Privacy Preserving Image Pixel Prefer based Secure Data Sharing in Cloud Using Façade Watermarking Least Algorithm

M. Venkatesh and Dr.P. Ramya

Abstract--- In public cloud anyone can see data of any user but in private cloud data is secured and limited to particular user. First, large-scale image recognition requires the storage of large image dataset, and the corresponding annotation contents are also huge in scale and size. Second, the image recognition task itself is computation intensive and consumes a lot of time and energy. Smart campus images can be stored in cloud servers and image features can also be extracted through cloud computing to alleviate storage and computation problems. Nevertheless, there are many challenges that need to be solved, such as the security problem of images in cloud computing in this paper, we proposed privacy-preserving image pixel selecting for secure cloud data sharing and cross domain access control is used to access the data efficiently. In this method the security system is based on selecting the accurate pixel in shown image. In cross domain access control to access the data efficiently the multi matching is performed. The façade watermarking least algorithm (FWL) used is to compress compressed bit stream to form a full resolution in the pixel domain. The proposed used to protect the data when process is while storing the data's in cloud with the image based on pixel which selected by the transmitter.

Keywords--- Image Pixel, Cross Domain Access, Façade Watermarking Least Algorithm, Pixel Domain.

I. INTRODUCTION

Watermarking of an image is a challenging task since any tempering or unauthorized access can lead to wrong diagnostic conclusions. At the same time any information loss is totally unacceptable during embedding and extraction process because every minor detail has valuable information. In image watermarking, data is embedded into the cover image such that perceptibility of the cover image does not change and the embedded data can be recovered faithfully by applying extraction procedure at receiver end. Robust watermarking schemes protect the watermarking information very well against various attacks such as compression, rotation and cropping etc. Generally watermarks in transform domain are more robust than spatial domain. Day by day digital area captures devices producing more images. Almost in every field image are use and those images placed in the database and as per the user need should to be retrieve from database for that purpose content base image retrieval system use it's the one of the quick solutions. Content-Based Image Retrieval is used for retrieving the corresponding images from the database based on their feature of images which derived the image itself like texture, color and shape and domain specific like human faces and fingerprints. The retrieval on the based on the content of an image is to be more effective than the text based which is called content-based image retrieval that are used for a various application like vision techniques of computer. Traditionally, search of the images is using text, tags or keywords or annotation assigned to the image while storing into the databases. Whereas if the image which is stored in the database are not uniquely or specifically tagged or wrongly described then it's insufficient, laborious and extremely time-consuming job for search the particular image in the large set of databases.

M. Venkatesh, Department of Computer Science & Engineering, Mahendra Engineering College, Namakkal Dt.
Dr.P. Ramya, M. Tech., Ph.D., Associate Professor, Department of Computer Science & Engineering, Mahendra Engineering College, Namakkal Dt.

Content-Based Image Retrieval (CBIR) is one instance of information retrieval that applies computer vision techniques to solve problems related to searching and managing large image databases. However, most CBIR systems that aims to manage digital collections in offline database tend to use image content as query rather than considering user preference in defining the image in question and this may not be convenient especially when it is the only available mode of search as future efforts of searching for the same query image may be redundant. Hence, this project aims to make image galleries more organized by introducing a combination CBIR and TBIR-based system for more convenient offline searches through automatic generation of textual metadata by using information obtained from user input and previous retrieval results.

The advent of high-speed network bandwidths and low-cost storage devices has led to inventing the enormous amount of data. The advent of cloud computing and big data technologies have enabled the users and commercials to gain insight among this huge data and helped to discover the hidden worth among data. The data in various forms such as text, image, audio and video have emerged as the source of unfiltered voice of users over the various social networks. The invent in latest technologies in hardware devices have enabled the growth of cost-effective electronic devices leading to generation of Enormous amount of multimedia data. The giant source of multimedia data is the various Social networking sites which gave rise to the multimedia images to be shared across the posts, profiles etc. The huge amount of data generated across sites and devices are required to store in their online repositories. Storing of data is no more challenge with advent of high capacity storage devices and third-party cloud storage providers such as Instagram, Flickr etc. The users are limited in storage on their devices and need to migrate their data on to cloud storage without compromising on their privacy. The way of storing the data for quicker retrieval is the major task. The conventional method of storing multimedia images with the keyword, tags are extinct and suffers with anomalies that the images may be stored with the wrong key words and description. One more major factor is to maintain the privacy of user’s data stored in those repositories.

