Privacy Protection in Interactive Content Based Image Retrieval

Yonggang Huang, *Member, IEEE,* Jun Zhang, *Member, IEEE,* Lei Pan, *Member, IEEE,*
and Yang Xiang, *Senior Member, IEEE.*

Abstract—Privacy protection in Content Based Image Retrieval (CBIR) is a new research topic in cyber security and privacy. The state-of-art CBIR systems usually adopt interactive mechanism, namely relevance feedback, to enhance the retrieval precision. How to protect the user's privacy in such Relevance Feedback based CBIR (RF-CBIR) is a challenge problem. In this paper, we investigate this problem and propose a new Private Relevance Feedback CBIR (PRF-CBIR) scheme. PRF-CBIR can leverage the performance gain of relevance feedback and preserve the user's search intention at the same time. The new PRF-CBIR consists of three stages: 1) private query; 2) private feedback; 3) local retrieval. Private query performs the initial query with a privacy controllable feature vector; private feedback constructs the feedback image set by introducing confusing classes following the K-anonymity principle; local retrieval finally re-ranks the images in the user side. Privacy analysis shows that PRF-CBIR fulfills the privacy requirements. The experiments carried out on the real-world image collection confirm the effectiveness of the proposed PRF-CBIR scheme.

Index Terms—CBIR, Relevance Feedback, Image Privacy, K-anonymity.

1 Introduction

With the pervasive connectivity and availability of the Internet and digital imaging technologies, massive digital images are created and used in various fields, including medicine, publishing, education, and so on. Facing big image datasets, effective retrieval becomes a fundamental requirement [1, 2]. Currently, two image retrieval approaches dominate [2]: text-based and content-based. In text-based image retrieval, images are searched based on the text descriptions associated with the images. In content-base image retrieval (CBIR) [1, 3], images are retrieved according to their visual similarities measured on low-level visual features. CBIR still works when textual annotations are not available, and has been implemented in the state-of-art image search engines, such as Google Image Search[1], and Bing Image Search[2].

Considering the rich sensitive information embedded in images, recently, image privacy has been drawing more and more attention. In May, 2016, a lawsuit alleged that the photo-tagging system of Facebook violates user privacy[3]. In fact, Facebook stopped the automatic application of facial recognition technology in Europe[4] three years ago, facing similar privacy concerns. To address serious privacy concerns, Google has also forbidden apps to use the face recognition feature on Google Glasses[5].

A few works have been reported on image privacy issues in CBIR.

- Most of existing research efforts [4–10] focused on privacy protection for CBIR in the cloud computing environment [11]. In this scenario, the privacy issue arises because of untrustworthy cloud. The main challenge is to leverage the tremendous computing power of the cloud for CBIR, whilst preventing the cloud from learning anything useful about the image dataset and query. While the images can be protected separately using image encryption technologies, the main concern is how to perform the similarity computation among the image features in a privacy-preserving way. Feature/index randomization [4–6] and encryption [7, 9, 10] are the commonly adopted methods.
- Very few works [12, 13] have been devoted to address the privacy issue in the viewpoint of the untrustworthy CBIR service provider. The major privacy concern in this situation is that the user's search intention could be learned by the service provider. Based on the search intention, service provider can infer the user's profile, such as user's interest, living place, health condition and even commercial secret [12]. For example, in the medical field, query with tumor images could leak the user's health condition to service provider. In the viewpoint of privacy protection for big data [6], user's search intention on CBIR is a primary source to profile user for potential malicious activities. Our work focuses on this privacy issue.

Because of the semantic gap [1] existing between visual features and semantic concepts, relevance feedback [14–16] is integrated in traditional CBIR systems [1, 3] to guarantee retrieval accuracy. Existing works [12, 13] overlook this point.

This work is to investigate the privacy issue in the Relevance Feedback based CBIR (RF-CBIR). Our goal is to develop a

Y. Huang is with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China (e-mail: yonggang.h@gmail.com)
J. Zhang, and Y. Xiang are with the School of Software and Electrical Engineering, Swinburne University of Technology, VIC, 3122, Australia (e-mail: junzhang@swin.edu.au; yxiang@swin.edu.au)
L. Pan is with the School of Information Technology, Deakin University, VIC, 3217, Australia (e-mail: l.pan@deakin.edu.au)
Manuscript received December 10, 2016; revised December 21, 2017.

[1] https://images.google.com

[2] http://www.bing.com/images

[3] http://www.theverge.com/2016/5/5/11605068/facebook-photo-tagging-lawsuit-biometric-privacy

[4] https://www.engadget.com/2013/02/07/facebook-european-facial-recognition-data-removal/

[5] https://developers.google.com/glass/policies

[6] https://www.theguardian.com/technology/2014/jun/20/little-privacy-in-the-age-of-big-data

new CBIR scheme, which can leverage the performance gain of relevance feedback and preserve the user's search intention at the same time. The major contributions of our work are summarized as follows.

- We propose a new Private Relevance Feedback based CBIR scheme (PRF-CBIR), which consists of three stages: private query, private feedback and local retrieval.
- We present a new private query method, which performs the initial query with a privacy controllable feature vector.
- We develop a new private feedback method, which introduces confusing classes into the feedback image set, to protect user intention.
- We conduct privacy analysis to confirm the new scheme can address all known attacks, even the attackers use smart learning technologies.

The rest of the paper are structured as follows. Some related work is briefly reviewed in Section II. Section III presents the system model and threat model of RF-CBIR. In Section IV, a new scheme is proposed to address the privacy issue in RF-CBIR. Section V provides a privacy analysis on the newly proposed approach. Section VI reports the experiments and results. Finally, the paper is concluded in Section VII.

2 Related work

To address privacy problems of CBIR, existing researches focus on the two categories based on the potential malicious attackers — an untrustworthy cloud provider or a untrustworthy CBIR service provider.

2.1 Untrustworthy Cloud Provider

Cloud provider holds a shared pool of configurable computing resources [11], and can provide CBIR outsourcing services. Facing a malicious cloud provider, the biggest challenge for user is how to compute the visual similarities using computing power of the cloud, while preventing the cloud from learning useful information about the image dataset and the query. Most of existing research efforts fall in this category.

Lu *et al.*'s work [4–6] are the first endeavors in this domain. In [5, 6], feature randomization technologies, including bitplane randomization, random projection, and randomized unary encoding, are used to scramble visual features whilst approximately preserving their distances. In [4, 6], Lu *et al.* randomized the search index using secure inverted index and secure min-Hash with approximate distance preserving property. Yuan *et al.* [7] proposed a lightweight secure image search scheme, namely SEISA. In SEISA, matrix based encryption is used to encrypt image features, which can achieve high search efficiency and accuracy. Xia *et al.* [8] presented a privacy-preserving CBIR approach which supports local-feature based CBIR with the earth movers distance (EMD) as similarity metric. In their approach, the EMD problem is transformed to a linear programming problem. The cloud provider then solves the linear programming problem without learning the sensitive information. Ferreira *et al.* [9] proposed a new Image Encryption Scheme with Content-Based Image Retrieval properties (ES-CBIR). Their work was motivated by the observation that texture feature is usually more relevant than color in object recognition. They encrypted texture information with a probabilistic scheme, and used deterministic encryption on image color information. This methodology allows CBIR based on color information with a high level of privacy protection. Zhang *et al.* [10] introduced a Privacy-preserving large-scale Image search system on Cloud (PIC). The core of PIC is employing multi-level homomorphic encryption protocols to conduct distance calculation for feature vectors.

