iJCEM
International Journal of Core Engineering & Management

ISSN:2348 9510

International Journal Of Core Engineering & Management(IJCEM)

Volume 1, Issue 1, April 2014

Image Retrieval using Re-Ranking Algorithms

Mayuri D. Joshi , Revati M. Deshmukh, Kalashree N.Hemke, Ashwini Bhake

Computer Technology Department , Yeshwantrao Chavan College Of Engineering, Nagpur, Maharashtra India.

mayuri1992joshi@gmail.com,rewtideshmukh@gmail.com,kalashreehemke@gmail.com, bhake.ashwini@gmail.com.

ABSTRACT

Our objective is to improve the performance of keyword based image search engines by re-ranking their original results. Present search engines face some limitations. First is that, many irrelevant results are obtained along with the demanded images. Secondly, duplication of images is normally seen in the image database retrieved from any search engine for the entered query word .Third is that , user has to surf along long pages for expected result or set of result images, leading to confusion many a times.

Re-ranking of images is a solution to these limitations. The objective of this work is to reduce user efforts and semi-automatically generate a large number of images for a specified object class which are more accurate as compared to the normally retrieved images. Candidate images are obtained by a text based web search querying on the object identifier (e.g., the word penguin). The WebPages and the images they contain are downloaded. The task is then to remove irrelevant images and re-rank the remainder. So the main motive of the proposed system is to complement the task of present search engines for the interest of user . This system is a combination of web and desktop application that is active on user or client side. This work will re-rank the images as well as form clusters of similar images and also notify the user to which cluster the entered query image will belong .

1. Introduction

Image retrieval is one of the major issues of user concern. The most common way of image retrieval is the text based image retrieval technique (TBIR) [6]. The working of TBIR needs to be understood. Suppose user enter a query in the search box of Google, how exactly Google returns the result?

When we enter a keyword in the search box, Google navigates through different pages looking for the occurrence of that word. Each image has a tagging associated with it. Google looks for the occurrence of the entered keyword in these tagging. As Google returns the result based on the search of text, it is text based image retrieval.

.

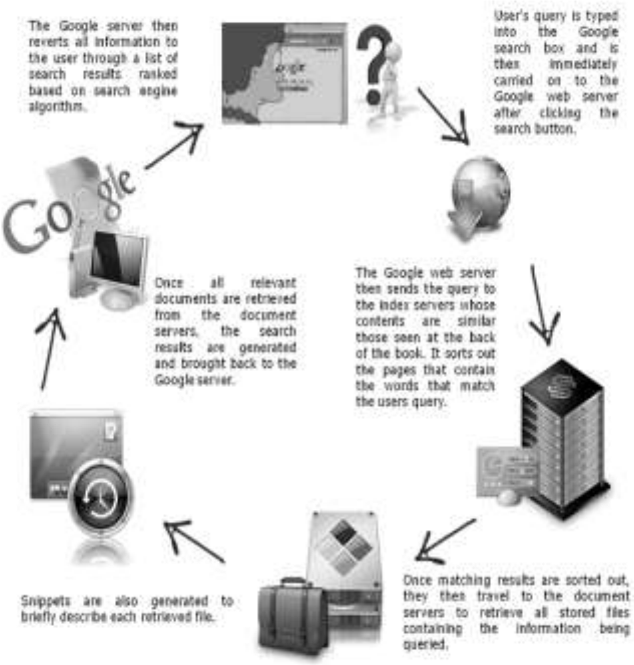

Figure 1. Working of Google search engine. [7]

After retrieving the images, Google then performs top-rank sequencing. In this, Google ranks the images on the basis of number of hits. The images are ranked in descending order of number of hits i.e. the image with maximum number of hits is ranked first. Thus Google performs text- based image retrieval for retrieving the images and top-rank sequencing to rank the retrieved images and then returns the result.

This technique of TBIR is not sufficient for accuracy hence the trend moves towards content based image retrieval (CBIR). Our work aims at extracting colour features of images and using histograms to re-rank the images to improve the performance.

