Content-Based Image Retrieval using Feature Extraction and K-Means Clustering

Geethu Varghese  
PG Student  
Department of Computer Science & Engineering  
Federal Institute of Science & Technology

Dr. Arun Kumar M N  
Associate Professor  
Department of Computer Science & Engineering  
Federal Institute of Science & Technology

Abstract

There are many methods to retrieve an image from an amassment of images in the database in order to meet users demand with image content kindred attribute, edge pattern homogeneous attribute, color homogeneous attribute, etc. An image retrieval system offers an efficient way to access or retrieve set of similar images by directly computing the image features from images by directly computing the image features from an image as reported by utilizing different kinds of techniques as well as algorithm. Content based image retrieval (CBIR) is most recently used technique for image retrieval from large image database. The reason behind content based image retrieval is to get perfect and fast result. There are many technique of CBIR utilized for image retrieval. A Block Truncation Coding technique is the famous method used for image retrieval. In the proposed system the advanced technique of BTC is used that is Ordered Dither Block Truncation Coding (OBDTC). ODBTC encoded data stream to construct the image features namely Color Co-occurrence and Bit Pattern features. After the extraction of this feature similarity distance is computed for retrieving a set of similar images. And to make the search more accurate K-means clustering method is used. The most similar images to the query image are selected and these features are being appended together and k-means clustering is applied. This method retrieves more similar images to the query image than the first search. The proposed scheme can be considered as very good in color image retrieval application. The process is implemented in a MATLAB 2014.

Keywords: Content-based image retrieval, bit pattern feature, color co-occurrence, K-means clustering, ordered dither block Truncation coding

I. INTRODUCTION

Recent years have visually perceived a rapid increase of the size of digital image amassments. Every day, both military and civilian equipment engenders giga-bytes of images. Large amount of information is out there. However, we cannot access to or make utilization of the information unless it is organized so as to allow efficient browsing, searching and retrieval. Image retrieval has been a very active research area since 1970’s. From early 90’s content based image retrieval has become a very active research area. Many Images retrieval system, both commercial and research, have been build. Most image retrieval system support one or more of the following options: random browsing, search by sketch, search by text, navigation with customized image categories. The most common approach to search on image collections are: based on textual metadata of image and based on content information of image. The commonest approache utilized is called Content-Based Image Retrieval. The main advantage of the Content Based Image Retrieval is the automatic retrieval process instead of the traditional keyword based approach, which is time consuming. The main application of Content Based Image Retrieval technology are fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medicine, historical research.

Content Based Image Retrieval (CBIR) is also known as Query by Image Content (QBIC) and Content Visual Information Retrieval (CVIR). In CBIR, content based means the searching of image based on actual content of image rather than its metadata. The Content Based Image Retrieval system is employed to extract the features, index those feature to appropriate structure and efficiently provide answers to the user’s question. To provide the satisfactory answer to the user question, CBIR provides some flow of work. Foremost CBIR system takes the RGB image as input, performs feature extraction, performs some similarity computations with the images stored database and retrieves the output image on the idea of similarity computation. Then select the similar images to the query image then the features of the selected similar images are extracted and appended together and k-means is applied which retrieve most similar images to the query image.

Relevance feedback (RF) could be a usually accepted technique to enhance the effectiveness of retrieval systems interactively. Basically, it's composed of 3 steps: (a) An initial search is created by the system for a user-supplied question pattern, returning a small number of images ;(b) The user then indicates that the retrieved pictures relevant;.(c) Finally, the system mechanically reformulates the initial question primarily based upon users relevance judgments. This method is continued till the user is satisfied. RF way facilitates to alleviate the linguistics gap drawback, since it permits the CBIR system to find out image perceptions. RF way sometimes manage small training sample, imbalance in training sample and real time demand. Another vital issue is concerned with the planning and implementation of learning mechanisms.