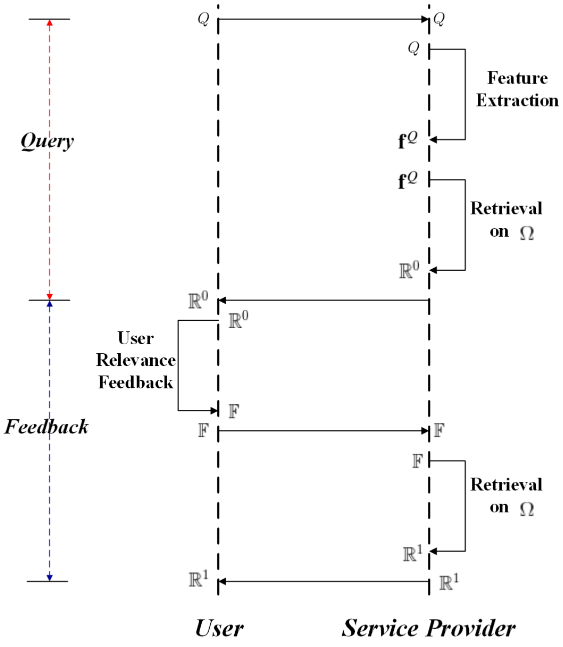


Fig. 1. The flow chart of RF-CBIR

2.2 Untrustworthy CBIR Service Provider

Service provider owns an enormously huge number of images and are willing to provide the CBIR service on the dataset. Service provider can provide CBIR service via its own servers or outsourcing to cloud provider. Facing an untrustworthy service provider, the problem is how to prevent it from inferring the user's search intention. To our best knowledge, literatures [12, 13] are the only two research efforts in this category.

Shashank *et al.* [12] are the pioneers in this direction. In their solution, the image features are organized to a hierarchical structure, and the retrievals are regarded as traverse over the hierarchical structure. In order to avoid revealing search intention, the user keeps the traverse path confidential to the service provider through exchanged messages based on quadratic residuosity assumption. However, this approach increases the interactions between the user client and the service provider, where both sides need interacting p times to fetch a node with p bits. Most recently, Weng *et al.* [13] studied this problem in the scenario of the near duplicate detection based on robust hashing and piece-wise inverted indexing. The basic idea is to randomly omit certain bits in the query to prevent the service provider from predicting the user's search intention accurately. However, this randomly query feature omitting approach is appropriate for near duplicate detection, but not suitable for general CBIR task.

3 Privacy Issues in Traditional RF-CBIR systems

3.1 RF-CBIR Model

Fig.1 presents the flow chart of the traditional RF-CBIR. In general, RF-CBIR consists of two stages, query and feedback.

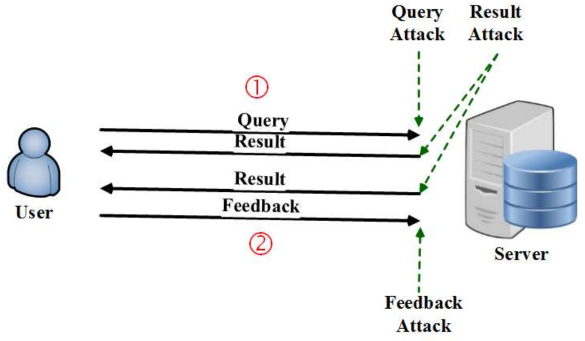


Fig. 2. Privacy Threat in RF-CBIR.

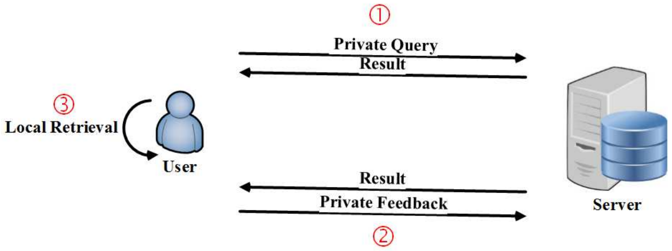


Fig. 3. The system model of PRF-CBIR.

In the stage of query, the user submits a query image Q to the service provider. The service provider extracts the low-level features of Q, denoted as $\mathbf{f}^Q$, and then retrieves from an image dataset Ω by computing the Euclidean distances with respect to the low-level features. The top N similar images of Q are then returned as initial result $\mathbb{R}^0$.

In the feedback stage, the user labels some relevant images as $\mathbb{F}$, and submits $\mathbb{F}$ to the service provider. The service provider regards the images in $\mathbb{F}$ as positive examples, and treats randomly selected images from $\mathbb{R}^0 - \mathbb{F}$ as negative examples. Machine learning based classifiers (e.g., Support Vector Machine (SVM) classifier [17]) are used to rank the images in Ω. The top N ranked images are returned as refined result $\mathbb{R}^1$.

3.2 Privacy Threat

We assume that the service provider behaves in a curious-but-honest way. That is, the service provider returns the retrieval result as accurately as possible. Meanwhile, it will infer the user's search intention curiously by deeply analysing the user related data.

The privacy threat of RF-CBIR is shown in Fig.2. Providing the data exchanged with the user, the service provider can learn the search intention through the following attacks:

- *Query Attack*: The service provider acquires the user's search intention directly with the query image Q. For example, if the user queries with a tumor image, the service provider can easily infer the user's search intention is tumor.
- *Result Attack*: By analysing the results returned to user, $\mathbb{R}^0$ and $\mathbb{R}^1$, the service provider may be very likely to infer the search intention. For instance, if the majority of the result are tumor images, it is likely that the user's search intention is tumor.
- *Feedback Attack*: The user's feedback image set $\mathbb{F}$, also can be utilized by the service provider to learn the search intention. Suppose the user labels some tumor images as relevant, the service provider can acquire that the search intention is tumor.

4 Proposed scheme

This section presents the details of the new PRF-CBIR scheme. Fig.3 and Fig.4 show the system model and the flow chart of PRF-CBIR respectively. PRF-CBIR has three stages — private query, private feedback and local retrieval.

