ISSN: 2321-7782 (Online)
Volume 3, Issue 3, March 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Design and Implementation of Content Based Image Retrieval Using Data Mining and Image Processing Techniques

Aboli W. Hole[1]
Computer Science & Information Technology
H.V.P.M's C.O.E.T
Amravati, India

Prabhakar L. Ramteke[2]
Computer Science & Information Technology
H.V.P.M's C.O.E.T
Amravati, India

Abstract: ***Now a days people are interested in using digital images. There is a great need for developing an efficient technique for finding the images. In order to find an image, image has to be represented with certain features. Color, texture and size are three important visual features of an image. We will implement an efficient image retrieval technique which uses dynamic dominant color, texture and size features of an image. As a first step, an image is uniformly divided into 8 coarse partitions. The centroid of each partition is selected as its dominant color after the above coarse partition. By using Gray Level Co-occurrence Matrix (GLCM), texture of an image is obtained. Color and texture features are normalized. A robust feature set for image retrieval is provided by using the combination of the color and texture features of an image in conjunction with the shape features. In retrieving the similar images, weighted Euclidean distance of color, texture and shape features is used.***

Keywords: Image retrieval, dominant color, Gray level co occurrence matrix, centroid, Weighted Euclidean distance.

I. INTRODUCTION

Image mining denotes combination of data mining and image processing technology to aid in the analysis and understanding in an image-rich domain. Data mining and image processing are interdisciplinary that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Clearly, image mining is different from computer vision and image processing techniques. This is because the focus of image mining is in the extraction of patterns from a large collection of images, whereas the focus of computer vision and image processing techniques is in understanding and/or extracting specific features from a *single* image. There is overlap between image mining and content-based retrieval, because both are dealing with large collection of images, image mining goes beyond the problem of retrieving relevant images.

In image mining, the goal is the discovery of image patterns that are significant in a given collection of images and the related alphanumeric data. In Broad sense, image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images and between image and other alphanumeric data. Image mining process involves some steps like refining image database, pre-processing, transformation and feature extraction, mining, interpretation and evaluation that based on knowledge. Here, images from an image database are first preprocessed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant patterns. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications. This paper gives brief overview of recently developed image mining techniques.

Due to the proliferation of video and image data in digital form, Content-based image retrieval (CBIR) has become a prominent research topic. Therefore an important problem that needs to be addressed is fast retrieval of images from large databases. To find images that are perceptually similar to a query image, image retrieval systems attempt to search through a

database. CBIR can greatly enhance the accuracy of the information being returned and is an important alternative and complement to traditional text-based image searching. For describing image content, color, texture and shape features have been used. Color is one of the most widely used low-level visual features and is invariant to image size and orientation. There are color histogram, color correlogram, and dominant color descriptor (DCD) as conventional color features used in CBIR. Without any other information, many objects in an image can be distinguished solely by their textures. Texture may describe the structural arrangement of a region and the relationship of the surrounding regions and may also consist of some basic primitives. We have used the texture features using gray-level co-occurrence matrix (GLCM) in our approach. Size feature has been extensively used for retrieval systems. We will implement CBIR system that is based on Dominant color and GLCM texture and size.

Need of Image Mining

Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. These images, if analyzed, can reveal useful information to the human users. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images. Image mining is more than just an extension of data mining to image domain. It is an interdisciplinary endeavour that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence.

Analysis of Problem

Content-based image retrieval (CBIR) has become a prominent research topic. To search for and browse through video and image databases located at remote sites, increased bandwidth availability will allow the users to access the internet in the near future. Therefore an important problem that needs to be addressed is fast retrieval of images from large databases. Image retrieval systems attempt to search through a database to find images that are perceptually similar to a query image. CBIR can greatly enhance the accuracy of the information being returned and is an important alternative and complement to traditional text-based image searching. It aims to develop an efficient visual content-based technique to search, browse and retrieve relevant images from large-scale digital image collections.

