Cluster Computing
https://doi.org/10.1007/s10586-018-1731-0

Content based image retrieval using deep learning process

R. Rani Saritha[1] · **Varghese Paul**[2] · **P. Ganesh Kumar**[3]

Received: 28 November 2017 / Revised: 26 December 2017 / Accepted: 5 January 2018

Abstract
Content-based image retrieval (CBIR) uses image content features to search and retrieve digital images from a large database. A variety of visual feature extraction techniques have been employed to implement the searching purpose. Due to the computation time requirement, some good algorithms are not been used. The retrieval performance of a content-based image retrieval system crucially depends on the feature representation and similarity measurements. The ultimate aim of the proposed method is to provide an efficient algorithm to deal with the above mentioned problem definition. Here the deep belief network (DBN) method of deep learning is used to extract the features and classification and is an emerging research area, because of the generation of large volume of data. The proposed method is tested through simulation in comparison and the results show a huge positive deviation towards its performance.

Keywords Image retrieval · Deep learning · Data analysis · Image extraction

1 Introduction

In this era, technology upgrades to its maximum with the help of creativity and innovation, with such an ideas in the field of ANN, the basic module is said to be image processing stream so that most of the systems will map the inputs to its outputs with varied uncertainty logic [1]. The image will be considered as the digital formation and it will be decimated to its corresponding bits. The classification of image or video in the existing systems seems difficult due to its methodology works with the file name search and not the content inside it [2]. Depending upon the query given by the user the ANN should have to classify the content with various attributes. Our proposed algorithm deal with Deep Learning methods, in which it confines each and every data, learns the contents by separating its features to the deep bottom. The database itself maintains a separate individual data centre that will contain a finite most significant amount of features [3]. Deep learning method shows its maximum performance to its extent and plays a smart extraction of the content from the data, which is on process [4].

Deep learning is one of the classifications of soft computing phenomenon in which extraction of data from millions of segregated images can be retrieved using this phenomenon [5]. The retrieval performance of a content-based image retrieval system crucially depends on the feature representation and similarity measurement, which have been extensively studied by multimedia researchers for decades Although a variety of techniques have been proposed, it remains one of the most challenging problems in current content-based image retrieval (CBIR) research, which is mainly due to the well-known "semantic gap" issue that exists between low-level image pixels captured by machines and high-level semantic concept perceived by humans. From a high-level perspective, such challenge can be rooted to the fundamental challenge of artificial intelligence (AI) that is, how to build and train intelligent machines like human to tackle real-world tasks [6–8] (Fig. 1).

Machine learning is one promising technique that attempts to address this challenge in the long term [9]. Recent years have witnessed some important advanced new techniques in machine learning. Deep learning is the part of machine

✉ R. Rani Saritha
ranisaritha3090@gmail.com

Varghese Paul
vp.itcusat@gmail.com

P. Ganesh Kumar
ganesh23508@gmail.com

[1] Department of Computer Applications, Saintgits College of Engineering, Kottayam, Kerala, India

[2] Department of CS/IT, TocH Institute of Science & Techology, Ernakulam, Kerala, India

[3] Department of Information Technology, Anna University of Technology, Coimbatore, Tamil Nadu, India

Published online: 07 February 2018

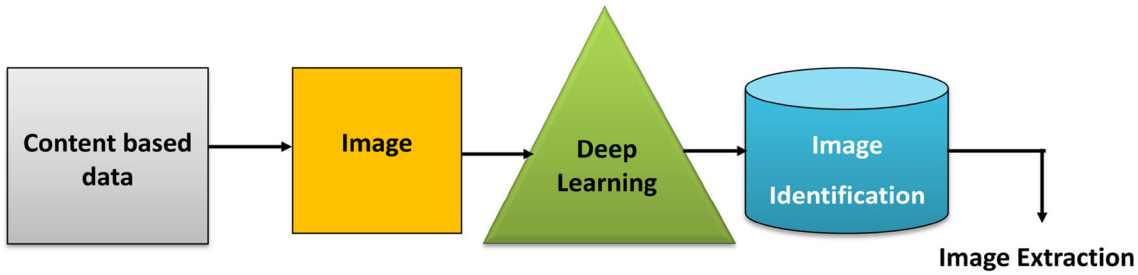

Fig. 1 CBIR system with deep learning

learning, which includes a family of machine learning algorithms that attempt to model high-level abstractions in data by employing deep architectures composed of multiple non-linear transformations [10]. Unlike traditional machine learning techniques that are often using "shallow" architectures, deep learning mimics the human brain that is organized in a deep architecture and processes information through multiple stages of transformation and representation [11]. By exploring deep architecture features at multiple levels of abstracts from data automatically, deep learning methods allow a system to learn complex functions that directly map raw sensory input datum to the output, without relying on human-crafted features using domain knowledge [12]. Many recent studies have reported encouraging results for applying deep learning techniques to a variety of applications, including speech recognition, object recognition, and natural language processing, among others. Inspired by the successes of deep learning, in this paper, we attempt to explore deep learning methods with application to CBIR tasks. Despite much research attention of applying deep learning for image classification and recognition in computer vision, there is a still limited amount of attention focusing on the CBIR applications. In the proposed method, we investigate deep learning methods for learning feature representations from the images and their similarity measures towards CBIR tasks [13–15].

1.1 Relevance of work

The incomplete annotation issue in text based image retrieval will degrade the retrieval performance of the searching process [6]. Query by Image Retrieval (QBIR) has evolved into a necessary [6] module in which the contents of the images are extracted to search the images from the database. But CBIR system also faces many challenging problems because of the large volume of the database constrain, the difficulty in both people and computer understanding the images, the difficulty of creating a query and the issue of evaluating results properly. In our method, we explore an alternative strategy for searching an image database [7] in which the content is expressed in terms of an image and its multiple features are extracted using different image feature extraction algorithms. These features are analyzed with the features of image database and the most similar images are retrieved using an efficient index based sorting algorithm [16].

1.2 Analysis

Specifically, we aim to address the following open research queries:

(i) Are deep learning techniques effective for learning good feature representations from images to tackle CBIR tasks?
(ii) How much improvement can be achieved by deep learning techniques when compared with traditional features crafted by experts in multimedia and computer vision?
(iii) How to apply and adapt an existing deep learning model trained in one domain to a new content based image retrieval task in another domain effectively?

In order to answer the above queries, we investigate a framework of deep learning for content-based image retrieval (CBIR) by applying a state-of-the-art deep learning method, that is, deep belief networks (DBNs) for learning feature representations from image data and conduct an extensive set of empirical studies for a variety of CBIR tasks. From the experimental studies, we obtain some encouraging results and reveal several important insights for addressing the open questions. As a summary, we formulated the following major contributions in this work:

- We introduce a deep learning framework for CBIR by training large-scale deep Deep belief Networks for learning effective feature representations of images.
- We conduct an extensive set of empirical studies for comprehensive evaluations of deep Deep belief Networks with application to learn feature representations for a variety of CBIR tasks under varied settings.

