Using Watermark-Based Protocol, Increase Robustness and Privacy of Content-Based Image Retrieval in Cloud Computing Environment

Shivaji R.Lahane[1] , Sonal H. Kunte[2]

[1] *Gokhale Education Society's, R. H. Sapat College of Engineering, Management Studies and Research, Savitribai phule pune university, Nashik - 422 005, (M.S.), INDIA*
shivajilahane@gmail.com

[2] *Gokhale Education Society's, R. H. Sapat College of Engineering, Management Studies and Research, Savitribai phule pune university, Nashik - 422 005, (M.S.), INDIA*
sonakuntes@gmail.com

ABSTRACT

Content-based image retrieval (CBIR) [1] are easy to manipulate and edit due to availability of powerful image processing and cloud computing environment. Searchable encryption (SE) scheme [11] allows image users to search over encrypted data collection. Nowadays, it is possible to do retrieval with preserving privacy [2] by encrypting it before outsourcing [5] and with copy-deterrence [3] so that unauthorized image users will not get access to it. Also tried to increase robustness of CBIR. Watermark certification authority (WCA) is a trusted agency who generate watermark-based protocol [4] for copy deterrence to avoid illegal distribution. Generating a unique watermark and directly Embedding it into the encrypted images by the cloud server before images are sent to the query user is one of the critical challenge WCA. The image integrity verification as well as generating trapdoor for retrieval in this environment is now days becoming the challenging for image users. Secure kNN algorithm [12] is used to protect feature vectors [2] (SCD, CSD, CLD, EHD) which are used to represent corresponding images. Recently hashing algorithm, auditing and alert generation are used at image users' side for verification purpose.Firstly, Zhangs' algorithm was used for encryption, decryption and generating watermark-based images. But now-a-days, watermark-based protocol is used, which improve robustness.

Keyword: - *Content-based image retrieval, copy-deterrence, Watermark certification authority, trapdoor, Secure kNN, feature vectors, watermark.*

1. INTRODUCTION

This paper describes the strategy followed by various techniques to retrieve required images. Several authors have been working in recent years on the efficient similarity search over encrypted data [5], preserving privacy [2] and copy deterrence [3] with the help of various visual descriptors [6] such as scalable color descriptor (SCD), Color structure descriptor (CSD), Color layout descriptor (CLD), Edge histogram descriptor (EHD). Indeed, it is well known that, given the different types of descriptors in practice, and availability of watermark-based algorithm, hashing algorithm and secure kNN algorithm [12], it becomes possible to obtain best result. Based on this consideration, we also used alert generation for verification. Unfortunately, the Zhangs' algorithm is not give

guaranteed that each watermark bit can be correctly extracted. So, watermark-based protocol is used to improve Zhang's watermarking algorithm to increase the robustness.

1.1 Visual descriptors in MPEG-7:

Set of descriptors are provided by description standard of multimedia content that is MPEG-7, which are used for multimedia data description. Following are MPEG-7 descriptors [1]:
1) Scalable color descriptor: In the huesaturation- value (HSV) color space, these descriptors are defined. To facilitates the scalability for the feature extraction, Haar transform encoding is used. This encoding is defined in SCD.
2) Color structure descriptor: In CSD, the small tructuring window is used to identify color distribution. To check interoperability, color structure histogram is built in the hue-min max difference (HMMD) color space.
3) Color layout descriptor: Distribution of spatial color information within images is provided by CLD. After that 64 blocks of image is created and according to discrete cosine transform, from each block CLD descriptor get extracted.
4) Edge histogram descriptor: The spatial distribution of image edges are captured by EHD. If underlying texture of the is not homogeneous then also for image matching, distribution of edges provide a good texture.

These descriptors can be represented as feature vectors. Between feature vectors Euclidean distance is used to measure image similarity.

