ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

A Study on Content Based Image Retrieval Systems

L. Haldurai[*, 1], V. Vinodhini[2]

[*] Research Scholar, Department of Computer Science, Dr.N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India

[1] Assistant Professor, Department of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore, Tamilnadu, India

[2] Assistant Professor, Department of Information Technology, Dr.N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India

ABSTRACT: A database is a collection of information that is organized for easy storage, retrieval and update. This information is represented in many forms like text, table, image, chart and graph etc. here we concentrate on information that is stored in the form of images. Content based image retrieval (CBIR) technique explores various methodologies in extracting implicit knowledge, patterns and relationships found in the images from the collection of images. This paper focuses on various techniques that were proposed in earlier literature.

KEYWORDS: Data Mining, CBIR, Image Mining, Feature Extraction, Image Retrieval

I. INTRODUCTION

In this era of information technology, all areas of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design, and historical research use images for efficient services. A large collection of images is referred to as image database. An image database is a system where image data are integrated and stored [1]. Image data include the raw images and information extracted from images by automated or computer assisted image analysis. The police maintain image database of criminals, crime scenes, and stolen items. In the medical profession, X-rays and scanned image database are kept for diagnosis, monitoring, and research purposes. In architectural and engineering design, image database exists for design projects, finished projects, and machine parts. In publishing and advertising, journalists create image databases for various events and activities such as sports, buildings, personalities, national and international events, and product advertisements. In historical research, image databases are created for archives in areas [1].

Content-based image retrieval technique uses visual contents to search images from large scale image databases based on users' interests. It becomes an active and fast advancing research area. Image content may include both visual and semantic content. Content-Based Image Retrieval (CBIR) is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape [2]. These techniques includes several areas such as image segmentation, image feature extraction, representation, mapping of features to semantics, storage and indexing, image similarity-distance measurement and retrieval which makes CBIR system development as a challenging task [3]. Several companies are maintaining large image databases, where the requirement is to have a technique that can search and retrieve images in a manner that is both time efficient and accurate [4].

Content-based image retrieval, also known as query by image content and content-based visual information retrieval is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. Content-based means that search makes use of the contents of the images themselves, rather than relying on human-input metadata such as captions or keywords. A *content-based image retrieval system* (CBIR) is a piece of software that implements CBIR. In CBIR each image that is stored in the database has its features extracted and compared to the features of the query image. There are three important feature components for content based image retrieval [5]. The most common are color [6, 7], texture [8, 9] and shape [10, 11] or combinations of these. These features are combined to achieve higher retrieval efficiency [7].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

II. METHODOLOGY

The existing methodology consists of following general steps.

Step1: **Pre-processing**

The image data is highly non-trivial. Reprocessing phase is applied to remove noise from the image. And it also consists of image segmentation. By applying noise removal filter, thinning, cleaning noise can be removed. Pre-processing phase also include object identification. Pre-processing determines the effectiveness of image mining application.

Step2: **Feature Extraction**.

Features, characteristics of the objects of interest, if selected carefully, are representative of the maximum relevant information that the image has to offer for a complete characterization of the lesion

Step 3: **Feature selection.**

Feature selection helps to reduce the feature space that improves the prediction accuracy.

Step 4: **Classification.**

Classification process involves two phases: Training and testing phase.

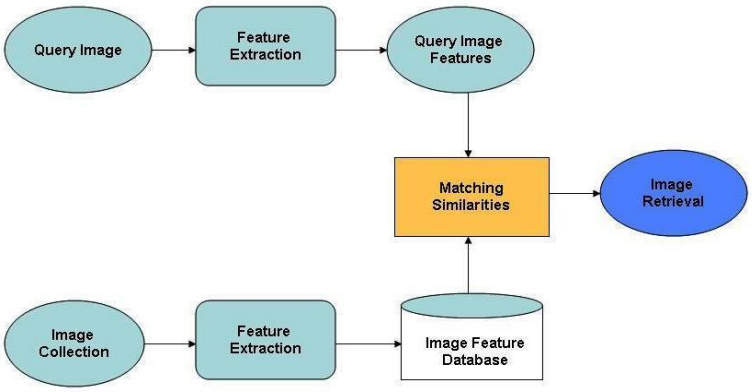

Fig.1 Content Based Image Retrieval Systems

A. SHAPE FEATURE

Shape is one of the important features and contains the most attractive visual information for human perception [12]. We use the term shape to refer to the information that can be deduced directly from images and that cannot be represented by color or texture; as such, shape defines a complementary space to color and texture [13]. Shape representations techniques used in similarity retrieval are generally characterized as being region based and boundary based [14].

