The design of content based image retrieval with a combination of visual content features

Chu Thinzar, Hlaing HtakeKhaung Tin

University of Computer Studies, Yangon, Myanmar, Email: hlainghtakekhaungtin@gmail.com

Article History
Received: 26 June 2019
Accepted: 19 August 2019
Published: August 2019

Citation
Chu Thinzar, Hlaing HtakeKhaung Tin. The design of content based image retrieval with a combination of visual content features. Science & Technology, 2019, 5, 189-194

Publication License
This work is licensed under a Creative Commons Attribution 4.0 International License.

General Note
Article is recommended to print as color digital version in recycled paper.

ABSTRACT
This paper describes the design and implementation of image retrieval and classification with a combination of visual content features, such as Color Moment, Color Auto-Correlogram and Gabor Wavelet features. When a new query image enters to the system as an input, it search and retrieving the relevant images from the image database. This system uses a combination of three types of features: Color Moment, Color Auto-Correlogram and Gabor Wavelet features to provide similar images of the user's query images. In order to improve the performance and accuracy of image retrieval, the Support Vector Machine (SVM) classifier is used with visual features. The experience of the system has shown that the CBIR using the SVM classifier with Color Moment, Color Auto-Correlogram and Gabor Wavelet features produced better results than the CBIR based on these features. This system is implemented with MATLAB programming language on window platform according to the processes of system design. This system accepts the input color image (RGB) with .jpeg extension. According to the feature extraction techniques, the color image is converted to grey scale and HSV color space.

Keywords: CBIR, Color Moment, Color Auto Correlogram, Gabor Wavelet features
1. INTRODUCTION

Numbers of digital images have increased dramatically due to the use of services of multimedia, digital cameras and the fact that the facade has become increasingly popular on the Internet. In many areas, the use of digital images is increasing. Given this situation, it is necessary to classify images into significant categories and objects such as tree, dog, humans, and planes to manage and organize the images in the database.

Digital images are used in multiple applications, e.g. Geography, medicine, architecture, advertising, military design and album. However, it is difficult for users to find and manage the maximum number of images in the database. In general, image retrieval system is divided into two methods such as Text-Based Image Retrieval (TBIR) and Content-Based Image Retrieval (CBIR).

The first method is text-based image retrieval system. This is easy to implement and the conventional database query technique where textual metadata to each image is added to retrieve image by keywords. This requires a manual annotation of the database images and consumes huge time and also a complex one. Moreover, the annotation process is not an effective format since users generally are not interested to make it in systematic way. This is because there are diverse users who tend to use different words to describe a same image characteristic and also more than one object can be referred by the same word, inefficiency. Hence, this method loses its popularity and performance due to lack of systematization.

The second one is content-based image retrieval system. Text-based image retrieval has some disadvantages, such as incompetence, data loss, time-consuming processing, and more complex tasks. This problem is solved using content-based image retrieval. “Content based” means searching and analyzing content of images that are not metadata [3].

Metadata is a keyword, tag, or description related to an image. The term “content” refers to color, shape, texture or other information that can be obtained from an image. CBIR is desirable because most web search engines use purely metadata information and result in large amounts of garbage. Also having humans manually enter image keywords in a large database, it may be inefficient, costly, and may not capture every keyword that describes the image. As a result, systems capable of filtering images based on content will allow better indexing and more accurate results.

There is a growing interest in CBIR because of the limitations inherent in metadata-based systems as well as the large range of possible uses for efficient image retrieval. The capability of present CBIR systems has been limited by their use of only primitive features, so they cannot satisfy most semantic-level query demands. Primitive features denote some general visual characteristic including color, shape, texture, and spatial relationships among objects, and these features can be used in most CBIR applications. The color feature which is widely used in CBIR systems.

2. CONTENT-BASED IMAGE RETRIEVAL OR CBIR

Image retrieval that is user-defined image from a large image database is a critical image processing technique. The method for large collections of multimedia and digital libraries has created a great need for the development of search engines to index and retrieve data. There are now a lot of good search engines that can extract texts in a format that the device can read. However, there is no quick tool to extract intensity and color images. The traditional way to search and index image is slow and expensive. Therefore, there is an urgent require to progress an algorithm to retrieve embedded content images. The digital image features (shape, texture, color, object structure, etc.) can be used as an index button to find and extract images from a large image database. Image retrieval based on the content of the image is called the content-based image retrieval [12]. Content-based Image Retrieval or CBIR involves two steps:

Feature Extraction
The first step in the feature extraction process is to extract the features of the image at different levels. These features are defined as one or more measurement methods, each specifying certain quantitative properties of the object and being calculated from the quantitative properties of the object. The extraction of features is a special form of dimensionality reduction.