2. REVIEW OF LITERATURE

Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun and Kui Ren all in their paper titled "A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing" focused on preserving privacy and doing copy deterrence while retrieving content based images in cloud computing environment. In this system, feature vectors are extracted to represent the corresponding images. For preserving privacy, access to unauthorized users get prevented. Watermark certification authority is provided in cloud computing environment for making images more secure [1].

Kui Ren, Zhihua Xia, Zhan Qin, and Yi Zhu, all in their paper titled Towards Privacy-preserving Content-based Image Retrieval in Cloud Computing focused on Content-based image retrieval (CBIR) applications which are developed very fastly along with the improvement in the availability, quantity and importance of images which are present in daily life. In this system, privacy is get preserved of retrieval process to control the access of images by authorized users only. There is data owner is present who send the CBIR service and image database to the cloud, without giving any idea about the original contents of the image database to the server [2].

Xinhui Wang, Zhihua Xia, Zhan Qin, Liangao Zhang, Kui Ren and Xingming Sun, all in their paper titled A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing focused on retrieval process of content based image with copy deterrence and preserving privacy in cloud computing. For preserving privacy, sensitive images, like personal and medical images, required to convert in encrypted form before outsourcing, because of this CBIR technologies present in plaintext domain becomes unusable. Moreover, secure kNN algorithm is used to protect the feature vectors, and standard stream cipher encrypt the image pixels [4].

Nasir Memon, K. Gopalakrishnan, Poorvi L. Vora, all in their paper titled Protocols for Watermark Verification focused on adding a watermark signal into the digital image which is later be detected or extracted for making an assertion about the particular image. There are two catagories of watermarks present: invisible and visible. Conspicuously company logos or visible messages are present in visible watermarks which indicates the image ownership. On the other hand, Invisible watermarks contains unobtrusive modifications to the image and the invisibly watermarked image which visually appears very similar to the original image [5].

Jens-Rainer Ohm, B. S. Manjunath, Akio Yamada and Vinod V. Vasudevan all in their paper titled Color and Texture Descriptors focused on presenting color overview and texture descriptors that get approved for the MPEG-7 standard Final Committee Draft. Histogram descriptor present in this color descriptors standard are get coded with the help of Haar transform, a dominant color descriptor, a color structure histogram, and a color layout descriptor.

These all texture descriptors contains the one which make characterization of homogeneous texture regions. It also contain another which provide representation of local edge distribution [7].

X. Wang, Z. Xia, Q. Wang and X. Sun, all in their paper titled A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data focused on a special index structure created for tree based images and uses a Greedy Depth-first Search algorithm for providing efficient ranked search of multi-keyword. For encrypting the index and query vectors, secure kNN algorithm comes in account, and this algorithm ensure accuracy of relevance score calculation in between query vectors and encrypted index [11].

P. T. Boufounos and S. Rane all in their paper titled Privacy preserving nearest neighbor methods: comparing signals without revealing them, focused on the privacy-preserving NN (PPNN) method, in which the reader will come to know that it convenient to make dividation of this in two different problems: privacy-preserving minimum finding method follow another method that is privacy-preserving distance computation. Under certain considerations, privacy model dictate that which mathematical tools should be useful for PPNN. It also define the complexity and structure of resulting protocols . These models makes assumptions upon requirement, behavior, sharing. These assumptions have main focus on participating entities behavior, the amount of possible information that get shared among participants and privacy requirements. [13].

3. SYSTEM ARCHITECTURE

Security keys and user authentication information is shared among image owner and users. Using these keys and information, encryption and decryption is done at image owner and users level. Robustness is considered while developing this system model. We are performing strong literature survey which helps in developing robustness and privacy of retrieval using various algorithms, at the same time deal with the problems. When image users get search results, hashing and auditing get performed for verifying that search results are satisfactory or not. If it is not then alert message is sent to appropriate image owner. For authorized query users, a trusted agency takes responsibility of generating watermarks and executing arbitration using watermark extraction algorithm. This trusted agency is nothing but watermark certification authority. These watermarks are get embedded in encrypted image by cloud server. Security keys and user authentication information is shared among image users and image owner which is responsible for maintaining privacy.