B. COLOR FEATURE

The color feature has widely been used in CBIR systems, because of its easy and fast computation [15]. Color is one of the visual attributes that can provide more information about the visual content of an image and the most widely used feature in CBIR [16, 17]. This is a compact representation of the color feature to characterize a color image [18-20]. Color is one of the important feature of an image, which depicts much of the information from the image. RGB color model do not correspond to the human way of perceiving the colors. And also RGB space do not separate the luminance component from the chrominance ones. Therefore, HSV color space is used in our approach. Each

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

component of HSV model contributes directly to visual perception, therefore it is commonly used in image retrieval systems [21], [22].

C. TEXTURE FEATURE

Texture is one of the most important defining features of an image [23]. In image classification texture provides important information as in many images of real world [24, 25]. Texture is another important attribute of an image and it refers to innate surface properties of an object and their relationship to the surrounding environment. For texture analysis we use a gray level co-occurrence matrix (GLCM), which is a simple and effective method for representing texture [26].

D. SIMILARITY MEASURE

The Similarity functions seek calculates the content difference between two images based on their features. One of the images is given as search parameter and another is stored in the database and had their features previously extracted [27]. A measurement of how close a vector to another vector is called similarity measurement [28].The query image features are used to retrieve the similar images from the image database. Instead of directly comparing two images, similarity of the query image features is measured with the features of each image in the database. Computing the distance between the feature vectors is the measure of similarity between two images. The retrieval systems return the first images, whose distance from the query image is minimum [29].

There are four major classes of similarity measures:
1. Color similarity,
2. Texture similarity,
3. Shape similarity,
4. Object and Relationship similarity.

III. LITERATURE REVIEW

Choras *et al.*,[30] proposed an integrated color, texture and shape feature extraction method in which Gabor filtration is used for determining the number of regions of interest (ROIs). They calculated texture and color features from the ROIs based on threshold Gabor features and histograms, color moments in YUV space, and shape features based on Zernike moments. The features presented proved to be efficient in determining similarity between images.

Jain and Singh [31] provided an overview of the functionality of content based image retrieval systems by combining advantages of HC and divide and conquer K-Means strategy. He proposed HDK method to use both advantages of HC and Divide and conquer K-Means by introducing equivalency and compatible relation concepts.

Singha and Hemachandran [32] presented the content based image retrieval using features like texture and color, called Wavelet Based Color Histogram Image Retrieval (WBCHIR). The texture and color features are extracted through wavelet transformation and color histogram and the combination of these features is robust to scaling and translation of objects in an image. He also demonstrated a promising and faster retrieval method on a WANG image database containing 1000 general-purpose color images.

Pinjarkar *et al.*,[33] discussed various methodologies used in the research area of Content Based Image Retrieval techniques using Relevance Feedback. To improve the retrieval performance of the CBIR the Relevance Feedback technique can be incorporated in CBIR system to obtain the higher values of the standard evaluation parameters used for evaluation of the CBIR system which may lead to better results of retrieval performance. He also discussed various relevance feedback techniques for Content Based Image Retrieval systems, the various parameters used for experimental evaluation of the systems and the analysis of these techniques on the basis of their results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

In [34] provides an overview of the technical achievements in the research area of Relevance Feedback (RF) in Content-Based Image Retrieval (CBIR). It also covers the current state of art of the research in relevance feedback in CBIR, various relevance feedback techniques and issues in relevance feedback.