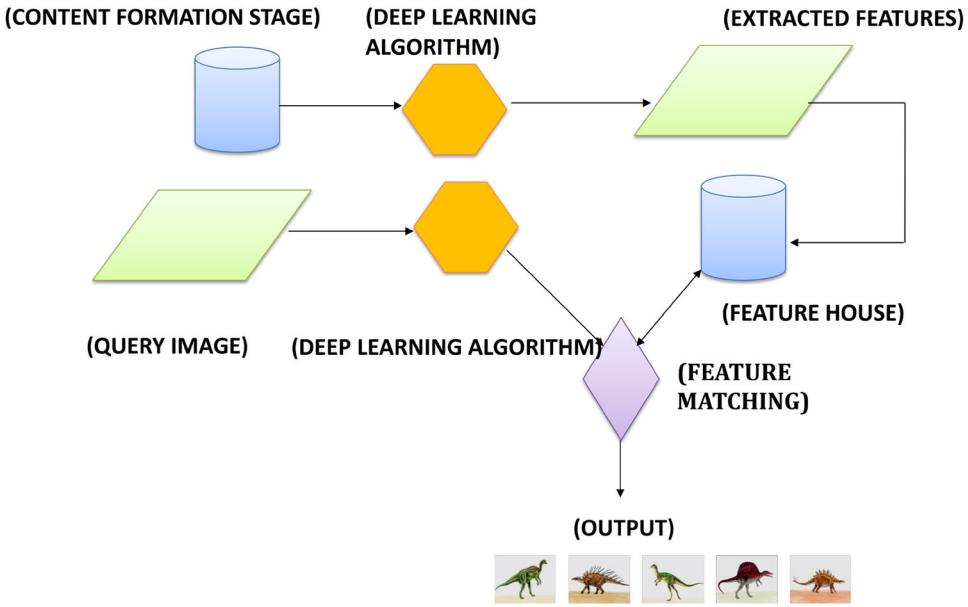

Fig. 2 CBIR results using deep learning

2 Literature review

For many previous content based approaches there were many local and global features to represent the image properties and content [17]. In [1], primitives and colony filter are used for color and texture feature extractions. In their work, an image is divided into many sub blocks and each block's color moments are extracted with respect to the algorithm which exists [18,19]. These moments are clustered into different classes by using a clustering algorithm and a specified color feature vector algorithm which is calculated from the query image and the images in the image database [20–24]. The distance between each digital image will be represented by each digital value but previously we cannot able to retrieve such an accurate value from each of the image that we searched particularly [25–28]. Some of the papers will define the mode of communication take place between the point nodes of the image [29]. The average precision is 59.61.

Object-based image retrieval systems retrieve images from a database by extracting the object features in the images. In this method database images are segmented and compare each segmented region against a region in the query image given by the user [30]. These types of image retrieval systems are generally successful for objects that can be easily separated from the background and that have distinctive colors or textures [21]. Xue and Wanjun proposed a model in which color histograms and color moment feature extraction methods are integrated. They stated that the index sorting was better than other approaches [31–33].

Huang et al. [23] work propose a model based on color and texture features [34]. They used color moments of the Hue, Saturation and Value (HSV) of the image and texture features are extracted by using Gabor descriptors [35–39]. They normalized the features and calculated the similarity by using Euclidean distance. They reported that the proposed method had a higher retrieval accuracy than the previous conventional approaches [40–42] (Fig. 2).

3 Proposed method

In our method, a multi-feature image retrieval method is introduced by combining the features of color histogram, edge, edge directions, edge histogram and texture features, etc. In this model, the content based image will be extracted from a collection of intended image groups. After performing some pre-processing steps like selection removal, its above features are extracted and are stored as small signature files. Similar images should have similar signatures. These signatures are compared with the content based signature. During the similarity measure, the distances between the different features are measured. Appropriate weights are applied to normalize the distance coefficients [1]. These normalized coefficients are sorted and indexed based on the distance values and their optimized state of functioning.

4 Content extraction

The digital image can able processed with two modes of operation i.e. spatial domain and the frequency domain in which transformation is required. In certain system the content extraction may take place with the help of Image segmentation. The extraction takes place in the spatial domain will provide a data on visual features such as color, radiance,

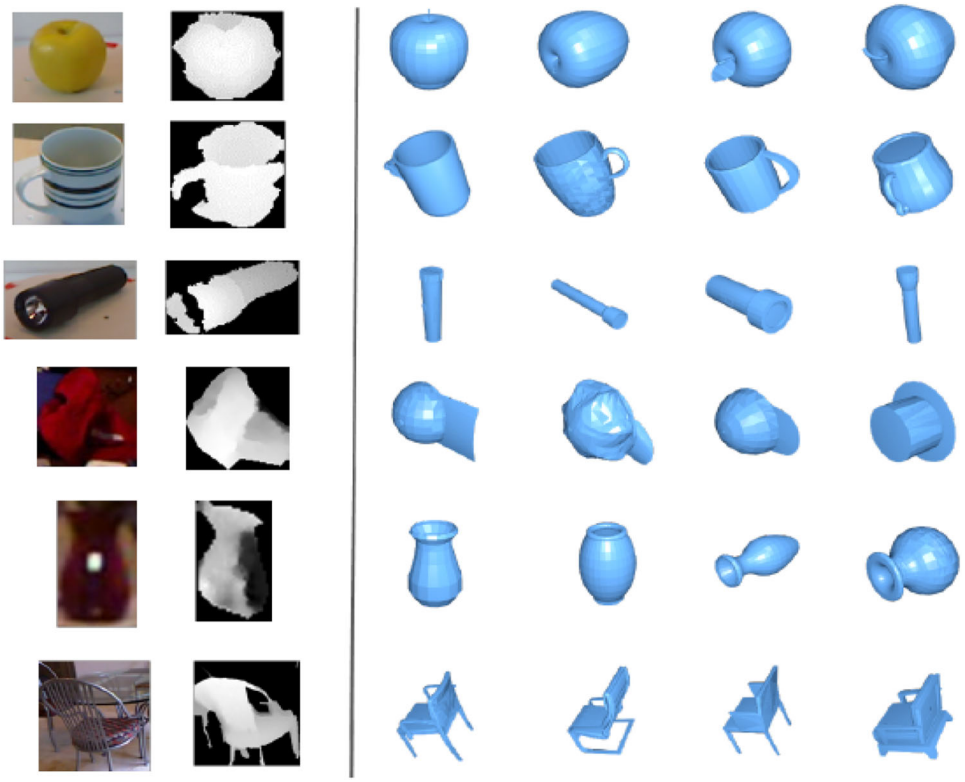

Fig. 3 Image extraction using deep learning

luminance, structures, etc., whereas the frequency domain features will provide the information related to the frequency, etc., The deep learning method used will have these both features which indeed a challenging due to the enormous amount of data. Out of this enormous data, significant features will be taken which makes the processing better (since it avoids maximum time complexity). This data will be computed with care to make classification and computer vision computation a better. The extraction of visual features such as color, texture, shape, spatial relationship, etc. is one of the major operations in designing a reliable and efficient image retrieval resource. By considering the above sequence of data or image extraction phenomenon each datum will be formed with some suitable content based text.

For every formation of data which is holding each and every image will be analysed by using the neural networking structure. The advanced level of such formation of data extraction will be majorly done by using the deep learning phenomenon. Each single character identification will be done in each and every content based text structure. Only through the help of such a feature extraction phenomenon major number of vital errors can be overcome and the précised output will be able to process. In each and every module of segmentation, each bit of information is utilized for comparing the inbuilt image of the respective data. Not only for the processing of feature, but also through this process property based extraction can also be done using the deep learning phenomenon (Fig. 3).

The retrieval performance of a content-based image retrieval system crucially depends on the feature representation and similarity measurements, which have been extensively studied by multimedia researchers for decades. A wide range of methods have been proposed due to the well-known "semantic gap" issue that exists between the low-level image pixels captured by machines and the high-level semantic concepts perceived by humans. From a high-level perspective, such challenge can be deep-rooted to the fundamental challenge of artificial intelligence (AI) that is, how to build and train intelligent machines like human to tackle real-world tasks. Machine learning is a system used to address this grand challenge in the long term. Recently, some important advances of new techniques in machine learning have developed. One important technique is known as "deep learning", which includes a family of machine learning algorithms that attempt to model high-level abstractions in data by employing deep architectures composed of multiple non-linear transformations. But conventional machine learning methods often uses "shallow" architectures, while deep learning mimics the human brain and is organized in a deep architecture and processes information through multiple stages of transformation and representation.

