INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 3, Issue 6)
Available online at www.ijariit.com

Multivariate Feature Descriptor based CBIR Model to Query Large Image Databases

Harkamal Kaur
Student
Department of ECE,
Ramgarhia Institute of Engineering and Technology, Phagwara, Punjab
kaurharkamal27@gmail.com

Er.Manit Kapoor
HOD
Department of ECE,
Ramgarhia Institute of Engineering and Technology, Phagwara, Punjab
hodece@riet.ac.in

Dr. Naveen Dhillon
Principal
Department of ECE,
Ramgarhia Institute of Engineering and Technology, Phagwara, Punjab
principal@riet.ac.in

Abstract: The content based image retrieval (CBIR) applications have grown their popularity in the past decade with the exponential growth in the image data volumes. The social networks have aggravated the size of image data on the internet. Social network enables everyone to upload the images of one's choice, which becomes the reason behind aggregation of millions of images on the cyber space. It's not possible to query these large image databases with the ordinary methods. Hence there was a strong requirement of a smart and intelligent method to discover the similar images, which has been accomplished by using the machine learning methods. In this paper, the multivariate feature descriptor method has been presented to extract the required and relevant information from the large image databases. The proposed multivariate method involves the image color and texture for the purpose of image matching to the query image (also known as a reference image). The most matching entities are returned as the final results by the image extraction method. There are five methods, which involves three singular feature and two multivariate feature based models, have been implemented. The multivariate model has been found much stable and returned the maximum accuracy under this model.

Keywords: Multivariate Feature, CBIR, Machine Learning, Image Synthesis, Texture Features.

INTRODUCTION

Now day's researcher works on the searching of similar images from the data base. A very promising approach is content based image retrieval (CBIR). In such systems, images are typically represented by approximations of their contents. This is called feature extraction aims to get similar information of images. CBIR will be a system to get similar pictures from picture database. The principle objective of CBIR is to give acceptable effectiveness through picture indexing and by retrieving. A standout amongst those principle assignments to CBIR frameworks will be comparability comparison; extracting characteristic and through this way, observing and stock arrangement of all instrumentation may be enhanced. These offers need aid really constantly utilized Likewise the representational of the image, which needs aid utilized for figuring the match for different portraits introduces in the database. A picture is matched based of the classification of the images.

CBIR Image Retrieval

To search similar images from the data base is now active research area. A very promising approach is content based image retrieval (CBIR). In such systems, images are typically represented by color, texture and shape. This so-called feature extraction aims at extracting information that is semantically meaningful. CBIR will be a system to get same image from database. The principle objectives of CBIR give acceptable effectiveness through picture indexing and by retrieving. A standout amongst those principle assignments to CBIR frameworks will be comparability comparison; extracting characteristic and so forth throughout this way, observing and stock arrangement of all instrumentation may be enhanced. A picture is matched on separate portraits by measuring the qualification between their comparing offers.

Motivation of This Work

CBIR frameworks would present an exploration of advanced pictures including two well-known problems: that semantic hole and the computational load on a deal with substantial document collections. The semantic hole may be the absence of fortuitous event between those data that you quit offering on that one can extricate starting with the visual information and the understanding g that that same information brings to a client for a provided for the circumstance. On the different side, that count load, when expansive picture collections would manage, might aggravate illogical utilization of CBIR frameworks.

The aim of the dissertation is to propose a new CBIR system; an important task of the system is:

1) To reduce the "semantic gap" between low-level image features.
2) To reduce the overall retrieval time.