1. Image owner encrypts original images, index and saves them on cloud server along with user authentication information.
2. Watermark certification authority (WCA) generate watermarks and send them on cloud server.
3. Cloud server embedded watermarks into encrypted images.
4. Image users generate trapdoors and then fire required query. According to query, server reply with proper results.
5. Hashing algorithm and auditing is used for verification purpose.
6. If the results are satisfactory to image users, process terminates otherwise alert message is sent back to image owner.

This system is useful in similar face finding or similar pattern finding mechanism, as kNN algorithm is used for similarity matching. K nearest neighbor algorithm is used to map which images form database are closely similar to input query image. It is also useful in finding geographical information as per image users requirements, and if no requirements match then image user have rights to send acknowledgement to cloud.

In medical image database, for grouping patients according to their health issues, this criteria used. That means the patients having similar health problems are grouped together. After some time if any doctor want to find that patients report, the he will fire the query along with its specification for getting results.

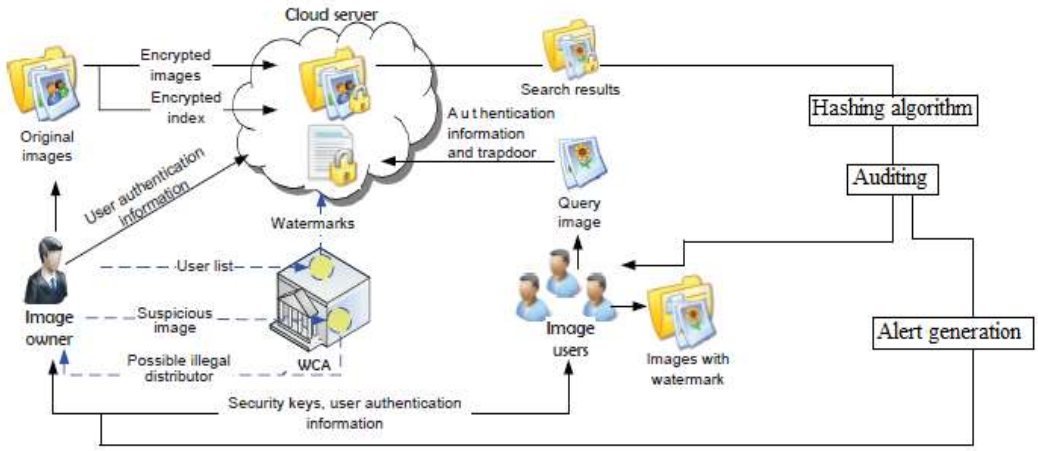

Fig -1: Framework of the content-based image retrieval using watermark-based protocol.

Steps to perform Image Encryption:
1. Generate the secret key with a one-way pseudorandom number generator.
2. For each pixel in image, denote the pixel value at position and compute bits.
3. Encrypt the pixel value using exclusive-or operator.

Steps to perform Image Decryption:
1. Obtain the secret key.
2. For each pixel in image, denote the pixel value and stream key at position and compute bits.
3. Decrypt the pixel value using exclusive-or operator.

Steps to perform Watermark Embedding:
1. Divide image into s sized non overlapping blocks.
2. For each watermark bit, the pixels in block is divided into two sets S0 and S1 according to a pseudorandom function.
3.If wi = 0, flip the bits of pixels in S0. Otherwise, flip the pixel bits in S1.

Steps to perform Watermark Extraction:
1. Divide image into s s sized non overlapping blocks.
2. Locate the set of blocks that carries the watermark bits according to the secret key.
3. For each image pixel, divide it into S0 and S1 according to secret key.
4. Fill the pixel and construct the blocks from original image.

Steps to perform Hashing Algorithm, Auditing and Alert generation:
1. Start.
2. Read data owner id(udoid).
3. If (doid udoid).
4. Stop.
5. Read file name from AWS.
6. Retrieve No. of blokes from TPA xml.
7. Select the blocks number the user want to verify.
8. Get the auxiliary information for block chal from TPA xml.
9. Based on Auxiliary information generate new root for MHT.
10. If (new root) file modified.
11. Else File not modified.
12. Stop.

