International Journal of Advanced Research in
Computer Science Engineering and Information Technology
Volume: 2 Issue: 2 08-Mar-2014,ISSN_NO: 2321-3337

Multi-Index Method for Geospatial Content Based Image Retrieval

Adlene Ebenezer P[#1], Suchithra S[#2], Poovaraghan R J[#3]
PG Student[#1], Assistant Professor [#2, #3]
Department of Computer Science Enginerring, [#1, # 2,# 3]
Easwari Engineering College, Ramapuram, Chennai. [#1, #2]
SRM University, Ramapuram Campus, Chennai. [#3]
adlenepackiadoss@gmail.com [#1], suchitras_2000@yahoo.com[#2] ,poova2003@gmail.com[#3]

Abstract— *In recent years, due to the enormous increase in image database sizes, there is a need for indexing and image retrieval system development. The problem for fast searching and retrieval of refined images has attracted tremendous attention. Content Based Image Retrieval is the most emerging field in the area of image search and indexing, finding similar images for the given query image from the image database. CBIR system focuses on retrieving images from the database, the system depends on how the indexing is being implemented. The proposed technique for indexing is weighted multi indexing; weights for the each index can be obtained dynamically for each query. The input to the search process is a multi object; a multi object search can be used to identify relevant groups of object which match a given set of query objects. In the area of satellite imagery retrieval, the images stored in the database are labeled by feature vectors, which are extracted from the images. CBIR indexes are built for each class of features.*

Keywords—Content Based Image Retrieval(CBIR), weighted indexing, multi-object, image database.

1.INTRODUCTION

Multimedia data have been increased in our life due to usage of digital device. This made us to face the challenges of retrieving contents from large database effectively. One of the most popular techniques used for retrieving is Content Based Image Retrieval (CBIR) [1]. Content-based image retrieval has attracted many in recent years. Due to large collection of images, efficiency is an important factor for Content Based Image Retrieval, therefore developing an efficient indexing method is great significance. Although indexing technique have been studied in database community; traditional indexing are still not efficient for image content. There are many techniques for extracting information from CBIR. After extracting the features from the image, these features are indexed into separate CBIR indexes. To address the issue of indexing, we have developed a dynamically weights for multiple CBIR indexes from relevance feedback. The input to a query is no longer a single object, it is multi object and user is interested in identifying groups of object which match those given in the query.

Satellite images are playing an important role in many applications, such as environmental study and homeland security. In many geospatial imagery search applications, it may be useful to query multiple CBIR data sources. For example, a system might search for images which match certain color, texture and shape characteristics. However, this requires that the result from multiple sources be merged in some intelligent fashion. We have developed techniques to enable rapid multi-index and multi-object geospatial CBIR searches. Multi-object search is a mechanism to provide users with the ability to require that certain objects be matched.

2. RELATED WORK

For CBIR indexing, a lot of traditional multidimensional indexing technique, such as k-d tree [2], R-tree [3], R^+-tree, and similarity indexing [4]. These indexing methods are designed specifically for the purpose of localizing objects of interest in an image. An indexing technique, locality-sensitive hashing, was proposed for solving the near neighbor search in high dimensional spaces efficiently [5].

The work presented by Matthew et al. [6] shows a method to estimate the missing score using Absence Penalty Method; this method is introduced to tune the performance of the retrieval. It gives balance estimation of missing distance values in order to increase efficiency.

D.Feng et al. [9] indexing method was inspired by image matching approach. By knowing the feature point correspondences, similarity between query image and dataset image are computed easily and retrieved quickly. The order of quantization is used to increase the distinction among the quantized feature vector. After determining the feature point correspondence, the multi-dimensional inverted index is developed to compute the number of feature point correspondences and approximates RANSAC is done for further estimation of the spatial correspondence of feature points.

Gong et al [8] focuses on a new feature matching strategy, match matrix, which describes the correspondence of the feature, to combine spatial relations and feature similarity, its global maximum is assumed to be reached if the retrieved image is same as the reference image. Thus, by calculating the maximum of function, the feature correspondence can be estimated. To solve the optimization problem we use two approaches. One is based on the branch-and-bound strategy to get a global optimal solution and the other uses an iterative algorithm that combine graduated assignment and variable metric methods to search for a local optimal solution with low computational complexity. The proposed method can work without the imitations of feature type, similarity criterion and transfer model, then its performance is evaluated using variety of real images. It is fast, robust and highest accuracy.

