

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.1274 – 1280

RESEARCH ARTICLE

Content Based Image Retrieval using Color Feature Extraction with KNN Classification

MS. PRAGATI ASHOK DEOLE¹, PROF. RUSHI LONGADGE²

M. Tech. (CSE) in G. H. Rasoni Academy of Engineering and Technology, Nagpur (MS) India
Pragatiac@gmail.com¹, rushilongadge@gmail.com²

Abstract

Image Retrieval system is an effective and efficient tool for managing large image databases. Content based image retrieval system allows the user to present a query image in order to retrieve images stored in the database according to their similarity to the query image. Content Based Image Retrieval (CBIR) is a technique which uses visual features of image such as color, shape, texture, etc. to search user required image from large image database according to user's requests in the form of a query. In this paper content based image retrieval method is used retrieve query image from large image database using three features such as color, shape, texture etc. The main objective of this paper is classification of image using K-nearest neighbors Algorithm (KNN).

Keywords— Image Retrieval; Content based image retrieval; Color Model; KNN Algorithm; Relative Standard Derivation

I. Introduction

Due to the advances in digital photography, storage capacity and networks speed, storing large amounts of high quality images has been made possible. Digital images are used in a wide range of applications such as medical, virtual museums, military and security purposes, and personal photo albums. However, users have difficulties in organizing and searching large numbers of images in databases, as the current commercial database systems are designed for text data and not well suited for digital images. Therefore, an efficient way for image retrieval is desired. (M. Vetterli, Feb. 2002). An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Image retrieval is classified into two types of retrieval are Text Based Image Retrieval and Content Based Image Retrieval. Text Based Image Retrieval is having demerits of efficiency, lose of information, more expensive task and time consuming. (T.Dharani, I.

Laurence Aroquiaraj,2013) Overcome these problems by using Content Based Image Retrieval (CBIR) system for image retrieval. The image retrieval system acts as a classifier to divide the images in the image database into two classes, either relevant or irrelevant. (V.Suryanarayana, N. R. Rao,2012).

II. Content Based Image Retrieval

In CBIR systems, image processing techniques are used to extract visual features such as color, texture and shape from images. The system uses a query model to convert the image into an internal representation of query, based on features extracted from input images. A retrieval model performs image retrieval by computing similarities between images in object and the query image. Content Based Image Retrieval (CBIR) is defined as a process to find similar image in the image database when a query image is given.

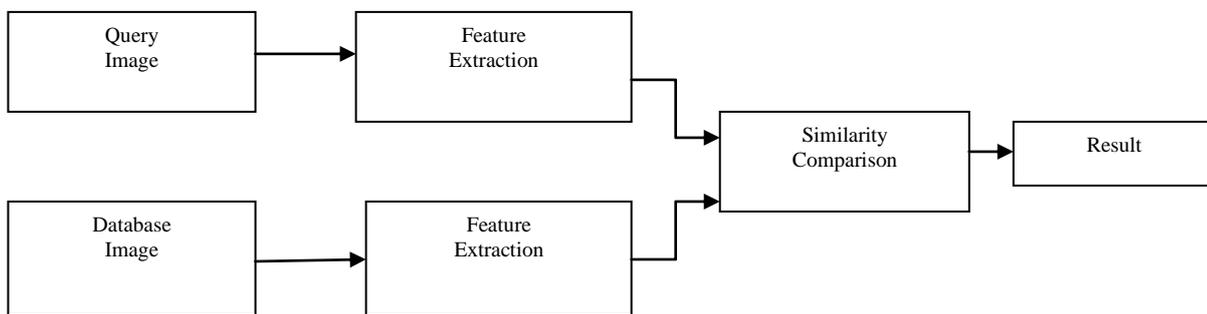


Fig 1. Block diagram of content based image retrieval

Figure 1 .show the basic block diagram of content based image retrieval. Textural features are extracted for both query image and images in the database. The distance (i.e., similarities) between the features vectors of the query image and database are then computed and ranked. The database images that have highest similarity to the query image are retrieved. Then the performance analysis is carried out using precision and recall. Content based image retrieval is working with different types of image database. All databases are having two types of images like labeled and unlabeled.