II. LITERATURE REVIEW

In [1] an efficient image retrieval technique which uses dominant color and texture features of an image is proposed. The proposed method yielded higher average precision and average recall with reduced feature vector dimension. A new and effective color image retrieval scheme for combining all the three i.e. color, texture and shape information, which achieved higher retrieval efficiency is presented in [2]. Trademark image retrieval (TIR) system is proposed in [3] to deal with the vast number of trademark images in the trademark registration system. The proposed approach commences with the extraction of edges using the Canny edge detector, performs a shape normalization procedure, and then extracts the global and local features. In [4], a further exploration and study of visual feature extraction is done. An image retrieval system is presented in [5], which used HSV color space and wavelet transform approach for feature extraction. A comprehensive survey, highlighting current progress, emerging directions, the spawning of new fields, and methods for evaluation relevant to the field of image retrieval is presented in [6]. It consider that the field will experience a paradigm shift in the foreseeable future, with the focus being more on application-oriented, domain-specific work, generating considerable impact in day-to-day life. Dominant color descriptor (DCD) is one of the color descriptors proposed by MPEG-7 in [7] that has been extensively used for image retrieval. A content-based image retrieval method based on an efficient combination of multi resolution color and texture features is proposed in [8]. In [9], a detailed evaluation of the use of texture features in a query-by-example approach to image retrieval is presented. Image retrieval mechanism is explored in [10], based on combination of color and texture features.

Color, texture and shape features have been used for describing image content. Color is one of the most widely used low-level visual features and is invariant to image size and orientation. As conventional color features used in CBIR, there are color

histogram, color correlogram, and dominant color descriptor (DCD). Color histogram is the most commonly used color presentation, but it does not include any spatial information. Color correlogram describes the probability of finding color pairs at a fixed pixel distance and provides spatial information. Therefore color correlogram yields better retrieval accuracy in comparison to color histogram. Color autocorrelogram is a subset of color correlogram, which captures the spatial correlation between identical colors only. Since it provides significant computational benefits over color correlogram, it is more suitable for image retrieval. Texture is also an important visual feature that refers to innate surface properties of an object and their relationship to the surrounding environment. Many objects in an image can be distinguished solely by their textures without any other information. Texture may consist of some basic primitives, and may also describe the structural arrangement of a region and the relationship of the surrounding regions. In our approach we have used the texture features using gray-level co-occurrence matrix (GLCM). Shape feature has been extensively used for retrieval systems. Shape signatures are computed from blurred images and global invariant moments are computed as shape features.

III. Overview of Image Mining Techniques

Basically, image mining techniques consist of object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network. Here, We briefly discuss these techniques and their application to image mining.

a) Object Recognition

Object recognition has been an active research area in field of image processing. Using object models that are known a priori, an object recognition system finds objects in the real world from an image. This is one of the major tasks in image mining. Automatic machine learning and meaningful information extraction can only be realized when some objects have been identified and recognized by the machine. An object recognition system typically consists of four components, namely, model database, feature detector, hypothesizer and hypothesis verifier. The model database contains all the models known to the system. These models contain important features that describe the objects. The detected image primitive features in the Pixel Level are used to help the hypothesizer to assign likelihood to the objects in the image. The verifier uses the models to verify the hypothesis and refine the object likelihood. The system finally selects the object with the highest likelihood as the correct object.

b) Image Retrieval

Image mining requires that images be retrieved according to some requirement specifications. The requirement specifications can be classified into three levels of increasing complexity [3]

(a) Level 1 comprises low level features of such as color, texture, shape or the spatial location of image elements.

(b) Level 2 comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons.