A novel indexing structure [13] that was developed to efficiently and accurately perform content based shape retrieval of objects from a large scale satellite imagery database. Objects of multiple scales are automatically extracted from satellite imagery and then encoded in to a bitmap shape representation. The entropy balanced bitmap tree, which exploits the probabilistic nature of bit values in automatically derived shape classes.

GeoIRIS [14](Geospatial Information Retrieval and Indexing System) which extract the feature automatically, mining the image from large scale database and indexing the database for fast retrieval. Here the complex queries that merge information from geospatial databases, retrievals of objects based on shape and visual characteristics, analysis of multi object relationships for the retrieval of objects in specific spatial configurations and semantic models to link low level image features with high level visual descriptors.

3.THE PROPOSED METHOD

To develop a method for weighting indexes [10], relevance feedback is used for producing relevant results. This method helps us to identify sets of feature within an image that correspond to similar sets of feature found in the result images. Using result marked relevant by users, mining techniques are applied for pairs of query and result images which have similar features. These features can be used for weighted indexing. Another algorithm, Absence Penalty Method is used to estimate missing distance value to increase efficiency to improve accuracy. The input to the system is multi-object so user desires that certain query object must be matched.

International Journal of Advanced Research in
Computer Science Engineering and Information Technology
Volume: 2 Issue: 2 08-Mar-2014,ISSN_NO: 2321-3337

Obligatory Object Query is used to address the issue; it allows query objects to be marked required based on the user's choice.

3.1 Weighted Indexing

To perform weighted indexing, we need to perform relevance feedback and itemset construction. To perform relevance feedback process, a group users evaluate the result of performing queries using equal weights for each index. The user marks the retrieved results as relevant or not-relevant.

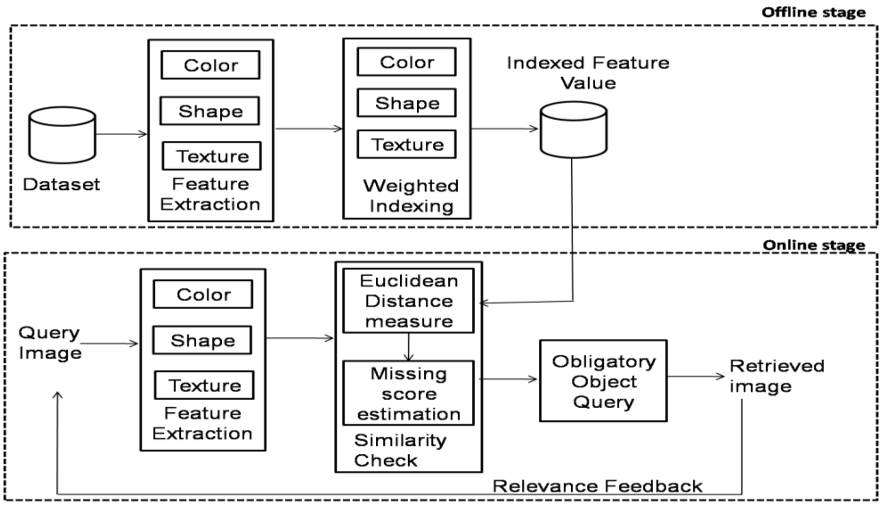

Fig 1. A Conceptual Framework for Content Based Image Retrieval

Using the relevance feedback, the features of the corresponding query and result images are analyzed. The features are discretized into bins using the statistical method of each feature individually. To identify the bin corresponding to the i^{th} feature, the following equation is used.

$$b_i = round\left(\frac{x_i - \mu_i}{k\sigma_i}\right) \qquad (1)$$

The constant k is used to control the width of the bins. For all pairs of query and result images in the relevance feedback list, the value of each feature is discretized according to Equation 1. The features are then partitioned according to indexes. For each query and result image, all of the features in a particular index are analyzed and if a given feature is in the same bin in both images, then it is placed into the list of *bins of interest* for that particular relevance feedback transaction. This process is repeated for each of the indexes, generating a distinct set of *bins of interest* for each class of features. These *bins of interest* correspond to ranges in feature space in which feature values for both the query and the result image exist.

By creating large itemsets of the *bins of interest*, it helps to determine the weights to apply to each index for a new query image. Each large itemset contain a set of bins that together were useful in identifying relevant imagery, because the feature values for both query and result were found in this bin. If a new query image has feature values found in these bins then these features are predicted to be useful for identifying relevant results. Accordingly, this index should be given extra weight to reflect each matched large itemset.

