International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 12, No.1 (Dec-2022)

https://dx.doi.org/10.12785/ijcds/1201106

A Comprehensive Study on Image Retrieval Algorithms of Cloud Storage for Information Extraction in Health Care System

Vasudeva.R[1] and Dr Chandrashekara S.N[2]

[1]*Department of Computer Science and Engineering, C Byregowda Institute of Technology, Kolar, Karnataka, India*
[2]*Department of Computer Science and Engineering, C Byregowda Institute of Technology, Kolar, Karnataka, India*

Received 17 Jul. 2021, Revised 21 Sep. 2022, Accepted 1 Dec. 2022, Published 8 Dec. 2022

Abstract: The world is full of digital data disseminated everywhere for creating, accessing, and storing information among the people's communication through media with network devices. The huge volume of data stored as a repository in a remote place where it is accessed in terms of datasets/databases. Such kinds of databases are extracted from the repositories in a short period based on its features and contents available inside the databases. The processing of transformation of data from one form to another using the above-said functions of a database plays a vital role in extracting information on large networks. Information Retrieval (IR) is the normal technique to derive the information from the warehouses through its features and functionalities. Cloud computing is the trending technique to store the accessed information and converted it into meaningful data as a database and stored in a remote place among the nodes of networks. The stored databases are collected and combined as a group of data based on the features present in the database then stored as clusters. Separate clusters are created in the cloud to store text, audio, video, and multimedia supported files remotely from different sources on the internet. From those clusters, extracting the accurate files needed should be processed through various classical algorithms and techniques with the supported tools. Content-Based Image Retrieval (CBIR) is the common popular technique used to get accurate results as an output from the cloud clusters in a short duration of time. The latency and efficiency will be very low to process the huge volume of data on the network. Classical approaches implemented innovative ideas such as query-based approach, features extraction, auto-encoders, and indexing on this CBIR technique and applied them through machine learning algorithms, CNN (Convolution Neural Networks), and deep learning. Since cloud computing is working with the internet as an important phenomenon all the techniques are needed for internet sources. The accuracy and latency of the CBIR technique on cloud storage servers is a challenging mission for people who are needed in different sectors, especially in healthcare. To find better solutions for the above-said problem this survey paper has been written from a technical perspective view, and helpful to the researchers get motivated to do their innovative works, and ideas on the CBIR model to emphasize in the competitive world.

Keywords: Information Retrieval, Cloud Computing, Content Based Image Retrieval, Machine Learning, Convolution Neural Networks.

1. Introduction

Content-Based Image Retrieval (CBIR) is the popular technique used in cloud computing for extracting the images from the repository on the cloud storage. The storage is done in a cluster form on cloud computing with special characteristics of the images collected from different internet sources. The images are generated from various users located at different places and all are connected through cloud performance via an internet source [1]. Cloud computing is a platform to communicate all the data stored on repositories and accessed them by sitting anywhere, anytime, anyone method. All the images are stored as databases that have their features and characters in nature for accessing them remotely with the cloud platform. When the size of the images is varied, then clusters working on a cloud are differentiating those while creating repositories. So the images that have been stored on the clusters of the cloud will act as a client-server architecture principle. The server is providing the access to all the clients on the request-response technique based, on the network modes [2]. The information Retrieval concept is used to extracting the information from the repositories based on the multimedia files supported to create data. Text-based information retrieval has developed to access data from the clusters but accuracy and latency constraints are not satisfied by them. Generally, CBIR is a concept used to extract images from the cloud servers using queries written on the Data Manipulation Languages (DML) such as Oracle, SQL techniques. The corresponding

E-mail address: vasudev.rama@gmail.com, snc_chandru@yahoo.co.in

images on the internet are extracted based on features of the images stored in the cloud servers. For that Feature Extraction techniques and algorithms are working on cloud storage [3]. The image selected for extraction, their features are identified and selected for wiring query-based approach by creating vectors. If the data sets are huge volumes in nature, and it is predicted through their values present on the vectors. When both are matched with a similarity of the images from the internet and repositories can be derived immediately to improve the latency on the network node computers. The similarity has founded in the images was done by features such as resolution, size, clarity, and type of the images created on the internet sources [4]. The entire process of CBIR has done using a lot of classical approach algorithms but the extraction time and accuracy are still ha challenges and; limitations to access them on a centralized huge cloud server on the network, The internet speed is also a factor for delay and latency problems in CBIR, but it reflects the contents available in the cluster's features [5]. The latest researches are said that CBIR quality can be improved with the help of image properties available on internet sources. CBIR is working with different approaches for extraction of the images [6].

The Figure 1 represents the components of CBIR technique to extract the information as images from the internet accurately.

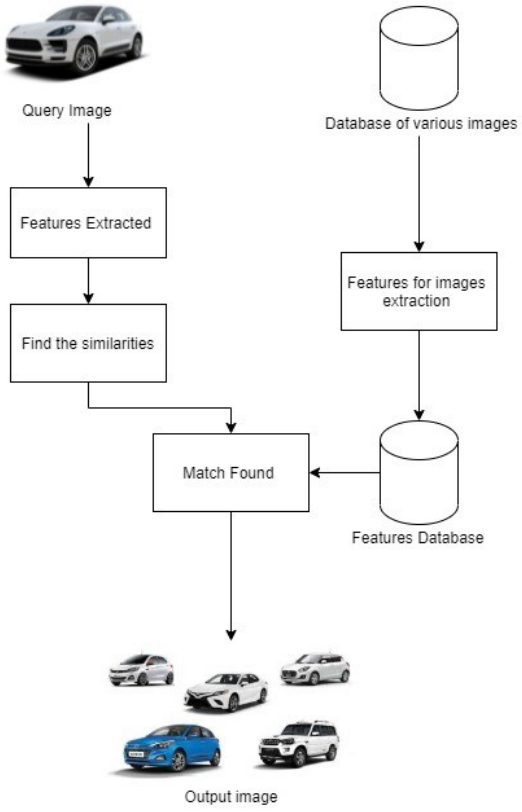

Figure 1. Components of CBIR

The Figure 1 represents the CBIR techniques phases to perform image retrieval from the repositories. Reference image as query image has to give as an input to find the number of related images from the repositories within a time. For that their features are extracted from the images like size, resolution, formats. In the repositories, algorithms are used for feature extractions through the images for finding the best matches among them. At last, the matching function has to be checked with several criteria of image and nearest feature images are classified at the initial level. Next level the exact match has found based on the features present in that image accurately. When the dataset size is huge in repositories more features have been selected for consideration of the retrieval process in less latency.

