Survey on Content Based Image Retrieval

Anuradha Shitole[1] and Uma Godase[2]

[1]Department of Information Technology, Pune University, Pune
[2]Department of Information Technology, Pune University, Pune

Abstract

Invention of digital technology has lead to increase in the number of images that can be stored in digital format. So searching and retrieving images in large image databases has become more challenging. From the last few years, Content Based Image Retrieval (CBIR) gained increasing attention from researcher. CBIR is a system which uses visual features of image to search user required image from large image database and user's requests in the form of a query image. Important features of images are colour, texture and shape which give detailed information about the image. CBIR techniques using different feature extraction techniques are discussed in this paper.

Keywords

CBIR; colour feature; texture feature shape feature.

1.Introduction

For efficient services in all fields such as government, academics, hospitals, crime prevention, engineering, architecture, journalism, fashion and graphic design use images. Due to the popularity of these types of digital image database becomes huge database, and to search and retrieve required image from the huge database becomes difficult and time consuming. To solve these problems traditionally text-based retrieval is used. In a computer system for browsing, searching and retrieving images from the huge database of digital images retrieval system is used. To search images, a user provides query terms of keyword and the system will return images similar to the query.

In text based image retrieval keywords, label, tag or any other information is associated with the image and using this metadata image retrieval is performed. In this method query is entered in terms of text. But, there are some drawbacks of this type of an image retrieval system. Annotation of each image in the database requires domain experts who add label or other information to the image. Use of different keywords for annotation of each image in very large database is a very time consuming process. It is also necessary to use unique keyword for annotation of each image, so this is a very complex task. Text descriptions are sometimes incomplete because they cannot depict complicated image features very well. Examples are texture images that cannot be described in the text. A language mismatch can occur when the user and the domain expert uses a different language.

Content-based image retrieval (CBIR) systems were introduced to address the problems associated with text-based image retrieval. To search and retrieve digital images CBIR uses content of the images. Content-based means that the search analyses the contents of the image not the metadata such as keywords or tags associated with the image. Here the term content means colours, shapes, textures or any other information that derived from the image. In CBIR systems input provide in terms of an image and based on image attribute matching the most similar images from database are retrieved.

2. Related Work

Feature extraction is the main task in the CBIR systems to retrieve the similar images from database. Similarity measurement decides how close a vector is to another vector. Selection of similarity metrics has a direct impact on the performance of content-based image retrieval. The type of feature vectors selected determines the type of measurement that used to compare their similarity. Euclidean distance is the most common metric used to measure the distance between two points in multi-dimensional space. To represent an image, in feature extraction features such as colour, texture or shape from image are extracted and creates a feature vector for each image.

2.1 Colour

Colour is one of the most important features visually recognized by humans in images. Colour features are the most widely used in CBIR systems [10]. To extract the colour features from the content of an image, a proper colour space and an effective colour feature extraction method are required.

2.1.1 Colour space:

For a certain combination of a colour model plus a mapping function colour space is a more specific term. The term colour space is used to identify colour models. Identifying a colour space automatically identifies the associated colour model. The purpose of a colour space is to facilitate the specification of colours. For different purposes many colour spaces, such as RGB, HSV, CMYK and CIE L*a*b have been developed.

RGB: Computer can only identify the RGB colour component of an image. RGB has three colour components: Red (R), Green (G) and Blue (B). RGB uses additive colour mixing. It describes what kind of light needs to be emitted to produce a given colour. For red, green and blue component RGB stores individual values. For representing colour images the RGB colour space is most widely used systems, but it is not suitable for CBIR because it is a perceptually non-uniform and device-dependent system.

CMYK: CMYK uses subtractive colour mixing and it describes what kind of inks needs to apply so the light reflected from the substrate and through the inks produces a given colour. CIELAB: The CIELAB (or L*a*b* or Lab) produce a colour space that is more perceptually linear than other colour spaces. Linear means that a change of the same amount in a colour value should produce a change of about the same visual importance.

HSV: To describe a specific colour the HSV colour space uses Hue (H), Saturation (S), and brightness values (V). In interactive colour selection and manipulation this colour system is very useful. It is easier to think about a colour in terms of hue and saturation than in terms of additive or subtractive colour components. HSV is a transformation of an RGB colour space. Its components are relative to the RGB colour space from which it derives. For human perceive HSV space is more suitable.

After selecting a colour space effective colour descriptor required. To represent the colour of the global or local areas colour descriptor needs. From various representation schemes several colour descriptors have been developed such as colour histograms, colour moments, colour edge and colour texture. Out of these colour histogram and colour moment are most commonly used methods for the colour feature extraction.