Clustering is a method that separate a given dataset into same teams based on specific requirements. The similar objects are measured in a cluster whereas dissimilar objects are in different clusters. Clustering plays a vital role in various fields including
image processing, mobile communication, computational biology, medicine and economics. K-Means is a well-known clustering algorithm which is popularly known as Hard C Means algorithm. This algorithm splits the given image into different clusters of pixels within the feature area, each of them defined by its center. First, each pixel within the image is allocated to the closest cluster. Then, the new centroids are computed within the new clusters. These steps are repeated until convergence. Primarily we need to determine the number of clusters K first. Then the centroids will be assumed for these clusters. We could assume random objects as initial centroids or first K objects in sequence could also serve as the initial centroids. Steps followed in the algorithm:

(a) determine the centroid coordinates. 
(b) Determine the distance of each object pixel to the centroids. 
(c) Group the object based on minimum distance with the centroid.

II. RELATED WORKS

There are many methods for content-based image retrieval. In [2] Roland Kwitt, Andreas Uhl proposed a framework of probabilistic texture retrieval in the wavelet domain from the view points of retrieval accuracy and computational performance. Then a novel retrieval method is proposed based on image representation in the complex wavelet domain and several statistical models for the magnitude of the complex transform coefficients. KL-divergences between the proposed statistical models are measured for similarity measurement. The problem with this method is that incorporation of color information is still an open problem. In [3] B Zhang, Y Gao, S Zhao, and Jia Liu, proposes a novel high-order local pattern descriptor, local derivative pattern (LDP), for face recognition. The proposed LDP operator labels the pixels of an image by comparing two derivative directions at two neighboring pixels and concatenating the results as a 32-bit binary sequence. The procedure applies a high-order local feature operator on each pixel to extract discriminative features from its neighborhood. Then model the distribution of high-order local derivative pattern by spatial histogram. Spatial histograms can be extracted from circular regions with different radiiuses. Many similarity measures for histogram matching have been proposed. In this paper, histogram intersection is used to measure the similarity between two histograms. The problem with this method is that the performance of LDP drops when it reaches to the fourth order.

In [4] Xiaoyang Tan and Bill Triggs propose a method for recognizing face under uncontrolled lighting condition. The overall process can be viewed as a pipeline consisting of image normalization, feature extraction, and subspace representation. This paper present a simple and efficient preprocessing chain that eliminates most of the effects of changing illumination while still preserving the essential appearance details that are needed for recognition. This paper focuses mainly on the issue of robustness to lighting variations. The problem with this method is that filter removes both the useful and the incidental information. In [5] J Chen, S Shan, Chu He, Guoying Zhao, M Pinen, propose a a simple, yet very powerful and robust local descriptor, called the Weber Local Descriptor (WLD). WLD consists of two components: differential excitation and orientation. The differential excitation component is a function of the ratio between two terms: One is the relative intensity differences of a current pixel against its neighbors, the other is the intensity of the current pixel. WLD features to compute a histogram by encoding both differential excitations and orientations at certain locations. The problem with this method is that it cannot be used for descriptor for the domain of face recognition and object recognition.

In [6] Chih-Chin Lai, Ying-Chuan Chen proposed a user oriented mechanism for CBIR method based on an an interactive genetic algorithm to infer which images in the databases would be of most interest to the user. Three visual features, color, texture, and edge, of an image are utilized in this approach. This method provides an interactive mechanism to bridge the gap between the visual features and the human perception. The color distributions, the mean value, the standard deviation, and image bitmap are used as color information of an image. The problem with this algorithms have many limitations when dealing with broad content image database. In [7] Subrahmanyam Murala, R. P. Maheshwari, R. Balasubramanian, proposed a novel image indexing and retrieval algorithm using local tetra patterns (LTrPs). This method encodes the relationship between the referenced pixel and its neighbors, based on the directions that are calculated using the first-order derivatives in vertical and horizontal directions. The magnitude of the binary pattern is collected using magnitudes of derivatives. The effectiveness of the proposed approach has been also analyzed by combining it with the Gabor transform (GT). This method does not take diagonal pixels for derivative calculation.