Content-Based Image Retrieval (CBIR) has received an intensive attention in the literature of multimedia information indexing and retrieval since this area started years ago, and consequently a broad range of techniques is proposed. Almost all kinds of image analysis techniques have been investigated in order to derive sets of meaningful features which could be useful for the description of pictorial information. First-generation content-based image retrieval systems were based on manual textual annotation to represent image content. However, this technique can only be applied to small data volumes and is limited to very narrow visual domains. In content-based image retrieval, images are automatically indexed by generating a feature vector) describing the content of the image. The similarity of the feature vectors of the query and database images is measured to retrieve the image. Retrieval can be characterized in following queries: query-by-visual sample and/or linguistic queries. In a query by visual sample, a sample image is submitted to the system, and the system searches database to find the images most 'similar' to the sample image. The similarity measure is usually taken from image features such as colour, texture and shape.

Edge detection is a fundamental tool used in most image processing applications to obtain information from the frames as a precursor step to feature extraction and object segmentation. This process detects outlines of an object and boundaries between objects and the background in the image. An edge-detection filter can also be used to improve the appearance of blurred or anti-aliased image streams. The basic edge-detection operator is a matrix area gradient operation that determines the level of variance between different pixels. The edge-detection operator is calculated by forming a matrix centered on a pixel chosen as the center of the matrix area. If the value of this matrix area is above a given threshold, then the middle pixel is classified as an edge. Examples of gradient-based edge detectors are Roberts, Prewitt, and Sobel operators.

All the gradient-based algorithms have kernel operators that calculate the strength of the slope in directions, which are orthogonal to each other, commonly vertical and horizontal. Later, the contributions of the different components of the slopes are combined to give the total value of the edge strength. The Prewitt operator measures two components.

Color Feature Extraction

Color, apart from being one of the most widely used visual features in content-based image retrieval, is relatively robust and simple to represent. Various studies of color perception and color spaces have been proposed, in order to find color-based techniques that are more closely aligned with the way that humans perceive color.

This section describes the entire indexing and retrieval process using the Discrete Cosine Transform. Here, we first extract the RGB components of the image. The RGB components separately are then divided into block sizes of 8*8each and DCT is applied on each of these blocks individually. Application of DCT causes the more significant details of the image to be transformed to the top-left corner of the DCT matrix. We then extract the 9 vectors from each block, create a resultant feature vector for that image. This process is repeated for all 1000 images in the database. During retrieval, the query image feature vector is matched with the 1000 database image feature vectors and the difference between the query and each resultant is calculated. The lesser the distance, the closer the similarity of the images.

II. RELATED WORKS

The information related to low level properties of image by physical and mathematical properties using a complex model. Images retrieval is increasing and crucial importance is given to domain specific information. Large and distributed collections of scientific, technical images are retrieved using sophisticated and precise measures of similarity and query-based semantics. Deal with images which contain complex information having dense form. Human eye can extract and understand such images after a year of training. Enhancing the power of public and the growing size in data in various forms including images have resulted in need of framework which can effectively store and process the data in parallel and also utilizing the benefits of CBIR and giant technologies such as big data and cloud computing. [1]

Content Based Image Retrieval is such a technique which uses features for searching a particular image from a database. Navigation-Pattern-based Relevance Feedback (NPRF) Approach is used. This Approach has high efficiency and effectiveness of CBIR in coping with the large-scale image data. In terms of efficiency, the iterations of feedback are reduced substantially by using the navigation patterns discovered from the user query log. It supports a large set of downloaded images. CBIR is such a technique that will ease the data handling and the user can easily access the data. The increasing amount of digitally produced images requires new methods to archive and access. The images can be retrieved using color, texture and shape. The most important feature in retrieving an image is color. There are so many methods to retrieve the color. They include color histogram, color moments etc. Color histogram is the widely used method for color feature extraction. Color histogram method doesn’t store the spatial information and also it is not invariant to scaling. [2]

There are two problems remain in this method. On the one hand, it brings too heavy workload. On the other hand, it still remains subjectivity and uncertainty. Because the image retrieval that is based on artificial notes still remains insufficiency. Content-based image retrieval (CBIR) is a technique that involves retrieving specific images from image databases primarily based on features that could be automatically derived. Typically, primitive features such as color, texture, object or shape are used, and could be combined with logical features such as the identity of the objects in the image, or even abstract attributes such as the significance or relevance of the image within a context. [3]