2.1.2 Colour Histogram

Colour histogram is a method which is used to represent colour feature of an image. A colour histogram is a type of bar graph. The height of each bar represents the amount of a particular colour used in the image. The bars in a colour histogram are known as bins and they represent the x-axis. On the number of colours in an image the number of bins are depend. Each bin contains the number of pixels and it denotes y-axis which shows how many particular colour pixels an image contains.

The colour histogram easily characterizes the global and local distribution of colours in an image. In colour histograms quantization process is used [8]. Quantizing reduces the space required to store the histogram information. Quantization reduces the information of the content of images. Colour histogram has two types, global colour histogram (GCH) and local colour histogram (LCH). A GCH takes colour histogram of the whole image and represents information of an image; it does not consider colour distribution of regions in the image. In the LCH an image divide into fixed blocks and takes the colour histogram of each of the blocks. LCH contains more information about an image.

2.1.3 Colour moment:

In the method of colour moment, colour information focus on the low-level colour moment of the image. For each colour component they calculate moment as first order, second-order and third-order moment. These colour moment as first-order (mean) and second (variance) and third-order (gradient), is very effective in colour distribution of images. Each pixel has three colour components and each component has three low-order moments so there are nine colour characteristics components. In the colour moment is only the initial colour characteristics extraction of image is done and the effect of extraction is very rough.

2.2 Texture

Texture is the natural property of all surfaces, which describes visual patterns each having the properties of homogeneity. A texture is characterized by intensity properties (tones) and spatial relationships (structural). This is a feature that describes the distinctive physical composition of a surface. The different texture properties as perceived by the human eye are, for example regularity, directionality, smoothness, and coarseness. Image textures have useful applications in image processing. They include: recognition of image regions using texture properties known as texture classification [3], recognition of texture boundaries using texture properties known as texture segmentation, texture synthesis, and generation of texture images from known texture models. For all types of textures there is no any single method which gives best result [9]. The commonly used methods for texture feature description are statistical, model-based and transform-based methods. The texture feature description categories are explained below.

2.2.1 Statistical Methods

Statistical methods analyse the spatial distribution of grey values by computing local features at each point in the image and derive set of statistics from the distribution of the local features. It includes the methods such as co-occurrence matrix; statistical moments, autocorrelation function, and grey level run lengths. The most commonly used statistical method is the Gray-level Co-occurrence Matrix (GLCM). It is a two-dimensional matrix of pairs of pixels which are separated by a distance in a given particular direction. It is popular in texture description and is based on the repeated occurrence of some grey level configuration in the texture. Gray-level co-occurrence matrix method of representing texture features has found useful applications in recognizing fabric defects [8], and in rock texture classification and retrieval.

2.2.2 Model Based Approaches

Model-based texture methods capture the process that generated the texture. In the model-based features, some part of the image model is assumed and estimation algorithm is used to set the parameters of the model. There are three major model based methods: Markov random fields, fractals, and the multi-resolution autoregressive features.

2.2.3 Transform Domain Features

There are several texture classifications using transform domain features such as discrete wavelet transforms, Fourier transform and Gabor wavelets. Transform methods analyse the frequency content of the image to determine texture features Wavelet analysis breaks up a signal into shifted and scaled versions of the original wavelet and which refers to decomposition of a signal into a family of basis functions obtained through translation and dilation of a special function. Moments of wavelet coefficients in various frequency bands are effective for representing texture.

2.3 Shape

The shape of an object is the characteristic surface configuration as represented by the outline. Shape recognition is one of the modes through which human perception of the environment is executed. It is important in CBIR because it corresponds to region of interests in images. Boundary-based and region-based are two types for the shape feature representations. In region based techniques all the pixels within a shape are taken into consideration to obtain the shape representation. Common region based methods use moment descriptors to describe shapes. Instead of providing information just at a single point moments combine information across an entire object. They capture some of the global properties which are missing from contour-based representations. Contour based shape representation is more popular than the region based shape representation. Simple contour-based shape descriptors include area, perimeter, compactness, eccentricity and orientation. Complex boundary-based descriptors include Fourier descriptors, grid descriptors, and chain codes.

2.3.1 Zernike moments (ZM)

The complex Zernike moments are derived from orthogonal Zernike polynomials so that Zernike Moment is known as orthogonal moments. Zernike polynomials are the complete set of complex valued functions. Zernike moments are invariant to the rotation and they are robust to noise. They have minimum information redundancy because they based on orthogonal. But, there are some problems with Zernike moments such as the continuous integrals approximated by discrete summations. This approximation responsible for numerical errors in the computed moments and affects the rotational invariance.

2.3.2 Eccentricity

The measure of aspect ratio is known as Eccentricity. It is simply ratio of the length of major axis to the length of minor axis. There are two methods to calculate the eccentricity principal axes method and minimum bounding rectangle method. Principal axes of a given shape defined as the two segments of lines that cross each other orthogonally in the centres of the shape and represent the directions with zero cross-correlation. Minimum bounding rectangle is the smallest rectangle which contains every point in the shape so that it known as minimum bounding box. For arbitrary shape eccentricity is calculated as the ratio of length and width of minimum bounding rectangle of the shape. Due to this a contour is look as an instance from a statistical distribution.