Several inputs are decided the features extracted quality when the number of outputs is similar to the maximum level. In other words, more features are present in the input image, which will give more output images from these features for CBIR. Feature extraction is the backbone of the CBIR process especially working on the clusters which are connected in the large network devices.

In this article, summarized and discussed the CBIR methodologies in Section 2, and then related works are done for CBIR in cloud computing storage and the research gap and solutions of CBIR techniques used in cloud computing storage have discussed elaborately in Section 3. Afterward, in Section 5 techniques used to perform CBIR approaches and their performances are explained and in Section 6 conclusion followed by Section 7 with future enhancement will discuss.

2. Methodologies used in CBIR

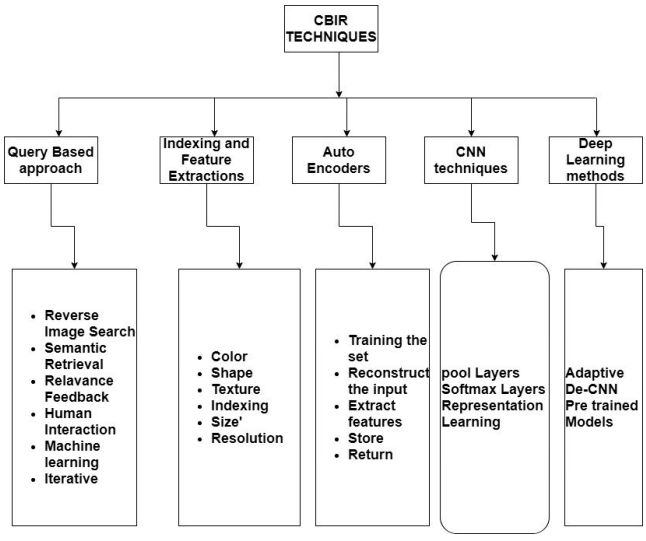

Figure 2. Techniques used in CBIR

CBIR is the familiar approach in cloud computing to retrieve images with high throughput and low latency. Unfortunately, the time which is used to extract the images

from the repositories is very high due to the size of the image or formats of the image vice versa. Classical approaches help in this CBIR method to increase the speed in the repositories using to find the features of the image and apply them with an algorithm for extraction. That was working in a small set of files effectively, but not for huge datasets. If multi-node clusters are present in the network then the time to retrieve the image from that cluster will make high latency and not accurate too. CBIR is the retrieval technique commonly used for extracting all multimedia files from the internet sources [7]. The Figure 2 represents the techniques used in CBIR approach elaborately.

The Figure 2 illustrates the techniques used in the CBIR approach and their methodologies also. It has to be classified into main five categories based on the concepts followed from the beginning itself. The sub-components of all the techniques are further classified into several types for developing the CBIR approach with high throughput and low latency levels. All methods are working with the metadata of the images created and generated from the internet sources; literally, they all connected with cloud server through internet technology. They are [8]

- Query based approach
- Feature Extraction and Indexing
- Auto encoders
- CNN techniques
- Deep learning

A. *Query Based Approach*

From Figure 1, the query image has to be selected as input to the CBIR, and the reference image from the database is retrieved from written a query to extract from it using DML / SQL. The accuracy and latency of this approach are not what the users expected on the level. But the output that they got is somewhat relevant to the input image matched on its features represented.

Reverse Image Search: In this method location of the source, image is identified, and find out the higher resolution of that image using any searching techniques. To do this effectively we have to find out the web page's where that images occur on the internet. Tracking the content creator of that image will help to identify the exact image during retrieval. All the information about that image is being collected and stored in a server for future classifications. Only relevant images are rectified and extracted from this type of approach. The best example for this is Google search images and Tin Eye search images.

Retrieval: In this method instead of letters, arguments have passed through the searching system and get the relevant images. Through the query-based searching mechanism, a written query is searching the images based on the color, shape, and texture of the image. The high–level concepts are found by the human feedback method in image retrieval techniques. The best example to explain this type of approach is "find the pictures of carrot" which is the argument passed to search the images.

Feedback and Human Interaction: Three categories got from the customer's point of view from the internet named as relevant, not relevant, and neutral. If the searching content was not identified by the keywords that we were used then type the new keyword until we get the exact image. This is the best method to get the users intent for extracting the accurate image.

Iterative and Machine Learning approaches: This approach is used to get the original image from the cluster on the higher networks using the repeated search or keywords available on the internet. The user got satisfied with the search and the last keyword is used to store and utilized for future searches would be easy in terms of time. Machine learning concepts are used to create a trained set of images for particular types that would be practiced to get the exact output from the system through training.

B. *Feature Extraction and Indexing*

The features of the image like color shape and texture have been taken into consideration of a selected image and their values are helped to find the exact image. These three factors are the main criteria if the image like RGB color components [9] in the image.

Color: Histograms are used to find the color quantization, and these will be used to make difference between two images based on the distance and number of histograms used. If two same images are taken for consideration to identify the exact image then will get confused due to the same color. At that time these histograms are used to make differentiate through distances in the color combinations through an indexing method. The entire image has to be divided into a number of subsets for easy recovery of color combo's and their distances.

Shape: The image edges and contours are noted for retrieval of the image in a cluster. Metric-based concepts like k-dimensional techniques are used to find the edges and contours as row and column. The skeletal representation of the image has to verify with all edges for accuracy. The extracted edges of the directional histograms are suggested for image matching concepts.

Texture: The visual patterns of the images are having homogeneity or the arrangements of their properties. The pixels and clarity of the images are calculated for the exact image retrieval. The nearest or closed images are predicted using these textures with the help of the brightness of the images stored. The query has written using these factors should retrieve the exact image.

