SURVEY ON MULTI QUERY CONTENT BASED IMAGE RETRIEVAL SYSTEMS

M.Praveen
PG Student
Department of Computer Science and Engineering,
Kumaraguru College of Technology,
Coimbatore,
India.
Praveen1993.4.18@gmail.com

K.Sivanarulselvan
Associate Professor
Department of Computer Science and Engineering,
Kumaraguru College of Technology,
Coimbatore,
India.
sivanarulselvan.k.cse@kct.ac.in

P.Betty
Assistant Professor
Department of Computer Science and Engineering,
Kumaraguru College of Technology,
Coimbatore,
India.
betty.p.cse@kct.ac.in

ABSTRACT

As the digital image collections are growing at a rapid rate, demand for efficient and effective tools for retrieval of query images from the database has increased significantly. Among them, content-based image retrieval systems (CBIR) have become very popular for retrieving images from a large database of digital images as it requires relatively less human intervention. Sometimes similar visual characteristics cannot always guarantee a satisfactory retrieval result. This is called Semantic gap between low level features and high level concepts. To overcome this issue many techniques have been employed in CBIR systems. Multi-query approach is one such technique. In this paper, different multi-query approaches for reducing semantic gap in CBIR systems are surveyed.

Keywords: Content based image retrieval (CBIR), Feature matching, Multiple query image retrieval (MQIR), semantic gap reduction techniques.

1 Introduction

The advancement of digital image capturing and producing technologies has rapidly increased the utilization of digital images in almost all the fields. Easy access of low cost cameras and mobile phones with built in camera also has added up to the increase of digital images in great rate at short time period. The increased use of internet and social media has added up to the accumulation of large amount of digital images in the web. According to Mary Meeker’s annual Internet Trends report [1], people have uploaded an average of 1.8 billion digital images each and every single day. That is 657 billion photos per year. When a large number of images are stored in the database it is difficult to annotate every single image manually for categorical retrieval and processing purpose. Hence, content based
image retrieval methods are used, where the similar images are retrieved based on the visual contents of the query or sample images such as colour, points, lines, textures or any other information that can be automatically extracted from the image. The retrieved images are later ranked according to the relevance between the query image and images in the database in accordance to a similarity measure calculated from the features. Hence the efficiency of a CBIR system greatly depends on the effectiveness with which the low level features extracted from the images represent the high level semantics of the image. To reduce the semantic gap here usage of multiple query images are proposed rather than the traditional single query approach. In this paper, different multi query content based image retrieval methods are studied.

1.1 Content Based Image Retrieval Systems

Previous methods were not generally based on visual features but on the textual annotation of images. The images were first annotated by text and then searched using the text based approach. However in many situations, text annotation method is inefficient. For a vast image data the vast amount of labour required in manual annotation. Also describing every visual feature within the images is consumes the time a lot and difficult. So instead of manual annotations by text based keywords, each image is indexed by its own visual features such as colour, texture, shape etc. A CBIR system [2] extracts the visual contents of the query image such as colour, texture, shape or any other information that can be automatically extracted from the image and use these information to retrieve content related images from the database. The retrieved images are later ranked according to the relevance between the query image and images in the database in accordance to a similarity measure calculated from the features. Hence the efficiency of a CBIR system greatly depends on the effectiveness with which the low level features extracted from the images represent the high level semantics of the image.

1.2 Principle of CBIR

Content-based retrieval utilizes the contents of images to represent as well as access the images from a large database. Content-based retrieval system can be divided into two types: off-line process of feature extraction and online process of image retrieval. Figure.1. shows the architecture of content-based image retrieval. In off-line stage, the system extracts visual attributes like colour, shape and texture of each image in the database based on the pixel values and stores them in feature vector database. The feature data also known as image features for every visual attributes of each image is very smaller in size compared to the original image data, thus the
feature database contains a compact form of the images in the image database. Using feature vector representation of image database over the original pixel values significant compression can be achieved. In on-line image retrieval, the user submits a query image to the CBIR system in search of desired images. The query image is represented with a feature vector in this system. The similarities among the feature vectors of the query example and those of the images in the feature database are later computed and ranked. Retrieval is computed by applying an indexing scheme which provides an efficient way for searching the image database. Finally, the retrieval results are ranked by the system and then the system returns the images which are most similar to the query images.