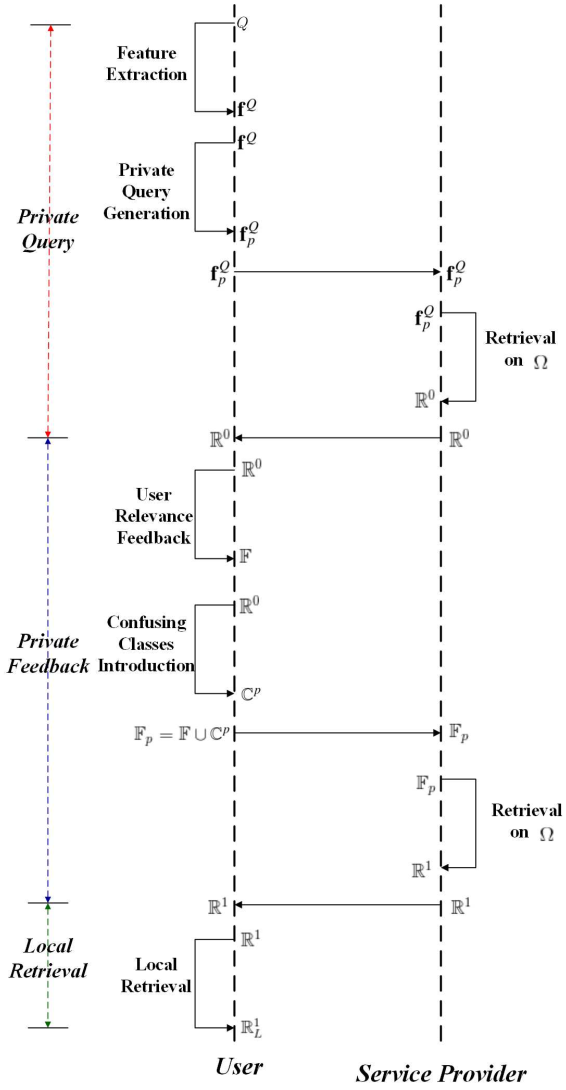


Fig. 4. The flow chart of PRF-CBIR.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2018.2793923, IEEE Transactions on Dependable and Secure Computing

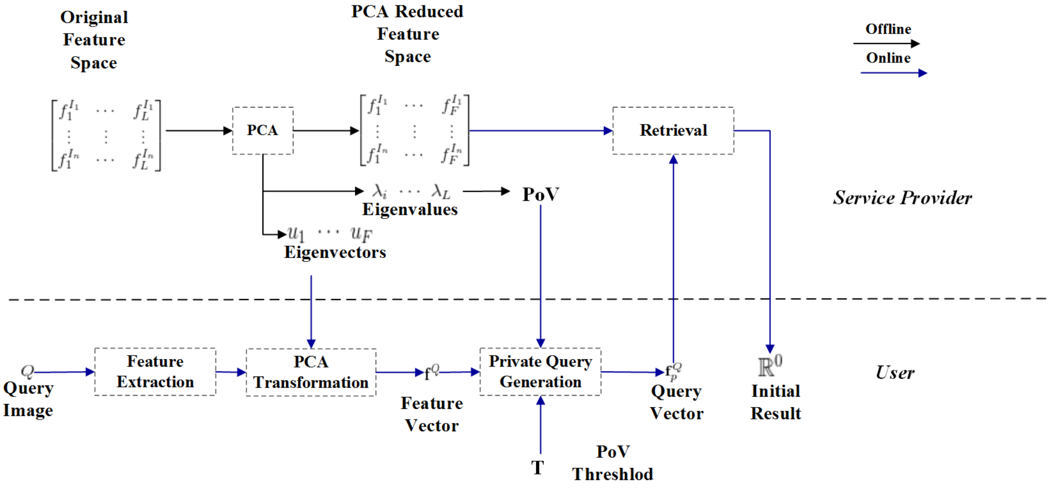


Fig. 5. The flow chart of private query.

4.1 Private Query

Private query is proposed to address the query attack and the result attack on $\mathbb{R}^0$. The basic idea is, instead of query image Q, we use a part of the Q's feature $\mathbf{f}_p^Q$ as query. In this way, the query attack can be avoided, and the result attack on $\mathbb{R}^0$ can be alleviated by adjusting the privacy information contained in $\mathbf{f}_p^Q$. Private query is developed by utilizing the Percentage of Variance (PoV) [18] defined in the Principal Component Analysis (PCA) [18] reduced feature space.

In this work, PCA is performed prior to retrieval for two goals. Firstly, PCA is adopted for dimension reduction [19] to avoid the curse of dimensionality. PCA can perform dimensionality reduction while preserving as much of the variance in the high-dimensional space as possible. Secondly, we make use of PoV, defined in the PCA reduced feature space, to measure privacy and generate query vector $\mathbf{f}_p^Q$. This is based on the observation that PoV can characterize the importance of different feature components.

Fig.5 shows the flow chart of the proposed private query method, which consists of offline stage and online stage.

4.1.1 Offline Stage

In this stage, the service provider performs PCA on the image dataset, and computes the PoV values of PCA feature components. Suppose the image dataset is $\Omega = \{I_1, \cdots, I_i, \cdots, I_n\}$, the feature vector of image $I_i \in \Omega$ can be denoted as $\mathbf{f}^{I_i} = \left(f_1^{I_i}, \cdots, f_j^{I_i}, \cdots, f_F^{I_i}\right)^T$. And the feature vectors of all images in Ω can be represented as a matrix $\mathbf{F} = \left(\mathbf{f}^{I_1}, \cdots, \mathbf{f}^{I_i}, \cdots, \mathbf{f}^{I_n}\right)^T$. PCA is then employed to learn a linear transformation as,

$$\mathbf{F}_L = \mathbf{F}(u_1, \cdots, u_i, \cdots, u_F), \tag{1}$$

where $\mathbf{F}_L$ keeps only the first L ($L < F$) important feature components. Those feature components are sorted in the descending order of importance. The learned weight vectors $u_1, \cdots, u_i, \cdots, u_F$ have two properties. Firstly, $u_1, \cdots, u_i, \cdots, u_F$ form an orthogonal basis for the L feature components. Secondly, they can preserve as much variability as possible in the original data. $u_1, \cdots, u_i, \cdots, u_F$ are also the top L eigenvectors of $\mathbf{F}$'s covariance matrix, and their eigenvalues $\lambda_1, \cdots, \lambda_i, \cdots, \lambda_F$ characterize the variance degree explained by the corresponding eigenvectors respectively.

For each PCA feature component f_i, we convert the eigenvalues to PoV as,

$$\mathsf{PoV}(f_i) = \frac{\lambda_i}{\sum_{j=1}^{L} \lambda_j}. \tag{2}$$

4.1.2 Online Stage

In this stage, the user generates the query vector $\mathbf{f}_p^Q$, and retrieves images from the service provider.

Suppose that the user's query image is Q, after feature extraction and PCA transformation, the PCA features of Q can be represented as $\mathbf{f}^Q$. The PoV threshold $\mathbf{T}$ and the query vector length l are set according to the user's privacy policy.