By exploring deep architectures to learn features at multiple levels of data abstracts, deep learning techniques allow a system to learn complex functions that directly map raw sensory input data to the output, without relying on human-crafted features using domain information. Recent studies

have reported encouraging results for applying deep learning techniques to a variety of applications, including speech recognition, object recognition, natural language processing and others. Because of this success history of deep learning, in this paper, we attempt to explore deep learning techniques with application to CBIR tasks. In computer vision, much research attention is given to apply deep learning algorithms for image classification and recognition, but still a limited amount of attention is given to focus on the CBIR applications.

In this paper, we investigate deep learning techniques for learning feature representations from images and their similarity measures towards CBIR tasks. In order to answer the queries, 1. will deep learning methods work for content based image retrieval, 2. how much deep learning is efficient for the improvement for content based image retrieval and 3. how to adopt the task of CBIR using deep leaning. The investigation on a framework of deep learning for content-based image retrieval (CBIR) was done by applying a state-of-the-art deep learning method, that is, deep belief networks (DBNs) for understanding learning and mapping feature extracted from an image from image data, and conduct an extensive set of empirical studies for a variety of CBIR tasks.

- We introduce a deep learning framework for CBIR by training huge dataset of data deep belief networks (DBN) for understanding effective feature significant features
- We conduct an extensive set of experiments for comprehensive evaluations of deep belief networks with application to map and understand feature representations for a variety of CBIR tasks under varied settings.

Feature extraction To extract the features of the high-frequency information, a set of feature maps are extracted from the input (LQ) image by way of a convolution. A high-dimensional vector is used to represent image patches extracted from the input image. These vectors are composed of a set of feature maps, and through the network, the feature map of the image patch is learned from the training data.

The feature extraction block consists of **nf** convolutional layers with a kernel size **sf** which outputs **df** features as follows: nf $\times$ Conv(sf, df).

Each variable should be determined in consideration of the following:

nf represents the low-level features, such as edges or corners, that can be extracted at the lower level of the layer, and more complex features, such as textures, can be extracted at the higher level [32];

sf indicates that a large size convolutional kernel can be replaced with multiple stages of a small size kernel to reduce both the number of parameters and the computational cost while maintaining the same receptive field [33];

df represents the number of LQ feature dimensions, which is a factor that influences the performance. Therefore, it is important to determine optimal values of all variables.

Mapping The features extracted from the previous block are non-linearly mapped by this block, which consists of three modules: shrinking, non-linear mapping, and expansion. It shows that these modules reduce the number of parameters and achieve better performance than a single convolution layer.

The shrinking module is meant to reduce the number of feature dimensions **df** to a shallow feature dimension **dm**. (dm $<$ df) by a 1×1 convolution that acts as a linear.

Zero padding We employed zero padding on all layers to avoid reducing the output size by the convolution of each layer so that the input and output feature maps have the same size.

PReLU Each convolutional layer, except for the last layer (the image reconstruction block), is followed by an activation function. We employed a parametric rectified linear unit (PReLU) instead of the more commonly used rectified linear unit (ReLU) as the activation function. In the ReLU, the negative part is zero, whereas the PReLU differs in that it has a learn-able parameter that adjusts the slope of the negative part during the learning process, which improves the accuracy at a negligible extra computational cost. Therefore, the PReLU is robust to the weakness of ReLU that may occur when the input value is less than zero.

Residual learning The LQ and HQ images are highly correlated except for the image details, and the difference between the details is very small. This means that it is sufficient to predict only the high-frequency components for HQ image generation. Therefore, we designed our network to predict the residuals. In addition, we achieve better performance with faster convergence by residual learning based on the brightness domain rather than other domains, including infrared (thermal-, far-, near-) and color-based (gray, lightness, intensity) images.

Loss function The training process of the network aims to minimize the loss between the predicted images and the corresponding high-quality images (ground truth). The HQ image is composed of a pixel-wise summation of the LQ and residual images, and thus the parameter learned in the network are the residuals R between the input LQ image X and the ground truth Y, R = Y-X.

4.1 Deep belief neural network

Deep learning is the kind of advanced Artificial Neural Network developed by many researchers to make the machine learning process to different level of frontier. The main role of this deep learning process is to extract the information in high level abstraction methodology. This algorithm usually consists of multistage nonlinear transformers which are like

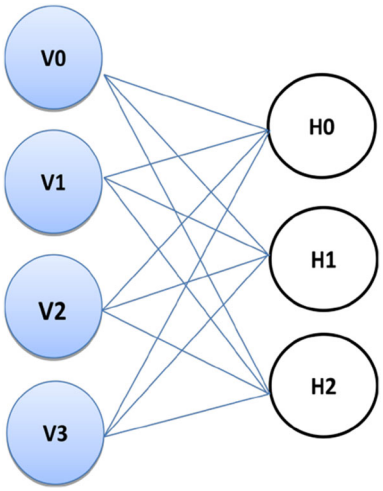

Fig. 4 Graphical structure of RBM

cascading many neural networks together. High level data abstractions are done with distributed representation i.e., data will be analyzed with different dimensions and parameters. Each abstractions re done through the hierarchical explanatory factors, in which the many sublevel information are generated with single previous level of generated information. Many traditional ANN methods failed to process the unlabeled data but the almost all deep learning algorithms have a capability of processing the unlabeled. The deep belief network (DBN) is one of the deep learning algorithms which is capable of processing the data unsupervised manner. This algorithm also have a capability classifying the invariant data which has the divergence of ranges i.e. the noise, displacement, smoothness etc. The necessity of the labelled dataset, inadequate parameter selection, and slow learning process makes the traditional neural networks inefficient, where the DBN stands to overcome the situation. So that the efficient local optimal solution is achieved. Since all the deep learning networks consists of multistage nonlinear transforms, DBN has several layers which includes restricted Boltzmann machine (RBM) stacked into multi-stages, which consists of only single hidden layers each to make the learning process faster proposed by Hinton et al. [43] (Fig. 4).

Restricted Boltzmann machine is designed based on the log-linear Markov random field (MRF) in which the energy function is linear in its free parameters. So that the learning features of one RBM will be given as input to the next RBM. The process of RBM depends on the probability distribution over visible variable; hidden unit via an energy function is given as (Fig. 5)

$$E\,(V, H) = -b'V - C'H - H'WV \tag{1}$$

The free energy formulas are derived is given by

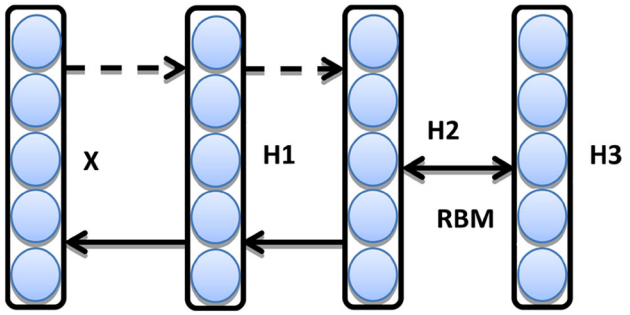

Fig. 5 Architecture of DBN

$$F\,(V) = -b'V - \sum_i log \sum_{H_i} e^{H_i(c_i + W_i V)} \tag{2}$$

where W is weights acts as connecting medium for hidden layer and visible layer, b and c are assigned as a offsets of visible as well as hidden layers respectively. The visible and hidden units in the RBM are independent to each other by condition. When RBM is used by probabilistic version of binary data the activation function would be sigmoidal in nature which gives

$$P\,(H_i = 1|V) = sigm\,(c_i + W_i V) \tag{3}$$

$$P\left(V_j = 1|H\right) = sigm\left(b_j + W'_j H\right) \tag{4}$$

So that the free energy of an RBM can be further simplified as

$$F\,(V) = -b'V \sum_i log\left(1 + e^{(c_i + W_i V)}\right) \tag{5}$$

Since it operates based on energy model the hidden layers are capable of capturing the higher order correlations such as high frequency components, directional information, edges of the data given. These stacked learning of data by multi stage RBM is responsible for using DBN unsupervised manner for data even if it has no labels.

