Abstract. The last decade is evidenced that there is a great interest in research on content-based image retrieval (CBIR). This has paved the way for a large number of new techniques and systems, and a growing interest in associated fields to support such systems. There are different system designs for content based image retrieval (CBIR) system. This paper propose a novel system architecture for CBIR system which combines techniques include content-based image and color analysis, as well as data mining techniques. To our best knowledge, this is the first time to propose segmentation and grid module, feature extraction module, K-means and k-nearest neighbor clustering algorithms and bring in the neighborhood module to build the CBIR system. Concept of neighborhood color analysis module which also recognizes the side of every grids of image is first contributed in this paper. The results show the CBIR systems performs well in the training and it also indicates there contains many interested issue to be optimized in the query stage of image retrieval.

Likewise, digital imagery has expanded its horizon in many directions, resulting in an explosion in the volume of image data required to be organized. In this paper, we discuss some of the key contributions in the current decade related to image retrieval and automated image annotation. We also discuss some of the main challenges involved in the adaptation of existing image retrieval techniques to build useful systems that can handle real-world data.

The results show the CBIR systems performs well in the training and it also indicates there contains many interested issue to be optimized in the query stage of image retrieval.

Keywords: content based images retrieval; K-means clustering; feature extraction;

1. Introduction

Image retrieval is the processing of searching and retrieving images from a huge dataset. As the images grow complex and diverse, retrieval the right images becomes a difficult challenge. For centuries, most of the images retrieval is text-based which means searching is based on those keyword and text generated by human’s creation.[1] The text-based image retrieval systems only concern about the text described by humans, instead of looking into the content of images. Images become a mere replica of what human has seen since birth, and this limits the images retrieval. This may leads to many drawbacks which will be state in related works.

For decades, text in a given language has been set to order, to categorize and to search from, be it manually in the ancient Bibliothek, or automatically Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

To overcome those drawbacks of text-based image retrieval, content-based images retrieval (CBIR) was introduced [2][3]. With extracting the images features, CBIR perform well than other methods in searching, browsing and content mining etc. The need to extract useful
information from the raw data becomes important and widely discussed. Furthermore, clustering technique is usually introduced into CBIR to perform well and easy retrieval. Although many research improve and discuss about those issues, still many difficulties hasn’t been solved. The rapid growing images information and complex diversity has build up the bottle neck.

Interpretation of what we see is hard to characterize, and even more so to teach a machine such that any automated organization can be possible. Yet, over the past decade, ambitious attempts have been made to make machines learn to understand, index and annotate images representing a wide range of concepts, with much progress. We, as pursuers of the ultimate goal to build intelligent machines capable of image management the way humans are, feel that it is now time for the community to move aggressively into making it a real-world technology. We sense a paradigm shift in the goals of the next-generation researchers in image retrieval. The need of the hour is to establish how this technology can reach out to the common man in the same way text retrieval techniques have. For example, Google and Yahoo! are household names today, primarily due to the benefits reaped through their use.

To overcome this dilemma, in this paper, we propose a novel CBIR system with an optimized solution combined to K-means and k-nearest neighbor algorithm (KNN). A creative system flow model, image division and neighborhood color topology, is introduced and designed to increase the clustering accuracy. The rest of this paper is organized as follows. Section 2 briefly describes the concepts of images retrieval, feature extracting, K-means and the structure of CBIR system. In Section 3, we provide a description of the Optimized Content Based Image Retrieval model with K-means and KNN combing with those module proposed in this paper. Experimental results and discussions are presented in Section 4. Finally, Sections 5 discusses our conclusion and future works.

2. Previous and Synonymous Work

This chapter introduces some important literatures review in this chapter. First come to content based image retrieval (CBIR). Before CBIR, the traditional image retrieval is usually based on text. Text based image retrieval has been discussed over those years. The text-based retrieval is easy to find out some disadvantages such as:

1. Manually annotation is always involved by human’s feeling, situation, etc. which directly results in what is in the images and what is it about.
2. Annotation is never complete.
3. Language and culture difference always cause problems, the same image is usually text out by many different ways.
4. Mistakes such as spelling error or spell difference leads to totally different results.