4. SYSTEM ANALYSIS

This system include six main entities, image owner, watermark certification authority, cloud servers, image users, hashing algorithm and auditing and alert generation. The local data, that is, the collection of images M = {m1,m2, ...,mn} get outsourced by image owner. C = {c1, c2, ..., cn} is the encrypted form of images which is provided by image owner to server, while enabling the ability to search over the encrypted images. Only the authorized image users can access the images by firing query to cloud server. For making request, firstly the image users needs to generate trapdoor TD. After that the image users submits the user authentication information and TD to server. The secrete key is shared among image owner and image users. Image users uses this secrete key to decrypt the resulting encrypted images [6]. Watermarks get generated by watermark certification authority [5] and forwarded to cloud server. The main concentration of this system is on supporting CBIR over encrypted images and detecting illegal distributions.

4.1 Mathematical Model:

Assuming the system as S={s,e,I,O,F} Where, Initial state(s):
At client side : Image owner have collection on images which he want to store at cloud in encrypted format.
Image users have query for accessing images.
At server side : nothing
End state(e):
At client side : Image users get results and do verification.
At server side: According to query, sends original images to image users.
Input(I):
Image owner send images to server which are further get encrypted over there. Image users fire query.
Output(O):According to query resulting images get provided by cloud server.
Functions(F):
Image encryption, Watermark generation and enbedding, Image decryption, Trapdoor generation, Alert generation.

5. EXPERIMENTAL SETUP AND RESULTS

5.1 Experimental Setup:

Visual Studio Operating Environment and Dotnet Operating Environment is used for performing operations on stored image dataset. For interfacing purpose, different SOAP protocols are used at developer as well as at client side.

5.2 Results:

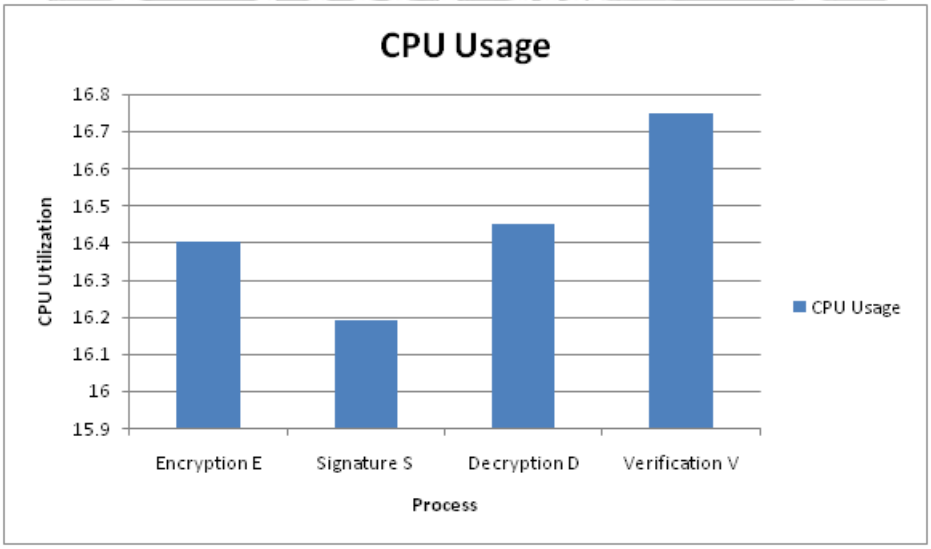

Fig- 2: CPU Utilization for Algorithms

In the above result, CPU utilization values according to step by step process for each algorithm is shown:
1. Encryption E: 16.4
2. Signature S: 16.19
3. Decryption D: 16.45
4. Verification V: 16.75

In the verification process, there is more CPU usage as the verification process contains hashing algorithm, Auditing and Alert generation. As this is the longer process compared to other algorithmm, it consumes more CPU utilization. The following figure shows, functional dependency result. If the various functions like Image Encryption, Feature Extraction, Indexing Images get performed on the original dataset, it will increase the result quality.