Size and Resolution: The feature vector representation is

used to find the features of the image from the repositories and help to create vectors. Using mathematical formulas in the vector-based calculation it is easy to find the exact images. To match the images fuzzy-based reasoning [10] is used. The Figure 3 will explain all the concepts used in feature extracting techniques.

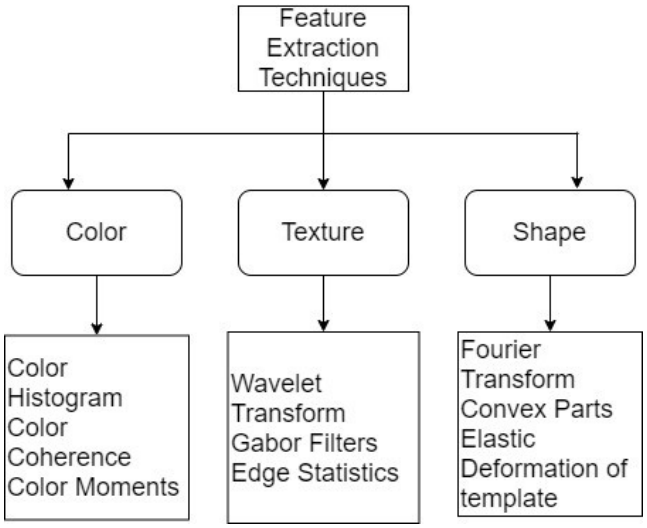

Figure 3. Feature Extraction techniques

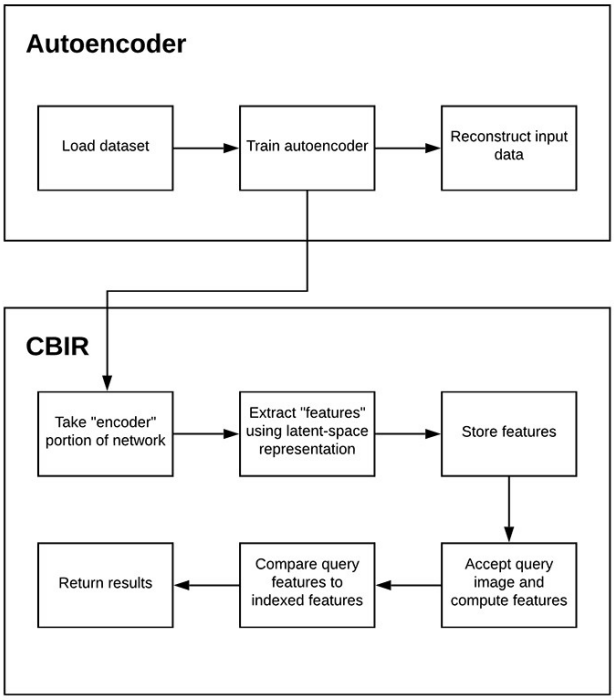

Figure 4. Auto encoder approach

Image Indexing: This could be achieved in the form of keywords related to descriptive meta data used. The retrieval keys are used to find the image on the cluster. New descriptors are added by the users whenever required.

C. Auto Encoders

Auto encoders are used to retrieve the image from the dataset using trained input dataset with dimensionally reduction and network trained system used for classifications. The input is reconstructed using the trained data set available on the network and it will be given to the CBIR approach for further processing. Their features are taken from various data sets of trained input and it will match the exact image on the network by indexed features. The final image features are stored as values and then returned from the latent-space representation technique. This technique is used to separate the network in to number of portions and then that portions are taken for consideration of retrieval processing. At the end, all the portions final result will be merged to get a original result based on its features on a huge databases [11] The Figure 4 will explain the methods of auto encoder approach in CBIR.

D. CNN Techniques

The latest trend ANN is used to get original images from the warehouse with input, hidden, and output layer concepts. The input image is taken for consideration and it will be accessed through CNN layers namely the pooling layer for finding the features of the images in real-time. The layers which are fully connected to the hidden layers give the trained set of all input images to find the exact one on the network accurately. To do the feature extraction unit successfully softmax layer is used to find the representations of the image. Finally, the exact pone will be retrieved from the huge database [12]. The Figure 5 will explain the process of CNN techniques used in CBIR approach.

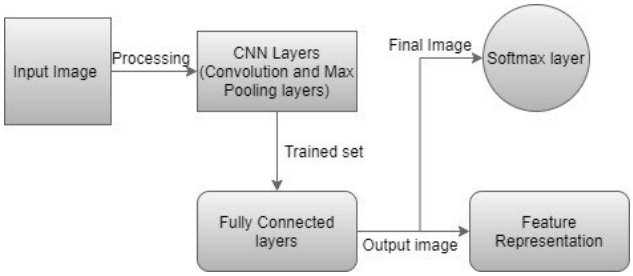

Figure 5. CNN used in CBIR

E. Deep Learning models

Artificial Neural Networks (ANN) are used to train the data sets of the collected images from the internet and stored

in a remote place for extraction. In CNN, pooling layers and soft-max layers are used to find the features of the images so that it could be easy to find the image [13]. In deep learning, more data sets are considered, and created several sub-data sets as a trained set. Different models are used to predict the exact output on the network for accurate output.

Adaptive De-CNN models: This method is used to predict more object classifiers for determining the feature extractions of the image. Multiple alternating layers used in CNN has taken into consideration as input layer images will be classified using object classifiers. More filters are used to rectify the features of the image to classify the right image on the network. It will introduce new concepts like deep layers for decision-making at the last layer. So if all images are extracted from the internet and will be taken for the final result is processed using this deep layer to avoid conflicts in the results.

Pre-trained Models:The pre-trained neural model is used to predict the weights of each layer such as input, hidden, pooling, soft-max, and output layer. The features extracted values are stored in a stack for verification and future process will be accessed from this trained data stored on the stack. Representation learning is the concept of using features of the images to put in the hidden layer when it is working online for the betterment of the results. Finally, user feedback is taken for the final output which is exactly connected to this learning. The Figure 6 is representing the deep learning representation learning.