In order to overcome these drawbacks, content based images retrieval (CBIR) was first introduced by Kato in 1992. The term, CBIR, is widely used for retrieving desired images from a large collection, which is based on extracting the features (such as color, texture and shapes) from images themselves. Content-Based Image Retrieval concerns about the visual properties of image objects rather than textual annotation. And the most popular and directly features is the color feature, which is also applied in this paper. In this work, color is selected as a primary feature in images clustering.

2.1. Data Clustering Method

Data Clustering is often took as a step for speeding-up image retrieval and improving accuracy especially in large database. In general, data clustering algorithms can be divided into two types: Hierarchical Clustering Algorithms and Non-hierarchical Clustering Algorithms. However, Hierarchical Clustering is not suitable for clustering large quantities of data. Non- hierarchical is suggested to cluster large quantities of data. The most adapted method for non-hierarchical clustering is the K-Means clustering algorithm proposed by James Mac Queen in 1967.

K-means clustering algorithm first defined the size of K clusters. Based on the features extracted from the images themselves, K-means allocates those into the nearest cluster. The algorithm calculates and allocates until there is little variation in the movement of feature points in each cluster. This paper applies k-means clustering algorithm for color clustering module based on this concept.

2.2. Role of Feature Extraction

In the image retrieval, feature extraction is the most import issue of the first step. The features extracted from the images directly lead to the results. Some preprocessing is needed to avoid retrieval noise. Steps such as removing the background, highlights the objects. All the preprocessing steps help in feature extraction [18]. Different topic images usually contain different features [19]. Deciding the features needed to be extracted is always popular issues, and then it still comes back to basic features of the images such as:

Color histograms are basic and fundamental feature of the images. It is the most commonly used and usually gets obvious effect result. Vary literatures apply color histograms as basic images comparison [20] [21].

Tamura Features. Tamura proposed six images features based on human visualization [22], coarseness, contrast, directionality, line-likeness, regularity, and roughness. Those are usually considered in many related research [3] [2]. Shape Feature. Besides color, shape is the most commonly used features. Some image retrieval applications require the shape representation to be invariant to rotation, translation, and scaling. Shape representation can be divided into two o categories, boundary-based and region-based [3]. MPEG-7 Features. The Moving Picture Experts Group (MPEG) defines some description of visual named MPEG-7. It includes Color Layout, Color Structure, Dominant Color, Scalable Color, Edge Histogram, Homogeneous Texture, Texture Browsing, Region-based Shape, Contour-based Shape, Camera Motion, Parametric Motion and Motion Activity features [23].

2.3. Importance of Contrast Context Histogram

Contrast Context Histogram (CCH) is first proposed for image descriptor [15]. CCH first uses histograms to measure the differences in intensity between various salient points around an object.

Through comparing the positive and negative histograms of the same object, it derives efficient and discriminative descriptors. The CCH method effectively resists image variability between photos of the same object. The CCH extract features in our system as fellow:

First, we assume that each photo has already been analyzed to derive several salient corners. For each center of a salient corner Pc, in the N*N pixel region R, we calculate the center-
based contrast $C(p)$ of point $p$ in the region $R$ as follows:

$$C(p) = I(p) - I(p_c)$$  \hspace{1cm} (1)

Then, for each central point $p$ in every region $R_i$, the positive contrast histogram and negative contrast histogram as shown in (2) and (3) respectively:

$$H_{p}^{+}(p) = \sum_{p \in R_i \text{ and } C(p) \geq 0} \frac{C(p)}{\#_{p}}$$  \hspace{1cm} (2)

$$H_{p}^{-}(p) = \sum_{p \in R_i \text{ and } C(p) < 0} \frac{C(p)}{\#_{p}}$$  \hspace{1cm} (3)

To combine the values of the positive and negative contrast histograms of all regions in $R_i$, the CCH descriptor of salient corners is defined as:

$$CCH(p) = (H_{p}^{+}, H_{p}^{-}, H_{p}^{+}, H_{p}^{-}, \ldots, H_{p}^{+}, H_{p}^{-})$$  \hspace{1cm} (4)

3. Proposed Method

In our CBIR system, it is divided into two parts:

- Learning
- Querying.