Matching
The second step is to match these properties to achieve similar results. Many visual information in the form of video and image data is scattered around the world and other irrelevant information such as numbers, numerical data, voice, audio, text, learning etc. Data mining get the important information from large data sets through data embedded models and knowledge to be discovered. Data mining involves the retrieval of beneficial data embedded in large images and videos. Most of the jobs in this area are limited especially for the improvement of content-based image retrieval systems (CBIRs). Retrieving images from a large image database is a vital task in terms of image processing and artificial vision.
Image retrieval with similarities is an elegant method applied developing CBIR. Basically, the CBIR system ought to automatically capture important data about the image for certain applications. The content of the image is a natural image and the interpretation of the information passed through the image is subjective and depended upon the basic human image system. Image data is applied for artificial vision banked on the desired image properties and interpreting these features for specific applications.

Most image retrieval activities are made up of searches and retrieval of images based on image equivalency analysis or features of the image database. The image retrieval system could be divided into two categories by the search form. For first group, image is explained based on user-defined text [4, 7]. The image is indexed and extracted according to the basic description; size, type, date and time of acquisition, owner’s identity information, keyword or textual description. For this reason, it is often referred to as image retrieval depend on the description or text. The image index is predefined based on descriptions and searches for these indices when the query “Search Images within the database is stored according to the specified set of descriptions”.

The text-based description is usually manually printed for each image by human operators, as creating automatic keywords for this image is difficult without having to combine visual information and extracting features. For this case, a process of producing energy that is not practical at the age of multimedia information. Because the image description is very subjective, the automotive process of creating a description of the text used to index the image can also be false and incomplete.

For second group, the query can point in following: “Search image is similar to the image of the specified query image”. The second group of an image retrieval process with similarity is CBIR [2-6]. In the CBIR system, searches and captures images depended on the content of these images and the desired image functions can be retrieved and used as a search for indexes or bases. Basically, the image recovery system includes three primitive contents such as: visual content or feature extraction, multidimensional indexing and retrieval.

The images in the database are indexed according to extract sections visual content (or features) such as color, texture, pattern, structure of the image, the shape of the object, the layout and position in the image, etc. The image can be acted as a multidimensional vector of features retrieved from the image. The feature vector performs by image signature. This feature vector can be assumed to be associated to a point in the multidimensional space. For example, the images can be displayed with the N-Dimensional feature vector, where the first element n1 can perform the color of the component. The following n2 can represent the shape of the component. The following n3 can represent the topology of some images and finally the element n4 can represent the surface of the images to contain the components N =n1+n2+n3+n4.

Image queries can be analyzed to distinguish visible and comparable functions to find matches in image indexes stored in databases. The extracted image function is saved as metadata and images indexed depended on this metadata. This metadata contains several dimensions of the extracted image function. The feature vector is to measure several different image functions. The feature vector with the same image is summarized in N-dimension space. The retrieval of an image similar to query image then decreases to find the index of the image in the search space. N-dimensions with vector features are in N-dimension space.

Architecture for the CBIR system is separated into two parts. In the first part, images from the database are preceeding offline. Each image features in the image database will be extracted to create image metadata describing the image using the image content features. This function then indexes the image and is stored with the image in the metadata database. The second part describes the image retrieval process. Image query is analyzed to retrieve visible features. These features take same images from the image database. Instead of directly compare both images, the visual properties of the query images are measured by the characteristics of image stored in the Meta database. Often, the similarity of both images is determined by calculating the distance between the vector vectors of two images. The system takes the image by returning the first image k, which has the distance from the same request image or less than the specified threshold.

Many image features can index images based on content. The most popular image retrieval system between them is color, texture and shape, image topology, color arrangement, interest region and so on.

3. PROPOSE SYSTEM DESIGN
The architecture of the content based image retrieval system with SVM classifier is described in figure 1. The system accepts input images from users. Then the Color Moment, Color Auto-correlogram and Gabor Wavelet features will be extracted from the query image, and then the system will extract the image by comparing the similarity of the measurement between the query image and the images of the database stored with the SVM classifier to efficient query results. Finally, the system will compare the similarities values of these features with the features of the image database. The same value for each feature for query image, and database images are calculated by Euclidean distance.
One of the most significant techniques for content-based image retrieval is color histogram. It is effective for calculation and good for searching results. For an \( m \times n \) image \( I \), the colors of the image is quantized to \( C_1, C_2, \ldots, C_k \). The color histogram \( H(I) = \{h_1, h_2, \ldots, h_k\} \), where the number of pixels is \( C_i \).

\[
\Pr(P \in C_i) = \frac{h_i}{m \times n} \tag{1}
\]

The color histogram also shows the possibility of any pixel in image \( I \), which is color \( C_i \). The weakness of the histogram method is that there is no space information in the color histogram. Color auto-correlogram techniques have been proposed to combine spatial information with color histograms. Auto-correlogram of image \( I \) for color \( C_i \) can be determined with distance \( k \)

\[
\gamma_{C_i}^{(k)} = \Pr(|p_1 - p_2| = k, p_1 \in I_{C_i}, p_2 \in I_{C_i}) \tag{2}
\]

The correlogram feature of image represent in figure 2 how the spatial autocorrelation of color changes with distance.

Figure 2 Sample color auto-correlogram feature

Auto-correlogram combines information about colors and spaces. For each pixel in the image, the auto-correlogram method must cross all the neighbors’ pixel. Correlogram method is more stable in color change than in histogram.