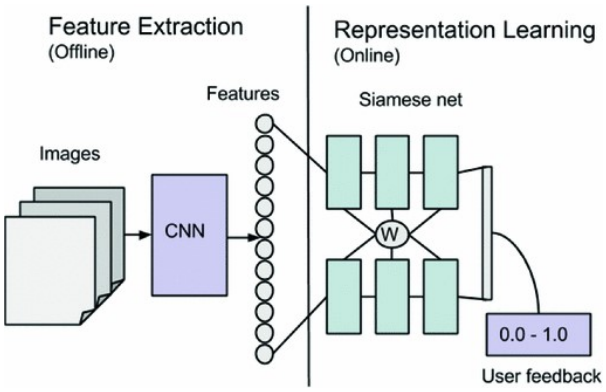

Figure 6. Deep Learning techniques for CBIR

3. Related Works in CBIR

Bunava M et al (2021) proposed a method Gray-Level Co- Occurrence-Matrix (GLCM) for finding the features of the frames from the video contents. Also, they proposed a Local Binary Pattern (LBP) to identify the features of the image as a pattern and stored them in a separate place for future representations. ANN is used to do all these works and it has improved the features based on Histogram of Oriented Gradients (HOG) techniques. Query by Image Content (QBIC) develops for implementing features of the images by ANN-based Pigeon Inspiration Based Optimization (PIO) classifiers. The measurements have been taken from the databases using f-measure concepts to declare the features of the image.

Zheng. P et al (2021) proposed a Hyper Spectral Image (HIS) cloud-based repository for the collection of unmixed information about the spectral images. It is mainly used to perform high scale images of big data repository where it was collected from the spatial-spectral repositories. Web interfaces have been created between the repository and databases for developing software as a service management platform to the users. This platform is used to retrieve the images from the repository based on HIS information on the clusters.

Li Y et al (2021) proposed a method for developing Remote Sensing (RS) image retrieval mechanism using the geo-localization concept and fusion-oriented processing of images. This method is helping to rescue the disaster-based problems from the images on the large-scale database. The mining process has been created for recovery purposes from the huge scale repositories of RS type of images on a cluster. This was developed based on mining algorithms by data pre-processing techniques followed on the repositories.

Singh. P (2021) proposed a method for feature reduction techniques used with deep learning concepts with the help of pooling layers worked on ANN. Feature representation learning is used to predict the features of the images from the layers. Repeated information is measured and feature mapping has to be developed for removing the repeated images on the clusters and similarity partial feature images are categorized by the classifiers available in the system. It is used to find the region of interest on the image. Image descriptors are used for image retrieval in the larger databases with the utilization of facial image classification. All the changes have incorporated the properties of the images to their original parameters.

Haq. N et al (2021) proposed the concept for image retrieval using a deep community-based automated medical images framework from the huge X-ray images available on the internet. It is used to develop the approach of image network generation and detection techniques with the help of deep learning concepts. This approach is used to extract similar images from the X-ray images repository through community clusters categorized on the network.

Lin. F et al (2020) proposed a method for video retrieval systems using deep learning based on cloud computing. It is working with accuracy and recognition time perfectly. For developing this deep neural networks are used to remove the blurred images and the alignment of the face in the videos is implemented. Several network architecture video sets like VGGFace, FaceNet have been used in this method to implement the significant variations of the face images. Normally facial expressions and their illuminations along with their conditions are the challenging tasks in video

retrieval whereas this deep learning neural network was implemented to extract the right information.

Yan H et al (2020) proposed a method to identify the video retrieval techniques using homomorphic encryption. The encryption in cloud storage is the biggest challenge in cloud computing technology only for video retrieval systems. The vector mechanism has been extended using the Order-Preserving Encryption (OPE) technique to retrieve the images from the video frames. Homomorphic Encryption (HE) gives security to the cloud-based videos is stored in a different cluster in real-time. A huge amount of volume is occupied by videos when compared with the other file systems available in a cloud.

Devaraj A et al (2020) proposed a network model for image retrieval system schemes Adagrad based CNN and Secure Multiple Share Creation (SMSC). It is used to find the image identification number from the feature extraction techniques. Image archival and query image patterns are used to do this with the help of deep learning concepts Core10k data set. In a cloud platform, minimum performance retrieval from large-scale IR has done using this approach with the help of SIARS. Multi-share creations are used to predict the exact output of the image from the various sources on the internet and properly trained them for the retrieval process.

Li J et al (2020) proposed a method called ASPE (Asymmetric Scalar Product Preserving Encryption) and the combination of Homomorphic Encryption for the image retrieval concepts. It has developed the concept of checking the search query to find whether the query is fake or copied from other sources. Search precision will be implemented through the semi-trusted computations and overhead in the clusters. This will ensure the cloud security on the databases with privacy demands secured encrypted data sets.

Kumar.S (2020) proposed a method for relocating the cloud data into suitable places on online through the distances of the clusters. The simulation process happened to find the distance of the data location available in clusters to provide secured transactions. The algorithms used to find the distance between the two clusters are K-means and KNN. It has to detect the Euclidian distance of the clusters connected in the networks and finding the distances for them to relocate the data from one place to another place for security purposes. In cloud data centers, storages are very high was not help to implement cluster security on networks.

Ramalhinho J et al (2020) proposed a method for CBIR, using a registration approach with Bayesian framework classification. Image guidance is the concept using in this method to collect multiple labeled object images from the sources randomly. Separate registration of all the objects will create the document details about the images. Patient information like Laparoscopic Ultrasound (LUS) images are collected from hospitals and will be classified using the registration approach with minimum details of the images. The size and metrics used in the LUS images have registered and monitored continuously for the changes that occurred in that values. It will be working during the patient health monitoring approaches applied on their reports.

Lu. H et al (2020) proposed a method to implement the speed of the CBIR approach, using Deep Fuzzy Hashing Network (DFHN) along with huge image datasets like CIFAR 10 and NUS-WIDE. Fuzzy logic systems and hashing indexing techniques are used to find the features of the images with the effective model of binary codes to remove the uncertainties in the images. Hamming distances and convolution layers are fully connected with fuzzy-based techniques to improve the retrieval quality of the clusters underlying the data. Since it used the fuzzy-based theory for image representations it will give the exact output to the users when convolution layers extracted modified images also.