(c) Level 3 comprises high level features of image. Commercially, there are many IRS like IBM's QBIC system, Virage , Photobook , Chabot, VisualSEEk, MARS, Surfimage and Synapse.

c) Image Indexing

To improve image retrieval speed, there is need of image data base with a fast and efficient indexing scheme. Typically, the image database to be searched is large and the feature vectors of images are of high dimension, search complexity is high. Two main approaches are: reducing dimensionality or indexing high dimensional data. Reducing the dimensions can be accomplished using two well-known methods: the Singular Value Decomposition (SVD) update algorithm and clustering. Current image systems retrieve images based on similarity. However, Euclidean measures may not effectively simulate human perception for certain visual content. Other similarity measures such as histogram intersection, cosine, correlation, etc., need to

be utilized. One promising approach is to first perform dimension reduction and then use appropriate multi-dimensional indexing techniques that support Non-Euclidean similarity measures. Develop an image retrieval system on Oracle platform using multi-level filters indexing. The filters operate on an approximation of the high dimension data that represents the images, and reduces the search space so that the computationally expensive comparison is necessary for only a small subset of the data. Present a new compressed image indexing technique by using compressed image features as multiple keys to retrieve images. Other proposed indexing schemes focus on specific image features such as color, shape and texture features

Classification and Image Clustering

Image classification and clustering are the supervised and unsupervised classification of images into groups. In supervised classification, we are given a collection of labelled (pre-classified) images, and the problem is to label a newly encountered, yet unlabeled images. Typically, the given labeled (training) images are used to do the machine learning of the class description which in turn are use to label a new image. In unsupervised classification (or image clustering), the problem is to group a given collection of unlabeled images into meaningful clusters according to the image content without a priori knowledge. The fundamental objective for carrying out image classification or clustering in image mining is to acquire content information the users are interested in from the image group label associated with the image. Intelligently classifying image by content is an important way to mine valuable information from large image collection. The classification module in the mining system is usually called classifier. The benefits of image classification and clustering include better image storage and management, and optimized image-indexing scheme for fast and efficient image retrieval, all of which are also important to the image mining systems.

d) Association Rule Mining

Association rule mining generate rules that have support and confidence greater than some user specified minimum support and minimum confidence thresholds. A typical association rule mining algorithm works in two steps. The first step finds all large item sets that meet the minimum support constraint. The second step generates rules from all the large item sets that satisfy the minimum confidence constraint. Association rule mining is frequently used in data mining to uncover interesting trends, patterns and rules in large datasets. Recently, association rule mining has been applied to large image databases. Although the current image association rule mining approach is far from mature and perfection compared its application in data mining field, there opens up a very promising research direction and vast room for image association rule mining. There are two main approaches. The first approach is to mine from large collections of images alone, and the second approach is to mine from a combined collection of images and associated alphanumeric data. Association mining from transaction database is a typical case of mining association rules from large database. In this case, an association rule can be generated by examining all the transaction data. The data is explicit and there is a specific and definite data item for each the component item and an individual customer transaction would include a subset of these items and in general a subset of all the items sold by the store. In image databases, manually labelling all the images is practically impossible, and we can only rely on automatic or semi-automatic analysis of the image content, before carrying out mining on the generated descriptions. The generated descriptions could be colour, texture, shape, size etc.

e) Neural Networks

Artificial neural network models have been studied for many years in the hope of achieving humanlike performance in several fields such as speech and image understanding. A neural network, by definition, is a massively parallel distributed processor made up of simple processing units, each of which has a natural propensity for storing experiential knowledge and making the knowledge available for use. Neural networks are fault tolerant and are good at pattern recognition and trend prediction. In the case of limited knowledge, artificial neutral network algorithms are frequently used to construct a model of the data. The conventional programming uses serial processing, while neural networks use parallel processing. Even though there

has been a lot of research work with regard to neural network and its applications, it is relatively new in the image mining domain.

f) Content Based Image Retrieval using Image Mining Techniques

In this section, we are presenting the notion of use of image mining techniques for effective content based image retrieval. Most of the CBIR techniques employ the color and texture features only to retrieve required images from the image database. Image mining process helps to determine different image patterns for each of the images in image database and query image as well. Similarity among target images with the query image is decided on the basis of the pattern which is similar to them. We can Euclidian distance measure or MLE i.e. Maximum likelihood Estimation. Following steps are involved in content based image retrieval using image mining techniques.