In [8] J J de Mesquita, P C Cortez, and A Ricardo Backes proposed a method for color texture classification using shortest path in graphs. First approach is to model a RGB color texture as a graph consider each color channel as an independent graph. As a second approach, create a single graph that represents the interaction among the three channels, connect the vertices from pixels sharing the same region, independent of the color channel. The shortest paths now represent not only the global information of the graph, but also hold information about the transitions between channels. Shortest paths computed by the Dijkstra’s algorithm. The main problem with this method is that window size must be small. In [9] Z Liu, H Li, W Zhou, R Zhao, and Qi Tian proposed to represent the spatial context of local features into binary codes, and implicitly achieve geometric verification by efficient comparison of the binary codes. In this approach, for each single feature, its surrounding features are clustered into different groups based on their spatial relationships with the center feature. The circle is divided into three fan regions by the blue line originating from the circle center. Those surrounding features that locate in the same fan region belong to the same group. Based on these grouped surrounding features, we generate a binary code to describe the center features spatial context. The problem with this method is that in this method a large number of features are take for retrieval so to reduce the number of features for retrieval purpose is still challenging.
In [10] Tanaya Guha,Rabab K. Proposed Image similarity sparse representation and compression distance . In this method it measuring the similarity between two given images is to quantify how well each image can be represented using the information of the other. A sparse representation-based approach to encode the information content of an image. It use the compactness of the representation of the image as a measure of its compressibility. Two quantities use are the dictionary learnt from the image itself, the dictionary extracted from the other image. Extract suitable features from the available training data. When a new query data is available, similar features are extracted from the query. The obstacle lies in evaluating and approximating the conditional compression.

III. PROPOSED METHODOLOGY

The main objective of this method is to retrieve similar images using feature extraction and k-means clustering. First a RGB color image of size MxN is taken. And this image is divided into multiple non-overlapping image blocks of size mxn, and each image block can be processed independently. Each image block b(i,j) is firstly converted into the inter-band average image 

\[ \overline{b_{k,l}}(i, j) = \frac{1}{3} \left[ b_{k,l}^{red}(i, j) + b_{k,l}^{green}(i, j) + b_{k,l}^{blue}(i, j) \right] \]

\[ k = 1,2,3,...m; \quad l = 1,2,...n, \]

Where (k,l) denotes the pixel coordinate on image block (i, j). The inter-band average computation is applied to all image blocks BTC approach performs the thresholding operation with a single threshold value obtained from the mean value of the pixels in an image block. A pixel of a smaller value compared to the threshold is turned to 0 (black pixel); otherwise it turns to 1(white pixel) to construct the bitmap image representation. The ODBTC employs the void-and-cluster dither array of the same size as an image block to generate the bitmap image. Let \( D_{(k,l)} \) denotes the dither array coefficient at position \( (k,l) \), where \( k = 1,2,...m \) and \( l = 1,2,...n \). Dither array can be easily computed as

\[ D_{d}(k,l) = d \frac{D_{(k,l)} - D_{min}}{D_{max} - D_{min}} \]

Where \( D_{min} \) and \( D_{max} \) denotes the minimum and maximum coefficient values in the dither array respectively. The variable \( d \) denotes the dither array index and is defined by

\[ d = \overline{b_{max}}(i, j) - \overline{b_{min}}(i, j) \]
The minimum $b_{\text{min}} (i, j)$ and maximum value $b_{\text{max}} (i, j)$ of inter band average image on image block $(i, j)$ can be computed as

$$b_{\text{min}} (i, j) = \min_{v} b_{v,i}(i, j),$$

$$b_{\text{max}} (i, j) = \max_{v} b_{v,i}(i, j).$$
Minimum and maximum quantizer from all image blocks is given as

\[
X_{\text{min}} = \{ x_{\text{min}}(i, j); i = 1, 2, \ldots, M; j = 1, 2, \ldots, N \},
\]