Roy s et al (2020) proposed innovative idea in CBIR approach to implement extracting technique using deep learning neural network hashing techniques. Retraining and fine-tuning steps involved in the intermediate representation of the DNN. It is mainly used to improve the RS images available on the internet. Target image features are optimized and retrieved using the semantic-based metric representation in the CBIR. The image retrieval is changed in terms of speed while using DNN for fine-tuning the image features with pre-trained models of deep learning. To improve the fast search option in this approach hashing codes are used to store the vector values present in the databases. More images are taken for consideration in feature extraction that may be eliminated using DNN with the help of hashing codes present in the network.

Maji. S et al (2020) proposed a method to implement CBIR, using adaptive Decentralized CNN and pre-trained models. Representation Learning is the new concept used in this approach to produce the trained dataset from the input images along with hidden, max pooling layer and softmax layer. These layers have measured the elements of the images as a feature in the cloud and their weights are noted during the retrieval time and will be considered for further processing. If any changes in the image and their weights are changes immediately will be notified through the classifiers, the object classifiers are the main role plays in the CBIR system, to implement the trained models of the recovery.

Kavitha P et al (2020) proposed a method for CBIR, with the features of Principal Component Analysis (PCA) along with genetic algorithms. DWT (Discrete Wavelet Transform) technique is used to predict the distances between the clusters with the help of KNN, K-means algorithm in the networks. It is used to increase the storage space, data processing capabilities for the expansion of the retrieval process.

Shen.M et al (2020) proposed a method to protect the multiple images on the databases with the help of privacy protection. MIPP (Multiple Image Owners with Privacy Protection) is used to predict the measurement of similarity images from the secured storage. Another technique called Multi party computation techniques are also used to protect the data sets on the network effectively.

Xu. Y et al (2019) proposed a method to provide security to large-scale images using Hamming Embedding algorithm as encryption techniques. The image descriptors are used to take the form of binary signatures and histograms. It is also used to predict the random sampling of images with hashing algorithms for retrieval. The inverted index is also used for working with image clustering in the form of making digital signatures. But the major disadvantage in this approach is providing low accuracy and efficiency before it outsourcing in the encryption stage.

Azeez N et al (2019) proposed a method HIPAA (Health Insurance Portability Accountability Act) along with Identity-Based Encryption (IBE) and the latest Re-encryption-based proxy services for new identity retrieval. Health card is issued in this approach to all users then the normal encryption algorithms are used such as DES, RSA. The storage encryption system is used to identify the differences in the encryption, especially for e-based health care systems.

Dong. B et al (2019) proposed a method in an AI-based CBIR system with the loop-based structure of human explainable approaches in x-ray images. The deep learning supervised algorithm model has been framed with a refinement strategy using binary classifiers trained in the online system. Relevance feedback has been collected from the users to find the methodology to retrieve the video contents from the large database. The self-supervised system will help the CBIR methodology to increase the speed of the process.

Tanioka. H et al (2019) proposed a method for fast calculation in index-based similarity check with L2 normalization to perform CBIR. VGG-16 pre-trained dataset is used to find the images by elastic search method and DCNN contribution is mainly adapted in this approach. It is mainly used to perform fast and secured image retrieval techniques. Cloud storage is the main problem of all approaches used in CBIR, where some of the encryption and feature extraction selection is also used for retrieval.

Asim M.N et al (2019) proposed a method for context-awareness to improve the IR techniques by the multi-dimensional concept of contexts used in this approach. Multiple sources have created the data over the internet and have to retrieve using the CBIR approach but the basic approach is IR. To overcome the problems of Ontology-based information retrieval schemes available on the internet for developing retrieval techniques.

M R Sundarakumar and G Mahadevan (2019) proposed a method called secured authentication in cloud computing using SHA 256 algorithm. The algorithm gives the DES, AES, RSA combination keys for 256-bit encryption under the encryption technology. All the other algorithms are working in a particular range whereas these SHA-256 algorithms are used to create the hash indexing values to store the contents of the images. It will help to select the exact image from the repository for accurate retrieval.

M R Sundarakumar and G Mahadevan (2019) proposed a method called FNN (Futuristic Nearest Neighbor) search and Elastic search for searching of the images with high accuracy and low latency. The Euclidian distance between the clusters has been calculated and fine the nearest image clusters on the network will give the exact result to the users while using the CBIR approach. For the improvement of the image retrieval concept, this modified algorithm is used and retrieved the exact image with high accuracy and speed.

Jain. M et al (2018) proposed new CBIR approach using the Hierarchical Divide and Conquer K Means (HDK) method to identify the distances between the cluster images. Data clustering is the approach as an unsupervised method and extracting the hidden patterns of the large datasets have identified and stored as an indexing concept. Color, shape, and texture values are stored in an indexing table to calculate the distances between them. The high dimensionality and optimal number problem were solved through this approach to perform CBIR accurately.

Shinde S et al (2018) proposed a method called Color Coherence Vector Pragmatic and Local Edge Histogram Descriptor (LEHD) for improving the CBIR approach by extended the indexing databases. Multi sequential approach color feature extraction is developed in this model for increasing the speed of the shape matching from the boundaries of the images. The ordering of imaging has to be considered for developing boundaries in the global, local edges.

M R Sundarakumar and Nithya Ganesan (2016) proposed a method called edge clustering mechanism to improve the low traffic management system on the network. Edge computing is the technique that is used to deliver the data to the next cluster without any hesitations even more resources are used in the network. All real-time information about the clusters is modified when new data entered into the clusters. Reinforcement algorithm is used to predict the next stages on the real-time clusters are used to predict the movement of the data continuously.

Mohan R et al (2015) proposed a method called NETRA, RETIN-based CBIR approach using histograms in the images based on color, shape, and texture. Minimum Probability of Error Retrieval (MPER) is the mechanism used to predict the features of the extraction based on the Supervised Multiclass Labeling (SML) algorithm. It is used to create segmentation and large cluster on the network.