In this paper, PoV is adopted to measure the privacy of the query. According to (2), the PoV of $\mathbf{f}_p^Q$ can be calculated as,

$$\mathsf{PoV}(\mathbf{f}_p^Q) = \sum_{f_i \in \mathbf{f}_p^Q} \mathsf{PoV}(f_i). \tag{3}$$

Given PoV threshold $\mathbf{T}$ and l, the goal of private query generation is to choose a continuous segment from $\mathbf{f}^Q$ satisfying $\mathbf{T}$, which can be formulates as,

$$\begin{aligned} \min \quad & \left\| \mathbf{T} - \mathsf{PoV}(\mathbf{f}_p^Q) \right\|, \\ \text{s.t.} \quad & \mathbf{f}_p^Q \in \left\{ \mathbf{f}^Q[i : (i + l - 1)] \mid 1 \leqslant i \leqslant (L + 1 - l) \right\}, \end{aligned} \tag{4}$$

where $\mathbf{f}^Q[i : (i + l - 1)]$ denotes a continuous segment from $\mathbf{f}^Q$, with the index from i to $(i + l - 1)$.

Consider that the length of $\mathbf{f}^Q$ is small, we traverse $\mathbf{f}^Q$ sequentially to find appropriate $\mathbf{f}_p^Q$ satisfying (4).

Finally, the service provider performs retrieval with respect to $\mathbf{f}_p^Q$, and returns the N most similar images as the result $\mathbb{R}^0$.

4.2 Private Feedback

Relevance feedback can significantly improve the retrieval performance. However, since user labels relevant images during feedback, the user's search intention is completely exposed to the service provider. In order to leverage the performance gain and preserve the users' privacy at the same time, private feedback is developed here. Private feedback can deal with the feedback attack and result attack on $\mathbb{R}^1$. The new idea of private feedback is introducing confusing classes into feedback image set $\mathbb{F}_p$.

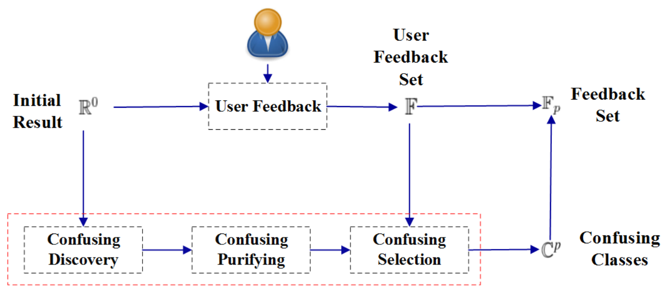


Fig. 6. The flow chart of confusing classes introduction.

$K-$anonymity [20] and differential privacy [21] are the two prevalent privacy preserving principles. $K-$anonymity is proposed in the scenario of data sharing. A release of data provides $K-$anonymity protection, if the information for each individual contained in the release cannot be distinguished from at least $K-1$ individuals. The motivating scenario for differential privacy is statistical database. The goal is to enable the user to query statistical information of the population, as represented by the database, while protecting the privacy of individuals. During private feedback, the user introduces confusing classes into feedback image set $\mathbb{F}_p$, and submits $\mathbb{F}_p$ to the service provider. This procedure resembles data sharing from the user to the service provider. K-anonymity better suits the scenario of preserving the privacy of relevance feedback than differential privacy. Therefore, we adopt $K-$anonymity as the privacy preserving principle. $K-$anonymity requires that each record is indistinguishable from at least $K-1$ other records. In this paper, we innovatively achieve $K-$anonymity feedback to protect the user's search intention. With the new technologies, the feedback images in $\mathbb{F}_p$ belong to K classes, making the target class of search indistinguishable from other $K-1$ classes. Accordingly, we define the $K-$anonymity feedback principle as follows.

Definition 1. ($K-$*anonymity Feedback Principle*) A relevance feedback is $K-$anonymity feedback, if in the feedback image set $\mathbb{F}_p$, the number of images belonging to the target class S is equal to the number of images belonging to $K-1$ other confusing classes $C_1, \cdots, C_{K-1}$. That is,

$$n(S, \mathbb{F}_p) = n(C_1, \mathbb{F}_p) = \cdots = n(C_{K-1}, \mathbb{F}_p), \tag{5}$$

where $n(C_i, \mathbb{F}_p)$ is the number of images in $\mathbb{F}_p$ belonging to class C_i.

During the relevance feedback, the user randomly labels some relevant images in $\mathbb{R}^0$ as $\mathbb{F}$. In order to make the feedback image set $\mathbb{F}_p$ satisfy the $K-$anonymity feedback principle in (5), a three-step approach is employed to introduce confusing classes, including confusing discovery, confusing purifying and confusing selection. The overall flow of confusing classes introduction is presented in Fig.6.

4.2.1 Confusing Discovery

The objective of this step is to discover confusing classes candidates in $\mathbb{R}^0$. In this paper, we use k-means [22] to cluster the images in $\mathbb{R}^0$, and regard the result clusters of k-means as confusing classes candidates.

The k-means relies on distances among images. Considering that $\mathbb{R}_0$ are retrieved with private query $\mathbf{f}_p^Q$, the distances among images on feature components in $\mathbf{f}_p$ are little. Therefore, we compute the distances using remained feature components in $\mathbf{f}$ as,

$$d(I_i, I_j) = \left\| \mathbf{f}_r^{I_i} - \mathbf{f}_r^{I_j} \right\|, \tag{6}$$

where $\mathbf{f}_r = \mathbf{f} - \mathbf{f}_p$ is the remained feature components in $\mathbf{f}$ after removing $\mathbf{f}_p$.

Given the initial retrieval result $\mathbb{R}^0$, the k-means clustering aims to partition it into k clusters ($k \leq |\mathbb{R}^0|$), $\mathbb{C} = \{\mathbb{C}_1, \cdots, \mathbb{C}_k\}$, to minimize the within-cluster sum of squares. That is,

$$\underset{\mathbb{C}}{\arg\min} \sum_{i=1}^{k} \sum_{I \in \mathbb{C}_i} \|I - \mu_i\|, \tag{7}$$

where μ_i is the mean of $\mathbb{C}_i$. The traditional k-means algorithm uses an iterative refinement technique to solve (7). Given an initial set of randomly selected k centroids, the algorithm proceeds by alternating between the assignment step and the update step. In the assignment step, each image is assigned to the cluster with the closest mean. In the update step, the new means are calculated to be the centroid of images in the cluster. The algorithm terminates until the assignments do not change.

The output of confusing discovery is k clusters, $\mathbb{C} = \{\mathbb{C}_1, \cdots, \mathbb{C}_k\}$, which will be used as confusing classes candidates in next step.

4.2.2 Confusing Purifying

The output of confusing discovery is k clusters, $\mathbb{C} = \{\mathbb{C}_1, \cdots, \mathbb{C}_k\}$. Ideally, each cluster should be in the same class. However, in practice, those clusters may be unclean. Therefore, we employ a consensus filters approach to further purify those clusters. The basic idea is, based on random subspace strategy [23], to construct a ensemble of SVMs [17] as filters. Then the outputs of SVMs are aggregated according to the consensus rules [24, 25].