Since the DBM is a stack of RBM networks the training process will be handled in stage by stage. initially the inputs (may be 1D or 2D) from visible layers are trained i.e X and then training of RBM is handled for H1 but in this case the data will be an outcome features of Stage-1 RBM, likewise the sequential training process will make a high level extraction of features.

We consider the privacy-preserving performance in terms of P2, because P3 is inversely proportional to P2, and P1 is decided by the system parameters. Privacy-enhanced retrieval is carried out according to different privacy policies. First, we randomly omit b = 1, . . ., 8 bits from each sub-hash value. We refer to these policies as the baseline policies. Since the server puts all candidates into a list, two metrics are used:

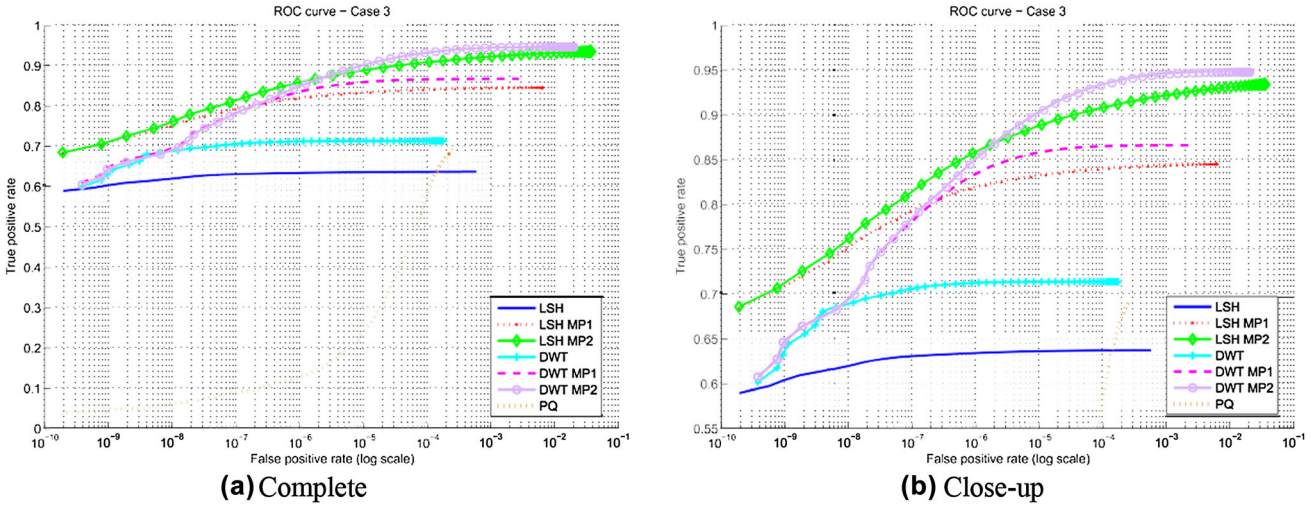

Fig. 6 ROC curve (Case 3, DBL). "MP x" means multi-probing within Hamming radius x. There is no significant difference from Fig. 6. The performance is not much affected by the increased database scale

- The number of candidates in the list;
- The entropy of the candidate categories in the list. Additionally, we also guess the query ID and the query category by majority voting and compute the success rate.

In the following experiments, Case 2 (DBM) is used as an example if not specified otherwise (Fig. 6).

Baseline privacy policies We first consider the case when multi-probing is not used. The average numbers of candidates per query are shown in Fig. 8. In the Fig. 8, "zero bit" per sub-hash corresponds to the non-privacy-preserving scenario. A more detailed plot is shown in Fig. 7 for LSH. In general, we can observe that the number of candidates increases with the number of omitted bits exponentially and the distribution of candidates among different sub-hash tables is quite even.

The entropy of candidate categories is shown in Fig. 8 in a similar way. This figure shows how difficult it is to guess the category of the query. Since there are 1000 categories in total, the maximum entropy of candidate categories in a list is approximately 10 bits. We can observe that the entropy increases linearly with the number of omitted bits (Fig. 9).

The entropy increases linearly with the policy and its estimation N2 defined in (2). We can see that most values are quite close to one, on average larger than 0.98. That implies we can estimate the computation and communication cost when defining a privacy policy, using Table II. When multi-probing is enabled, the load of the server is tremendously increased. We choose $r = 1$, so that each sub-hash of the query is expanded to $np = 17$ items (Fig. 10).

The average numbers of candidates are shown in Fig. 11. We can observe trends that are similar to the non-multi-probing case. Note that when 8 bits are omitted from each sub-hash value, the candidate list covers more than 25% of the whole database, which is probably not desirable

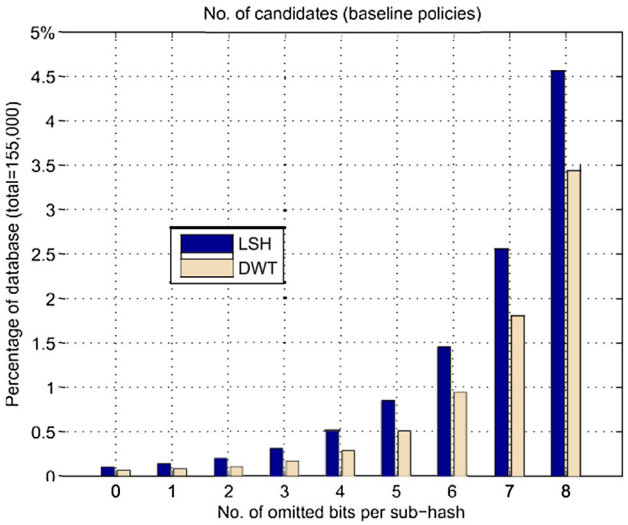

Fig. 7 Average number of candidates per query (Case 2, baseline policies). The candidate number increases exponentially with the policy

in practice. The corresponding empirical entropy values of candidate categories are shown in Fig. 12. Due to the large amount of candidates, the entropy stays high from the beginning.

Other retrieval policies Given a fixed number of bits to omit, which sub-hash value(s) should be chosen? We have tested some more privacy policies where not all sub-hash values are involved in bit omission. The results are shown in Fig. 13. Basically, the figure tells that for 8-bit omission, 8×1 schemes (omit 8 bits in one sub-hash value) are more effective than 2×4 schemes (which are more effective than 1×8 schemes). They are even more effective than 4×4 schemes (16-bit omission) which are more effective than 2×8

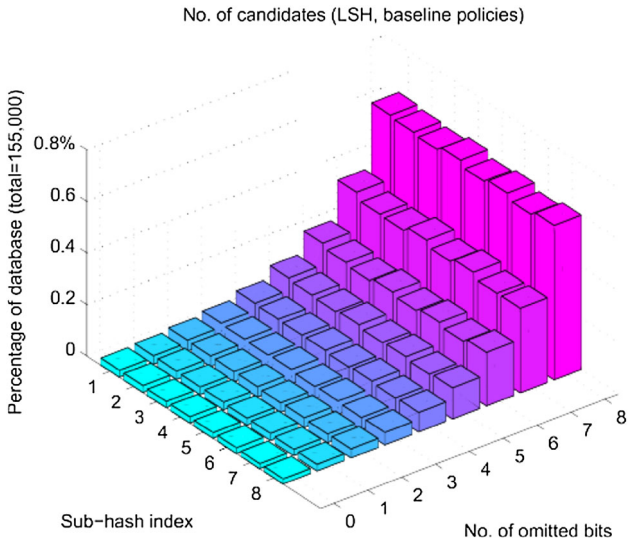

Fig. 8 Average no. of candidates per query (Case 2, LSH, baseline policies). The distribution of candidates among different sub-hash tables is quite even