2. Literature Survey

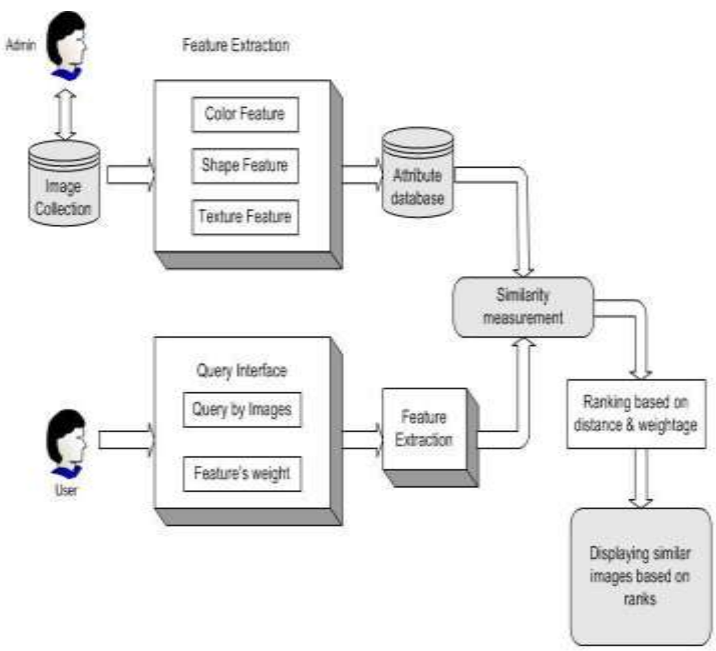

Figure 2. Architecture of image harvesting and re-ranking system [5]

Content based image retrieval uses content of image for retrieval. The architecture diagram (Fig. 2) [5] gives an overview of Content Based Image Retrieval (CBIR). Content of image refers to the colour, shape and texture of the image.

ISSN:2348 9510

International Journal Of Core Engineering & Management(IJCEM)

Volume 1, Issue 1, April 2014

This system is based on the following functionalities and features:

a) Extraction

Extraction means deciding the feature based on which re-ranking will be performed and extracting the same. If the entered query is ‘sunset’ colour will be the appropriate feature for re-ranking. But if we need to differentiate between ‘cotton’ and ‘snow’ texture will be the appropriate feature for discrimination.

b) Distance calculation and similarity measurement:

This step calculates the difference between the images in terms of corresponding chosen feature. The images are plotted in feature space and then the distance between them is calculated using any of the following formulae.

Given two feature vectors A and B such that

A= a_1 B= b_1

a_2... b_2...

Euclidean distance is given by:

$$\sqrt{\sum_{i=1}^{n}(a_i - b_i)^2}$$

City block is another approach for distance measurement. [3]

$$\sum_{i=1}^{n}|a_i - b_i|$$

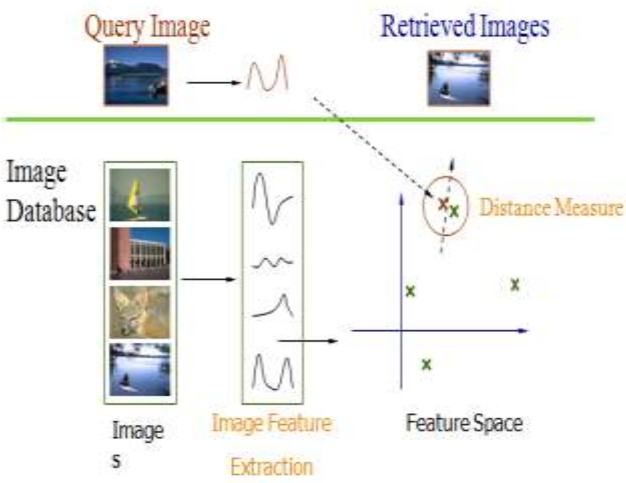

Figure 3. Distance calculation and measurement

Lesser the distance more similar the images will be. As mentioned in [5], for CBIR implementation, image classification should be fast and efficient. CBIR emphasises on use of visual content of image like colour, texture, shape etc. for image comparison and retrieval rather than textual query. We have considered colour feature for comparison.

c) Following the distance calculation and similarity measurement, Re-ranking is performed. Various algorithms can be used to perform re-ranking. Our implementation uses K-mean and hierarchical algorithm.

3. Implementation and Experimental results

Details of Completed Modules with supporting results

The completed modules are

1. Web Module
2. Desktop Module

ISSN:2348 9510
International Journal Of Core Engineering & Management(IJCEM)
Volume 1, Issue 1, April 2014

1. The Web Module

This module is used for downloading the images from search engine on the client machine where further processing will be done. This module downloads bulk of images for a given query on a single click. For experimental purpose, we have downloaded 100 images for keyword 'Rose'.

Figure 4. Part of dataset used

2. The Desktop Module

The downloaded images are further processed in this module. There are two sub-modules-

I. Indexing Module

This module performs clustering by using K-means and hierarchical clustering algorithms in combination.

K-means Algorithm

- Input
  - k: the number of clusters
  - D: a dataset containing *n* elements
- Output: a set of *k* clusters
- Method

(1) arbitrarily choose *k* elements from *D* as the initial cluster mean values

(2) **repeat**

(3) assign each element to the cluster whose mean the element is *closest* to

(4) once all of the elements are assigned to clusters, calculate the *actual* cluster means

(5) **until** there is no change between the new and old cluster means

Hierarchical clustering Algorithm

This algorithm is used in bottom up fashion for comparing images and assigning the clusters based on threshold and computed distances.