International Journal of Advanced Research in

Computer Science Engineering and Information Technology

Volume: 2 Issue: 2 08-Mar-2014,ISSN_NO: 2321-3337

To determine the weight for an index given a query image, all large itemset for the index are compared to the feature bins identified from the query image. The set of large itemsets, L_d, for index d can be defined as

$$L_d = \{L_d^1, L_d^2, \dots, L_d^N\} \qquad (2)$$

Where L_d^n is the n^{th} itemset for index d are defined as q_d, then the group of itemsets for index d that match the query can be found by:

$$L'_d = \{q_d \cap L_d^n | L_d^n \epsilon L_d\} \qquad (3)$$

These matched itemsets for index d can then be used to calculate the weight that will be applied to the index. The following equation sums over all matched rules giving extra weight to matched itemsets of cardinality greater than 1.

$$w_d = \frac{1}{|L_d|} \sum_{L_d^{n'} \epsilon L'_d} \left(1 + ln\left|L_d^{n'}\right|\right) s\left(L_d^{n'}\right) \qquad (4)$$

The $s\left(L_d^{n'}\right)$terms correspond to the support of the given itemset. The weight computed in Equation 4 is normalized by the total number of large itemsets from the index.

The Absence Penalty Method [6] method is used to retrieve the missing distance. The input to the algorithm is query object and desired number of results. This method will over retrieve from each individual CBIR index and aggregate the results, estimating missing values where needed. First, a parameter is set to populate with object keys and aggregate scores. This algorithm will over retrieve from each index and estimate the missing values where needed. The set of aggregate result does not grow unbounded and only contain up to k objects, which correspond to the best aggregate results. The algorithm performs random access to determine the actual distance value for a result if the number of missing scores is less than missing score count threshold. The random access operations is used to refine scores for result which were found in the result sets of many other indexes but were missing in a small number of result sets. It performs random access operations for results whose known scores are highly ranked but still having missing scores after the first pass of random access. APM is more suitable for real time online applications due to the performance guarantee despite the use of some random access operations. The Obligatory Object Query [6] helps the user to desires certain object must be matched. This algorithm allows query objects to be marked required based on the user's preference. The algorithm will strive to match the required objects first, at the expense of identifying a match which might otherwise be globally optimal.

3.2 Feature Representation

Feature representation is a key challenging step for building CBIR system. Here, three types of visual features are used: color, shape and texture.

For color, we use block truncation method [8]. Each image is separated into red, green and blue components. The average of each component is calculated and grouped in RL, RH, GL, GH, BL and BH. RH is obtained by taking only red component of all pixels in the image which are above red average and RL is obtained by taking only red component of all pixels in the images which are below red average. Similarly GH, GL, BH and BL can be obtained. Apply color moments to each splitted component and apply clustering algorithm to find the cluster. Thus the color feature of an image is extracted.

For texture, we adopt Gray Level Co-occurrence Matrix [11]. Each image is converted to gray scale image. Construct GLCM of four directions respectively. A GLCM whose angle of adjacent pixels is (0^0, 45^0, 90^0, 135^0). After constructing the co-occurrence matrix of four directions, calculate texture parameters energy, contrast, entropy, correlation, and local balance of the four co-occurrence matrix and then calculate the mean and standard deviation of each parameter in order to form the various components of the texture feature vector. After extracting texture parameters, each component represents a different physical meaning and the range is not the same. When the Euclidean distance is as a measure, the small weight compared with the large weight is easy to ignored, so they need to feature internal normalization, which has many methods such as: linear normalization method, uniform distribution of the normalization method, sorting, normalization, and Gaussian normalization method. And Gaussian is the best and most commonly used whose advantage is that a small element value in the whole normalized distribution and played a better balance role.

For shape, we employ Canny Edge Detection [12]. It is inevitable that all images have some amount of noise, to prevent that noise images are smoothed by applying a Gaussian filter. This algorithm finds edges where the grayscale intensity of the image changes the most. These areas are determined gradients of the image. Gradients at each pixel in the smoothed image are determined by applying Sobel-operator. Then the image is blurred edges in the image of the gradient magnitudes to sharp edges. Each pixel's gradient direction is round to nearest angle, corresponding to the use of an 8-connected neighborhood. Compare the edge strength of the current pixel with the edge strength of the pixel in the positive and negative gradient direction, if the edge strength is largest; preserve the value of the edge strength. If not suppress the value. The edge pixel remaining after suppression step, most of them are true edges, but some may be caused by noise or color variations due to rough surface. The Canny algorithm uses double thresholding. Finally strong edges are determined by suppressing all weak edges. Thus the features of the image is extracted and indexed.