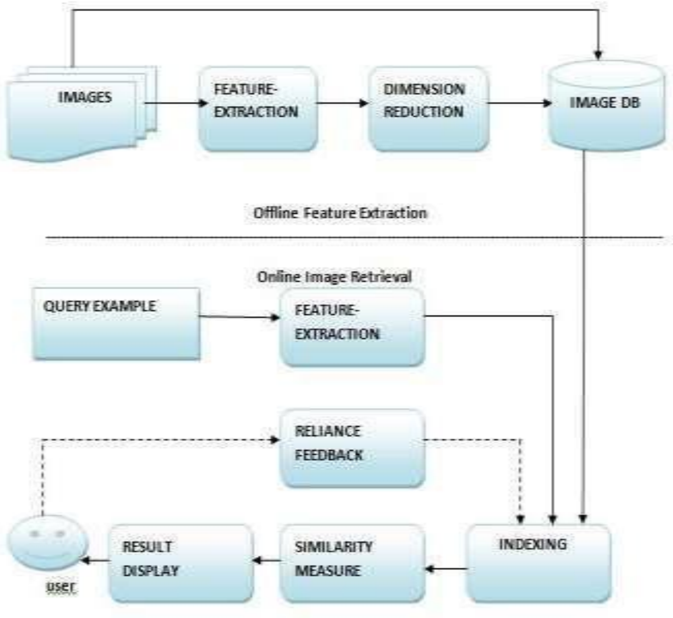

Figure1 Framework of Content based Image Retrieval

LITERATURE SURVEY

Gudivada et al. [8] proposed that images are being generated at an ever-increasing rate by sources such as defense, civilian satellites, military reconnaissance, surveillance flights, fingerprinting, scientific experiments, biomedical imaging, and home entertainment systems. Song et al. [22] proposed the cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Our techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the un-trusted server cannot learn anything about the plaintext when only given the cipher-text; they provide query isolation for searches, meaning that the un-trusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching. Curtmola et al. [5] presented a per-keyword index construction, where each entry of the whole hash table index contains the trapdoor for a keyword and an encrypted set of file identifiers. According to this searchable symmetric encryption allows a party to outsource the storage of its data to another party in a private manner while maintaining the ability to selectively search for it. Bellare et al. [3] presented as-strong-as-possible definitions of privacy, and constructions achieving them, for public-key encryption schemes where the encryption algorithm is *deterministic*. Khapli et al. [12] suggested that the emergence of multimedia, the availability of large digital archives, and the rapid growth of World Wide Web (WWW) have recently attracted research efforts in providing tools for effective retrieval of image database. Perronnin, Florent et. al. [16] has worked on the large-scale image retrieval with compressed fisher vectors.

The authors have proposed to use as an alternative the Fisher kernel framework. We first show why the Fisher representation is well-suited to the retrieval problem: it describes an image by what makes it different from other images. One drawback of the Fisher vector is that it is high-dimensional and, as opposed to the BOV, it is dense. Xia et al. [27] described that the results could return not only the exactly matched files but also the files including the terms semantically related to the query keyword. In the proposed scheme, a corresponding file metadata is constructed for each file. Then both the encrypted metadata set and file collection are uploaded to the cloud server. With the metadata set, the cloud server builds the inverted index and constructs semantic relationship library (SRL) for the keywords set. Cao et al. and Yang et al. [28] proposed that scheme for multi-keyword ranked search, where

"Inner product similarity" is used for result ranking. This paper for the first time defines and solves the challenging problem of privacy preserving multi-keyword ranked search over encrypted cloud data. Ahmad et al. [1] described the detailed view of texture, color and shape descriptors for CBIR. A comparison of different challenges in this field is discussed. There is an exponential growth of images around and the information which they carry needs to be effectively utilized. So the authors have extracted the low level features from the image such as texture, shape, and color.

EXPERIMENTAL DESIGN

The feature extraction and classification has been performed by using the following methodology. The following methodology has been defined for the extraction of the features from the dataset images. The content based image retrieval program utilizes the extracted features for the feature classification purposes to calculate the similarity of the images to create the semantic ranking. The feature extraction algorithm works in the following manner:

Algorithm 1: SIFT Algorithm

1. Acquire the image matrix, (I) If image I am the 3-D image then covert to 2-D image

3. Compute the difference of Gaussian over the input image matrix I, dG = diffOfGaussian (I).

4. Compute the minima matrix from the difference of Gaussian matrix dG, minima = findMinima (dG)

5. Find the maxima points in the minima component, maxima = findMaxima (minima)

6. Obtain all pixel values in the maxima point locations and convert all maxima points to features of length 128 byte and return the feature matrix.

Algorithm2: Freak Algorithm

1. Acquire the image matrix, (I) If image I am the 3-D image then covert to 2-D image.
2. Mark the point of interest
3. Compute the maximum matrix point from the required point of interest.
4. Obtain all pixel values in the maxima point locations and return the feature matrix.

