A Review on Feature Extraction Techniques for CBIR System

Kavita Chauhan¹, Shanu Sharma¹
¹Department of Computer Science & Engineering, ASET, Amity University, Uttar Pradesh, Noida, India

ABSTRACT: Ongoing expansion of digital images requires improved and efficient methods for sorting, browsing and searching operations through ever-growing image databases. Content Based Image Retrieval (CBIR) systems are search engines for image databases, which perform indexing on images according to their content and features. This paper presents the systematic review of various existing CBIR systems and their feature extraction techniques. Further the performance analysis and limitations of these systems have been discussed.


I. INTRODUCTION

The advancement in computer technologies produces huge volume of multimedia data, specifically on image data. The greatest challenge of the World Wide Web is that the more information available about a given topic, the more difficult it is to locate accurate and relevant information. Generally users know that which information they need, but are unsure where to find it. Search engines can facilitate the ability of the users to locate such relevant information.

Content Based Image Retrieval is a technique which uses visual content to search and compares images from large scale image databases according to the interest of the users [1]. In this process firstly, the user submits a query image or a series of images and the system is required to retrieve images from the database as similar as possible. It also includes another task which is a support for browsing through large image databases, where the images are supported to be grouped or organised in accordance with similar properties [2]. During the past few years, CBIR has gained much devotion for its potential application in multimedia management. The term ‘content’ in this context might refer to colors, shapes, texture, or any other information that can be derived from the image itself. Basically, there are two ways of image retrieval in CBIR- Query by tag and Query by example. In the former method, the query is submitted in the form of tag like square is used for searching square shape in image, and in the later method the query is given with the image example such that the results resemble the given query [2]. CBIR is also known as Query by Image Content (QBIC) and Content-Based Visual Information Retrieval (CBIR).

This paper is divided into four sections. Section I contains the introduction of the CBIR system and their challenges. Section II discusses the result of the various feature extraction techniques in the form of a table defining the various attributes and how they differ from each other. Section III concludes this paper followed by future scope.

This paper is prepared exclusively for International Conference on Information Engineering, Management and Security 2015 [ICIEMS] which is published by ASDF International, Registered in London, United Kingdom. Permission to make digital or hard copies of part or all 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. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

2015 © Reserved by ASDF.international

II. RELATED WORK

A thorough and systematic literature has been done on various CBIR systems. The origin of major studies vary from listed repositories (IEEE, ACM Library, IJCA, Science Direct etc.) to common purpose search engines like Google. Various research papers with the search string ‘Feature Extraction Techniques and CBIR System’ has been searched and finally 65 papers have been downloaded out of which 17 papers were considered of most relevant to the objective. Review for the selected paper is further presented: Ligade et al. [3] proposed a review on techniques of Neural Network, Interactive Genetic Algorithm and Relevance Feedback where the characteristics of the above 3 techniques has been described along with their current achievement, uses and advantages in image retrieval. The experimental evaluation is done by them using the convergence ratio, precision and recall parameters. Walia et al. [4] proposed a new similarity measure that improves the efficiency of the CBIR system by using the dominant color descriptor. Their work compares two images and consider their no. of dominant colors and their distance and thus improves the performance using the colors and also the results were verified on two different image databases. Sarangi et al. [5] proposed an automatic contrast enhancement technique using differential evolution for gray-scale images. The technique attempts to demonstrate the methods adaptability and effectiveness for searching global optima solutions to enhance the contrast and detail in gray scale images. Agarwal et al. [6] proposed a novel feature descriptor for CBIR system by integrating Cooccurence of Haar like Wavelet Filter (CHLWF) with Color Histogram. It extracts the image properties from different visual perspectives to give the image representation almost similar to the human interpretation and hence improves the effectiveness of retrieval system. Jeyabharathi et al. [7] analyzed the performance of the most useful feature extraction techniques which includes Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Independent Component Analysis (ICA) and also the performance of the most renowned classification techniques, i.e., Support Vector Machine (SVM) and Nearest Neighbour (NN). The performance metric used by them were Recognition Rate and F-Score. Based on the performance evaluation models, they concluded that PCA with SVM provide more recognition accuracy than others. Ezekiel et al. [8] proposed a CBIR technique based on multi-scale feature extraction scheme. They designed a Pulse Coupled Neural Network (PCNN) based fusion of a fast wavelet transformation and Contourlet transformation coefficients applied on Rescaled Range(R/S) analysis techniques. The method highlights the edges, segment edges and finds control points to answer the image retrieval query. Zhang et al. [9] introduced a method of color principal feature extraction called ColorPCA which works in color image space extracting the principal features directly from the color images. It considers only one parameter known as reduced dimension to estimate the projection axes. Syam et al. [10] proposed a CBIR system based on GA for Medical image retrieval using the feature extraction of color, texture and shape. They used Squared Euclidean Distance (SED) for the computation of similarity measure for efficient retrieval of images. Their work assures the benefit of the shape feature in addition to the other features. Ligade et al. [11] proposed an image extractor method on multi-feature similarity synthesis using the GA for efficient image retrieval in CBIR systems. The method extends the GA algorithm by using methods of relevance feedback for magnified retrieval performance, also the method uses both the implicit and explicit feedback technique. Ho et al. [12] proposed a novel system architecture for CBIR system in which the well-known techniques are combined like content-based image, color analysis and data mining techniques for better performance and efficiency. It combines the segmentation and grid module, feature extraction module, K-means clustering and neighbourhood module to build the CBIR system. Chadha et al. [13] proposed an improved technique of image retrieval by incorporating the Query modification through image cropping. This feature identifies the user’s interest region in a particular image and thus resulted into more precise and personalized search results. This technique results into 28% improvement in accuracy. Rashedi et al. [14] proposed a method for the CBIR system improving its precision by Feature selection using the Binary Gravitational Search Algorithm. It selects the most relevant features from the query image thus leading to more accurate results by reducing the semantic gap. The results are examined in the Corel database. Also they compared the work of GA and BPSO and found BGSA to be the best among them. Pighetti et al. [15] proposed a new architecture combining the multi-objective interactive genetic algorithm and Support Vector Machine. The multi-objective IGA is used for its capability to converge towards global optima and SVM for its capabilities to learn user evaluations required by IGA. Madugunki et al. [16] described the detail classification of the CBIR system and also about their efficiency. The work compares the CBIR and TBR technique and discusses about the effect of different matching techniques on retrieval process. Euclidean Distance Method, City Block Distance and Canberra Distance Method are used to calculate the matching distance and found Canberra Distance Method best among others. Selvarajah et al. [17] introduced a descriptor called Combined Feature Descriptor for the CBIR system to enhance the retrieval performance. It uses the concept of Haar Wavelet and color histogram moment. The descriptor works for the application which includes traditional color moments, 2D-Discrete Wavelet Transform. Abubaker et al. [18] proposed a CBIR technique based on query and extracts the most vital attributes, i.e., color, shape and texture. It includes the automatic extraction of spatial based color feature using invariant Fourier descriptor and texture feature using the Gabor filter. The distance metrics were used for distance calculation and their weightage were determined. Based on the data, the output images are sorted and ranked so that most similar images can be displayed to the user. Omar et al. [19] proposed a WhatAreYouLOOKing4(WAY-LOOK4) system using the Local descriptor and Image Signatures. It contains 3 system components: feature extraction, image database indexing and similarity retrieval. The system maintains reasonable storage and computational costs. The system is simple because of no iterations for clustering or complex wavelet transformation.