Purifying is performed on clusters with the size a little larger than $\mathbb{F}$. That is, $|\mathbb{C}_i| \geq (1+\alpha)|\mathbb{F}|$, where α is an adjustment factor.

Regarding each cluster $\mathbb{C}_i$ as positive training dataset, and remained examples $\mathbb{R}^0 - \mathbb{C}_i$ as negative training dataset, an ensemble of SVMs are constructed by using the random subspace method. Random subspace method is an ensemble learning method. It constructs base classifier on random sampled features instead of the entire feature set. Given the original feature space $\mathbf{f}$, a base SVM classifier C_t is constructed on the subspace $\mathbf{f}_t$ randomly sampled from $\mathbf{f}$. Suppose T SVM classifiers are constructed, those SVMs are then used to classify the images in $\mathbb{C}_i$. According to the consensus rule, only the images labelled as positive by all SVMs are retained in $\mathbb{C}_i$.

Algorithm 1 presents the procedure of consensus filters based confusing purifying.

4.2.3 Confusing Selection

The goal of this step is to select $K-1$ clusters to form image set $\mathbb{F}_p$ satisfying the $K-$anonymity feedback principle.

Suppose the cluster set after purifying is $\mathbb{C}^p$, the valid clusters firstly should be of certain size, that is, $|\mathbb{C}_i| \geq (1+\alpha)|\mathbb{F}|$, where α is the adjustment factor. Secondly, the clusters should not contain any images in $\mathbb{F}$, as $\mathbb{C}_i \cap \mathbb{F} == \phi$. We randomly select $K-1$ clusters from those valid clusters. For each of those selected clusters, we randomly choose $(1+\alpha)|\mathbb{F}|$ images. Finally, those $K-1$ clusters with the size of $(1+\alpha)|\mathbb{F}|$ are added to the user feedback image set $\mathbb{F}$, forming the private feedback set $\mathbb{F}_p$.

Finally, regarding $\mathbb{F}_p$ as positive examples, and randomly selected images from $\mathbb{R}^0 - \mathbb{F}_p$ with the same size of $\mathbb{F}_p$ as negative

```
Input  : The clusters set : $\mathbb{C} = \{\mathbb{C}_1, \cdots, \mathbb{C}_k\}$,
         the initial retrieval result : $\mathbb{R}^0$.
Output: The purified clusters set : $\mathbb{C}^p$.
$\mathbb{C}^p \leftarrow \emptyset$;
// Choose the clusters with certain size.
$\mathbb{C}^{temp} \leftarrow \{\mathbb{C}_i | (\mathbb{C}_i \in \mathbb{C}) \wedge (|\mathbb{C}_i| \geq (1+\alpha)|\mathbb{F}|)\}$;
foreach Cluster $\mathbb{C}_i \in \mathbb{C}^{temp}$ do
    // Regard $\mathbb{C}_i$ as positive training dataset,
       and $(\mathbb{R}^0 - \mathbb{C}_i)$ as negative training
       dataset.
    $\mathbf{S}^+ \leftarrow \mathbb{C}_i$;
    $\mathbf{S}^- \leftarrow (\mathbb{R}^0 - \mathbb{C}_i)$;
    // Train $T$ SVMs using the random subspace
       strategy.
    for $t \leftarrow 1$ to $T$ do
        // Bootstrap $\mathbf{f}_t$ of length $L$ from
           original feature space $\mathbf{f}$.
        $\mathbf{f}_t \leftarrow$ Bootstrap$(\mathbf{f}, L)$;
        $C_t \leftarrow$ TrainSVM$(\mathbf{f}_j, \mathbf{S}^+, \mathbf{S}^-)$;
    end
    // Classify images in $\mathbb{C}_i$ using the
       constructed SVMs.
    for $t \leftarrow 1$ to $T$ do
        $L_t \leftarrow$ SVMPredict$(C_t, \mathbb{C}_i)$;
    end
    // Aggregate the outputs of SVMs with
       consensus rule.
    $L \leftarrow$ ConsensusAggregation$(\{L_i\}_{i=1}^T)$;
    // Purify the cluster based on aggregation
       result.
    $\mathbb{C}_i \leftarrow \{I_m | (I_m \in \mathbb{C}_i) \wedge (L(I_m) ==$ positive$)\}$;
    $\mathbb{C}^p \leftarrow \mathbb{C}^p \cup \mathbb{C}_i$;
end
return $\mathbb{C}^p$.
```

Algorithm 1: Consensus filters based confusing purifying

examples, the service provider constructs SVM to rank the images in Ω, and returns the top N ranked images, as result $\mathbb{R}^1$.

The overall process of private feedback is summarized in Algorithm 2.

4.3 Local Retrieval

Since some confusing classes are introduced for private feedback, the feedback retrieval result $\mathbb{R}^1$ would not be satisfactory. Therefore, we re-rank the images in $\mathbb{R}^1$ in the user side locally, namely local retrieval, to enhance the retrieval performance.

In local retrieval, the user feedback image set $\mathbb{F}$ is regarded as positive training dataset, and the introduced confusing classes are regarded as negative training dataset. The SVM is then trained to re-rank the images in $\mathbb{R}^1$, producing the final result $\mathbb{R}_L^1$ to the user.

5 Privacy Analysis

In this section, we formally analyse the privacy preserving performance of the proposed PRF-CBIR, in comparison with traditional RF-CBIR.

As presented in Section 3.2, traditional RF-CBIR scheme suffers from the query attack, result attack and feedback attack. We gauge the success of the attack with success probability $P(S|O,T)$,

```
Input  : The initial retrieval result : $\mathbb{R}^0$.
Output: The feedback retrieval result : $\mathbb{R}^1$.
// The user side.
// The user randomly labels some relevant
   images in $\mathbb{R}^0$ as $\mathbb{F}$.
$\mathbb{F} \leftarrow$ UserFeedback$(\mathbb{R}^0)$;
// Introduce confusing classes through three
   steps: discovery, purifying and
   selection.
$\mathbb{C} \leftarrow$ Discovery$(\mathbb{R}^0)$;
$\mathbb{C}^p \leftarrow$ Purifying$(\mathbb{C})$;
$\mathbb{C}^p \leftarrow$ Selection$(\mathbb{C}^p)$;
// Construct the feedback image set
   following the $K-$anonymity feedback
   principle.
$\mathbb{F}_p = \cup_{\mathbb{C}_i \in \mathbb{C}^p} \mathbb{C}_i \cup \mathbb{F}$;

// The service provider side.
// Construct SVM using $\mathbb{F}_p$ and randomly
   selected $\mathbb{N}$ to rank the images in $\Omega$.
$C \leftarrow$ TrainSVM$(\mathbf{f}, \mathbb{F}_p, \mathbb{N})$;
$\mathbb{R}^1 \leftarrow$ SVMPredict$(C, \Omega)$;
return $\mathbb{R}^1$.
```

Algorithm 2: Private feedback

where S is the sensitive search attention (target class), O is the attack object, and T represents the attack method.

5.1 Attack Types

Result attack and feedback attack are both performed by analysing the image set of result or feedback respectively. In our research, we take into account two attack types on the image set.