Algorithm 3: Training; SVM Classification with the Background Feature Extractor CBIR

1. Read the query image and load in the memory as the 3-D image matrix and convert it to the greyscale to produce the 2-D image matrix. The pre-processing methods are applied to remove/reduce the noise in the image. Perform the feature extraction over the query image to fetch the known features form the image matrix.
3. Reshape the feature descriptor data for the validation of the training set containing the feature descriptors.
4. Create the group matrix with the appropriate group ids denoted to the training samples and run the SVM over the training dataset to match the testing set to obtain the training matrix in the form of similarity and bias.
5. Then the SVM classifier is applied over the SVM for classification of given data. Return the matching sample on the basis of SVM classification information in the classification matrix and return the decision logic

RESULT ANALYSIS

SIFT: The scale invariant feature transform (SIFT) based CBIR model shows the accuracy of nearly 60%, the precision of 40% and recall of 20%. The overall accuracy is acceptable, but the precision and recall values are lying under 50% barrage, which shows the underperformance.

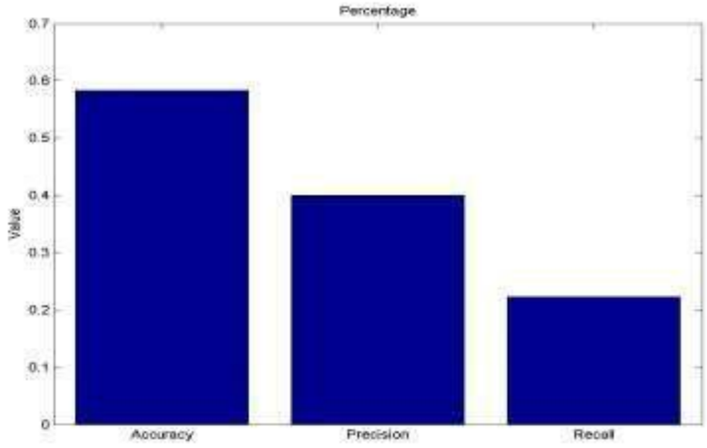

Figure 1: The Accuracy Based Analysis Of Sift Based Classification

HOG, GLCM, FREAK and AREA Calculation: It shows the accuracy of nearly 100%, the precision of 100% and recall of 100%.of assembled the overall accuracy and precision values are highly acceptable; under 50% value of recall, parameter shows the underperformance. Figure also shows that HoG has been recorded with nearly 0.90 accuracy, 0.90 precision, and 1 recall values. This shows that there is a complete absence of false negative cases under HoG, whereas the problem is aggravated by the false positive cases indicated by the extremely lower precision value. . GLCM has been recorded with nearly 0.4 accuracy, 0.35 precision, and 0.20 recall values. Freak feature based model has been observed with 0.85 accuracies, 0.90 precision, and 0.90 recall values, which clearly shows the reason for underperformance. AREA calculation has been recorded with nearly 0.90 accuracy, 1 precision, and .90 recall values