\[
X_{\text{max}} = \{ x_{\text{max}}(i, j); i = 1, 2, \ldots, M; j = 1, 2, \ldots, N \},
\]

where \( x_{\text{min}}(i, j) \) and \( x_{\text{max}}(i, j) \) denotes the minimum and maximum value respectively, over red, green, and blue channels on the corresponding image block \((i, j)\). The two values can be formally formulated as

\[
x_{\text{min}}(i, j) = \left[ \min_{v \in k, l} b_{v, l}^{\text{red}}(i, j), \min_{v \in k, l} b_{v, l}^{\text{green}}(i, j), \min_{v \in k, l} b_{v, l}^{\text{blue}}(i, j) \right]
\]

\[
x_{\text{max}}(i, j) = \left[ \max_{v \in k, l} b_{v, l}^{\text{red}}(i, j), \max_{v \in k, l} b_{v, l}^{\text{green}}(i, j), \max_{v \in k, l} b_{v, l}^{\text{blue}}(i, j) \right]
\]

At the end of the ODBTC encoding, the bitmap image, \( b_m \), the minimum quantizer, \( X_{\text{min}} \), and maximum quantizer, \( X_{\text{max}} \), are obtained. The decoder simply replaces the elements of value 0 in the bitmap by the minimum quantizer, and elements of value 1 in the bitmap by the maximum quantizer.

Fig. 2: Block diagram of ODBTC encoding.

The ODBTC employed in the proposed method decomposes an image into a bitmap image and two color quantizers which are subsequently exploited for deriving the image feature descriptor. Two image features are introduced in the proposed method to characterize the image contents, i.e., Color Co-occurrence Feature (CCF) and Bit Pattern Feature (BPF). The CCF is derived from the two color quantizers, and the BPF is from the bitmap image.

**A. Color Co-occurrence Feature (CCF):**

Color Co-occurrence Feature (CCF) can be derived from the color co-occurrence matrix. CCF is computed from the two ODBTC color quantizers. The minimum and maximum color quantizers are firstly indexed using a specific color codebook. The color co-occurrence matrix is subsequently constructed from these indexed values. The CCF is derived from the color co-occurrence matrix at the end of computation. The color indexing process on RGB space can be defined as mapping a RGB pixel of three tuples into a finite subset (single tuple) of codebook index. The color indexing process of the ODBTC minimum quantizer can be considered as the closest matching between the minimum quantizer value of each image block \( x_{\text{min}}(i, j) \) and the codebook \( C_{\text{min}} \) which meets the following condition.

\[
\bar{i}_{\text{min}}(i, j) = \arg \min_{q = 1, 2, \ldots, N_c} \| x_{\text{min}}(i, j) - C_{\text{min}}^q \|_2
\]
For all \( i = 1,2,...,\frac{M}{m} \) and \( j = 1,2,...,\frac{N}{n} \). Similarly for maximum quantizer can be calculated. The color co-occurrence matrix (i.e., Color Co-occurrence Features (CCF)) for a given image can be directly computed as

\[
CCF \left( \mathbf{t}_1, \mathbf{t}_2 \right) = \Pr \left[ \frac{i}{m_{\max}} \left( i, j \right) = \frac{i}{m_{\max}} \left( i, j \right) = \frac{j}{n_{\max}} \mid i = 1,2,...,\frac{M}{m} ; j = 1,2,...,\frac{N}{n} \right]
\]

For \( \mathbf{t}_1, \mathbf{t}_2 = 1,2,...,N_c \). The color co-occurrence matrix is a sparse matrix, in which the zeros dominate its entries. To reduce the feature dimensionality of the CCF and to speed up the image retrieval process, the color co-occurrence matrix can be binned along its columns or rows to form a 1D image feature descriptor.