Content-Based Image Retrieval is also known as Query by Image Content and Content-based Visual Information Retrieval. The main goal in CBIR system is searching and finding similar images based on their content. In CBIR, retrieval of image is based on similarities in their contents like textures, colors, and shapes etc. Which are considered as the lower level features of an image. These conventional approaches for image retrieval are based on the computation of the similarity between the user's query and images. In CBIR each image stored in the database, has its features extracted and compared to the features of the query image. [4]

This method uses geometrical properties of connected component analysis to group the text components. It is observed from recent works in on scene text extraction from natural scenes and camera images that these methods assume big fonts such as in caption texts in video and high contrast texts with clear character shapes for text detection. These assumptions are valid for camera based images, but not necessarily valid for video-based images due to unfavorable properties of videos such as variation in contrast, complex background, different fonts and sizes, different orientations, perspective deformation, color bleeding, background movements. [5]

A new segmentation method is proposed by Bhowmik and Roy to segment Bangla words. They traced the lower contour of each connected component in hand-written word-image anticlockwise. During this process, relevant features are extracted and their vectors are nor-malized. Another segmentation method was proposed by Lorgio and Govin-daraju to segment Arabic handwriting. The algorithm over-segments each word and then it removes extra breakpoints using knowledge of letter shapes. Comparing methods of internal and external segmentation approaches mentioned in the literature, experiments show that external segmentation provides greater interactivity, savings of computation, and simplifies the job of the recognizer. [6]

Rhetorical structure theory posits a hierarchical structure of discourse relations between spans of text. This structure is richer than hierarchical topic segmentation, and the base level of analysis is typically more fine-grained – at the level of individual clauses. Unsupervised approaches based purely on cohesion are unlikely to succeed at this level of granularity. Elsner propose the task of conversation disentanglement from internet chat room logs. Unlike hierarchical topic segmentation, conversational threads may be disjoint, with unrelated threads interposed between two utterances from the same thread. [7]

This technique depends exclusively on word repetition to find tight regions of topic similarity. Instead of focusing on strict lexical repetition, Kozima uses a semantic network to track cohesiveness of a document in a lexical cohesion profile. This system computes the lexical cohesiveness between two words by \activating" the node for one word and observing the \activity value" at the other word after some number of iterations of \spreading activation" between nodes. The network is trained automatically using a language-specific knowledge source (a dictionary of definitions). Kozima generalizes lexical cohesiveness to apply to a window of text, and plots the cohesiveness of successive text windows in a document, identifying the valleys in the measure as segment boundaries. [8]

Bayesian topic models to text segmentation was investigated first, using HMM-like graphical models for linear segmentation. Eisenstein and Barzilay extend this work by marginalizing the language models using the Dirichlet compound multinomial distribution; this permits efficient inference to be performed directly in the space of segmentations. All of these papers consider only linear topic segmentation; we introduce multi-scale lexical cohesion, which posits that the distribution of some words changes slowly with high-level topics, while others change rapidly with lower-level subtopics. This gives a principled mechanism to model hierarchical topic segmentation. [9]

The average of pixel contours of the text box is closer to the average of the pixels marked as background on the text image. The problem of text embedded in complex background is not addressed, and performance results for segmentation or recognition are not reported. in have presented a simple text segmentation method. To agree with the possible polarity of text in an image, two segmented text regions are generated. A connected-component method is applied to the segmented result to remove the components that do not fulfill the specified aspect ratios. Finally, a score is assigned to each of the segmented images based on their text-like characteristics. [10]

Novel watermarking algorithm for medical images by randomizing the information according to the secret password key and then embedding it into the cover medical image followed by embedding of cover image into another secret key image. The proposed scheme aims to protect sensitive information transfer and cloud storage from unauthorized access. Extraction of the embedded information at the receiver end is done in two steps. Firstly, non-blind detection is carried out in which cover medical image is extracted with help of secret key image. In the second step, binary image watermark is extracted by analyzing the DCT (Discrete Cosine Transform) coefficient values. [11]

Cluster-based database watermarking technique which first applies cluster theory to the database watermarking technology. The cluster theory is used to cluster the source data and the clustering results determine the quantity of embedded watermark information and embedded position. This method makes the watermark information more disperse and hidden. This paper also introduces an odd-even modifying method to denote the watermarking information to decline the modification of original database. [12]