3. Techniques of CBIR

In the CBIR features are most important content for indexing and retrieval of the image. Colour, Shape and Texture are most important features of image. The feature extracted are then form a vector and this vector will be used for indexing of particular image.

3.1 Colour & Texture Features for CBIR:

Biragle and Doye focus on colour and texture features for content based image retrieval [7]. For improved retrieval performance they use the decomposition scheme based on colour planes in combination with histogram. For each plane three level decomposition performed. Each image divided into non-overlapping sub images and analysed using standard wavelet and histogram. Standard Wavelets used for texture feature extraction and colour histogram for colour feature extraction.

To create the feature vector first level energy and computed standard deviation of each sub-band is used. Then to find similarity between images Euclidean distance metric is used. The average percent retrieval efficiency using this method is up to 75%. The main advantage of three-colour plane wavelet decomposition is that it yields a large number of sub bands and which improves the retrieval accuracy.

3.2 CBIR using colour, texture & shape feature:

Hiremath and Pujari presents a framework for combining information of three features such as colour, texture and shape information to achieve high retrieval efficiency [6]. The purpose of this method is to capture local colour and texture descriptors in a segmentation framework of grids and shape describable in terms of invariant moments computed on the edge image. In this method, an image is partitioned into non overlapping tiles. These tiles are local colour and texture descriptors for the image. This grid framework is extended across resolutions to capture different image details within the same size tiles. A matching procedure based on an adjacency matrix of a bipartite graph. For the query image and the target image tiles bipartite graph is built. The labelled edges of the bipartite graph indicate the distances between tiles.

Image similarity computation based on most similar highest priority (MSHP) principle. In this method first and second order statistical moments of Gabor filter used to extract the texture feature, colour moment used for the colour feature extraction and Gradient vector flow fields(GVF) are used to extract the shape of the object. In per resolution 52 features are computed for every image tile. This method allows to tile from query image is matched to any tile in the target image. The main advantage of this method is that it creates robust feature set for image retrieval.

3.3. CBIR Using Feature Combination & Relevance Feedback (RF):

Zhao and Tang, Perform CBIR Using Optimal Feature Combination and RF [4]. In this approach the comparison of the retrieval performance using different combinations of features is done. The combination is done in two levels. One is the combination of colour and texture features and the other is the combination of two textures extracted by two different texture feature extraction methods. By comparing different combinations of visual features they choose optimal combination of Gabor and wavelet transform texture features with the colour moment colour feature which has a better retrieval performance. Retrieval is based on the similarity of visual features combination of query image and for the similarity measurement Euclidian distance method used.

From the retrieval result user selects the relevant images and the irrelevant images. This feedback is provided as training set to the Support Vector Machine (SVM) classifier to train the SVM [5]. This approach focus on to minimize the gap between low-level features representation of images and the user's high-level semantic concepts, for that it uses SVM based on RF(Relevance Feedback) to learn user's query concepts [6]. SVM and feature similarity based relevance feedback using best feature combination improves the retrieval precision. As number of feedback increases the retrieval accuracy improves correspondingly. But in the relevance feedback for the same output different users have different views about the similarity. So it becomes complex process.

3.4. Semantic Image Retrieval by combining three Features:

To overcome the disadvantage of relevance feedback Singh, Dubey, Dixit and Gupta experimentally evaluated two phase method for the extraction of semantic information [2]. In the first phase feature database of images is created. The feature database contains information about the colour, shape and texture of the image. In the feature extraction colour histogram for the colour feature, for the texture feature coarseness, contrast, energy and directionality used and for the shape feature Zernike moments and edge used. Feature database is created and stored according to the highest three values of the hue histogram to reduce the processing time and reduce dataset.

In the second phase, images which are relevant to the query image are retrieved. Then on reducing dataset feature matching is done and extracts similar images to the query image considering colour, texture and shape feature individually. For each feature set of images are obtained. Finally combine all the features which obtain in a set of images and retrieve images which are semantically most similar to the query image.

This method retrieves images which are semantically similar with the query image and it improves the precision and recall of the image retrieval system. It retrieves all the similar images based on the each feature individually so there is a no chance of missing relevant images, this is the main advantage of this method. This approach is time consuming cause in this approach similarity matching applied two times first using each feature individually and then jointly and image retrieval process to perform two times first image retrieve using each feature individually and then from that images most similar images are retrieved.