Maximum Frequency Inferring (MFI). Suppose the image set is $\mathbb{I}$, and the images in $\mathbb{I}$ belong to M classes, $C_1, \cdots, C_i, \cdots, C_M$. The numbers of images in each class can be represented as $n(C_1, \mathbb{I}), \cdots, n(C_i, \mathbb{I}), \cdots, (C_M, \mathbb{I})$, respectively. The classes can be sorted in descending order according to the numbers of images, and the result class sequence is $C_1', \cdots, C_i', \cdots, C_M'$. The most frequent class C_1' will be inferred as S, namely *Maximum Frequency Inferring* (MFI). For example, if the images in $\mathbb{I}$ belong to three classes: A, B and C. The number of images in A, B, C are 10, 20, 15 respectively. MFI will infer class B as the user's target class.

MFI can be easily prevented by introducing a larger confusing class into the image set of feedback or result. For example, during the feedback, the user labels 5 images from target class S as relevant. When introducing 6 images from confusing class F to the feedback image set, MFI will infer the user's target class as F incorrectly.

In this paper, we consider the service provider as a *intelligent adversary*. That is, the service provider can choose the optimal frequency position of classes in the image set to attack.

Optimal Frequency Inferring (OFI). Suppose that the images in $\mathbb{I}$ belong to M classes, $C_1, \cdots, C_i, \cdots, C_M$, and the sorted class sequence is $C_1', \cdots, C_i', \cdots, C_M'$. The *intelligent adversary* can learn the optimal class C_s' at the s^{th} frequency position s, based on the testing query set $\mathbb{Q}$ produced by itself. For the queries in $\mathbb{Q}$, the optimal class C'_s is at the frequency position s with the maximum probability that C'_s is the target class.

$$C'_s = \arg\max_{C'_i,\cdots,C'_M} P(C'_s == S|\mathbb{Q}), \tag{8}$$

With the learned optimal frequency position s, the s^{th} frequent class C'_s in $\mathbb{I}$ will be inferred as S. We name this inferring scheme as *Optimal Frequency Inferring* (OFI).

5.2 Query Attack

In RF-CBIR, the service provider can accurately reproduce the user's search intention S based the query image Q. While in PRF-CBIR, the query image is not provided to the service provider, and query attack becomes infeasible. That is,

$$\begin{aligned} P_{RF}(S|Q) &= 1, \\ P_{PRF}(S|\emptyset) &= 0. \end{aligned} \tag{9}$$

5.3 Result Attack on $\mathbb{R}^0$

Consider that $\mathbb{R}^0$ is retrieval by $\mathbf{f}^Q$ without introducing confusing classes, the MFI attack is equivalent to OFI, and we analyse the MFI attack type only.

In RF-CBIR, $\mathbb{R}^0$ is retrieved with the total feature vector of $\mathbf{f}^Q$. In comparison, $\mathbb{R}^0$ is retrieved with the private query $\mathbf{f}_p^Q$ in PRF-CBIR. Since $\mathbf{f}_p^Q$ is a part of $\mathbf{f}^Q$, the number of images belonging to target class S in $\mathbb{R}^0$ is less than that of RF-CBIR, $n_{PRF}(S,\mathbb{R}^0) < n_{RF}(S,\mathbb{R}^0)$. Accordingly, the probability that S is the most frequent class in PRF-CBIR is smaller than in RF-CBIR, $P_{PRF}(C'_1 == S) < P_{RF}(C'_1 == S)$. We have,

$$P_{PRF}(S|\mathbb{R}^0, \mathsf{MFI}) < P_{RF}(S|\mathbb{R}^0, \mathsf{MFI}). \tag{10}$$

According to (4), given a length l, $\mathbf{f}_p^Q$ is generated from $\mathbf{f}^Q$, corresponding to the PoV threshold $\mathbf{T}$. PoV characterizes the importance of feature components. With the increase of $\mathbf{T}$, more important feature components will be introduced into $\mathbf{f}_p^Q$, and the number of images belonging to S in $\mathbb{R}^0$ will increase. Consequently, the attack success probability will increase.

$$\mathbf{T}\uparrow,\ n_{PRF}(S,\mathbb{R}^0)\uparrow,\ P_{PRF}(C'_1 == S)\uparrow,\ P_{PRF}(S|\mathbb{R}^0,\mathsf{MFI})\uparrow. \tag{11}$$

5.4 Feedback Attack

5.4.1 RF-CBIR

In RF-CBIR, the feedback image set $\mathbb{F}$ contains only the relevant images, which are all from the target class. Therefore,

$$P_{RF}(S|\mathbb{F}, \mathsf{MFI}) = P_{RF}(S|\mathbb{F}, \mathsf{OFI}) = 1. \tag{12}$$

5.4.2 PRF-CBIR

In PRF-CBIR, the feedback image set $\mathbb{F}_p$ is generated following the $K-$anonymity feedback principle (5). From (5), the numbers of images in $C_1,\cdots,C_i,\cdots,C_K$ are the same,

$$n(C_1,\mathbb{F}_p) = \cdots = n(C_i,\mathbb{F}_p) = \cdots = n(C_K,\mathbb{F}_p). \tag{13}$$

Assuming that the sorted class sequence is $C'_1,\cdots,C'_i,\cdots,C'_K$, and the target class is $S = C'_s$ $(1 \le s \le K)$. Since the size of S is equal to other classes, the probabilities that S ranked at position $1,\ldots,K$ will be the same value of $\frac{1}{K}$.

$$P(S == C'_1) = \cdots = P(S == C'_i) = \cdots = P(S == C'_K) = \frac{1}{K}. \tag{14}$$

We show that the success probabilities of attack on $C'_1,\cdots,C'_i,\cdots,C'_K$ are with the same value of $\frac{1}{K}$. Therefore,

$$P_{PRF}(S|\mathbb{F}_p, \mathsf{MFI}) = P_{PRF}(S|\mathbb{F}_p, \mathsf{OFI}) = \frac{1}{K}. \tag{15}$$

5.5 Result Attack on $\mathbb{R}^1$

5.5.1 RF-CBIR

In RF-CBIR, $\mathbb{R}^1$ is retrieved with the feedback image set $\mathbb{F}$ with the SVM classifiers. The retrieval accuracy is improved through feedback, therefore, the success probability of attack with MFI increases compared to that on $\mathbb{R}^0$. On the other side, because of the classification error of SVM classifier, irrelevant images will exist in $\mathbb{R}^1$. Thus, the success probability of attack with MFI is smaller than that on $\mathbb{F}$.

$$P_{RF}(S|\mathbb{R}^0, \mathsf{MFI}) < P_{RF}(S|\mathbb{R}^1, \mathsf{MFI}) < P_{RF}(S|\mathbb{F}, \mathsf{MFI}) = 1. \tag{16}$$