TABLE I. Research Gaps and solutions for CBIR approach

Author and Year	Research Gap identified and their Disadvantages	Methodology Used	Solutions
Buvana,M et al,2021 [1]	Video frames features are increased	• Query by Image Content (QBIC) • Pigeon Inspired based Optimization (PIO) • Artificial Neural Network(ANN)	• Gray-Level Co-Occurrence Matrix (GLCM) • Local Binary Pattern (LBP) Histogram of oriented gradients (HOG) features.
Zheng,P et al,2021 [2]	Meta-data from large-scale HSI big data	• SaaS with HSI storage, management, through web interfaces.	• Hyper spectral image (HSI) repository
Li.Y et al 2021 [3][3]	Large-scale mining and image processing	• Geo-localization of RS images and processing	• Remote Sensing image retrieval
Singh.P et al,2021 [4]	Image region of interest finding	• Classification • Partial Feature Mapping	• Max-pooling
Haq,N et al 2021 [5]	X-ray large scale database handling	• Deep Learning • Network community based image detection	• X ray database medical image automated retrieval
Li.J et al 2020 [9]	Personal privacy demand, Secure encrypted data sets	• Image retrieval based on Privacy-preserving • Semi-trusted computation	• Asymmetric Scalar-product-Preserving Encryption (ASPE) • Homomorphic Encryption (HE)
Yan,H et al,2020 [7]	Video image retrieval from the images	• Homomorphic Encryption (HE) • Order-Preserving Encryption (OPE) • Vector extending mechanism	• Secure Video Retrieval (SVR)
Devaraj,A et al,2020 [8]	Cloud platform large scale image retrieval performance	• Secure Image Archival and Retrieval System (SIARS) • Creation schemes for Multiple shares	• Adagrad CNN (AG-CNN) • Secure Multiple Share Creation (SMSC) schemes
Lin.F et al 2020 [6]	Facial expressions,illumination conditions and significant variation	• Deep neural networks (DNN) • CNN models (VGGFace) for face recognition	• Deep learning Recognition
Kumar,S. 2020 [10]	Locations of Cloud cluster data center	• KNN • K-Means	• Euclidean Distance • Reinforcement algorithm

Ramalhinho,J et al 2020 [11]	Laparoscopic Ultrasound (LUS) handling	• Bayesian framework	• Registration approach based Content-Based Image Retrieval (RCBIR)
Lu,H et al 2020 [12]	Speed of the Image retrieval	• Deep hashing -deep neural network (DNN) • Fuzzy logic technique	• Deep Fuzzy Hashing Network (DFHN)
Roy,S et al2020 [13]	Remote Sensing image retrieval	• Pre-trained DNN as an intermediate • Semantic-based metric space • Binary hash codes	• Deep hashing -deep neural network (DNN)
Kavitha,P et al 2020 [14]	Extended storage space, and flexibility	• Discrete wavelet transform technique (DWT) • K-Means algorithm and SVM	• Principle Component Analysis (PCA)
Xu,Y, 2019 [15]	Low efficiency and accuracy	• Binary signatures • Random sampling • Min-Hash algorithm • Secure inverted index	• Large-scale secure image retrieval • Hamming embedding algorithm
Azeez,N et al,2019 [16]	Encryption on storage	• DES and AES • Electronic Health Card (EHC) • Secured Hash Index • SHA 256 instead of 128	• Health Insurance Portability and Accountability Act (HIPAA) • Identity Based Encryption (IBE) • IB-Proxy Re-Encryption(IBPRE)
Dong.B et al,2019 [17]	Large data - Video retrieval	• Self-supervised deep learning model • Refinement strategy • Relevance feedback	• Explainable AI (XAI) system • XAI-CBIR
Tanioka.H, 2019 [18]	Fast and scalability in CBIR	• Deep Convolution Neural Network (DCNN)	• Elastic search • Pre-trained model
Asim,M N et al,2019 [19]	IR- ontology based schemes	• Multi-dimensional context	• Context awareness
Shen,M et al 2020 [20]	Multiple image extraction problem	• Secure multi-party computation	• Multiple Image owners with Privacy Protection • MIPP

Jain,M, et al 2018 [21]	High dimensionality	• Unsupervised data clustering Extracting from hidden patterns	• HDK (Hierarchical Divide and Conquer k means)
Shinde,S et al,2018 [22]	Indexing of database image	• Multi-sequential approach • Color Feature Extraction	• Color Coherence Vector • Local Edge Histogram Descriptor (LEHD) • Global Edge Histogram Descriptor [GEHD]
Mohana,R, et al,2015 [23]	Large databases retrieval time extension	• Features vectors • Image segmentation	• NETRA CBIR RETIN CBIR • Minimum Probability of Error Retrieval (MPER) • Supervised Multi class Labeling (SML) Algorithm
Xia, Z et al 2015 [24]	Feature Extraction of images	• Earth mover's distance (EMD) • Linear programming (LP) • Local sensitive hash (LSH)	• Privacy-preserving CBIR

Xia. Z et al (2015) proposed a method called Privacy-preserving CBIR schemes for easy retrieval. The Earth Mover's Distance (EMD) has been calculated based on the Linear programming techniques and will be used for doing features extraction. The new concept called Local Sensitive Hash (LSH) is used to save the values in the form of vectors and everything is controlled by LP. This method is used to classify the distance between the image parameters which are decided by the features of that.

4. Research Gap and Solutions

CBIR approach extract the images from the large datasets were stored in repositories of the cloud storage. This approach has done various applications of real world problems in different sectors especially in health care. The patient's database is very huge and dynamic in nature for maintaining them with cloud computing has challenges and limitations using internet sources. The Table I will summarize the research gaps and their solutions used in previous technologies

5. Techniques used in CBIR and their Performances

CBIR approach is used in cloud computing to retrieve the images from the huge repository rapidly. The challenges faced in CBIR are the retrieval time of images is very high due to the size of the storage. Most of the techniques used in CBIR approach in cloud computing has given their performance for improve the accuracy and reduce the latency time. But when size of the database or repository is huge classical approaches are not working efficiently to retrieve the images in fast search. Query based image searching is used to get the images frequently in CBIR approach where as feature extraction techniques are used to find the exact match of the image searched by the users. Different modern algorithms and techniques are used to improve the accuracy of the image retrieval in CBIR [25], [14], [20], [15], [16], [17], [18], [19], [26], [27], [21], [22]. The Table II will explain all the techniques followed in CBIR and their performance levels.