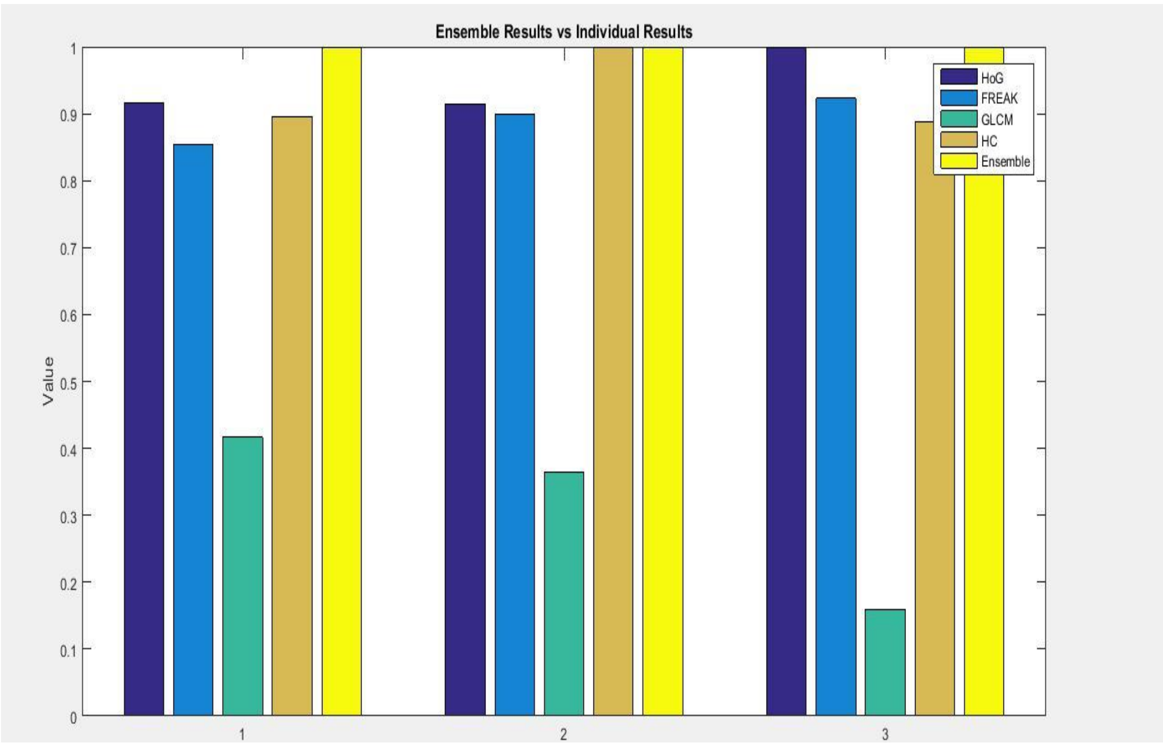

Figure 2: The Accuracy Based Analysis of Freak Based Classification

.

CONCLUSION

The proposed model has been designed by using the multivariate features, which involves the color and texture based features in order to realize the stable and accurate CBIR model. The multivariate feature combination involves the HoG, GLCM, SIFT, AREA calculation and Freak features, where HoG, GLCM implements the color oriented features, whereas SIFT,AREA & Freak gives the texture features. In the independent evaluation, the results are obtained in the form of precision, recall, and accuracy. HoG has been recorded with nearly 0.90 accuracy, 0.90 precision, and 1 recall values. This shows that there is a complete absence of false negative cases under HoG, whereas the problem is aggravated by the false positive cases indicated by the extremely lower precision value. . GLCM has been recorded with nearly 0.4 accuracy, 0.35 precision, and 0.20 recall values. The SIFT based evaluation gives the 0.60 precision and 0.40 recall, which clearly indicates the presence of both false positive and false negative cases. Freak feature based model has been observed with 0.85 accuracies, 0.90 precision, and 0.90 recall values, which clearly shows the reason for underperformance. AREA calculation has been recorded with nearly 0.90 accuracy, 1 precision, and .90 recall values. It is due to the presence of higher number false negative cases. In the proposed model, precision (0.1) and recall (1.00) proves the efficiency, which is certainly outperformed all other CBIR models under this research.

REFERENCES

1. Ballard, L., Kamara, S., & Monrose, F. (2005). Achieving efficient conjunctive keyword searches over encrypted data. In Information and Communications Security (pp. 414-426). Springer Berlin Heidelberg.

2. Bellare, M., Boldyreva, A., & O'Neill, A. (2007). Deterministic and efficiently searchable encryption. In Advances in Cryptology-CRYPTO 2007 (pp. 535-552). Springer Berlin Heidelberg.

3. Boneh, D., Di Crescenzo, G., Ostrovsky, R., & Persiano, G. (2004, January). Public key encryption with a keyword search. In Advances in Cryptology-Eurocrypt 2004 (pp. 506-522). Springer Berlin Heidelberg.

4. Boneh, D., Di Crescenzo, G., Ostrovsky, R., & Persiano, G. (2004, January). Public key encryption with a keyword search. In Advances in Cryptology-Eurocrypt 2004 (pp. 506-522). Springer Berlin Heidelberg.

5. Bringer, J., & Chabanne, H. (2012). Embedding edit distance to enable private keyword search. Human-centric Computing and Information Sciences, 2(1), 1-12

6. Cao, N., Wang, C., Li, M., Ren, K., & Lou, W. (2014). Privacy-preserving multi-keyword ranked search over encrypted cloud data. Parallel and Distributed Systems, IEEE Transactions on, 25(1), 222-233.

