Magnitude and Modified Block Truncation Coding. IEEE computer society, 2010.

5. Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu and Vincent S. Tseng. Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns. IEEE Transactions On Knowledge And Data Engineering, Vol. 23, No. 3, 1041-4347/11, March 2011.

6. Arti Khaparde, B L Deekshatulu, M.Madhavilatha, Zakira Farheen, Sandhya Kumari. Content Based Image Retrieval Using Independent Component Analysis. IICNS International Journal of Computer Science and Network Security, VOL.8 No.4, April 2008.

7. K. Arthi, J. Vijayaraghavan , Content Based Image Retrieval Algorithm Using Colour Models. International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, March 2013.

8. R. Chaudhari, A. M. Patil. Content Based Image Retrieval Using Color and Shape Features. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 1, Issue 5, November 2012.

9. Y. D. Shinde, S. D. Thepade. Content based Image Retrieval using Sectorisation of Self Mutated Hybrid Wavelet Transforms. International Journal of Computer Applications, Volume 120 –No.24, June 2015.

10. R. Madana Mohana , A. Rama Mohan Reddy. CCBIR: A Cloud Based Implementation of Content Based Image Retrieval. WSEAS TRANSACTIONS on COMPUTERS, Volume 14, 2015.

11. A.Ramesh Kumar, D.Saravanan. Content Based Image Retrieval Using Color Histogram. (IICSIT) International Journal Of Computer Science and Information Technologies,Volume 4, 2013.

12. Suchismita Das, Shruti Garg, G. Sahoo. Comparison of Content Based Image Retrieval Systems Using Wavelet and Curvelet Transform. The International Journal of Multimedia & Its Applications (IJMA), Vol.4, No.4, August 2012.

13. Vinutha.B.S, Savitha.T. A Novel Approach For CBIR Using Cloud Computing. International Conference on Computer Science, Electronics & Electrical Engineering, 2015.

14. Microsoft Corporation “Server & Cloud Blog “, 2015.” http://blogs.technet.com/b/servercloud/archive/2015/05/04/announcing-microsoft-azure-stack.aspx”.

15. Krunali Vikas Vartak, Vinayak Ashok Bharadi. Cloud Based CBIR Software as a Service. International Journal of Applied Information Systems (IJAIS), Volume 2, February 2012.