Bulo *et al*.,[35] proposed a novel approach to content-based image retrieval with relevance feedback, which is based on the random walker algorithm introduced in the context of interactive image segmentation. The idea is to treat the relevant and non-relevant images labeled by the user at every feedback round as "seed" nodes for the random walker problem. The ranking score for each unlabeled image is computed as the probability that a random walker starting from that image will reach a relevant seed before encountering a non-relevant one.

In [36] put forth an image mining technique using wavelet transform. The author proposed an image mining approach using wavelet transform. It uses common pattern identification and data mining models. They have constructed a prototype system for identification using DWT +PCA system. Handling large number of images at the source machine is a crucial task and hence gives rise to memory management issue. The database should be so efficient to handle images efficiently.

Lionel Gueguen and Mihai Datcu [37] addressed the problem of extracting relevant information from Satellite image time series (SITS) based on the information-bottleneck principle. The method depends on suitable model selection, coupled with a rate–distortion analysis for determining the optimal number of clusters. They presented how to use this method with the Gauss–Markov random fields and the auto binomial random fields model families in order to characterize the spatio-temporal structures contained in SITS. In this approach spectral or geometrical information was not taken into account.

In [38] proposed a new image mining technique using directional spatial constraints. Retrieval performance was evaluated by the author using precision and recall using a ground truth that was constructed by manually identifying the objects satisfying each query. It needs manual interpretation to improve processing time.

P.Rajendran and M.Madheswaran [39] discussed an image mining technique. It combines low level features extracted from images and high level knowledge from specialist. It does not addressed feature redundancy, image noise and time complexity.

In [40] proposed a framework focuses on color as feature using color Moment Block Truncation Coding (BTC).To extract feature for image dataset. Then K-Means clustering algorithm is conducted to group the image dataset into various clusters. This approach does not addressed time and speed of the algorithm.

In [41] proposed two methods for discovering the underlying associations between text and images. The first method based on transformation measures the information similarity between visual features and textual features. Another method uses a neural network to learn direct mapping between visual and text features by incrementally summarizing associated features into a set of information template. It needs to perform batch learning on a fixed set of training data.

Maybin Muyeba *et al.,* [42] proposed a system for image mining using fuzzy rule. It relates the property of composite attributes. They partitioned the property value into fuzzy property sets. In this approach fuzzy measures and correlation association is not described.

Surya S. Durbha *et al.,* [43] proposed feature selection and feature transformation based on wrapper based approach. They adopted region based system. They have argued that selecting a relevant feature subset increases the rate of correctly identifying a semantic class. This system currently used imagery from only one sensor.

Dubey [44] illustrated about an image mining methods which is dependent on the color Histogram, texture of that image. The query image is considered, then the COLOR Histogram and Texture is created and in accordance with this the resultant image is found. In this approach computing time for RGB color space not considered.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

IV. CONCLUSION

This paper elucidates different proposed methods and techniques used by the researchers. It propagates the significance of content based image retrieval systems. The ultimate aim of CBIR systems is to extract the features like shape, color and texture of the image from the database and compare it with the query image in order to retrieve the desired image. More methods and techniques are in progress to make content based image retrieval system more effective and efficient. This overview focuses on content based image retrieval implementations, usability and challenges. It also delivers conceptual overview of methodology.