The important issues in content based image retrieval system are: 1. Selection of the image database, 2. Calculating similarity measurement, 3. Evaluation of the performance of retrieval process and 4. Low-level image features extraction [3]. Evaluation of retrieval performance is a very crucial problem in content-based image retrieval (CBIR). (Refer Figure 1 Here) for architecture of CBIR system.

1.3 Classification of CBIR systems

Content-based retrieval systems can be classified into classes depending on the features they use such as colour, texture, and shape [4]. Each features class is further divided into subclasses by the type of the algorithm used for constructing the feature vector. Colour features are further divided into histogram based and statistical features. Texture features are further divided into spectral features and statistical features. Shape features are further divided as model based, boundary and region based feature extraction methods.

Colour feature is a significant one in searching the collections of colour images of different subject matter. Colour plays an important role in the human visual perception mechanism. All methods for representing colour feature of images can be classified into colour histograms and statistical methods of colour representation. The most popularly used colour spaces are RGB (that is red, green, and blue), CMY (that is cyan, magenta and yellow), CMYK (that is cyan, magenta, yellow, and black), Lab (that is CIE L*a*b, lightness, a and b are considered to be two colour dimensions, from green to red and from blue to yellow) HIS and HSV (hue, saturation, and value).

Texture provides us information on the structural arrangement of surfaces and objects in the image. Texture cannot be defined for a separate pixel; where as it depends on the distribution of intensity over the image. Texture possesses properties like periodicity and scalability; it can
be described mainly by directions, contrast, and sharpness. Texture analysis plays a major role in the comparison of images supplementing the color feature. The most frequently used statistical features are general statistical parameters which are calculated from pixels’ intensity values, parameters which are calculated based on the co-occurrence matrices and the texture histograms which are built upon the Tamura features.

Along with colour and texture characteristics, the shape of objects is also often used for the comparison of images. Shapes can be represented and described in two methods: external methods, by which the region is represented in term of its external characteristics (that is its boundary), and internal ones, in which the region is represented in terms of its internal characteristics (that is the pixels comprising the region). Shape features can be classified in boundary descriptors and region descriptors. Further they can be classified as structural and global. The global boundary descriptors include different signatures, Fourier descriptors and wavelet descriptors.

1.4 Feature matching

In the matching process, the system compares the image signature of query image with the image signatures of the other images stored in the database. For matching, the distance between image signatures of the query example and database images is computed and then the images are ranked according to their distance threshold. Typically Euclidean distance is considered as the most important measure. Then according to ranking, system returns the results that have visual similarity to the query. Feature matching can be done in two ways. One way is by region in which regions can be obtained by segmentation and then distance between two regions is measured based on their low level features. Other is by image which consists of number of regions. The Euclidean distance between points \( p \) and \( q \) is the length of the line segment connecting those points. In the Cartesian coordinates, if \( p = (p_1, p_2, ..., p_n) \) and \( q = (q_1, q_2, ..., q_n) \) are two points in the Euclidean \( n \)-space, then distance \( (d) \) from \( p \) to \( q \), or from \( q \) to \( p \) is given by the Pythagorean formula:

\[
d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}
\]

(1)

\[
= \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}
\]

(2)

In recent years several other distance measures have been developed for histograms like city block distance and the Minkowsky distance.

2 Semantic gap in CBIR and reduction techniques
Although great interest and a large number of new techniques and systems are emerged in content-based image retrieval in the last decades, the gap between low-level visual features and high-level semantic understanding of images, which is also known as the semantic gap problem exists. The gap between the object in the world and the information in a computational description derived from a recording of that scene is the bottleneck to further improvement of the performance of a content-based image retrieval system. Therefore, in order to solve the problems and improve CBIR performance, Region Based Image Retrieval approaches, image annotation, and relevance feedback [5] have been received more attention in the recently years.

2.1 Image Annotation

The problem of narrowing semantic gap lies in estimating the context of the search requests or the usage contexts for retrieved images. For example, in content based approaches a user query (usually in the form of a sample image), is matched against low-level features (e.g. colors, shapes), extracted from the image, thereby providing no insight into why the user wants to retrieve the specified image or of how the retrieved image will actually be used [6]. In order to show image semantics at a higher level, one of the approaches uses different models and machine learning methods to identify the relation between image visual features and semantics, then label the image with keywords, this is image annotation. Most existing image annotation approaches can be classified into two categories, probabilistic modeling based methods and classification based methods. Classification based methods treat keywords as classes and trained classifiers are employed to annotate an input image based on classification results. Probabilistic modeling based methods attempt to obtain the correlations or joint probabilities between images and annotations. The disadvantage in this approach is that it requires large amount of human labor in the manual annotation when the image collection is very large. It is hard to precisely annotate the very rich content of an image by humans due to the perception subjectivity.

