RELEVANT IMAGE SEARCH BASED ON HASH CODE OF THE QUERY IMAGE

S Arunkumar, V Subramaniyaswamy, M Prabhu and S Rajarajan,
School of Computing, Sastra University, Tamilnadu, India

ABSTRACT

Traditional image search engine provides images based on textual words associated with them. Multiple images have textual words in common. Therefore, the search engine displays images which are also not relevant to user’s needs. Hence, this search provides some irrelevant images which lead to non-optimal image search. So, image search based on textual words remain major disadvantage of the existing search engine. Aiming at reduced memory usage and more likely to provide an ideal and better image search, this paper employs the concepts of image feature extraction, hashing techniques, hamming space calculation and relevancy retrieval of images. Query image and all other images in the database are subjected to feature detection and extraction. Features are manipulated to binary hash code, where they can be made as reference to each image. Images are ordered by the hamming distance calculated from hash code. The most relevant images are then retrieved as per user’s query image. It can be employed for various applications such as entry check, that uses images rather than finger print recognition or even in guest detector in houses or offices and much more.

Keyword: Hamming distance, hash code, query image.


1. INTRODUCTION

Use of images across globe, has been exponentially increased in recent years. Images are searched, gathered and utilized for various purposes by different varieties of people. The explosion of images on the internet has also led to the growth of retrieval mechanisms. Nowadays, the need for efficient search mechanism plays an essential role. The role of the internet and various search engines embedded are useful for retrieving images, but lack in terms of relevancy.

Traditional search engines, search for requested image in the form of textual word associated with them. Generally, images in the database servers are stored with some textual words indexing[7] them. So, whenever there is a request from client side which is in the form
of text, the searching is done in a way, by matching those image-indexing texts. Large images are stored in the database server, so, many images may have textual words in common [8, 9, 10]. These common textual words contribute to retrieval of relevant and irrelevant images.

There are two main problems in traditional search mechanisms. At first, an improper reference to images has to be corrected. Secondly, efficient search mechanisms contributing to relevancy must be employed. Dealing with these two problems, this work aims at three specific modules – image feature extraction, distance calculation, ordering, and retrieval [11, 12].

Image is a form of characterizing human vision. These images are then stored with some related textual words, for easy reference. But, this methodology does not contribute to various applications. The sprouting idea to index these images is through some features of these images. Basic features such as color, boundary, brightness, shape and many others can be used. There also exists, scale invariant feature which does not change for images even after transformations are applied. These remain unique to a particular image. Features extracted from the images are then utilized to represent or index these images. Image features are transformed to another form for easy representation. Distance calculation between query image and image in the database helps in organizing and retrieving relevant images.

2. LITERATURE SURVEY

Traditional image search engine provides images based on textual words associated with them. Multiple images have textual words in common. Therefore, search engine displays images which are also not relevant to user’s needs. Hence, this search provides some irrelevant images which lead to non-optimal image search. So, image search based on textual words remain major disadvantage of the existing search engine.

In [1] Information retrieved from repository images are effectively and efficiently used by Content based image retrieval method. This method helps those who are unfamiliar with usage of the database to retrieve related images using the image content.

In [2], the privacy of the images which are uploaded in cloud is discussed. Usually before uploading, information are encrypted and sent to cloud. But once encrypted, searching on the information in cloud are not easy. To avoid this, image is encrypted using a searchable encryption where descriptor of an image is sent in plaintext and the encrypted form of image is sent. But in this paper, descriptor is also encrypted. For finding a relevancy, features are extracted from the query image which is then encrypted and matched against the descriptor in the cloud environment.

In [3], method for reducing the search space is discussed. If more images are there in a database in particular or system or server at large, then time required to find the relevant image among them is always more and this may not be easy also. Instead of treating all the images as one group, images can be categorized or clustered based on some image attributes by particle swarm optimization algorithm and k-mean clustering algorithms. For finding relevant images, these cluster attributes or features can be matched against the feature extracted from the query image. This reduces the search time.

In [4], method is described where an image is segmented into many nodes which are then connected to form a graph. Comparing the feature of a query image with the nodes of this graph is easier than comparing original raw image. This method produces more efficient relevant images based on the query image.