The analysis of different techniques is presented in the following table.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Author</th>
<th>Year</th>
<th>Source</th>
<th>Major Findings/Limitations/Future Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ligade et al.</td>
<td>2014</td>
<td>IJCSIT</td>
<td>Reviews NN, IGA and RF using the convergence ratio, precision and recall and provides the various CBIR system based RF techniques.</td>
</tr>
</tbody>
</table>
| 2     | Walia et al.    | 2014 | IEEE  | 1. Proposed a new similarity measure that improves the efficiency of the CBIR system by using the dominant color descriptor.  
 2. In future, it can be used for color based retrieval of videos. |
| 4     | Agarwal et al.  | 2014 | IEEE  | Proposed a novel feature descriptor for CBIR system by integrating Cooccurrence of Haar like Wavelet Filter (CHLWF) with Color Histogram |
| 5     | Jeyabharathi et al. | 2013 | IEEE  | Analyses the performance of feature extraction techniques (PCA, LDA & ICA) and classification techniques (SVM & NN) using the Recognition Rate and F-Score and finds PCA with SVM to the best. |
| 7     | Zhang et al.    | 2013 | IEEE  | 1. Proposed a method of color principal feature extraction called ColorPCA.  
 2. This method contains three main problems:  
a. Nonlinear structures remains unclear on how to extend the ColorPCA algorithm.  
b. To extend the ColorPCA idea in manifold structures or discriminant information of samples for locality preservation  
c. To determine the optimal reduced dimensions for dimension reduction algorithms using ColorPCA. |
| 8     | Syam et al.     | 2013 | IEEE  | Proposed a CBIR System using the GA and SED with feature extraction process. |
| 9     | Ligade et al.   | 2013 | TJPRC | 1. Proposed an image extractor method on multi-feature similarity synthesis using the Genetic Algorithm for efficient image retrieval in CBIR systems  
 2. They consider only the occurrence frequencies of image in result and not the location of the retrieval result which directly reflects the similarity of it and the query image. |
| 10    | Ho et al.       | 2012 | IEEE  | Proposed an architecture for CBIR system combining content-based image, color analysis and data mining. |

III. CONCLUSION

In this work, overview of various feature extraction techniques for Content Based Image Retrieval System is explained briefly. This paper classifies the current methods and summarizes their features. After having a review, it can be considered that no single method can be resulted as best or very good for all types of images or all the methods uniformly good for a specific image type. Considering all these limitations and major findings, CBIR system remains to be a challenging problem in image processing. Feature extraction techniques for CBIR system is still a pending problem in the world and more research need to be carried out for better estimation.

REFERENCES