2.2 Region-based Image Retrieval

RBIR is an image retrieval approach which concentrates on contents from regions of images and not the content from the entire image as in early CBIR [7]. For RBIR, it first segments the images into a number of regions, and extracts a set of features, which are identified as local features, from segmented regions. A similarity or the distance measure determining the similarity between the target regions in a query and a set of segmented regions from other images is utilized
later to determine relevant images to the query based on local and regional features. The motivation of RBIR approaches are based on the fact that a high-level semantic understanding of the images can be better reflected by local features of images, rather than global features. Since users are often more concerned with interested objects in images when using an image retrieval system, considering images based on region (RBIR) allow users to pay more attention to regional properties that may a better characterize objects that are also made up of local regions. This strategy is able to better reflect the characteristics of images from the perspective of image region and objects, and in turn further improve the image retrieval performance. One of the drawbacks of such systems is that it is difficult to determine the significance of different regions. Also they are computationally costly and time consuming. These systems also require efficient region matching algorithm to effectively find the similarity between images.

2.3 Relevance Feedback

Relevance feedback (RF) is a method where interactive supervised learning technique has been used to bridge the semantic gap between the low-level features of the image and the semantic content of the images and, thus, to improve the retrieval results. In specific, RF attempts to insert a subjective human perception of image similarity into a CBIR system and hence for this to be accomplished, the users are required to assess, in each RF round, the retrieved images as relevant or the image as irrelevant to the initial query and to submit their assessment as a feedback to the system. Then, the system takes into account of their feedback and updates in the appropriate way the image ranking criterion. Here speed of the performance is the major drawback. Satisfactory results can be achieved only after several iterations.

Often a single image used as query may not be able to reflect what the user has in mind and hence many not able to retrieve relevant images according to the user’s requirement, widening the semantic gap. Multi-query approaches have been proposed to overcome limitations of traditional CBIR systems to improve the retrieval efficiency and reduce the semantic gap. Various studies show multi query approaches provides better precision value than the traditional single query CBIR.

3 Multi-Query CBIR Systems

A general Multi-Query Image Retrieval (MQIR) system which is implemented in the CBIR is illustrated in figure 2.

(Refer Figure 2 Here)

The features of every single image in the database are stored in the index table. The MQIR system
extracts features of images in multi-query. Every image in the query set has its own representations for the query processing. For each query image in multi-query, the proper weight for similarity measurement is set. And finally, the MQIR system will express the retrieval results in similarity ranking order. The main focus of following literature survey is to study various methods to retrieve similar images from database by combining low level features of multiple query images in CBIR systems.

4 Related work

4.1 Maximum of multiple queries and Average of multiple queries

Elena Stojanova et al (2013) [8] proposed two single-group methods are used. In the first method, maximum of the multiple queries (MQ-Max), each query is executed separately and separately from the other queries in the same set. When the retrieval phase is finished, the retrieved ranked lists are combined so that the rank of every image in the database is determined as the maximum of the individual scores extracted from each query. The other method, Average of multiple queries (MQ-Avg) is very similar with the first method, and belongs to the single group multi query methods as well. The main difference lies in the phase where the ranked lists are combined. In this case, rather than calculating the maximum of each score, the image score in final ranked list is calculated as the average of the individual scores. The results were compared based on the Mean Average Precision. The best value of MAP was obtained with MQ-Avg method and ten images in the query set. The drawback of this approach is that the bigger number of images in the query set not always leads to better results.