Color Moments

Color moment is a measurement of the color distribution in an image. It is primarily used for color indexing purposes because of the features of image retrieval applications to compare how similar two images are based on color. It encodes both shape and color information which is a good feature to be used under changing lighting conditions and can be calculated for any color model. This system uses two types of color moments. These are:

a) Mean

The first color moment can be interpreted as the average color in the image, and it can be calculated by using the following formula:

\[
E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij} \tag{3}
\]
where $N$ is the number of pixels in the image and $p_{ij}$ is the value of the j-th pixel of the image at the i-th color channel.

**b) Standard Deviation**

The second color moment is the standard deviation, which is obtained by looking for the square root of the variance of the color distribution.

$$\sigma_i = \sqrt{\frac{1}{N} \left( \sum_{j=1}^{N} (p_{ij} - E_i)^2 \right)}$$  \hspace{1cm} (4)

Where $E_i$ is the mean value, or first color moment, for the ith color channel of the image.

**Gabor Wavelets**

This plot is defined as a surface structure formed by repeating certain elements in different relative spatial positions. Repetition includes a local variation of scale, orientation of the elements. Image textures are defined as images of natural textured surfaces. It contains important information about the structural arrangement of the surface: clouds, leaves, bricks, etc.; and also describes the relationship between the surface and the surrounding environment.

This feature describes the distinctive physical composition of a surface. Gabor wavelet is widely used to extract texture of the images. Each wavelet capturing energy at a specific frequency and specific orientation. Gabor filter can be shown with the equation 3.5.

$$\Psi(x, y, \lambda, \theta) = \frac{1}{2\pi S_x S_y} e^{-\frac{x^2}{2S_x^2} + \frac{y^2}{2S_y^2}} e^{j2\pi \frac{x}{\lambda} \theta}$$  \hspace{1cm} (5)

where ($x$, $y$), the pixel position in the spatial domain
- $\lambda$, Wavelength of frequency of pixels
- $\theta$, Orientation of a gabor filter
- $S_x$, $S_y$, Standard deviation of the x & y directions

**Support vector machines**

The concepts of image classification and Content Based Image Retrieval have been combined in the proposed system is to classify the images to identify and retrieve the formation of query image. Classifier is generally used to classify the category of each A supervised learning model involving learning algorithms that analyzing data. It works by accepting the complete set and then reading it so that each of the relevant input, relevant outputs is extracted. The whole process is considered a classification. Classifying the data by looking for the best hyperplane separates all the data points of a class from the other classes.

Support vector machines depend on the concept of the decision plan that determines the boundaries of the decision. The decision plan is a plan that separates sets of objects with different members in the class. Support vector machine learning algorithm is used to produce the classification parameters according to the calculation feature.

Results can be received in two discrete or continuous formats. The classifier assigns the input space and the feature space. Feature spaces are defined as stored space to calculate similarities using kernel functions.

Support vector machine (SVM) formerly separated binary classes ($k = 2$) with the maximum margin criteria. However, problems in the real world often require more than two categories of discrimination. Therefore, the recognition of multi-class has a wide range of applications, including optical character recognition, inclusion detection, speech recognition and bioinformatics. In practice, the classification problems of multi classes ($k> 2$) are generally divided into a set of binary problems, so that the standard SVM can be directly used. The set of two schemes are one-versus-rest (1VR) and one-to-one (1V1) guidelines.

**a) One-Versus-Rest Approach**

One-versus-one (1VR) method creates a different binary classifier k for k-class classifications. The m-th binary classifier is trained using the m-th class data as a positive example and the remaining k-1 classes is a negative example. During the test, the class label is determined by the binary classifier that gives the highest output value. Imbalanced training is a major problem of 1VR approach. Suppose that every class has an equal training size. The proportion of positive and negative samples in each individual classifier is k-1. In this case, the symmetry of the original problem is lost.
b) One-Versus-One Approach

Another classical method of multi-class classification is a one-to-one d (1V1) or pairwise decomposition. It evaluate all possible pairwise classifiers and thus induces k(k−1)/2 individual binary classifiers. Applying each classifier to a test example would give one vote to the winning class. A test example is labeled to the class with the most votes. The size of classifiers created by the one-versus-one approach is much larger than that of the one-versus-rest approach. However, the size of Quadratic Programming (QP) in each classifier is smaller, which makes it possible to train fast. In addition, compared with the one-versus-rest approach, the one-versus-one method is more symmetric.

3. CONCLUSION

The concepts of image classification and Content Based Image Retrieval have been combined in the proposed system is to classify the images to identify and retrieve the formation of query image. Classifier is generally used to classify the category of each image. CBIR is the process of searching for relevant images in image database when the user sets a new image or query image. The proposed system presented: the fusion of auto-correlogram, color moments, gabor wavelet features and the combination of these three features and support vector machine technique, in similarity image retrieval system. SVM classifiers can be learned from relevant and irrelevant user-generated image for training data. The experimental results demonstrate that there is considerable increase in retrieval efficiency when the three types of visual features and SVM classifier are combined.

REFERENCE