REFERENCES

[1] Chi Kuo Chang, "Image Information Systems", Proc. of IEEE Pattern Recognition, vol. 73, no 4, pp.754 - 766, April 1985.
[2] Singh, J., J.S. Kaleka and R. Sharma., "Different approaches of CBIR techniques", Int. J. Comput.Distributed Syst., 1: 76-78, 2012.
[3] Chauhan, N. and M. Goyani., "Enhanced Multistage Content Based Image Retrieval', Int. J.Comput. Sci.Mobile Comput., 2: 175-179, 2013.
[4] Xiaoling, W., "A novel circular ring histogram for content-based image retrieval", Proceedings of the 1st International Workshop on Education Technology and Computer Science, pp: 785-788, 2009.
[5] H.J. Lin, et al. "A Study of Shape Based Image Retrieval", 24th IEEE International Conference on Distributed Computing Systems Workshops, 2004.
[6] Y. Rui, T.S. Huang., S.F. Chang SF, "Image Retrieval: Current Techniques, Promising Directions and Open Issues", J. Vis. Commun. Image Rep., Vol. 10, No. 4, pp.39-62, 1999.
[7] P. Hiremath., J. Pujari., "Content Based Image Retrieval Using Color, Texture and Shape Features", IEEE International Conference on Advanced Computing and Communications, ADCOM, 2007.
[8] M.M. Islam., D. Zhang., G. Lu., "Automatic Categorization of Image Regions Using Dominant Colour Based Vector Quantization", Digital Image Computing: Techniques and Applications, pp.191-198, 2008.
[9] N. Rasiwasia., N. Vasconcelos., and P.J. Moreno., "Query by Semantic Example", 15th International Conference on Image and Video Retrieval, Lecture Notesin Computer Science, Vol. 4071, pp. 51-60, 2006.
[10] V.A. Tanase., "Content Based Image Retrieval Systems: A Survey", Dept. of Computing Science, Utrecht University, Technical Report, 2000.
[11] W. Zhang., et al., "Shape Based Indexing in a Medical Image Database", IEEE Workshop on Biomedical Image Analysis, 1998.
[12] D. Zhang., G. Lu., "Shape Retrieval Using Fourier Descriptors", 2nd IEEE Pacific Rim Conference on Multimedia, Citeseer, 2001.
[13] C.B. Akgul, et al., "Content Based Image Retrieval in Radiology: Current Status and Future Directions", Journal of Digital Imaging, Vol. 24, No. 2, pp. 208-222, 2011.
[14] A.A.A. Youssif., A.A. Darwish., R.A. Mohamed., "Content Based Medical Image Retrieval Based on Pyramid Structure Wavelet", International Journal of Computer Science and Network Security, Vol. 10, No. 3, 2010.
[15] Y.K. Chan, C.Y. Chen, "Image Retrieval System Based on Color Complexity and Color Spatial Features", Journal of Systems and Software, Vol. 71, No. 1, pp. 65-70, 2004.
[16] D. Shi, L. Xu, L. Han, "Image Retrieval Using Both Color and Texture Features", Journal of China Universities of Posts and Telecommunications, 2007.
[17] H. Muller, et al., "A Review of Content Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions", International Journal of Medical Informatics, Vol. 73, No. 1, pp. 1-24, 2004.
[18] M.A. Stricker., M. Orengo., "Similarity of Color Images", IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology, International Society for Optics and Photonics, pp. 381-392, 1995.
[19] J. Huang, et al. "Image Indexing Using Color Correlograms", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997.
[20] H. Zhang, et al. "Image Retrieval Based on Color Features: An Evaluation Study", Photonics East'95, International Society for Optics and Photonics, 1995.
[21] H.-W. Yoo, H.-S. Park, and D.-S. Jang, "Expert system for color image retrieval," Expert Syst. Appl., vol. 28, no.2, pp. 347–357, 2005.
[22] H. Nezamabadi-pour and E. Kabir, "Image retrieval using histograms of uni-color and bi-color and directional changes in intensity gradient," Pattern Recognit. Lett., vol. 25, no. 14, 2004.
[23] P. Suhasini, K.S.R. Krishna, I.M. Krishna, "CBIR Using Color Histogram Processing", Journal of Theoretical and Applied Information Technology, Vol. 6, No. 1, pp. 116-122, 2009.
[24] J.R. Smith, C.S. Li, "Image Classification and Querying Using Composite Region Templates", Computer Vision and Image Understanding, Vol. 75, No.1, pp. 165-174, 1999.
[25] M. Rehman, et al., "Content Based Image Retrieval: Survey", World Applied Sciences Journal, Vol. 19, No. 3, pp. 404-412, 2012.
[26] R. M. Haralick and L. G. Shapiro, "Computer and Robot Vision", Volume I, Reading, MA: Addison-Wesley, 1992.
[27] L.C.C. Bergamasco, F.L.S. Nunes, "Content Based Retrieval for 3D Medical Models: A Study Case Using Magnetic Resonance Imaging", Data Inicio, 2012.
[28] C. Chin Chen ∗and H. Ting Chu, "Similarity Measurement between Images", 2005.
[29] T. Acharya , Ajoy K. Ray, "Image Processing Principles and Applications", Book, Published by John Wiley & Sons, Inc., Hoboken, New Jersey. 2005