TABLE II. Techniques followed in CBIR and Performances

Techniques followed	Limitations	Data sets taken and their Performances (in Percentage)
• Discrete Wavelet Transform (DWT) • K-Means • Euclidean Distance	Processing speed and accuracy is less	82.5 accuracy in large data sets
• Genetic Algorithm • HARP	Lower arithmetic time	Precision 89.1 , Recall 69.8
• Principal Component Analysis(PCA) • Linear Discriminate Analysis (LDA)	Limited with database capacity and complex	For PCA -ORL 77.68 , and UMIST 64.9 For LDA ORL 79.68 and UMIST 67.5
• Deep Neural network (DNN)	Big degree of matching with feature extractions	High recognition probability
• Encrypted Difference Histogram(EDH)	Reduced user accounts in cloud server	Block size is varied based on data centre allocation
• Local Binary Pattern(LBP) • Local Neighborhood Difference Pattern(LNDP) • Centre Symmetric LBP	Accurate precision and recall graph evaluation measure	Corel-10K , Precision -41.61, Recall – 16.91, Precision-78.77
• Support Vector Machine (SVM) Classifiers	Limited parameters of images have taken	Accuracy level is only 0.834
• Hybrid Fuzzy C-Means (HFCM)	Higher recovery compared with K-means	Precision -0.71, Recall- 0.61
• Bhattacharya Distance (BC)	Statistical data of images are very complex to store	Only limited for higher databases
• GLCM(Grey Level Co –occurrence Matrix)	Better features finding	Mean of 82.6 precision
• Kerkre transform	Distinct images of large set speed is slow	Min cost and accuracy
• SIFT and SURS classifiers	Reduce power consumption GPU utilization	Deep learning trained sets will give high accuracy
• Shear let Transform	Depth and shape of object notified	Precision -68.634, Recall- 80.236

6. Conclusion

CBIR is an important approach in cloud computing for extracting images from repositories effectively. When all the data stored in a repository need an online platform to maintain in different places cloud computing is playing a vital role in that domain. Cloud computing has done storage mechanisms with the help of cloud centralized servers available everywhere in the form of clusters. Virtual machines are used to perform cloud server storage concept effectively and the data which are stored in that were derived with different algorithms and mechanisms. In this article, several research papers and survey papers were taken for the study to improve the performances of the CBIR in the cloud especially retrieved the health care systems. Classical approaches like K-means, and KNN data mining algorithms are used to find the distances between the images in the cluster by calculating Euclidean distance and the same will be rectified using the query-based approach of the customer. Another approach features extraction has done using the color, shape, and texture of the image then finally it will match the parameters for finding the exact image. Autoencoders are doing a repeated search to find the exact image on the internet iteratively using trained input images. CNN methods are used to derive the image from the cluster by training the input image to proceed into several trials in object classifiers. These classifiers have extracted the features from the images and mismatching founds are omitted from the search to identify the exact final output. Finally, deep learning algorithms are used to find the exact image from the internet using representation learning with encryption also. The features extracted from the images were stored in indexing patterns and that will be encrypted with the latest encryption techniques. When the intruder comes and enters the database for extracting the images from the cluster it will provide security over the network system to avoid conflicts. The various other techniques and patterns were listed in the above tables 1, 2 for deciding to suit which technique is faster in the CBIR approach. This paper gives a distinct opinion about the CBIR approach in cloud computing for fast retrieval of images from the health care system information storage accurately.

To illustrate the contribution for this paper, the authors are summarized the CBIR techniques used in the health care system approach as following points:

1) The authors have taken different algorithms and techniques from many recent research articles on the topic of CBIR; every paper is discussing the image retrieval techniques from the huge datasets on the large network clusters. The challenges and limitations are notified as less accuracy and low latency in healthcare system images database suitable to the huge volume on the repository
2) The author has reviewed this paper for the CBIR approach based on classical approaches and recent algorithms with the methodologies and tools used for that. Compared with previous methods were used in the CBIR approach.
3) The author has found the techniques and model for the CBIR approach using the deep learning method. It will admire the researches in future those who are developing research projects in health care information system extraction.

7. Future Enhancement

CBIR approach deals only with the limited amount of storage in cloud computing storage systems. In the future, data sets from big data and IoT frameworks will be handled to retrieve the images from the various high-volume repositories remotely. Virtual machines will be used to create the storage in clusters for increasing the size of the storage [28]. Day by day number of sources created data will be increased rapidly through internet sources, that also be handled for accurate results outcome and low latency. Health care systems are the leading sector in the digital world for creating data from remote locations. An enormous amount of storage systems are created to handle this but sufficient techniques are not possible to access them properly [23]. This paper suggests the best methods are available to retrieve the images from internet sources accurately. Health care sector information is stored in various places for accessing through a centralized server concept which will lead to less latency and high accuracy for the researchers who are doing researches in this domain [24].

References

[1] M. Buvana, K. Muthumayil, and T. Jayasankar, "Content-based image retrieval based on hybrid feature extraction and feature selection technique pigeon inspired based optimization," *Annals of the Romanian Society for Cell Biology*, pp. 424–443, 2021.

[2] P. Zheng, Z. Wu, J. Sun, Y. Zhang, Y. Zhu, Y. Shen, J. Yang, Z. Wei, and A. Plaza, "A parallel unmixing-based content retrieval system for distributed hyperspectral imagery repository on cloud computing platforms," *Remote Sensing*, vol. 13, no. 2, p. 176, 2021.

[3] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing big data: A survey," *Information Fusion*, vol. 67, pp. 94–115, 2021.