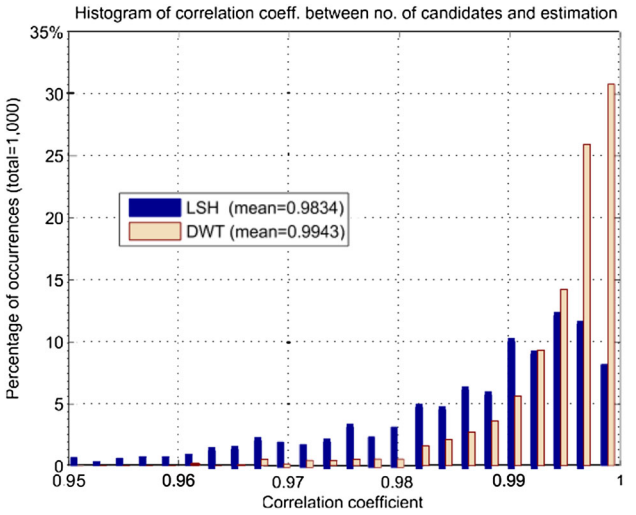

Fig. 10 Histogram of the correlation coefficient between the no. of candidates and estimation (Case 2, baseline policies, close-up). Good correlation implies that the system behaviour can be predicted in advance

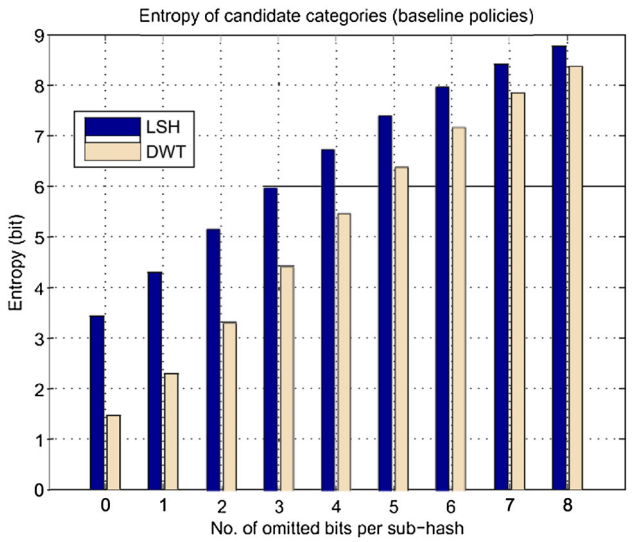

Fig. 9 Entropy of candidate categories per query (Case 2, baseline policies)

schemes. Therefore, we can conclude that in order to achieve a larger number of candidates, one should:

- Concentrate the omitted bits in fewer sub-hash values rather than spread them in more sub-hash values.

This is consistent with the definition of N2 in (2), because the number of candidates increases exponentially with b, but linearly with n. Another observation is that there is no significant difference between sub-hash values (thanks to dimension reduction). This is consistent with the outcomes in Fig. 8.

Another question is that, within a sub-hash value, which bits should be chosen first? According to the design of robust hashing, the bits in a sub-hash value are (ideally) equally important. This is verified by experiments. A single bit with a fixed position is omitted from each sub-hash value. The results are shown in Fig. 13. One can see that the bit position within a sub-hash value makes no big difference. This is consistent with the assumption.

Influence on retrieval one can imagine that privacy enhancement actually forces the server to behave like multi-probing. Therefore, privacy enhancement should improve retrieval performance and will get a better outcome. Figure 14 shows a ROC curve comparison of LSH with and without using privacy enhancement. In this figure, "private x" means x bits are

The same trend can be observed. By increasing the search radius, we may get more irrelevant items in the candidate list. The retrieval performance is not degraded because we use threshold-based decision making. Since the threshold stays the same, most irrelevant items are filtered out.

Majority voting attack The majority voting attack has been applied to estimate the query's category and ID. Specifically, the most frequent category or ID in the candidate list is considered as the one of the query. The rate of success is shown in Fig. 15 for query category estimation with DWT, and in Fig. 16 for query ID estimation with LSH. Since the attack is unlikely to succeed in Case 1, we only consider Case 2 and 3. The baseline privacy policies are used.

Several results are worth noticing. First, majority voting indeed works to some extent when there are near-duplicates. That means the majority of a candidate list is likely to be the near-duplicates of the query. On the one hand, this is

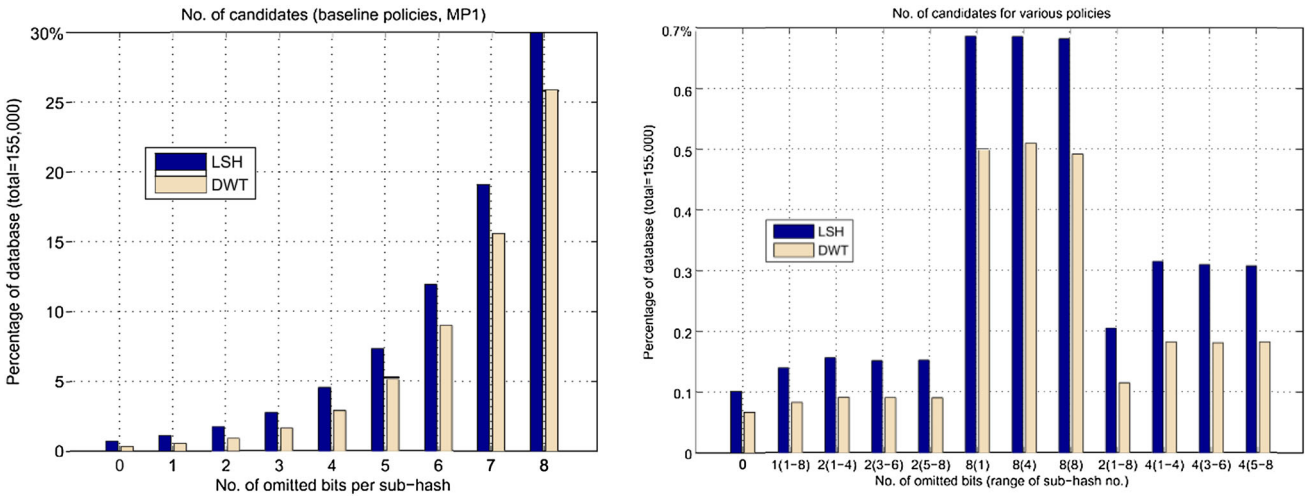

Fig. 11 Average no. of candidates per query (Case 2, baseline, MP1). Multi-probing makes privacy protection much stronger, but it is not always desired

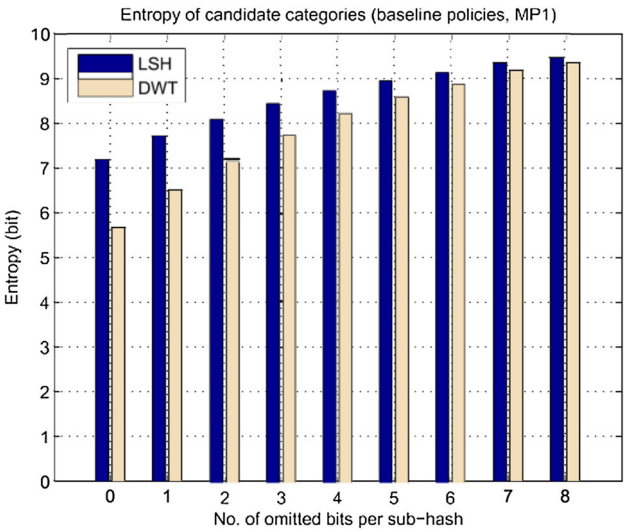

Fig. 12 Entropy of candidate categories per query (Case 2, baseline, MP1). Multi-probing makes privacy protection much stronger

perfectly reasonable, because an effective CBIR algorithm is supposed

To behave like that; on the other hand, this implies a threat to the privacy-preserving mechanism. Second, note that the success rate decreases when the number of omitted bits increases. Therefore, in order to prevent majority voting, the number of omitted bits should not be too small. Also note that multi-probing actually reduces the success rate. This is because probing more hash buckets makes the candidate list more noisy. In practice, the attack may not be so straightforward, because each item may have a different number of relevant items, and the database composition can be much more complex than in our cases. For a particular query, the success rate of majority voting depends on the number of relevant items in the database. Majority voting is likely to work for those items that have most relevant items, e.g., the most popular ones.