	Fn1	Fn2	Fn3	Fn4	Fn5	Fn6
Fn1	1	0	0	0	0	0
Fn2	0	1	0	0	0	0
Fn3	0	0	1	0	0	0
Fn4	0	0	0	1	0	0
Fn5	0	0	0	0	1	0
Fn6	0	0	0	1	1	1

Fig -3: Security and copy-deterrence increased from Fn1 to Fn6

Here:
Fn1: Datasets
Fn2: Image Upload
Fn3: Image Preprocessing
Fn4: Image Encryption
Fn5: Feature Extraction
Fn6: Indexing Images.

In this system, as the number of operations increase which are defined in mathematical model, it will provide better result. If only image uploading and image extraction is used then it will not preserve privacy because image encryption and watermark embedding is not present there. On other side, if all features get used server will provide better result.

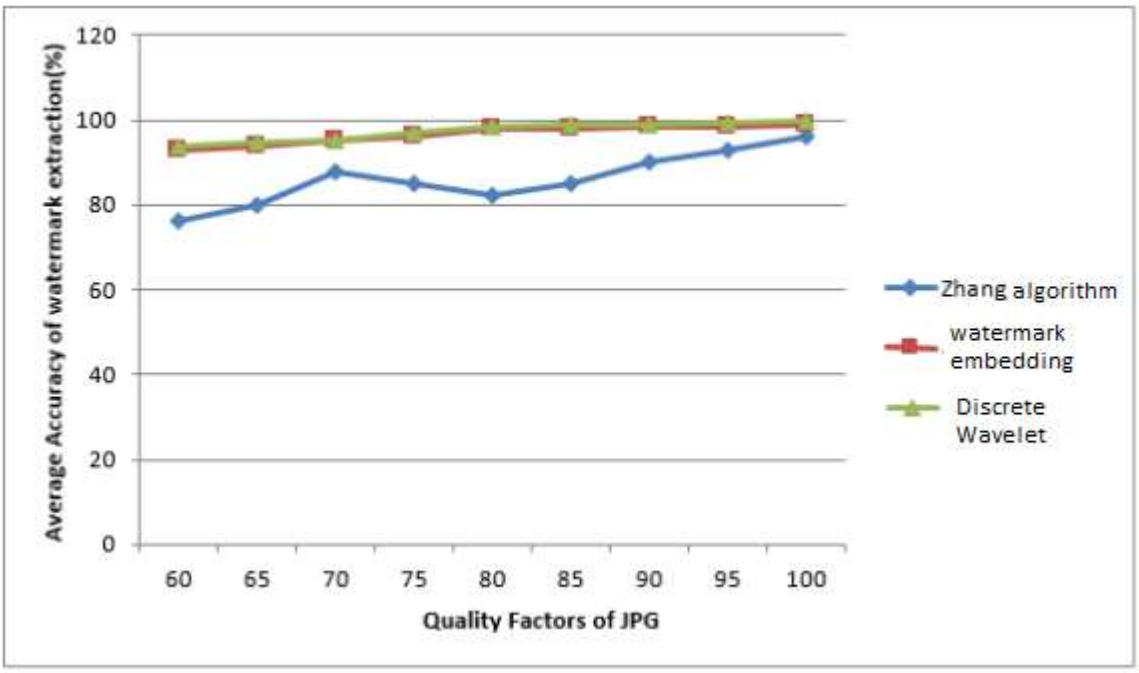

Fig -4: Average accuracy of system

Watermark size	Zhang algorithm	Watermark embedding	Discrete wavelet(proposed)
60	76	93	94
65	80	94	94.8
70	88	95	95.3
75	85	96	96.9
80	82	98.1	98.6

Fig -5: Average accuracy values of system

Fig. 4 and Fig. 5 shows the average accuracy of watermark extraction. As compared to the previous Zhang algorithm and watermark embedding, proposed system gives more accuracy. This results in improving the performance of the system.