The learning part tells about the training process which a huge mount sample images are input in the first step, then the images’ features are extracted for the clustering. K-means algorithm is selected to cluster the training data because of it is easy to implement, efficient and well developed in the recent 50 years. Finally the training output the clustering result as a learning code book.

The query part describes the images searching process. Inputting the query images and matches to the training result. The output shows the most similar images for user’s query. Figure 1 shows the overview of the CBIR system.

The main system architecture contains four modules both in learning stage and query stage:

- Segmentation and grid module
- The K-Means clustering module
- The feature extraction module
- The neighborhood concept module

The flow chart of the system architecture is shown in Figure 2. Each module is described as follow.

- **Fig. 1. Basic Overview of the CBIR system**

3.1. Image segmentation and grid module

The data images input into the system will be first processed in this module. In the images retrieval, larger images usually decrease the retrieval accuracy. Small images grids help in feature extraction and images processing. Therefore, this module first divides the images into $F*F$ grid and every grid will divided again into $S*S$ sub-grids while during the feature extraction module. In this stage, an input images will finally divided into $(F*S)^2$ grids which the $F*F$ grids uses both in feature extraction module and neighborhood module. The inside $S*S$ grids are only used in color
feature extraction. Figure 3 shows the sample of images segmentation.

3.2. Feature Extraction Module

The input images, including the training and query stage, are all processed in this module. It is also the most important in image retrieval. Since color is the most popular and intuitive feature based on human visualization, it is applied in the system. In order to get more powerful features, the CCH method is also applied for extracting the important feature point. The two feature extractions are described as below:

1. Color Feature Extraction

   Input images will be divided into F*S grids before this stage. All grids are input to extract the color feature. First, the module compute the average RGB value of the F*F grids. Second, the inside S*S grids in every F*S grids will also be input to calculate the average RGB value. The S*S grids’ detail RGB information is append after the F*S grids’ color feature information. All those are prepared for first K-means clustering. Figure 3 illustrates the color feature extractions of this stage.

2. CCH feature extraction

   The system utilizes CCH to find out the important feature points. All the points are detected for preparing the input data of the neighborhood module and K-means clustering or KNN classifying. The information of CCH feature points, including the 64 dimensions data, combines with the neighborhood module result. Taking it as the input for the second round K-means clustering, the K-means clustering results in a fragment-based database, call the Code book. As the same implementation in query step, K-means is replaced by KNN algorithm. Query data inputed will be classified to improve the training code book, also correct classified result helps for quickly retrieval. Figure 5 shows the sample for CCH feature extraction.

3.3. The Neighborhood module

In this module, the input data is from the CCH feature points. Feature points of every image will consider as an index to build up the neighborhood table. Also the first K-means clustering result of every F*F grids are imported to denote the value of the neighborhood table. The steps of every detail are described as follow:

1. Input the CCH feature point Y of the X picture, represented as PICXY.
2. Get the first K-means clustering result based on the CCH feature point’s coordinate.
3. Get the neighborhoods’ first K-means clustering results.
4. Appending the results from step 3 according with the order left to right then top to bottom. If there is no neighborhood, then the value will be “0”, which stands for the side of the pictures.

5. Appending the CCH information into the neighborhood table.


Figure 5 shows the data presentation and processing of the neighborhood module. Taking the same format, the query step apply KNN algorithm instead of K-means.

3.4. The K-means/KNN module

K-means clustering is applied twice in our system. The K-means clustering helps generate the code book. In order to keep those input (include the training and query stage) being clustered in the same standard, our architecture keeps the cluster central points in the training stage. The central points are imported into the K-means clustering in the query stage. Finally, the code book can be mapped in the same comparing standard.

K-nearest neighbor algorithm (KNN) is also involved into our CBIR system. Based on the training result, KNN is applied for the query data images. KNN helps to classify the input data; also it fixes the code book which means the training result can be self-adapted.