Additionally, the attack is also affected by the database density. Since the success rate in Case 3 is much lower than in Case 2, we conclude that for a fixed indexing scheme, the success rate decreases with the number of distinct items. In the next section, we provide a more formal security analysis from both the client's and the server's points of view.

5 Analysis and secured retrieval of data

Attacks on a retrieval system can be categorized according to the opponent's resources, computing capability, and the application scenarios. The security notions below are defined in a similar way as in MAC algorithms. First we define the power of an attack.

An attack is verifiable if the attacker knows beforehand that the attack will succeed with a high probability. In a PCBIR service, a client is virtually free to perform all the attacks against the server. As long as the service policy allows, he can send as many (kinds of) queries as he wants. The server, on the other hand, is not able to perform (adaptive) chosen-text attacks, because the protocol is not interactive and must be initiated by the client.

Basically, our protocol works as follows: the server receives a reduced hash value and a privacy policy from the client, and returns many hash values and IDs (Meta data). Since no actual content is sent, the confidentiality of the query and of the database are both preserved to a certain extent. The hash value does reveal some information about the input content. However, this information leak is not sufficient to create a verifiable attack. A multimedia object (pictures,

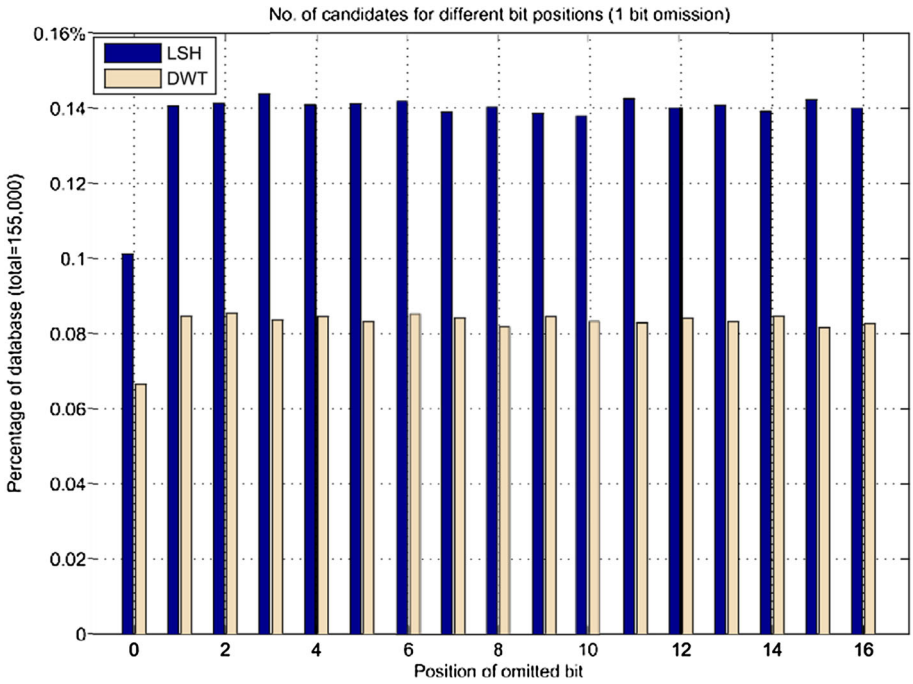

Fig. 13 No. of candidates for different 1 bit omission policies (Case 2). Within a sub-hash value, the position of a bit does not matter

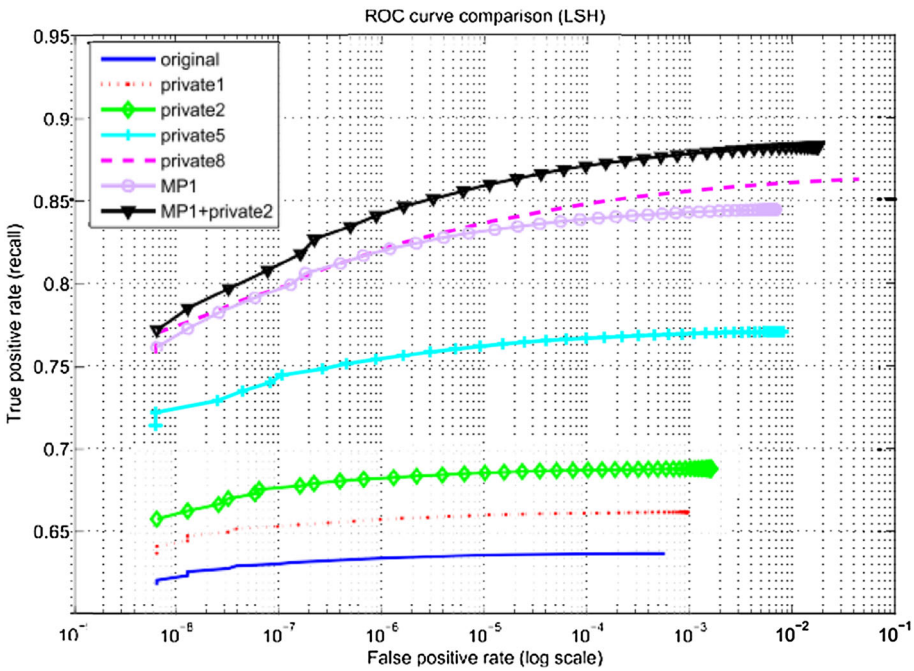

Fig. 14 ROC curve comparison (Case 2, LSH). "Original" means no privacy enhancement. "Private x" means x bits are randomly omitted from each sub-hash value. "MP x" means multi-probing within Hamming radius x

audio- visual clips) typically takes hundreds of kilobytes, while a hash value only takes hundreds of bits. The strong lossy compression makes it unlikely to invert a hash rate and value. The latest results show that image reconstruction from descriptors with the help of existing knowledge (known-text attack) is not really satisfactory [31,32]. Hash values in fact can be considered as extremely compressed versions of descriptors. To the best of our knowledge, so far there is no existing work on content reconstruction from hash values. The one-wayness of random projection based approaches has been proved in terms of content indistinguishability, which is adapted from the indistinguishability definition widely used in crypt- analysis. This property basically says an adversary cannot distinguish two different inputs with sufficient confidence by observing their hash values.