The working can be understood from the figure

ISSN:2348 9510
International Journal Of Core Engineering & Management(IJCEM)
Volume 1, Issue 1, April 2014

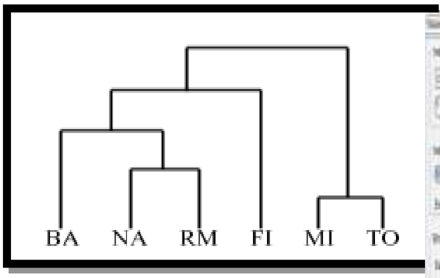

Figure 5. Hierarchical clustering

The indexing module uses these algorithms to cluster the images based on color feature of image. Following steps are followed-

1. The Red, Green and Blue content of an image are extracted and plotted as histograms and histogram matrices are being generated.
2. The indexing module takes the number of clusters k as input.
3. It then applies K-means for the iterations and hierarchical clustering for image classification.
4. The distance calculation in K-means is done by using the histogram matrices
5. The indexing modules returns number of clusters containing re-ranked images.

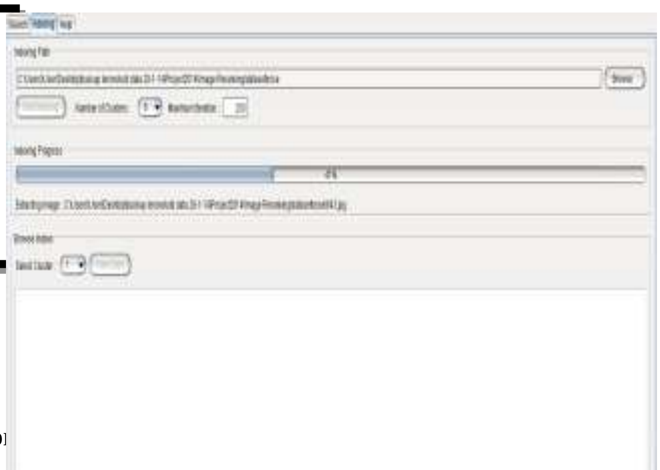

Figure 6. Indexing module home page. Shows indexing in progress

For experimental purpose we have obtained result for four clusters. They are as follows.

Figure 7. Cluster1 of 4 obtained after indexing.

ISSN:2348 9510
International Journal Of Core Engineering & Management(IJCEM)
Volume 1, Issue 1, April 2014

Figure 8. Cluster2 of 4 obtained after indexing.

Figure 9. Cluster 3 of 4 obtained after indexing.

Figure 10. Cluster 4 of 4 obtained after indexing.

II. Search Module

This module is used to find the relevant cluster of the given input image. It takes an image as input and returns a cluster containing relevant images based on distance matching obtained in indexing module.

Figure 10. Result of search module for input image.

4. Applications

This project finds application in various fields like medical sciences for disease diagnosis , in astronomy , mechanical engineering and other fields where image clustering and detection is required.

5. Future Scope

- Different features like texture, edges can be used for Re-Ranking.
- An intelligent Re-Ranking System Capable of identifying which feature should be used and Re-rank accordingly

6. Conclusion

Basic thing reviewed from this paper is that the text-based image retrieval is not sufficient for obtaining precise images for a given query. Thus techniques based on CBIR are found to be more vibrant and are likely to be adopted for such applications. This implemented work is

ISSN:2348 9510
International Journal Of Core Engineering & Management(IJCEM)
Volume 1, Issue 1, April 2014

efficient for image clustering and classification with considerable accuracy.

The domain of image harvesting, retrieval and re-ranking offers a vast scope for exploration as well as innovation and has a broad future scope.

7. References

[1] Venkat N.Gudivada, Vijay V. Raghavan "Content-Based Image Retrieval Systems" IEEE Transaction 0018-9162, 1995 .

[2] Edward Remias, Gholamhosein Sheikholeslami, Aidong Zhang." Block-Oriented Image Decomposition and Retrieval in Image Database Systems". IEEE Transaction 0-8186-7469-5, 1996.

[3]Szabolcs Sergy´an, Budapest Tech, John von Neumann ,Faculty of Informatics." Color Histogram Features Based Image Classification in Content-Based Image Retrieval Systems".6th International IEEE Symposium on Applied Machine Intelligence and Informatics-2008.

[4] Yihun Alemu, Jong-bin Koh, Muhammed Ikram, Dong-Kyoo Kim." Image Retrieval in Multimedia Databases: A Survey". Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing ,IEEE-2009.

[5] K.A. Shaheer Abubacker, L.K. Indumathi." Attribute Associated Image Retrieval and Similarity Re-ranking". Proceedings of the International Conference on Communication and Computational Intelligence – 2010, Kongu Engineering College, Perundurai, Erode, T.N.,India.27 – 29 December,2010.pp.235-240.

[6] Lixin Duan, Wen Li, Ivor Wai-Hung Tsang, and Dong Xu, Member, IEEE. "Improving Web Image Search by Bag-Based Re-ranking".IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 11, NOVEMBER 2011.

[7] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *Proc. ICCV*, 2003.