For all \( i = 1,2,...,\frac{M}{m} \) and \( j = 1,2,...,\frac{N}{n} \). The symbol \( \delta_H \) denotes the hamming distance between the two binary patterns.

\[
\bar{b}(i, j) = \arg \min_{q=1,2,...,N_c} \delta_H (b_m(i, j), Q_q)
\]

For all \( i = 1,2,...,\frac{M}{m} \) and \( j = 1,2,...,\frac{N}{n} \). The binary vector quantization produces a representative bit pattern codebook from a set of training bitmap images obtained from the ODBTC encoding process. These bit pattern codebooks are generated using binary vector quantization with soft centroids, and many bitmap images are involved in the training stage. The bitmap of each block \( b_m(i, j) \) is simply indexed based on the similarity measurement between this bitmap and the codeword \( Q_q \) which meets the following criterion.

The similarity measure of the feature:

The similarity between two images can be measured using the relative distance measure. The similarity distance plays an important role for retrieving a set of similar images. The query image is firstly encoded with the ODBTC, yielding the corresponding CCF and BPF. The two features are later compared with the features of target images in the database. A set of similar images to the query image is returned and ordered based on their similarity distance score, i.e., the lowest score indicates the most similar image to the query image.
\[ \delta(\text{query}, \ t \ \text{arg} \ et) = \alpha_1 \sum_{i=1}^{N_\text{c}} \left| \frac{CCF^{\text{query}}(t) - CCF^{\text{arg} \ et}(t)}{CCF^{\text{query}}(t) + CCF^{\text{arg} \ et}(t) + \varepsilon} \right| + \alpha_2 \sum_{i=1}^{N_\text{b}} \left| \frac{BPF^{\text{query}}(t) - BPF^{\text{arg} \ et}(t)}{BPF^{\text{query}}(t) + BPF^{\text{arg} \ et}(t) + \varepsilon} \right| \]

Where \( \alpha_1 \) and \( \alpha_2 \) denote the similarity weighting constants, representing the percentage contributions of the CCF and BPF in the proposed image retrieval system. A small number \( \varepsilon \) is placed at the denominator to avoid the mathematical division error. After the similarity computation a set of images are retrieved from the database related to the query images. These images may also contain some irrelevant images also, so in order to make the search more accurate K-Means Clustering is used. We provide a feedback form in which if we want to search some more image which are similar to the query image then press 1 or 0 to exit. Then select the images more similar to the query image from the retrieved images. The CCF and BPF features of all selected images are taken and appended together, then k-means clustering is applied and centroid is calculated and again similarity computation is performed with rest images in the database and this search provide some more accurate result and this search can be repeated three times to provide more accurate image retrieval.

IV. EXPERIMENTAL RESULTS

The method was implemented in a MATLAB 2014 prototype and tested with image collected from the database of 1000 images. All the images are stored in JPEG format with size 384 x 256 or 256 x 384. It was applied to RGB images. Both the image were retrieved and test were performed on a desktop PC with the following characteristics: Intel Core i3 CPU, 3.4 GHz, 4 GB RAM.

The database was color images of different varieties. There are different categories which include 100 people, 100 rose flowers, 100 beeches, 100 old building, 100 horses, 100 dinosaur, 100 elephant, 100 mountain, 100 food, 100 buses.

A color image of size M x N is taken as a query image. This image is divided into image blocks of size m x n. ODBTC encoding is performed. The image blocks are converted into inter-band average image. It is applied to each block independently. BTC approach performs thresholding operation with a single threshold value obtained from the mean value of the pixel in an image block. A pixel of a smaller value compared to the threshold is turned to 0 otherwise it is turned to 1 to construct the bitmap image. Also maximum and minimum quantizer is computed. ODBTC transmit image bitmap, and also two extreme color quantizer (minimum and maximum quantizer) to the decoder module over transmission channel. Figure 5 below shows minimum and maximum quantizer of the query image.

Fig. 5: Minimum and Maximum Quantizer of the query image.
Figure 6 shows the bitmap image of the query image obtained in ODBTC encoding. After the maximum and minimum quantizer and bitmap image is computed then value 0 in the bitmap is replaced with value in minimum quantizer and value 1 in bitmap is replaced by value in the maximum quantizer and reconstructed image is formed. Figure 7 below shows the reconstructed image of the query image.