3.5. The image retrieval query module

This paper provides a query method based on modules mentioned before. Considering the grid fragment, color feature, and CCH feature points, all images input for query will be divided into pieces. Then KNN is applied to classify those images which maps to the training result (code book). The query grid images are compared with those picture grids in the same cluster, the system then calculates the difference based on the color feature which is introduced in 3.2. All fragments are tagged and linked with one grid in the code book. By calculate the most amount of the grids, the CBIR finally output the query and retrieval result.

4. Implementation and Experimental Results

Based on the method provided in this paper, the experiments are designed to verify the architecture of the CBIR system. Also the experiments shows the modules proposed in this article perform good and well organized with the CBIR system architecture.

First the CBIR system splits and calculates the average RGB, we have implemented and figure 6 shows samples of the color feature and split result.
Fig. 6. Samples of the color feature result.

The color features extracted from \(X\) pictures divided into \(Y\) fragments is denoted as table 1 and pictures training clustered result is visualized as shown in figure 7.

In the code book, the color features and the CCH features are included. Based on the fragments, image can be retrieval in detail and the clustering helps decrease the computing cost. This paper has verified several settings for proposed CBIR system architecture. The results indicates the system perform well for training and querying.

In order to verify the CBIR system, the Wang’s dataset which includes 386*254 pixel images is applied as the testing and training data [25]. The 1,000 images from the Wang’s dataset are applied into the codebook. 50 images are randomly selected as query images. With the CCH feature points, the CBIR system proposed in this paper successfully retrievals the correct images for those query images. Furthermore, landscaped images are selected from 10,000 images as the training and querying images [25]. The results indicates the CBIR system proposed retrievals correctly when it is well trained. Table 2 shows the query sample of the Wang’s dataset.

Table 2. Retrieval results for the Wang’s dataset.

<table>
<thead>
<tr>
<th>Data Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Picture Number</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
| ... | ... | ... | ...
| X | Y | [8,132,24] | [99,1,24]...[89,256,1] |

Fig. 7. Cluster NO.13 derived by K-Means in the code book.

In the code book, the color features and the CCH features are included. Based on the fragments, image can be retrieval in detail and the clustering helps decrease the computing cost. Several images are input for images retrieval. To query for similar images, the system operates as section 3 described. Table 3 shows the image retrieval results. The CBIR system proposed in this paper successfully find out the pictures which are included in the code book. For those which are not in the code book, the similar pictures are also figured out as table 3 shown. Also the retrieval processes which are generated by the grid module is shown in table 3, the grid puzzle images. The experiment indicates that the CBIR system retrievals the easy-to-tell similar image, even though the image inputted for query is never trained in the codebook.
5. Conclusions and Future Work

The proposed system is designed to operate the content based image retrieval system. It has been verified with the photos of places of interest in Taiwan and the Wang’s dataset [24][25]. Our experimental results demonstrate that our CBIR system architecture not only works well for image retrieval, but also improves its precision.

In our knowledge, this paper first combines segmentation and grid module, feature extraction module, K-means clustering and neighborhood module to build the CBIR system. Furthermore, the concept of neighborhood module which recognizes the side of every grids of image is first contributed in this paper. Applying the concept of fragment based code book into the content based image retrieval system also contributes in our system architecture. The experimental results confirm that the proposed CBIR system architecture attains better solution for image retrieval. Our model represents the first time in which combine new modules and techniques proposed in the paper have been integrated with CBIR system.

Images can be retrieval correctly through the proposed CBIR system. For those images which are contained in the code book, all of them can be searched as the most similar result. Also for general images selected randomly, the query results are similar to the input data. Since the CBIR system is based on the color feature, the retrieval results are directly and easy to tell the performances. In the future work, we hope to build a generalized query method which increase the system searching ability and provide more accurate content descriptions of places of interest places by performing color feature analysis and CCH image extraction simultaneously. As a result, the CBIR system will be able to suggest more relevant annotations and descriptions.

Furthermore, we hope to optimize the system architecture and modules proposed in this paper. There exists some detail setting can be discussed and optimized with the images retrieval issues.

References