7. Chang, Y. C., & Mitzenmacher, M. (2005, January). Privacy preserving keyword searches on remotely encrypted data. In Applied Cryptography and Network Security (pp. 442-455). Springer Berlin Heidelberg.

8. Chuah, M., & Hu, W. (2011, June). Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data. In Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on (pp. 273-281). IEEE.

9. Chuah, M., and W. Hu. "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data." In Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on, pp. 273-281. IEEE, 2011.

10. Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2006, October). Searchable symmetric encryption: improved definitions and efficient constructions. In Proceedings of the 13th ACM conference on Computer and communications security (pp. 79 -88). ACM.

11. Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2006, October). Searchable symmetric encryption: improved definitions and efficient constructions. In Proceedings of the 13th ACM conference on Computer and communications security (pp. 79 -88). ACM.

12. Goh E-J Secure indexes. Cryptology ePrint Archive, Report 2003/216

13. Golle, P., Staddon, J., & Waters, B. (2004, January). Secure conjunctive keyword search over encrypted data. In Applied Cryptography and Network Security (pp. 31-45). Springer Berlin Heidelberg.

14. Hwang, Y. H., & Lee, P. J. (2007). Public key encryption with conjunctive keyword search and its extension to a multi -user system. In Pairing-Based Cryptography–Pairing 2007 (pp. 2-22). Springer Berlin Heidelberg.

15. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., & Lou, W. (2010, March). Fuzzy keyword search over encrypted data in cloud computing. In INFOCOM, 2010 Proceedings IEEE (pp. 1-5). IEEE.

16. Liu, C., Zhu, L., Li, L., & Tan, Y. (2011, September). A fuzzy keyword search on encrypted cloud storage data with small index. In Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on (pp. 269-273). IEEE.

17. Ren, K., Wang, C., & Wang, Q. (2012). Security challenges for the public cloud. IEEE Internet Computing, (1), 69 -73.

18. Song, D. X., Wagner, D., & Perrig, A. (2000). Practical techniques for searches on encrypted data. In Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on (pp. 44-55). IEEE.

19. Song, D. X., Wagner, D., & Perrig, A. (2000). Practical techniques for searches on encrypted data. In Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on (pp. 44-55). IEEE.

20. Song, Dawn Xiaodong, David Wagner, and Adrian Perrig. "Practical techniques for searches on encrypted data." In Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on, pp. 44-55. IEEE, 2000.

21 Stefanov, E., Papamanthou, C., & Shi, E. (2013). Practical Dynamic Searchable Encryption with Small Leakage. IACR Cryptology ePrint Archive, 2013, 832.

22. Wang, C., Cao, N., Li, J., Ren, K., & Lou, W. (2010, June). Secure ranked keyword search over encrypted cloud data. In Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on (pp. 253-262). IEEE.

23. Wang, C., Cao, N., Ren, K., & Lou, W. (2012). Enabling secure and efficient ranked keyword search over outsourced cloud data. Parallel and Distributed Systems, IEEE Transactions on, 23(8), 1467-1479.

24. Wang, C., Ren, K., Yu, S., & Urs, K. M. R. (2012, March). Achieving usable and privacy-assured similarity search over outsourced cloud data. In INFOCOM, 2012 Proceedings IEEE (pp. 451-459). IEEE.

25. Witten, I. H., Moffat, A., & Bell, T. C. (1999). Managing gigabytes: compressing and indexing documents and images. Morgan Kaufmann.

26. Yang, C., Zhang, W., Xu, J., Xu, J., & Yu, N. (2012, November). A fast privacy-preserving multi-keyword search scheme on cloud data. In Cloud and Service Computing (CSC), 2012 International Conference on (pp. 104-110). IEEE.

27. Zerr, S., Demidova, E., Olmedilla, D., Nejdl, W., Winslett, M., & Mitra, S. (2008, March). Zerber: r-confidential indexing for

distributed documents. In Proceedings of the 11th international conference on extending database technology: Advances in database technology (pp. 287-298). ACM.

28. Zerr, S., Olmedilla, D., Nejdl, W., & Siberski, W. (2009, March). Zerber+ r: Top-k retrieval from a confidential index. In Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology (pp. 439-449).