Considering that relevance feedback usually can achieve satisfactory retrieval performance, the target class will be the major class for most queries. In this case, the optimal frequency position for inferring is the most frequent class. That is,

$$P_{RF}(S|\mathbb{R}^1, \mathsf{OFI}) = P_{RF}(S|\mathbb{R}^1, \mathsf{MFI}). \tag{17}$$

5.5.2 PRF-CBIR

In PRF-CBIR, $\mathbb{R}^1$ is retrieved with the feedback image set $\mathbb{F}_p$. According to (13), the numbers of images in $\mathbb{F}_p$ belonging to $C_1,\cdots,C_i,\cdots,C_K$ are the same, $n(C_1,\mathbb{F}_p) = \cdots = n(C_i,\mathbb{F}_p) = \cdots = n(C_K,\mathbb{F}_p)$. Ideally, the retrieved result $\mathbb{R}^1$ will have the same class distribution as $\mathbb{F}_p$,

$$n(C_1,\mathbb{R}^1) = \cdots = n(C_i,\mathbb{R}^1) = \cdots = n(C_K,\mathbb{R}^1). \tag{18}$$

However, because of the classification error of SVM classifier, some images from noisy classes will be presented in $\mathbb{R}^1$. For simplicity, we assume that the noisy classes $V_1,\cdots,V_v$ are with the approximate size with C_i. In this situation, the sorted class sequence will be $\bar{C}'_1,\cdots,\bar{C}'_i,\cdots,\bar{C}'_{K+v}$, and the probabilities that S ranked at position $1,\ldots,(K+v)$ will be,

$$P(S == \bar{C}'_1) \approx \cdots \approx P(S == \bar{C}'_i) \approx \cdots \approx P(S == \bar{C}'_{K+v}) \approx \frac{1}{K+v}. \tag{19}$$

We can see that the success probabilities of attack on $\bar{C}'_1,\cdots,\bar{C}'_i,\cdots,\bar{C}'_{K+v}$ are with the same value of $\frac{1}{K+v}$. That is,

$$P_{PRF}(S|\mathbb{R}^1, \mathsf{MFI}) = P_{PRF}(S||\mathbb{R}^1, \mathsf{OFI}) \approx \frac{1}{K+v} \le \frac{1}{K}. \tag{20}$$

5.6 Summary

From (9, 10, 12, 16, 17), we can see that the success probability of attack in RF-CBIR is,

$$\begin{aligned} P_{RF}(S) &= \max(P_{RF}(S|Q), P_{RF}(S|\mathbb{R}^0), P_{RF}(S|\mathbb{F}), P_{RF}(S|\mathbb{R}^1)) \\ &= \max(1, P_{RF}(S|\mathbb{R}^0), 1, P_{RF}(S|\mathbb{R}^1)) = 1. \end{aligned} \tag{21}$$

From (9, 10, 15, 20). we observe that the success probability of attack in PRF-CBIR is,

$$\begin{aligned} P_{PRF}(S) &= \max(P_{PRF}(S|\emptyset), P_{PRF}(S|\mathbb{R}^0), P_{PRF}(S|\mathbb{F}), P_{PRF}(S|\mathbb{R}^1)) \\ &= \max(0, P_{PRF}(S|\mathbb{R}^0), \frac{1}{K}, \frac{1}{K+v}) \\ &= \max(P_{PRF}(S|\mathbb{R}^0, \mathsf{MFI}), \frac{1}{K}). \end{aligned} \tag{22}$$

Additionally, according to (11), $P_{PRF}(S|\mathbb{R}^0, \mathsf{MFI})$ can be controlled by setting a proper $\mathbf{T}$, satisfying $P_{PRF}(S|\mathbb{R}^0, \mathsf{MFI}) \le \frac{1}{K}$. In this situation, the success probability of attack in PRF-CBIR is,

$$P_{PRF}(S) = \frac{1}{K}. \tag{23}$$

From (21, 23), we see that the success probability of attack is decreased from 1 in RF-CBIR to $\frac{1}{K}$ in PRF-CBIR.

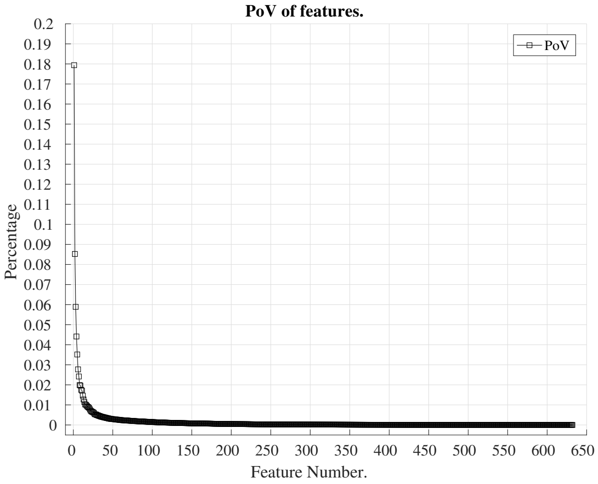


Fig. 7. The PoV of features.

6 Experiments and Results

A large number of experiments were carried on a real-world image collection to evaluate the performance of the proposed scheme. This section reports the experiments and results.

6.1 Experiment Setting

6.1.1 Dataset

In this paper, the Caltech-256 image dataset [26] is used for the empirical study. The Caltech-256 was created by the California Institute of Technology to facilitate computer vision research. We used a subset of Caltech-256, which consists of 50 classes and each class includes 105 images. For the image retrieval task, 5 query images are randomly selected from each class and the rest 100 images in that class are regarded as the ground truth. The privacy performance and retrieval performance are measured on the 250 queries.

6.1.2 Image features

We used seven low-level features to represent images as follows.

- Scalable color descriptor [27] (18-dimension).
- Color layout descriptor [27] (120-dimension).
- Color and edge directivity descriptor [28] (144-dimension).
- Edge histogram descriptor [29] (80-dimension).
- Fuzzy color and texture histogram [30] (192-dimension).
- Tamura feature [31] (18-dimension).
- Gabor feature [32] (60-dimension).

The features were extracted with the open source tool, LIRE [33]. All together, those features form 632-dimension vector. After PCA feature reduction, the feature components are in the descending order of importance, and their importance can be characterized by corresponding PoV. As shown in Fig.7, the top 100 features represent the 86% PoV of all features. Therefore, we choose to keep the top 100 features for balancing dimension reduction and variance preservation.

6.1.3 Evaluation metrics

We use success rate to measure the performance of privacy, and precision to measure the performance of retrieval.

Success rate is used to measure the service provider's attack success, which is the ration of the number of successful attacks over the number of all the attacks.

$$\text{Success Rate} = \frac{\text{\# successful attacks}}{\text{\# all the attacks}}. \tag{24}$$

Average precision is used to measure the accuracy of image retrieval. For a query q, precision is defined as the fraction of retrieved images that are relevant.

$$\text{Precision } (q) = \frac{\text{\# relevant images}}{\text{\# retrieved images}}, \tag{25}$$

Average precision is defined as the average of the precisions over all queries.

$$\text{Average Precision } (\mathbb{Q}) = \frac{1}{|(\mathbb{Q}|} \sum_{q \in (\mathbb{Q}} \text{Precision } (q), \tag{26}$$

where $\mathbb{Q}$ is the set of queries. In the following experiments, we report the average precision of top 5, top 10, top 15, top 20, top 30, top 100, and top 200.