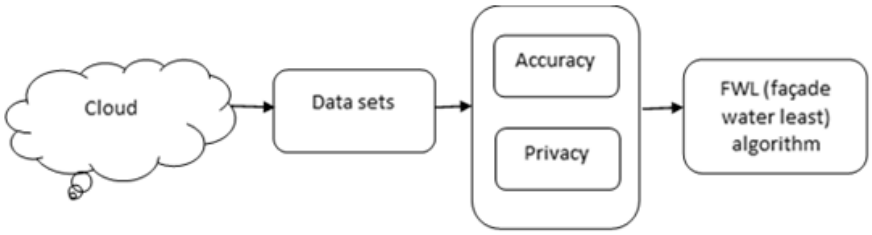
The Clients can store, transfer or exchange their data using public cloud model. The encryption method for public cloud and also the cloud service provider's verification mechanism using the third party auditors with framework model. The Cloud Data Storage is one of the mandatory services which are acquiring in this rapid development business world. [13]

Hybrid of DWT-SVD technique has been proposed to embed watermark in the image. The properties of DWT and SVD were found to be complementing each other as this method reveals good transparency and robustness. During experimenting, the PSNR values for all test images above 30dB indicate good perceptual transparency of algorithm. This technique is robust against challenging image processing attacks with NCC closer to 1 and BER closer to 0. [14]

Double digital watermarking system which is capable of embedding two different digital? Private images into one ordinary image. This new system has two main features. First, two digital private images can be? Reproduced from one digital watermark image, so twice the number of digital private images can be transmitted simultaneously as confidential information. Second, since this double digital watermark is constructed using two different characteristics of orthogonal functions, it is also possible to detect whether or not someone has altered the digital file. [15]

III. IMPLEMENTATION OF PROPOSED SYSTEM

In particular, we propose a new, simple, and efficient approach that offers the same level of data security and confidentiality in the cloud without the process of reversible data hiding. The façade watermarking least algorithm (FWL) used to compress the image via a lossless image coder in order to create space used to protect the data when process is while storing the data’s in cloud with the image based on pixel which selected by the transmitter. The proposed algorithm determines improved privacy preserving in images. By implementing the proposed algorithm we can improve the efficiency in image transferring process and also provide more security in data



Façade Watermarking Least Algorithm

STEP-1: Cover Image to Watermarked Image

Input:

1- Cover Image

2- Watermark text.

Output:

Watermarked Image

Begin

1- Check the length of the watermark text to know how many copies will be embedded in the first LSB and if it will embed in the second LSB.

2- Embedding the length of the watermark text in the first LSB.

3- Convert the watermark text from characters to bits.

4- Inverse the watermark bit.

5- Check the coordinate of X, if it is odd, the algorithm will add 1 to X, and if it is even, the algorithm will subtract 1from X.

6- Embed the watermark bit in the first LSB.

7- Go to 4 until finishing all the watermark.

8- Go to 4 if we need to embed another copy of the watermark text.

9- Save the Image as bitmap image

End

STEP-2: Watermarked Image to Watermark text

Input:

Watermarked Image.

Output:

Watermark text.

Begin:

1- Get the length of the watermark text from the first LSB.

2- The user can choose which copy he wants if there is morethan one copy.

3- Check the coordinate of X, if it is odd, the algorithm willadd 1 to X, and if it is even, the algorithm will subtract 1from X.

4- Get the bit from the first LSB.

5- Converse the bit and save it in array.

6- Go to 3 until finishing all the watermark text.

7- Convert the array to characters.

End

Advantages

- It improved more privacy.
- It provides highly effective and highly reliability.

IV. RESULTS AND DISCUSSIONS

This simulated result is shown by the response time reference for the identity process time and the façade watermarking least algorithm of security. From this, the proposed method shows the updated time of accessing data from the server and the time delay.

4.1 Comparison of Cloud Security

It has cloud data and the proposed project has been tested. Data Sharing System All files are adjusted to adjust the data processing.