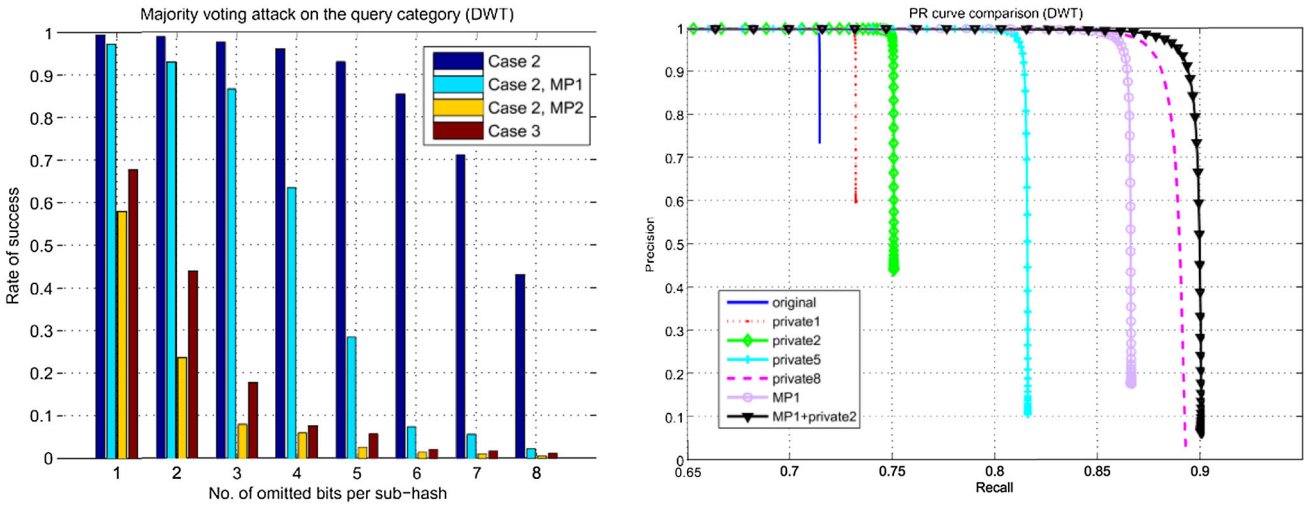

Fig. 15 Majority voting attack on query category (DWT). The effect decreases with the privacy policy

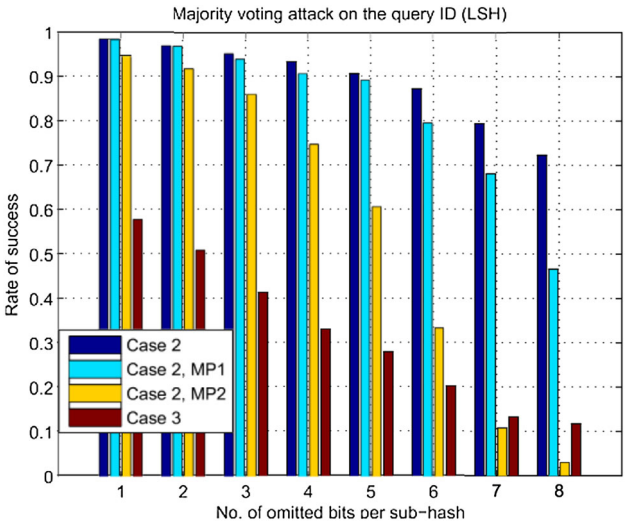

Fig. 16 Majority voting attack on query ID (LSH). The effect decreases with the privacy policy

5.1 Client privacy (P1, P2)

The server can guess the query content from its hash value. This is computationally difficult as explained above. Therefore, P1 is always guaranteed. If there is a match in the database, then the scope of the query is narrowed down to the candidates. The success rate of a brute-force attack is inversely proportional to the number of candidates. In order to quantify the privacy protection P2 at this stage, we can use k-diversity, which links to the notion of k-anonymity. A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k − 1 individuals whose information also appear in the release. K-diversity requires that for each query, the server should return at least k distinct answers, which can be controlled by the privacy policy.

In a known-text attack, the server already knows some of the client's past queries and/or query hash values. In general, a new query is more likely to be related to past ones. This principle helps to further narrow down the query scope. The effectiveness of this strategy depends on the behaviour of both parties. If the server always updates the client's profile by observing new queries, then the client can lead the server to build a wrong profile by sending fake queries. If the client always tends to query similar things, then he is more vulnerable to profiling. To conclude, it is good practice to make the k-diversity sufficiently large and sometimes send fake queries. In addition, clients sometimes have the option to be anonymous by using e.g. T or, 2 which makes profiling even less effective.

5.2 Server privacy (P3)

We assume that the client's interest is to know what is in the database. It is computationally difficult for the client to guess the database content from received hash values. Another possible attack is to pre-compute the hash values of many multimedia objects offline, and compare them with the hash values returned by the server during past queries. This is negligibly more efficient than directly querying the server online, because the computational difficulty mainly lies in figuring out the content. In a known-text attack, the client knows some content in the database. Since related content is likely to exist together, he can give priority to verify the existence of related content. In a (adaptive) chosen-text attack, the client can adapt his search in a more efficient way. For example, he can send online queries to the server, and meanwhile perform the offline search. However, recall that

we assume the server has much higher computing power than the client, which is often true in practice. The attacks mentioned in this section are only feasible for clients which are as powerful as or even more powerful than the server. They are somewhat beyond the setting of a curious-but-honest model. On the other hand, these attacks can be thwarted by making the hash generation dependent on a key. The server should periodically update the hash generation key and re-compute hash values of the database. The key should be communicated to the client before a query session. Note that the LSH algorithm is actually key dependent. The Gaussian vectors for projection are generated by a pseudo-random number generator (PRNG). The seed to the PRNG acts as the key. The DWT algorithm is not keyed, but there are keyed alternatives based on wavelets.

6 Conclusion and discussion

For the small dataset with 1000 images the accuracy rate would be 98.6% but with a large data set (> 10000 images) the accuracy would be 96% without losing the time complexity requirement. The content features extraction seems to be reliable compared to the existing algorithms, the DBN generates a huge data set for learning features and provides a good classification to handle the finding of the efficient content extraction. The framework has been implemented and extensively evaluated in different scenarios. As a future enhancement this same mention can be forward to real time extraction.

References

1. Lew, M.S., Sebe, N., Djeraba, C., Jain, R.: Content-based multimedia information retrieval: State of the art and challenges. ACM Trans. Multimed. Comput. Commun. Appl. **2**(1), 1–19 (2006)
2. Bringer, J., Chabanne, H., Patey, A.: Privacy-preserving biometric identification using secure multiparty computation: an overview and recent trends. IEEE Signal Process. Mag. **30**(2), 42–52 (2013)
3. Aghasaryan, A., Bouzid, M., Kostadinov, D., Kothari, M., Nandi, A.: On the use of LSH for privacy preserving personalization. In: Proceedings of the 12th IEEE International Conference Trust, Security, Privacy in Computing and Communications (TrustCom). pp. 362–371 (2013)
4. Fanti, G., Finiasz, M., Ramchandran, K.: One-way private media search on public databases: the role of signal processing. IEEE Signal Process. Mag. **30**(2), 53–61 (2013)
5. Acar, G., et al.: FPDetective: dusting the web for fingerprinters. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS), pp. 1129–1140 (2013)
6. Balsa, E., Troncoso, C., Diaz, C.: OB-PWS: Obfuscation-based private web search. In: Proceedings of the IEEE Symposium on Security and Privacy, pp. 491–505 (2012)
7. Erkin, Z.: Protection and retrieval of encrypted multimedia content: when cryptography meets signal processing. EURASIP J. Inf. Secur. **2007**, 20 (2007)
8. Lagendijk, R.L., Erkin, Z., Barni, M.: Encrypted signal processing for privacy protection: conveying the utility of homomorphic encryption and multiparty computation. IEEE Signal Process. Mag. **30**(1), 82–105 (2013)
9. Shashank, J., Kowshik, P., Srinathan, K., Jawahar, C.V.: Private content based image retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
10. Sabbu, P.R., Ganugula, U., Kannan, S., Bezawada, B.: An oblivious image retrieval protocol. In: Proceedings of the IEEE International Workshops of Advanced Information Networking and Applications (WAINA), pp. 349–354 (2011)
11. Erkin, Z., Franz, M., Guajardo, J., Katzenbeisser, S., Lagendijk, I., Toft, T.: Privacy-preserving face recognition. In: Proceedings of the 9th International Symposium on Privacy Enhancing Technologies (PETS), pp. 235–253 (2009)
12. Sadeghi, A.-R, Schneider, T., Wehrenberg, I.: Efficient privacy-preserving face recognition. In: Proceedings of the 12th International Conference on Information Security and Cryptology (ICISC), pp. 229–244 (2009)
13. Osadchy, M., Pinkas, B., Jarrous, A., Moskovich, B.: SCiFI—a system for secure face identification. In: Proceedings of the IEEE Symposium on Security and Privacy (SP), pp. 239–254 (2010)
14. Barni, M., Failla, P., Lazzeretti, R., Sadeghi, A., Schneider, T.: Privacy-preserving ECG classification with branching programs and neural networks. IEEE Trans. Inf. Forensics Secur. **6**(2), 452–468 (2011)
15. Hsu, C.-Y., Lu, C.-S., Pei, S.-C.: Image feature extraction in encrypted domain with privacy-preserving SIFT. IEEE Trans. Image Process. **21**(11), 4593–4607 (2012)
16. Annamalai, R., Srikanth, J.: Integrity and privacy sustenance of shared large scale images in the cloud by ring signature. Int. J. Comput. Appl. **114**(12), 13–18 (2015)
17. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the ACM SIGMOD International Conference on the Management of Data, pp. 439–450 (2000)
18. Duchi, J.C., Jordan, M.I., Wainwright, M.J.: Advances in Neural Information Processing Systems 25. Privacy aware learning, pp. 1430–1438. Curran Associates, Red Hook (2012)
19. Rane, S., Boufounos, P.T.: Privacy-preserving nearest neighbor methods: comparing signals without revealing them. IEEE Signal Process. Mag. **30**(2), 18–28 (2013)
20. Lu, W., Swaminathan, A., Varna, A.L., Wu, M.: Enabling search over encrypted multimedia databases. In: Proceedings of the SPIE, Media Forensics and Security, vol. 7254, pp. 725418-1–725418-11, (2009)
21. Hoiem, D., Sukthankar, R., Schneiderman, H., Huston, L.: Object-based image retrieval using the statistical structure of images? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2004)
22. Xue, B., Wanjun, L.: Research of Image Retrieval Based on Color. In: Proceedings of the IEEE International Forum on Computer Science-Technology and Applications, pp. 283–286 (2009)
23. Huang, Z.C., Chan, P.P.K., Ng, W.W.Y., Yeung, D.S.: Content-based image retrieval using color moment and Gabor texture feature, in Proceedings the IEEE Ninth International Conference on Machine Learning and Cybernetics, Qingdao. pp. 719-724 (2010)
24. Lu, W., Varna, A.L., Swaminathan, A., Wu, M.: Secure image retrieval through feature protection. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP). pp. 1533–1536 (2009)
25. Annamalai, R., Srikanth, J.: Accessing the data efficiently using prediction of dynamic data algorithm. Int. J. Comput. Appl. **116**(22), 39–42 (2015)