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

[30] R.S. Choras, T. Andrysiak, M. Choras, "Integrated color, texture and shape information for content-based image retrieval", Pattern Anal Applic. 10: 333-343, 2007.
[31] Jain, M. and S.K. Singh, "A survey on: Content based image retrieval systems using clustering techniques for large data sets", Int. J. Manage. Inform. Technol., 3: 23-39, 2011.
[32] Singha, M. and K. Hemachandran., "Content based image retrieval using color and texture". Signal Image Process. Int. J., 3: 39-57, 2012.
[33] Pinjarkar, L., M. Sharma and K. Mehta., "Comparison and analysis of content based image retrieval systems based on relevance feedback", J. Emerg. Trends Comp. Inform. Sci, 2012.
[34] Patil, P.B. and M.B. Kokare, "Relevance feedback in content based image retrieval: A review", J. Appli. Comp. Sci. Math., 10: 41-47,2011.
[35] Bulo, S.R., M. Rabbi and M. Pelillo., "Content based image retrieval with relevance feedback using random walks", Patt. Recogn., 44: 2109-2122, 2011.
[36] Sanjay T.Gandhe, K. T. Talele and Avinash G. Keskar , "Image Mining Using WaveletTransform", *Knowledg-Based Intelligent Information and Engineering Systems, Springer link bookchapter, pp. 797-803,2007.*
[37] Lionel Gueguen and Mihai Datcu, " Image Time-Series Data Mining Based on the Information-Bottleneck Principle", *IEEE Transactions On Geoscience and Remote Sensing, Vol. 45, no. 4,827, April 2007.*
[38] Selim Aksoy and R. Gokberk Cinbi , "Image Mining Using Directional Spatial Constraints", *IEEE Ggeosciences and Rremote Sensing letters, vol. 7, no. 1, pp. 33-37,Aug 2009.*
[39] P. Rajendran and M. Madheswaran , "An Improved Image Mining Technique For Brain TumourClassification Using Efficient classifier*", International Journal Of Computer Science and Information Security, Vol. 6, no. 3, pp. 107-116, Jan 2009.*
[40] Sanjay Silakari, Mahesh Motwani and Manish Maheshwari., "Color Image Clustering using Block Truncation Algorithm*", Ijcsi International Journal Of Computer Science issues, Vol. 4, no. 2,Aug 2009*
[41] Tao Jiang and Ah-Hwee Tan,Learning, "Image-Text Associations", *IEEE Transactions On Knowledge and Data Engineering, Vol.21, no. 2, February 2009*
[42] Maybin Muyeba, M. Sulaiman Khan, and Frans Coene, "A Framework for Mining Fuzzy Association Rules from Composite Items", *Springer-Verlag Berlin Heidelberg, pp. 62–74, May 2009.*
[43] Surya S. Durbha, , Roger L. King, and Nicolas H. Younan, "Wrapper-Based Feature Subset Selection for Rapid Image Information Mining", *IEEE GeoScience and Remote Sensing letters, vol. 7, no. 1, January 2010.*
[44] Rajshree S. Dubey , "Image Mining using Content Based Image Retrieval System", International Journal On Computer Science and Engineering Vol. 02, no. 07, pp. 2353-2356, October 2010.

BIOGRAPHY

L. Haldurai working as an Assistant Professor in Department of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore. He received his Master of Computer Applications (MCA) degree from Bharathiar University. His research interests are Data Mining, Data Communications and Artificial Intelligence & Expert Systems.

V. Vinodhini has completed her Post Graduation and M.Phil degree from Bharathiar University. At present working as an Assistant Professor in Department of Information Technology, Dr.N.G.P. Arts and Science College, Coimbatore. Her research interests are Data Mining & Warehousing.