In [5], method to design efficient image search engine is described. It may not be easier to map the query key word to find a optimal relevant image match based on user perception and intent. Apart from query keyword, some additional information like surrounding and other
attributes are also the input for the search engines. Based on these attributes, images are ranked and using this ranking relevant images are found.

In [6], basic principle for finding a relevant image for the query image is discussed. Feature is extracted from the query image and is represented by a descriptor. The same process is done on the images available in a directory or databases as well. Distance between descriptor of query image and sample images are used for finding the relevancy between them.

3. METHODOLOGY

This work aims at its contribution to relevancy retrieval of images. The main focus is to overcome the problems on image retrieval where traditional systems are bound to retrieve images which are partially irrelevant. Then two main objectives are, at first, representing images by their features and the other is to calculate distance between the images. This helps in identification of relevancy between them. The comparison takes place between the query image and all other images in the database server. On ordering the images in the database, retrieving most relevant images become easy.

Searching top relevant images from the database server is the ultimate goal of this work piece. To overcome the problems faced by the traditional image search engine, this work renders good service of retrieving images based on user’s need. It focuses on three main methodologies.

This work includes the following modules for development of this method. These are as follows: -

3.1. Preprocessing

Fundamentally, for image search, handful of images are required. Images of query are also required. Query images may either be present or absent in database. The images in the database are already pre processed, in way, that features are extracted and converted to hash codes. Three efficient algorithms which are SIFT algorithm, Hashing algorithm and Hamming space are considered for further processing.

3.2. SIFT feature extraction

Query image and all other images in the database are subjected to feature detection and extraction. There are various features that could represent an image. They are color, shape, contour, edge, brightness and regions of interest, points of interest and many others.

Points of interest remain best suitable for this paper. These points are identified in these images based on the priority of remaining unchanged, even on multiple transformations. Consider two images - at first, a dog faces the camera and barks, while the other, similar dog turns right and barks. In both the cases, it is the dog that barks. Points of interest are selected, such that these points are invariant to the transformations. With the help of such interesting points, similarity between images can be identified.

Features of the image, thus corresponds to scalable invariant features so that any transformation applied on the image does not affect its representation. These features will be much helpful in comparison of the images in database with the query image.

3.3. Hashing of image features

Multiple features are available for a single image. So it is impossible to refer an image using a specific feature point. So multiple feature points, transformed to feature vector is used to represent the image.
For efficient search mechanism, feature vector is manipulated to another form where it can be made as reference to each image. There are various mechanisms such as inverted file, tree based indexing and hashing. Since an image holds high dimensional feature, hashing provides a better mechanism. Features are converted to hash codes for convenience.

3.4 Hamming space calculation and relevancy retrieval
Hash codes are then utilized for hamming distance calculation. This hamming space indicates the similarity variation between the query image and images in database. The hamming distance calculated act as one of the reference to the corresponding image. This module is much helpful to identify the most relevant images to the query provided.

Images are ordered and grouped by hamming distance already calculated. Either relevant images are retrieved as top ranking elements, followed by other images in the database or with pre-assumed acceptance level of hamming distance, only relevant images are retrieved. The most relevant images are then retrieved as per user’s query image. The most identical image is also retrieved. This relevancy retrieval of images contributes to efficient search mechanism

4. RESULTS AND DISCUSSIONS
To assess the performance of this work, small set of query images are considered. Database is constructed which contains limited set of input images.

Work done in this paper remains as a fundamental application for retrieving relevant images using image itself as query. This paper can be further improved by employing this piece of work in various applications. This proves to be a good method when embedded in recognition softwares, login checks, organizational identity and many others. Capturing images and cross checking it with other set of images in the back end helps in those applications. Relevancy can also be checked in terms of feedback from end users by counting the number of clicks per image. This methodology when employed in search engines, shall contribute to performance by meeting user’s expectation and elegance. Fig -1 is the query images whose features are detected and extracted using SIFT and fig -2 shows the resultant relevant images based on the given query image.

Figure 1 Input Image
5. CONCLUSION

This paper is thus concluded with relevancy retrieval of images with respect to query image. This paper remains advantageous in reduced memory space occupancy and reduces the complexity of processing. In various applications such as entry check, login, image matching and many others, this paper might play a useful role. Scalability of this work can be expanded in terms of multiple query images.

REFERENCES