III. Color Feature Extraction

Color feature is most common feature of image. The color images are having the standard Color is RGB color. Color histograms are commonly use content based image retrieval. Feature means characteristics of object. Feature extraction is refers that dimensionality reduction of that object. It plays an important role in image processing. Features are classified into three types in image processing, that is low, middle and high. Low level features are color, texture and middle level feature is shape and high level feature is semantic gap of objects. These three features are extracted by different techniques. Color feature is extracted by Color Histogram and Color correlogram. Shape and texture feature is extracted by wavelet Transform.

Color Histogram

Color is an important attribute to describe the content of image. A color histogram which represents the proportion of specific colors in an image, has been widely used among color representation methods. Color histogram provides HSV color space and

RGB color space. colour histogram is computed which shows the proportion of pixels of each colour within the image. Then this colour histogram for each image is stored in the database. During the search time, the user can either specify the desired proportion of each colour or submit a reference image from which a colour histogram is calculated. The matching process then retrieves those images whose colour histograms match those of the query most closely .

1. RGB Color Space

The most popular color space is RGB which stands for Red-Green-Blue. This space consists of the additive primary colors of light Red, Green and Blue This histogram is the most used histogram in computer graphics and it uses its red, green and blue components to create a new color. This color histogram combines red, green and blue colors to form a new color. To form a new color it is necessary to increase the values of one or more of the components of the RGB components.

2. HSV Color Space

The HSV color space are defined in terms of three constituent components; Hue, Saturation and Value. Hue varies from 0 to 1.0, the corresponding colors vary from red through yellow, green, cyan, blue, magenta, and back to red, so that there are actually red values both at 0 and 1.0. As saturation varies from 0 to 1.0, the corresponding colors (hues) vary from unsaturated (shades of gray) to fully saturated (no white component). As value, or brightness, varies from 0 to 1.0, the corresponding colors become increasingly brighter.

Color Correlogram

Color correlogram are the feature of color information. The color correlogram has the advantages that includes the spatial correlation of colors, can be used to describe the global distribution of local spatial correlation of colors and is simple to compute.

Color Moments

Color moments are used differentiate images based on their features of color., These moments provide a measurement for color similarity between images. These similarity values can be compared to the values of images indexed in a database for image retrieval .The color histogram, color moments and color set only contain the color information of each pixel in an image Color moments are measures that can be used differentiate images based on their features of color. Once calculated, these moments provide a measurement for color similarity between images. These values of similarity can then be compared to the values of images indexed in a database for tasks like image retrieval.

Wavelet transform

The wavelet transform transforms the image into a multiscale representation with both spatial and frequency characteristics. It calculates the frequency. This allows for effective multi-scale image analysis with lower computational cost. Wavelet is popular tool in image processing and computer vision. Many applications, such as compression, detection, recognition, image retrieval have been investigated. Wavelet transform has features of space-frequency localization and multiresolutions. Wavelet transform are use for both shape and texture feature.

IV. Similarity Comparison

Relative standard Derivation

For similarity comparison, we have used Relative standard Derivation, *RSD* using equation,

$$SD = \sqrt{1/N \sum (X_i - X)^2}$$

$$RSD = \text{stdev}/\text{mean} * 100$$

The *relative standard deviation* (RSD) gives the precision as a percentage of the mean (average). More precise data yield a smaller RSD.

V. K-Nearest Neighbors Algorithm

K-Nearest Neighbors Algorithm also known as KNN, is a supervised machine learning method, which classifies the data with this classifier a set of data can be classified in order to discover which elements are from the set of data. The purpose of this algorithm is to classify a new object based on attributes and training samples. KNN algorithm classify query image to relevant image in image database. KNN calculate distance between training vector and test vector.

VI. Performance Measure

CBIR performance is analyzed by computing the values of precision and recall. Content Based Image Retrieval includes the standard measures. Those measures are the initial measures of the image retrieval process. The measures are Recall, Precision. These measures are used to improve the image retrieval process with various measures.

Precision:

Precision is the fraction of retrieved images that are relevant to the input image

$$\text{Precision} = \frac{\text{Total no. of Retrieval Relevant image}}{\text{Total no. of Retrieval image}}$$

Recall:

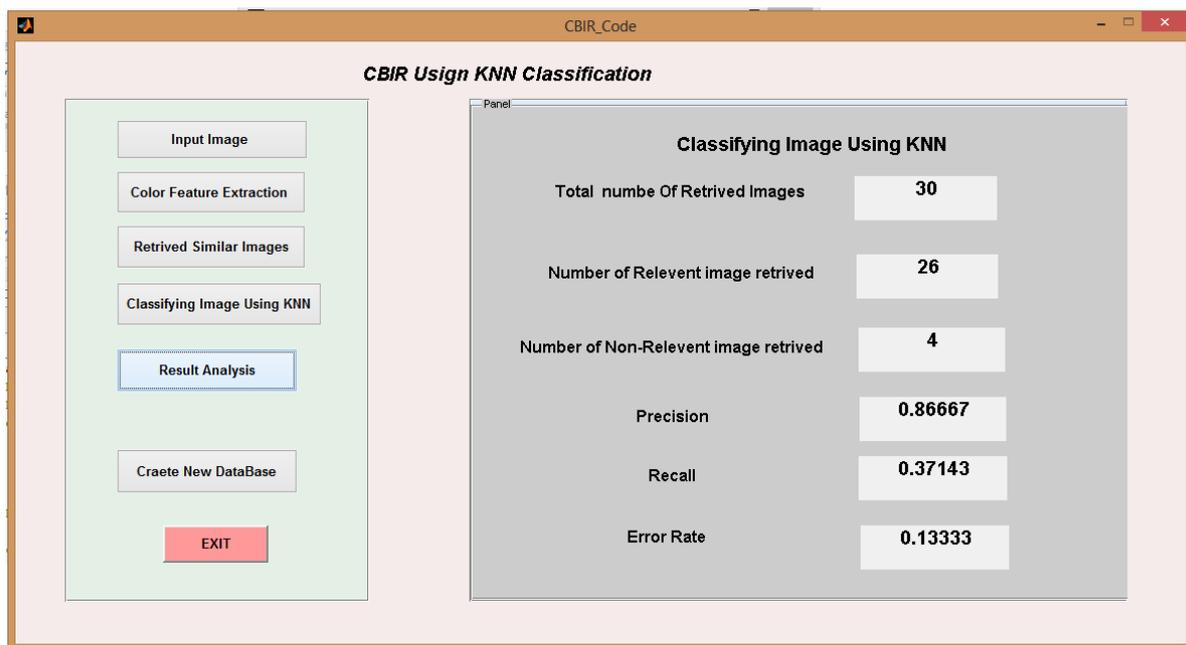
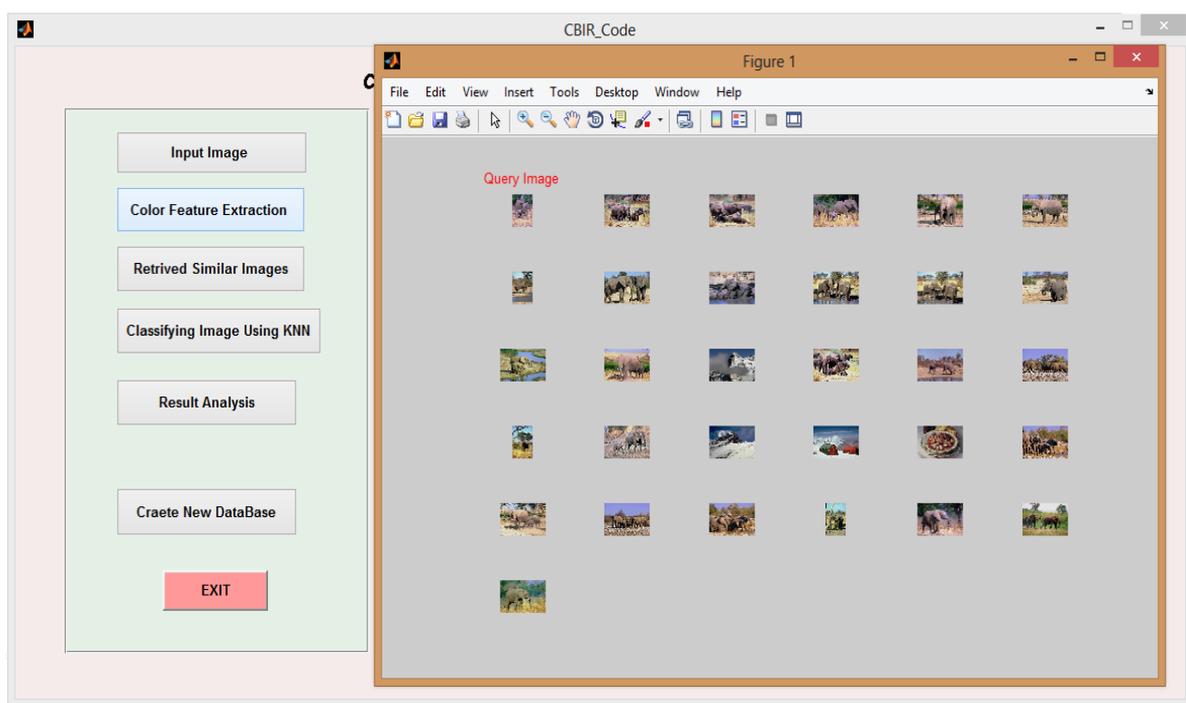
Recall is the fraction of the images that are relevant to the query that are successfully retrieved.