Table 4.1: Comparison of Data Partition

Algorithm	Partition in %
FWL	70
RSA	50
DES	35

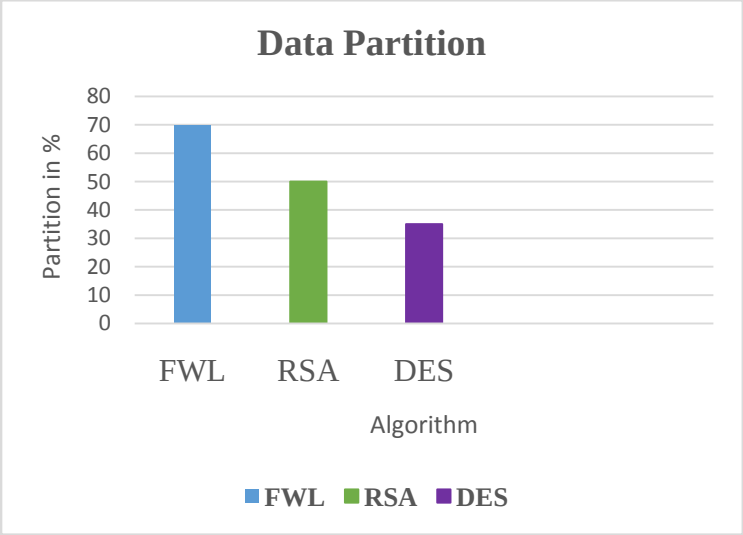


Chart 4.1: Comparison of the Data Partition

Chart 4.1 describes the shows Comparative results can be generated by different methods of network sharing security in cloud. This comparison results show that the proposed method creates very little differentiating than other ways.

4.2 Processing Time

Data processing time is named the run time, which is reading all parameters to the end result of the information. This parameter suggested the operating system run time ratio.

Table 4.2: Processing Time

	Data 1 Time (ms)	Data 2 Time (ms)	Data 3 Time (ms)
FWL	6	4.4	5
RSA	4	3.8	3.9
DES	5	4.9	4.8

In the above table shows the values of processing time

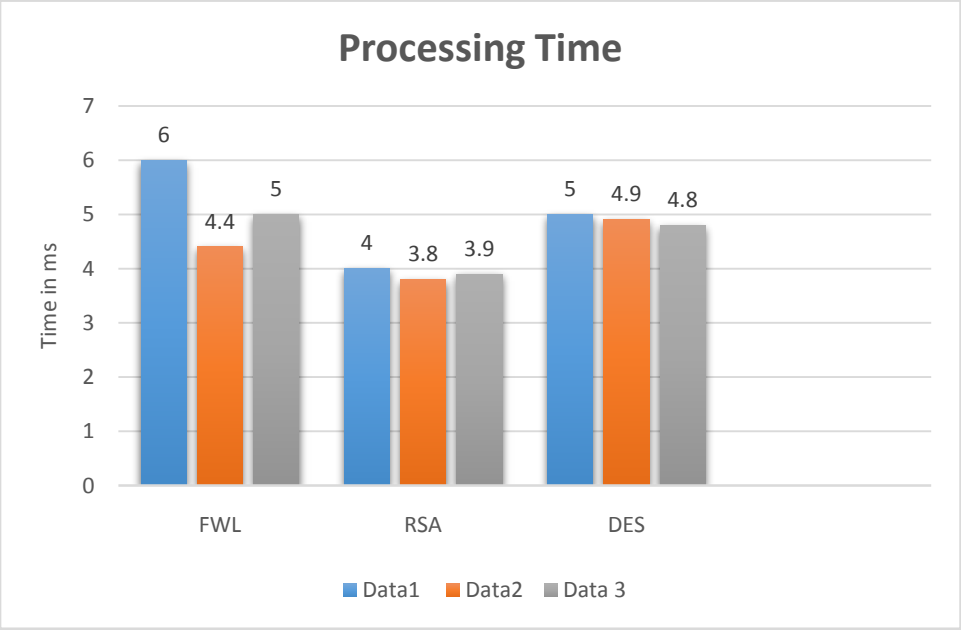


Chart 4.2: Processing Time

Chart 4.2 describes Continuously arranged for organizational processing time ratios.

4.3. Evaluation of Data

The determination of the useless data currently interrupted or ended up differently. It is possible to use the rating for more information on the use of the frame to re-install the application.

Table 4.3: Evaluation of Data

Time Interval (MS)	FWL in %	RSA in %	DES in%
5	14	21	35
10	28	38	44
15	38	48	56
20	47	63	67
25	63	64	72

Table 4.3 describes the particulars of dataset recycled to an evaluation of the projected method.