4. CONCLUSIONS

In this paper, encryption on images is done before outsourcing and with copy deterrence, aceess from unauthorized users are get prevented. The trusted agency is used to generate watermark based protocol for avoiding illegal distribution. Unlike zhangs algorithm, which have been proved to be effective in encryption and decryption, my method is to retrieve appropriate images that are saved on cloud server by image owners. The main focus in my paper in on hashing algorithm, auditing and alert generation.

The Performance of this system based on factors like how efficiently the watermark is get embedded and extracted from the image, how smoothly the encryption and decryption process carried out without loosing the image pixels, and the rate of verification. Comparitively, previous systems does not have verification and audit generation. Because of this unavailability, the accuracy of result is less. This system focus on improving the result accuracy with minimum CPU utilization. This factors will contribute in increasing the effeciency of the system.

In short, Efficiency of project can be calculated based on

1. Time consumption of the index construction.

2. Time consumption of the trapdoor generation.
3. Time consumption of the search operation.
4. Time consumption of the watermark embedding.
5. Storage consumption of the index.
6. Watermark extraction accuracy.

5. ACKNOWLEDGEMENT

I would like to express greatfulness to P.G.Dept. of Computer Engineering, GESs R.H. Sapat C.O.E.M.S and R, Nashik.

6. REFERENCES (Font-11, Bold)

[1] Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun and Kui Ren, A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing, IEEE TRANSCATION ON INFORMATION FORENSIC AND SECURITY, vol.11, 2016, pp. 2594 - 2608.

[2] Z. Xia, Y. Zhu, X. Sun, Z. Qin, and K. Ren, Towards privacy-preserving content-based image retrieval in cloud computing, IEEE Transactions on Cloud Computing, vol. PP, no. 99, 2015, pp. 1-1.

[3] Z. Qin, J. Yan, K. Ren, C. W. Chen, and C. Wang, Towards efficient privacy-preserving image feature extraction in cloud computing, in ACM International Conference on Multimedia. ACM, 2014, pp. 497-506.

[4] Zhihua Xia, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun, and Kui Ren, A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing, IEEE TRANSCATION ON INFORMATION FORENSIC AND SECURITY, VOL. , NO. , SEPTEMBER 2016.

[5] K. Gopalakrishnan, N. Memon, and P. L. Vora, Protocols for watermark verification, IEEE MultiMedia, no. 4, pp. 66-70, 2001.

[6] M. Kuzu, M. S. Islam, and M. Kantarcioglu, Efficient similarity search over encrypted data, in Proc. of 28th International Conference on Data Engineering. IEEE, 2012, pp. 1156-1167.

[7] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 6, 2001, pp. 703-715.

[8] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, Public key encryption with keyword search, in Advances in Cryptology-Eurocrypt. Springer, 2004, pp. 506-522.

[9] M. Kuzu, M. S. Islam, and M. Kantarcioglu, Efficient similarity search over encrypted data, in Proc. of 28th International Conference on Data Engineering. IEEE, 2012, pp. 1156-1167.

[10] Z. Xia, X. Wang, X. Sun, and Q. Wang, A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data, IEEE Transactions on Parallel and Distributed Systems, vol. PP, no. 99, 2015, p. 1.

[11] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, Privacy-preserving multikeyword ranked search over encrypted cloud data, IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, 201, pp. 222-233.

[12] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, Searchable symmetric encryption: improved definitions and efficient constructions, in Proc. of 13th ACM conference on Computer and communications security. ACM, 2006, pp. 79-88.

[13] S. Rane and P. T. Boufounos, Privacy-preserving nearest neighbor methods: comparing signals without revealing them, IEEE Signal Processing Magazine, vol. 30, no. 2, 2013, pp. 18-28.