$$\text{Recall} = \frac{\text{Total no. of Retrieval Relevant image}}{\text{Total no. of Relevant image}}$$

VII. EXPERIMENTAL RESULT

The simulations were taken place in MATLAB7.10.0 .The proposed approach, the image database containing 500 image. In which 150 image are used for testing and remaining 350 image are used for training, so image database containing 350 image. First of all we have given number of image in database, user have to Select input image in test image then Open input image i.e query image. query image are display on screen ,after that query image retrieve no. of relevant image in large image database using color, shape and ,texture. Color feature are extracted by using color correlogram, color moment and hsv histogram. KNN Algorithm classify query image to relevant image in image database then compute the value of precision & recall .Our framework for CBIR is built in matlab. Test results for some objects are shown in figures given below where only .retrieved images are shown for the query image on the right.





VIII. Conclusion

Content based image Retrieval System is a process to find the similar image in image database when query image is given. In this paper, we use color feature extraction, color feature are extracted by using three technique such as color Correlogram, color moment ,HSV histogram. The main focus of this paper, the KNN algorithm and relative standard derivation .Here we use KNN classifier for the classification of image and Relative standard derivation to measure similarity between two image.and finally compute value of Precision and Recall.

References

- [1] T. Dharani, I. Laurence Aroquiaraj, "A Survey on Content Based Image Retrieval," IEEE-2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), PRIME 2013, 978-1-4673-5845-3/13/2013.
- [2] Archana B. wagnare et al , 'Low Level Feature Extraction for Content based image retrievalL 201231 ' Volume I – NO 2, April 2012.
- [3] International Journal of Future Computer and Communication, Vol. 1, No. 4, December 2012.
- [4] Suchismita Das, ShrutiGarg, G. Sahoo, "Comparison of Content Based Image Retrieval Systems Using Wavelet and Curvelet Transform," The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.4, (August 2012).
- [5] Manish chodhury 'Novel CBIR System based on Ripplet Transform Using interactive technique vol, jan 2012.
- [6] Harikrishna Narasimhan, Purushothaman Ramraj, "Contribution-Based Clustering Algorithm for Content-Based Image Retrieval".2013
- [7] Y. Luo, D. Tao, C. Xu, C. Xu, H. Liu, Y. Wen, "Multiview Vector-Valued Manifold Regularization for Multilabel Image Classification," IEEE Trans. Neural Netw. Learning Syst., vol. 24, no. 5, pp. 709-722, 2013.
- [8]T. Mei, X.-S. Hua, W. Lai, L. Yang, Z. Zha, Y. Liu, Z. Gu, G. Qi, M. Wang, J. Tang, X. Yuan, Z. Lu, and J. Liu, "MSRA-USTC-SJTU at TRECVID 2007: High-level feature extraction and search," in Proc. Nat. Inst. Standards Technol. TREC Video Retrieval Evaluation, 2007, pp. 1–12.
- [9] Y. Liu, T. Mei, X.-S.Hua, J. Tang, X. Wu, and S. Li, "Learning to video search rerank via pseudo-preference feedback," in Proc. IEEE Int. Conf. Multimedia Expo, 2008, pp. 297–300.
- [10]. W. Hsu, L. Kennedy, and S.-F.Chang, "Video search reranking through random walk over document-level context graph," in Proc. Assoc. Comput.Machinery Int. Conf. Multimedia, 2007, pp. 971–980.
- [11]L. Kennedy and S.-F.Chang, "A reranking approach for context-based concept fusion in video indexing and retrieval," in Proc. Int. Conf. Image Video Retrieval, 2007, pp. 333–340.
- [12] R. Yan, A. Hauptmann, and R. Jin, "Multimedia search with pseudorelevance feedback," in Proc. Int. Conf. Image Video Retrieval, 2003, pp. 238–247.