3.5. CBIR using Multiple SVM's Ensemble:

Yildizer, Balci, Hassan and Alhajj proposed efficient content-based image retrieval using Multiple Support Vector Machines Ensemble [1]. The main purpose of this method is to find a good similarity measure between images. Similar images belong to predefined classes with close probabilities and to find the class probabilities this approach uses Support Vector Regression (SVR) model. The main steps in this approach are first, Feature extraction: in this step before feature extraction resize the images and transform images from RGB data space to another space. After that Daubechies wavelet transformation applied several times to extract features from images. Second, Feature reduction: SVR ensemble constructs for feature reduction. It creates new low dimensional feature vectors in reduced size due to which distance calculation reduced. Third, construct a new SVR model to find the class probability estimates of the test Images. Fourth, Query phase: in this phase Euclidian distance measurement uses as distance measure and return similar images having lowest distance.The main purpose of this CBIR technique is that to handle large image databases and by reducing the dimensionality of the feature vectors reduce the cost of distance measurements due to which the accuracy of the classifier increase. Average classification accuracy of this approach is 62% and average precision is 64%.

4. Discussions

In the table Comparison of different CBIR techniques is given. It shows which features used in to the particular CBIR technique to extract the image information.

Table 1: Comparison of methods

Sr. No.	Paper name	Feature extraction method	Performance evaluation parameter	Advantages	Disadvantages
1.	Colour & Texture Features for CBIR	Colour histogram, Standard wavelet	Retrieval accuracy	Improves the retrieval accuracy	Insufficient feature set
2.	CBIR using colour, texture & shape feature	Colour moment, Gabor filter, GVF	Retrieval efficiency	Create robust feature set	High semantic gap
3.	CBIR Using Feature Combination & RF	Colour moment, Gabor, wavelet, co-occurrence matrix	Precision	Minimize the semantic gap using RF with SVM	It is time consuming to label negative examples
4.	Semantic Image Retrieval by Combining three Features	Colour histogram, Tamura, Zernike moment & edge	Precision and recall	1.Reduce dataset 2. All similar images of related features are retrieved.	Similarity measurement and image retrieval perform two times so it increases calculations
5.	CBIR using Multiple SVM's Ensemble	Daubechies wavelet	Precision, classification accuracy	1.Narrow down search space 2.Handle large image database	Feature sets not sufficient

5. Conclusion

In this paper we surveyed the field of content-based image retrieval, by presenting an overview of the most important aspects of images. We compare various methods used in the CBIR techniques using different feature extraction techniques. The Semantic Image Retrieval method reduce dataset but in this method image retrieval perform two times so it becomes complex. CBIR uses multiple SVM's ensemble method narrow down search space and handle large image database using SVM. In CBIR system SVM is also used for the classification images. It proves CBIR using multiple SVM's ensemble method has better performance than other methods. In future, we expect more efficient method for image retrieval which has high retrieval accuracy and precision.

REFERENCE

[1] Ela Yildizer, Ali Metin Balci, Mohammad Hassan, Reda Alhajj, “Efficient content-based image retrieval using Multiple Support Vector Machines Ensemble”, Expert Systems with Applications 39, 2385–2396, 2012

[2] Nishant Singh, Shiv Ram Dubey, Pushkar Dixit, Jay Prakash Gupta, “Semantic Image Retrieval by Combining Colour, Texture and Shape Features”, International Conference on Computing Sciences,2012.

[3] Jipsa Kurian, V.Karunakaran, “A Survey on Image Classification Methods”, International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 1, Issue 4, October 2012.

[4] Lijun Zhao, Jiakui Tang, “Content-Based Image Retrieval Using Optimal Feature Combination and Relevance Feedback”, International Conference on Computer Application and System Modeling,volume-4, v4-436 – v4-442, 2010.

[5] Francesca Bovolo, Lorenzo Bruzzone and Lorenzo Carlin, “A Novel Technique for Subpixel Image Classification Based on Support Vector Machine”, IEEE transactions on image processing, vol. 19, no. 11, november 2010.

[6] Bingxin Xu, Qian Yin, Guangjun Lv, “Using SVM to Organize the Image Database”, International Conference on Computational Intelligence and Security, 2009.

[7] P. S. Hiremath , Jagadeesh Pujari, “Content Based Image Retrieval using Colour, Texture and Shape features”, 15th International Conference on Advanced Computing and Communications,2007.

[8] Lenina Birgale, Manesh Kokare and Dharmpal Doye, “Colour and Texture Features for Content Based Image Retrieval”, Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation, 2006.

[9] Howarth, P., & Rüger, S, “Evaluation of texture features for content-based image retrieval”, In International conference on image and video retrieval (pp. 326–334), 2004.

[10]James Z. Wang, Jia Li and Gio Wiederhold, “SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries”, IEEE transactions on pattern analysis and machine intelligence, vol. 23, NO. 9, september 2001.