The receiver decodes this data stream to reconstruct the image. The decoder replaces the element of value 0 in the bitmap with value of minimum quantizer and element of value 1 in the bitmap by value of maximum quantizer. Two image features are extracted to characterize the image contents i.e color co-occurrence feature and bit pattern feature. The CCF uses the two color quantizer and BPF uses bitmap image. Then similarity between two images i.e query image and set of images in the database as target image are measured using relative distance measure. And set of images are retrieved, this also contain irrelevant images. So select the relevant images from the retrieved images. Then the CCF and BPF features of all selected images are extracted and appended together. K-Means Clustering is applied. Then again similarity is computed and it retrieve more similar image to the query image and it can be repeated three times to yield better retrieval. Figure 8 shows a query image of people is tested and similar images are been retrieved.
Content-Based Image Retrieval using Feature Extraction and K-Means Clustering

Fig. 8: Retrieving images similar to query image

Fig. 9: Providing relevant feedback

Fig. 10: Selecting the image more similar to the query image.

Fig. 11: More relevant images but providing relevant feedback

Fig. 12: More similar images to query image

Fig. 13: Retrieving images similar to images of elephant.
Content-Based Image Retrieval using Feature Extraction and K-Means Clustering

Content-based image retrieval using halftonic block truncation coding is compared with k-means clustering method. The k-means clustering shows good retrieval of images than the halftonic block truncation coding. Here each image of people, beach, elephant, horse,ower, dinosaur etc are taken as query image and tested using halftonic based block truncation coding and k-means clustering. And the accuracy by which correct images are retrieved using halftonic based block truncation coding is
calculated and then more relevant images are selected and its features are extracted and k-means is applied and the accuracy of retrieving correct image using k-means is also calculated.

### Table 1

<table>
<thead>
<tr>
<th>Category</th>
<th>Using Halftone BTC</th>
<th>Using K-Means Clustering Iteration 1</th>
<th>Using K-Means Clustering Iteration 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>People</td>
<td>79.1</td>
<td>87.5</td>
<td>91.6</td>
</tr>
<tr>
<td>Beaches</td>
<td>50</td>
<td>66.6</td>
<td>79.1</td>
</tr>
<tr>
<td>Building</td>
<td>79.1</td>
<td>83.3</td>
<td>87.5</td>
</tr>
<tr>
<td>Bus</td>
<td>66.6</td>
<td>91.6</td>
<td>100</td>
</tr>
<tr>
<td>Dinosaur</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Elephant</td>
<td>66.6</td>
<td>91.6</td>
<td>95.8</td>
</tr>
<tr>
<td>Flower</td>
<td>79.1</td>
<td>91.6</td>
<td>95.8</td>
</tr>
<tr>
<td>Horse</td>
<td>54.1</td>
<td>75.0</td>
<td>91.6</td>
</tr>
<tr>
<td>Mountain</td>
<td>66.6</td>
<td>83.3</td>
<td>91.6</td>
</tr>
<tr>
<td>Food</td>
<td>83.3</td>
<td>91.6</td>
<td>95.8</td>
</tr>
</tbody>
</table>

V. CONCLUSION

This work presented a novel method for image retrieval using feature extraction and k-mean clustering, which is easy to implement while shows promising performance. First ODBTC encoding is done then maximum and minimum quantizer and bitmap image is obtained. Then color co-occurrence and bit pattern features are extracted from the query image. The similarity between two images (i.e., a query image and the set of images in the database as target image) can be measured using the relative distance measure. The similarity distance plays an important role for retrieving a set of similar images. Then select the images that are more relevant to query image. Then features of the selected images are appended together and k-mean is applied, which provide more similar images to the query image.

The advantages of the proposed method are that it can effectively retrieve more similar images to the query image.

REFERENCES