[4] P. Singh, P. Hrisheekesha, and V. K. Singh, "Cbir-cnn: Content-based image retrieval on celebrity data using deep convolution neural network," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 14, no. 1, pp. 257–272, 2021.

[5] N. F. Haq, M. Moradi, and Z. J. Wang, "A deep community based approach for large scale content based x-ray image retrieval," *Medical Image Analysis*, vol. 68, p. 101847, 2021.

[6] F.-C. Lin, H.-H. Ngo, and C.-R. Dow, "A cloud-based face video retrieval system with deep learning," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 8473–8493, 2020.

[7] H. Yan, M. Chen, L. Hu, and C. Jia, "Secure video retrieval using image query on an untrusted cloud," *Applied Soft Computing*, vol. 97, p. 106782, 2020.

[8] A. F. S. Devaraj, G. Murugaboopathi, M. Elhoseny, K. Shankar, K. Min, H. Moon, and G. P. Joshi, "An efficient framework for secure image archival and retrieval system using multiple secret share creation scheme," *IEEE Access*, vol. 8, pp. 144 310–144 320, 2020.

[9] J.-S. Li, I.-H. Liu, C.-J. Tsai, Z.-Y. Su, C.-F. Li, and C.-G. Liu, "Secure content-based image retrieval in the cloud with key confidentiality," *IEEE Access*, vol. 8, pp. 114 940–114 952, 2020.

[10] S. Kumar, "Relocating big data among information centres over the online cloud environments," *Available at SSRN 3561400*, 2020.

[11] J. Ramalhinho, H. F. Tregidgo, K. Gurusamy, D. J. Hawkes, B. Davidson, and M. J. Clarkson, "Registration of untracked 2d laparoscopic ultrasound to ct images of the liver using multi-labelled content-based image retrieval," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 1042–1054, 2020.

[12] H. Lu, M. Zhang, X. Xu, Y. Li, and H. T. Shen, "Deep fuzzy hashing network for efficient image retrieval," *IEEE transactions on fuzzy systems*, vol. 29, no. 1, pp. 166–176, 2020.

[13] S. Roy, E. Sangineto, B. Demir, and N. Sebe, "Metric-learning-based deep hashing network for content-based retrieval of remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 2, pp. 226–230, 2020.

[14] P. Kavitha and P. Vidhya Saraswathi, "Content based satellite image retrieval system using fuzzy clustering," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–12, 2021.

[15] Y. Xu, X. Zhao, and J. Gong, "A large-scale secure image retrieval method in cloud environment," *IEEE Access*, vol. 7, pp. 160 082–160 090, 2019.

[16] N. A. Azeez and C. Van der Vyver, "Security and privacy issues in e-health cloud-based system: A comprehensive content analysis," *Egyptian Informatics Journal*, vol. 20, no. 2, pp. 97–108, 2019.

[17] D. R. Chittajallu, B. Dong, P. Tunison, R. Collins, K. Wells, J. Fleshman, G. Sankaranarayanan, S. Schwaitzberg, L. Cavuoto, and A. Enquobahrie, "Xai-cbir: Explainable ai system for content based retrieval of video frames from minimally invasive surgery videos," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 66–69.

[18] H. Tanioka, "A fast content-based image retrieval method using deep visual features," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 5. IEEE, 2019, pp. 20–23.

[19] M. N. Asim, M. Wasim, M. U. G. Khan, N. Mahmood, and W. Mahmood, "The use of ontology in retrieval: a study on textual, multilingual, and multimedia retrieval," *IEEE Access*, vol. 7, pp. 21 662–21 686, 2019.

[20] M. Shen, G. Cheng, L. Zhu, X. Du, and J. Hu, "Content-based multi-source encrypted image retrieval in clouds with privacy preservation," *Future Generation Computer Systems*, vol. 109, pp. 621–632, 2020.

[21] M. Jain and S. Singh, "An efficient content based image retrieval algorithm using clustering techniques for large dataset," in *2018 4th International Conference on Computing Communication and Automation (ICCCA)*. IEEE, 2018, pp. 1–5.

[22] S. Shinde, M. Desai, M. Gupte, and D. Racha, "Content based image retrieval using multi-sequential search," in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE, 2018, pp. 973–978.

[23] R. M. Mohana and A. R. M. Reddy, "Ccbir: A cloud based implementation of content based image retrieval," *WSEAS transactions on Computers*, vol. 14, pp. 333–346, 2015.

[24] 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. 6, no. 1, pp. 276–286, 2015.

[25] S. Maji and S. Bose, "Cbir using features derived by deep learning," *ACM/IMS Transactions on Data Science (TDS)*, vol. 2, no. 3, pp. 1–24, 2021.

[26] M. Sundarakumar and G. Mahadevan, "Authorization for secured cloud storage through sha-256," *International Research Journal of Engineering and Technology (IRJET)*, 2019.

[27] M. Sundarakumar and G. Mahadevan, "Improving speed and accuracy of image retrieval using elastic search and features nearest neighbor search," *International Journal of Recent Technology and Engineering*, 2019.

[28] M. Sundarakumar and N. Ganesan, "Real time traffic management system using cloud infrastructure with edge clustering mechanism," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 11, 2016.

Vasudeva R Vasudeva R completed his Master of Technology from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. Currently pursuing Ph.D on Cloud Computing (Domain) in VTU, He is currently working as Assistant Professor in department of Computer Science and Engineering, CBIT, Kolar, India Since 2010. His research interest includes Cloud Computing, Big Data Analytics. He is a member of IEI

Dr. S N Chandrashekara Dr. S N Chandrashekara, is currently working as Professor and Head of the Department of Computer Science and Engineering at CBIT, Kolar, India. He received his Bachelor of Engineering from Bangalore University and Masters of Technology in Computer Science and Engineering from NITK, Suratkal. He was awarded Ph.D in Network and Devices from IISc, Bangalore. His area of research includes Computer Networks, Cloud Computing, Cyber Security and IOT Medical Image Processing. He is a Life Member of Indian Society for Technical Education, Computer Society of India. He is a Fellow of Institute of Engineers and Life member of ISSE.