4.2 Query expansion methods

Yanzhi Chen et al (2012) [9] proposed a multi-group method to improve the content based image retrieval performance using multi-query. A novel discriminative ranking criterion is proposed for multiple queries. This method can work with a single query image as well as exploits the extra information in multiple queries when available. These query instances are used to automatically collect the set of query-dependent positive and negative data samples as used by discriminative query expansion (DQE). In contrast to average query expansion (AQE), which uses positive samples to improve query recall, the dataset images is re ranked in DQE by a data-dependent weight vector learnt from both positive and negative samples. Unlike the linear ranking models, in this proposed method constructs a nonlinear ranking model using an ensemble of linear support vector machines (SVM) which are adaptively weighted by boosting. The
disadvantage of this method is the requirement of negative images, which may not be available in all the cases. Basura Fernando (2013) [10] proposed a new method based on pattern mining in the multiple query CBIR systems. Using the minimal description length principle, the most suitable set of patterns is derived to describe the query object from the set of query images, with patterns corresponding to local feature configurations. The archive dataset can be obtained even without costly re-ranking based on geometric verification. Then be searched in efficient manner for similar images based on this representation, using a combination of two inverted file systems. Based on the obtained results, the query can be expanded and the process can be repeated.

4.3 Aggregation of query images

Cai-Zhi Zhu et al (2014) [11] studied how aggregating multiple images, in query or in database, impacts the performance of visual object retrieval in a Bag-of-Words framework. Five different multi-image aggregation methods has been compared, and suggested selecting the average pooling method in mostly for its very superior advantages in accuracy, speed, and memory footprint. Though current aggregation methods are not very optimal in certain cases, where aggregation of the entire given query images would not present a better result, in this case selective aggregation approach is proposed.

4.4 Early and late fusion strategies

Fatih Calisir (2015) [12] proposed an approach to improve the mobile image retrieval performance using multiple query images. Here bag-of-visual-words approach is used to represent the images, and employ early and late fusion strategies to utilize the information in multiple query images. With extensive experiments on an object image dataset with a single object per image, the study shows that multi image queries result in higher average precision performance than single image queries. The major drawbacks of the above methods is that they are more suitable only for single semantic multiple query images. They cannot support query images with different semantics, where features of every query image is required to be present in the retrieved images.

4.5 Pareto front method

Considering the content-based image retrieval problem for multiple query images corresponding to different image semantics, Ko-Jen Hsiao (2015) [13] propose a novel multiple-query information retrieval algorithm that combines the Pareto front method with efficient manifold ranking. The goal is to find images related to all queries. This algorithm can retrieve samples that
are not easily retrieved by other multiple-query retrieval algorithms. Drawback of this method is that it is hard to incorporate the entire features of every image and may discard some important features in a particular image.

4.6 Query replacement method

Vimina E R (2015) [14] proposed a multi query system using query replacement algorithm which utilizes the statistical features of a query image set to determine the similarity of the candidate images in the database for ranking and retrieval. In this work a novel query replacement algorithm is proposed for boosting the efficiency of a multi-query CBIR system. The algorithm is based on the principle that if an element in set X is to be replaced with an element in set Y, it will cause minimum information change if the replaced element has high similarity with the element being replaced. Using this method with smaller number of query images, high retrieval precision rate is obtained but the major drawback is high computation cost during the run time. (Refer Table 1 Here) for Comparison of techniques.

5 Conclusion

In this paper, essential concepts of single-query content-based image retrieval systems are studied and limitations of various traditional semantic gap reduction techniques for CBIR are examined. Multi-query approach is proposed for efficient semantic gap reduction in CBIR. Survey is done on different multi-query content based image retrieval systems. Techniques used in these systems increases the efficiency of content based image retrieval systems by reducing the spatial gap between low level features and high level concepts and thus increases the precision rate of image retrieval. Usage of SIFT feature descriptor for feature extraction produces better results than other feature extraction algorithms. Pattern based query expansion method can be used in multi query approach to obtain higher precision rate.

References


List of Figures:

Figure 1: Content based image retrieval system
List of Tables:

Table 1: Comparison of Multi query techniques
<table>
<thead>
<tr>
<th>Paper</th>
<th>Method</th>
<th>Feature Extraction Algorithm</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yanzhi Chen (2012) [9]</td>
<td>discriminative query expansion</td>
<td>SIFT feature descriptor</td>
<td>Suitable for multi group multi queries, DQE produces better results than Average query expansion methods. Requirement of large query set with positive and negative image groups is the major drawback.</td>
</tr>
<tr>
<td>Vimina E R (2015) [14]</td>
<td>Query replacement method</td>
<td>Colour and Edge directivity descriptor.</td>
<td>Using this method with smaller number of query images, high retrieval precision rate is obtained but the major drawback is high computation cost during the run time.</td>
</tr>
</tbody>
</table>