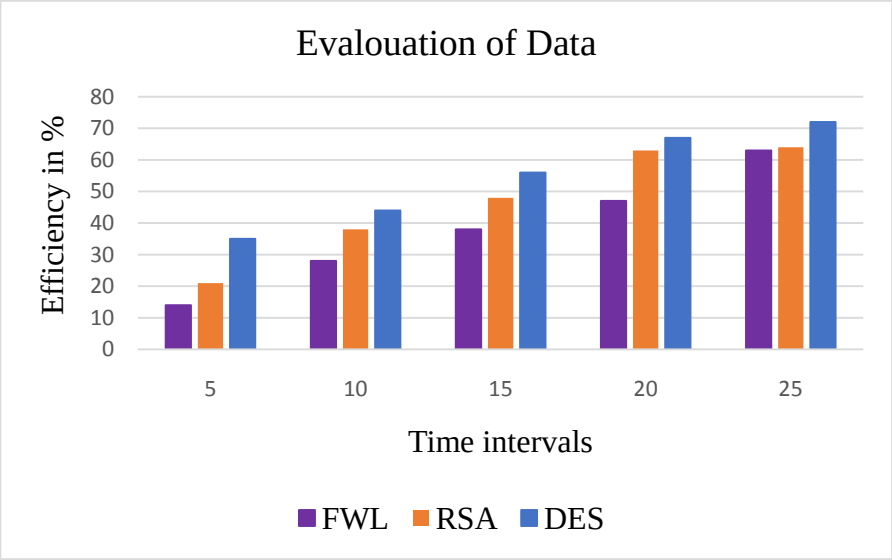


Chart 4.3: Evaluation of Data

Chart 4.3 describes the evaluation of data is prepared in different formats, indicates that it has reduced the buffer position at the time interval of the proposed project.

REFERENCES

[1] R. Shetty, "A Big Data Approach towards Web Mining Using CBIR", International Journal of Emerging Technology in Computer Science & Electronics, Vol. 23, Pp. 1-3, 2013.

[2] J. Salomia Brigitha, "Survey on Content Based Image Retrieval Technique", International Journal Of Advanced Research In Computer Science And Software Engineering, Vol. 6, Pp. 1-5, 2016.

[3] B. Kusuma and M.P. Arakeri, "Survey on Content Based Image Retrieval Using Map Reduce Over Hadoop", International Journal of Advances In Electronics And Computer Science, Pp. 1-3, 2015.

[4] A. Patel and S. Tanna, "Improving Performance in Terms of Computation Time In CBIR Using Hadoop Map-Reduce", International Journal of Advance Research and Innovative Ideas in Education, Vol. 2, Pp. 1-8, 2013.

[5] J. Kwang I. Kim and A.K. Jain, "Text Information Extraction In Images And Video: A Survey", Elsevier, Pattern Recognition, Vol. 37, No. 5, Pp. 977–997, 2004.

[6] P.J. Sushma, "Text Detection in Color Images", International Conference on Intelligent Agent & Multi-Agent Systems, Pp. 1-6, 2009.

[7] X.W. Zhang, X.B. Zheng and Z.J. Weng, "Text extraction algorithm under background image using wavelet transforms", International Conference on Wavelet Analysis and Pattern Recognition, Vol. 1, Pp. 200-204, 2008.

[8] X. Liu and J. Samarabandu, "Multiscale Edgebase Text Extraction from Complex Images", IEEE Multimedia And Expo, International Conf., Tronto, Canada, Pp. 1721-1724, 2006.

[9] R. Farhoodi and S. Kasaei, "Text Segmentation from Images with Textured and Colored Background", 13th Iranian Conference on Electrical Engineering. Zanjan, Iran.

[10] C.W. Liang and P.Y. Chen, "DWT based text localization", International Journal of Applied Science and Engineering, Vol. 2, No. 1, Pp. 105-116, 2004.

[11] D.S. Chauhan, A. Adarsh, B. Kumar, R. Gupta and J.P. Saini, "Double secret key based medical image watermarking for secure telemedicine in cloud environment", 40th International Conference on Telecommunications and Signal Processing (TSP), Pp. 626-631, 2017.

[12] K. Huang, M. Yue, P. Chen, Y. He and X. Chen, "A cluster-based watermarking technique for relational database", In First International Workshop on Database Technology and Applications, Pp. 107-110, 2009.

[13] D. Boopathy and M. Sundaresan, "Data encryption framework model with watermark security for data storage in public cloud model", International Conference on Computing for Sustainable Global Development (INDIACom), Pp. 903-907, 2014.

[14] A. Siddiqui and A. Kaur, "A secure and robust image watermarking system using wavelet domain", 7th international conference on cloud computing, data science & engineering-confluence, Pp. 599-604, 2017.

[15] T. Sasaki and T. Nagase, "Constructing Digital Watermark Based on Orthogonal Functions", 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), Pp. 140-143, 2018.