26. Andoni, A., Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. Commun. ACM **51**(1), 117–122 (2008)
27. Boufounos, P., Rane, S.: Secure binary embeddings for privacy pre- serving nearest neighbors. In: Proceedings of the IEEE International Workshop on Information Forensics and Security, (WIFS), pp. 1–6 (2011)
28. Gasarch, W.: A Survey on Private Information Retrieval, in Bulletin of the EATCS, vol. 82, pp. 72–107. EATCS, Rio (2004)
29. Ostrovsky, R., Skeith, W.E.: A survey of single-database private information retrieval: techniques and applications. In: Proceedings of the 10th International Conference on the Practice and Theory in Public-Key Cryptography, pp. 393–411 (2007)
30. Danezis, G., Gürses, S.: A critical review of 10 years of privacy technology. In: Proceedings of the 4th Surveillance and Society Conference, (2010)
31. Weinzaepfel, P., Jégou, H., Perez, P.: Reconstructing an image from its local descriptors. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 337–344 (2011)
32. Khelifi, F., Jiang, J.: Perceptual image hashing based on virtual watermark detection. IEEE Trans. Image Process. **19**(4), 981–994 (2010)
33. Özer, H., Sankur, B., Memon, N., Anarim, E.: Perceptual audio hashing functions. EURASIP J. Appl. Signal Process. **2005**, 1780–1793 (2005)
34. Varna, A.L., Wu, M.: Modeling and analysis of correlated binary fin- gerprints for content identification. IEEE Trans. Inf. Forensics Secur. **6**(3), 1146–1159 (2011)
35. Cano, P., Batlle, E., Kalker, T., Haitsma, J.: A review of audio fingerprinting. J. VLSI Signal Process. Syst. **41**(3), 271–284 (2005)
36. Andoni, A., Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 459–468 (2006)
37. Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
38. Mu, Y., Shen, J., Yan, S.: Weakly-supervised hashing in kernel space. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3344–3351 (2010)
39. Cao, L., Li, Z., Mu, Y., Chang, S.-F.: Submodular video hashing: a unified framework towards video pooling and indexing. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 299–308 (2012)
40. Lv, Q., Josephson, W., Wang, Z., Charikar, M., Li, K.: Multi-probe LSH: Efficient indexing for high-dimensional similarity search. In: Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB), pp. 950–961 (2007)
41. Joly, A., Buisson, O.: "A posteriori multi-probe locality sensitive hashing," in Proc. 16th ACM Int. Conf. Multimedia. pp. 209–218 (2008)
42. Zhang, W., Gao, K., Zhang, Y.-D., Li, J.-T.: Data-oriented locality sensitive hashing. In: Proceedings of the ACM International Conference on Multimedia, pp. 1131–1134 (2010)
43. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)

R. Rani Saritha is an Assistant Professor at the Department of Computer Applications, SAINTGITS College of Engineering which is affiliated to Kerala Technological University. Her research interests include Digital Image Processing, Image Retrieval, Big Data, Information Security, etc. Rani Saritha R has taken the PG Degree in Information and Communication Engineering from Anna University, Chennai and PG Degree in Computer Applications from the University of Madras. She has published papers in International Journal of Scientific & Engineering Research and International Journal of Innovative Research in Computer and Communication Engineering. She is a member of Indian Society for Technical Education (ISTE), Computer Society of India (CSI), International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT).

Varghese Paul is currently working as Dean CS/IT in the Department of Computer Science and Information Technology, TocH Institute of Science and Technology, Ernakulam and research Supervisor of Cochin University of Science and Technology, M G University Kottayam, Anna Technical University Chennai, Bharathiar University Coimbatore, Bharathidasan University Trichy and Karpagam University Coimbatore. His research areas are Data Security using Cryptography, Data Compression, Data Mining, Image Processing and E_Governance. He has developed a TDMRC Coding System for character representation in computer systems and encryption systems. He has published 23 papers in International as well as national journals. He has worked as an Industrial Engineer with O/E/N India Ltd., Cochin, Communication Engineer with KSE Board, SCADA Engineer in Saudi Electricity Department and Head of IT Department CUSAT. He is the Life Member of Indian Society for Technical Education (ISTE), Information Systems Audit and Control Association USA (ISACA) and National Geographic Society, USA

P. Ganesh Kumar is currently working as an Assistant Professor in the Department of Information Technology, Anna University Regional Campus, Coimbatore. He is the recipient of student scientist award from TNSCST in 2003, IEEE best paper award in 2007, IET best paper award in 2011, Korean IISE best paper award in 2015, travel grant as Young Scientist from DST in 2013 to Singapore, Workshop Grant from DBT in 2014, travel grant as from DBT and CSIR in 2014 to USA. His research interests include Application of soft computing techniques for data mining based problems in Bio-informatics, Wireless Sensor Networks, Smart Grid, Image Processing, Cloud Computing and Big Data Analytics. He has published his research work in 17 International SCI/Scopus journals, 37 International Renowned Conferences and 58 National Conferences. He is the resource person for delivering technical talk in UGC, AICTE, TEQIP, ICMR, IETE, and IEEE sponsored seminars in various technical institutions. He is the Life Member of ISTE, IEEE Senior Member, Professional member of IEEE-Computational Intelligence Society, Intelligent Transportation Society, ACM and CSI. He is a Post Doctoral Fellow during April 2015 to March 2016 in Kyungpook National University, South Korea.