- Pre-processing of Query/Database images
- Feature Extraction
- Pattern Mining
- Evaluation Criterion (Distance metric)
- Display of retrieved images

IV. CONCLUSION

In this paper, we have focused on recently developed image mining techniques. The purpose of image mining techniques is discovering meaningful correlations and formulations from previously collected image data. Many different application areas utilize image mining as a means to achieve effective usage of semantic information about images. Image mining is becoming progressively more widespread in both the private and public sectors. Sector such as biomedical, space research organization, remote sensing, fashion, crime prevention, publishing, medicine, architecture, commonly use image mining to reduce costs, enhance research, and increase sales. As image mining is still not fully focused, there is a huge scope for its development. Future research should highlight on development of powerful query language, devise automated image mining techniques based on image retrieval techniques based on its content.

Due to the proliferation of video and image data in digital form, Content-based image retrieval (CBIR) has become a prominent research topic. Therefore an important problem that needs to be addressed is fast retrieval of images from large databases. To find images that are perceptually similar to a query image, image retrieval systems attempt to search through a database. CBIR can greatly enhance the accuracy of the information being returned and is an important alternative and complement to traditional text-based image searching.

Acknowledgment

My thanks to the Guide, Prof. P. L. Ramteke and Principal Dr. A. B. Marathe, who provided me constructive and positive feedback during the preparation of this paper.

References

1. M.Babu Rao, Dr. B.Prabhakara Rao, Dr. A.Govardhan, “Content based image retrieval using Dominant color and Texture features”, International Journal of Computer science and information security, Vol.9 issue No: 2, February 2011.pp:41-46
2. X-Y Wang et al., “An effective image retrieval scheme using color, texture and shape features”, Comput. Stand. Interfaces (2010), doi:10.1016/j.csi.2010.03.004
3. Chia-Hung Wei, Yue Li, Wing-Yin Chau, Chang-Tsun Li, “Trademark image retrieval using synthetic features for describing global shape and interior structure”, Pattern Recognition, 42 (3) (2009) 386–394.
4. FAN-HUI KONG, “Image Retrieval using both color and texture features”, proceedings of the 8th international conference on Machine learning and Cybernetics, Baoding, 12-15 July 2009.

5. JI-QUAN MA, "Content-Based Image Retrieval with HSV Color Space and Texture Features", proceedings of the 2009 International Conference on Web Information Systems and Mining

6. .Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, "Image retrieval: ideas, influences, and trends of the new age", ACM Computing Surveys, 40 (2) (2008) 1–60.

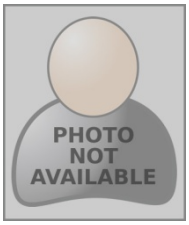
7. Nai-Chung Yang, Wei-Han Chang, Chung-Ming Kuo, Tsia-Hsing Li, "A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval", Journal of Visual Communication and Image Representation, 19 (2) (2008) 92–105.

8. Young Deok Chun, Nam Chul Kim, Ick Hoon Jang, "Content-based image retrieval usin multiresolution color and texture features", IEEE Transactions on Multimedia, 10 (6) (2008)1073–1084.

9. P. Howarth and S. Ruger, "Robust texture features for still-image retrieval", IEEE Proceedings of Visual Image Signal Processing, Vol.152, No. 6, December 2005.

10. S. Liapis, G. Tziritas, "Color and texture image retrieval using chromaticity histograms and wavelet frames", IEEE Transactions on Multimedia 6 (5) (2004) 676–686.

AUTHOR(S) PROFILE



Aboli Wasudeorao Hole, received the B.E. degree in Computer Science from, P. R. Pote(Patil) College Of Engineering And Technology, Amravati in 2013 and is currently persuing Master's Degree in Computer Science and Information Technology from H.V.P.M's College of Engineering And Technology, Amravati.

Prof. Prabhakar L. Ramteke, received the B.E. degree in Computer Science from, Government College Of Engineering And Technology and M.E. degree from Prof. Ram Meghe Institute of Technology, Badnera, respectively. His field of specialization is Mobile Computing. He is currently working as Associate Professor at H.V.P.M's college of Engineering and Technology, Amravati.