4.EXPERMENTAL RESULT

In this section tells about the experimental result of Inverted Indexing and Weighting Indexing in our CBIR System. For our satellite image data set, there are many feature extraction algorithm that shows the information in the image. The database consists of color, shape and texture feature corresponding to image. To evaluate the performance of CBIR queries, a group of users was assigned to check whether the retrieved images are relevant or not relevant. The user first evaluates the result performance with each index weighted equally. After retrieved result the relevancy score is calculated for relevance feedback. It is observed that our weighted indexing method improves the efficiency in the large scale image database.

5. DISCUSSION

In this paper we have developed a weighted indexing for Geospatial CBIR. Our method is efficient and robust, making it applicable for real-time retrieval systems. The weights generated while indexing are not static it can be varied based on the features of the query image. Multi object help us to identify the group of objects that were similar to set of objects presented as the query.

REFERENCES

[1] Micheal S. Lew, Nicu Sebe, Chabane Djeraba and Ramesh Jain, "Content-based multimedia information retrieval: State of art and challenges," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1-19, 2006.

[2] G.Scott and C.-R. Shyu, "Knowledge-driven multidimensional indexing structure for biomedical media database retrieval," *IEEE Trans. Inf. Technol. Biomed.*, vol.11, no. 3, pp. 320-331, May 2007.

[3] Ben Kao, Sau Dan Lee, Foris K.F. Lee, David Wai-lok Cheung and Wai-Shing Ho, "Clustering Uncertain Data Using Voronoi Diagram and –Tress Index " *IEEE Trans. Knowl. Data Eng.*, vol 22, no.9, pp. 1219-1233, Sept. 2010.

[4] White, D.A., Jain.R , Similarity indexing with the ss-tree. *Proc. 12th Intl. Conf. on Data Engineering*, (1996) 516-523.

[5] Mayur Datar, Nicole Immorlica, Piotr Indyk and Vahab S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions" , in *Proc. 20th Annual Sysmposium on Computational Geometry*, New York, NY, USA, 2004, PP. 253-262.

[6] Matthwe N. Klaric, Grant J. Scott and Chi-Ren Shyu, "Multi-Index Multi-Object Content-Baesd Rerieval" *IEEE Transaction on Geoscience and Remote Sensing*, Vol.50, No.10, Oct. 2012.

[7] G.J. Wen, J.Lv and W.Yu, "A high-performance feature-matching method for image registration by combning spatial and similarity information" *IEEE Trans. Geosci. Remote Sens.*, Vol. 46, No. 4, pp. 1266-1277, Apr. 2008.

[8] Sanjay Silakari, Mahesh Motwani and Manish Maheshwari, "Color Image Clustering Using Block Truncation Algorithm" *International Journal of Computer Science Issues*, vol. 4, No. 2, 2009.

[9] Deying Feng, Jie Yang and Congxin Liu, "An Efficient Indexing Method for Content-Based Image Retrieval" *Institute of Image Processing and Pattern Recognition*, China, 2013.

[10] M.Klaric, G.Scott and C.R. Shyu, "Mining visua associations from user feedback for weighting multiple indexes in geospatial image retrieval," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2006, pp. 21-24.

[11] Guangwen Zhang, Lei Yang and Fan Zhang, "An Intergrated Color and Texture Feature Extracion Algorithm," *International Conference on Computer Science and Network Technology*,2012.

[12] Ping Zhou, Wenjun Ye, Yaojie Xia and Qi Wang, "An Improved Canny Algorithm for Edge Detection," *Journal Of Computational Information System*, 2011.

[13] G.Scott, M.Klaric and C.R. Shyu, "Entropy balanced bitmap ree for shape based object rerieval from large-scale satellite imagery database," *IEEE Trans. Geosci. Remote Sens.,* Vol. 49, No. 5, pp. 1603-1616, May 2011.

[14] C.R. Shyu, M.Klaric, G. Scott, A.Barb, C.Davis and K. Palaniappan, "GeoIRIS: Geospatial Information Retrievall and Indexing System-Content mining, Semantic Modeling and Complex Queries," *IEEE Trans. Geosci. Remote Sens.*, Vol. 45, No. 4, pp. 839-852, Apr. 2007.