6.1.4 Other setting

We used the LIBSVM [34] to solve SVMs. Gauss kernels were applied in our experiments, and we applied the grid search method [35] to tune parameters for SVMs.

6.2 Evaluation of Privacy Performance

In this section, we evaluate the privacy preserving performance of the proposed PRF-CBIR, in the viewpoint of attacks.

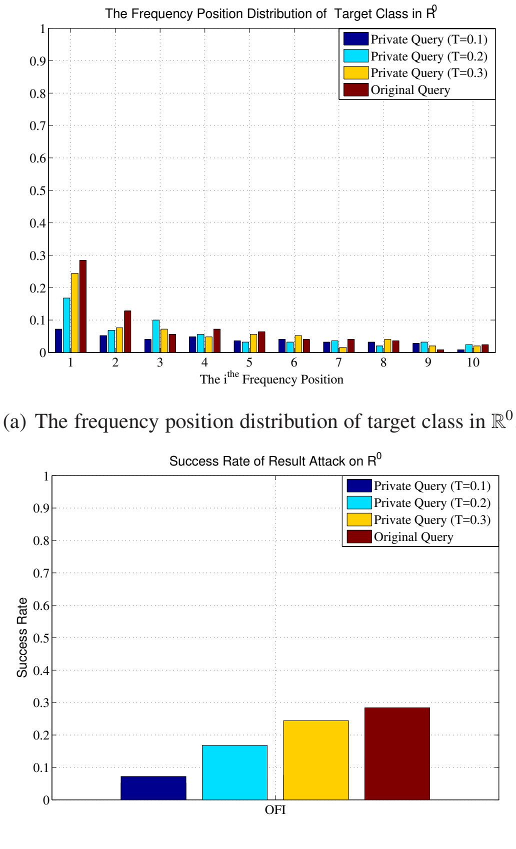
6.2.1 Result Attack on $\mathbb{R}^0$

In this experiment, we evaluate the result attack on $\mathbb{R}^0$. In PRF-CBIR, $\mathbb{R}^0$ is retrieved with private queries. The private queries are generated based on (4) with the length $l = 40$ and PoV threshold $\mathbf{T} = 0.1, 0.2, 0.3$. The original query in RF-CBIR with the full feature vector is implemented as baseline.

Fig.8 reports the frequency position distribution of target class and the success rate of OFI attack. Fig. 8(a) shows that the target classes of private query distribute more homogeneous in frequency positions than that of original query. Fig. 8(b)shows that the attack success rate of private query is lower than that of original query. The reason is that a part of feature vector of query image is used in private query, while the full feature vector are utilized in original query. For private query, with the increase of $\mathbf{T}$, the success rate increases as well. For example, when $\mathbf{T} = 0.1$, the success rate is 7%. When $\mathbf{T} = 0.3$, the success rate increases to 28%. This experimental observation is consistent with theoretical analysis of (11) in Section V. The reason is that more important feature components are included in private query with the increase of $\mathbf{T}$.

6.2.2 Feedback Attack

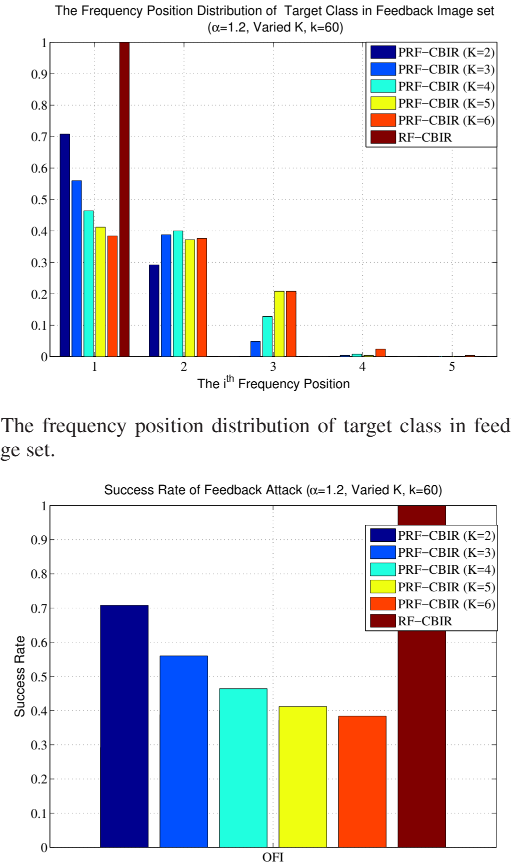
These experiments were carried out to evaluate the effectiveness of private feedback for feedback attack. The initial queries are performed using the private query with the PoV threshold $\mathbf{T} = 0.2$ in Section 6.2.1. The top 300 images are returned as result $\mathbb{R}^0$. The simulated user feedbacks are performed by randomly selected 5 relevant images in $\mathbb{R}^0$ as user feedback image set $\mathbb{F}$. In this experiment, we set the cluster number of k-means for confusing



(a) The frequency position distribution of target class in $\mathbb{R}^0$.

(b) Success rate of result attack on $\mathbb{R}^0$.

Fig. 8. Analysis of result attack on $\mathbb{R}^0$.



(a) The frequency position distribution of target class in feedback image set.

(b) Success rate of feedback attack.

Fig. 9. Analysis of feedback attack.

discovery k as 60, and the adjustment factor for confusing selection α to 1.2. Since the proposed private feedback follows the K-anonymity feedback principle, we used different K values. The traditional RF-CBIR scheme is implemented as baseline. In RF-CBIR, the user's labelled relevant images are regarded as positive examples, the negative examples are randomly selected from the remained images from the initial result. The positive and negative training set are with the same size.

The privacy preserving performance is reported in Fig.9. We can see the attack success rate of PRF-CBIR is significantly lower than that of RF-CBIR. For example, the success rate of feedback attack with OFI for RF-CBIR and PRF-CBIR (K=5) are 100% and 41% respectively. The reason is that confusing classes are introduced into the feedback image set in PRF-CBIR, while in RF-CBIR, the feedback image set only contains the user labelled relevant images. For PRF-CBIR, with the increase of K, the success rate of feedback attack decreases. For example, when K=2, the success rate of OFI is 71%. When K=6, the success rate of OFI decrease to 37%. This is because that more confusing classes are introduced with the increase of K, and the target classes disperse within the 1^{th} to K^{th} frequency position. The success rate of feedback attack is not as ideal as that in the theoretical analysis of (15) in Section V. The reason is that introducing $K - 1$ confusing classes is a really challenging task because of the clustering efficiency. We will try to improve the three-step confusing classes introducing method in our future work.

6.2.3 Result Attack on $\mathbb{R}^1$

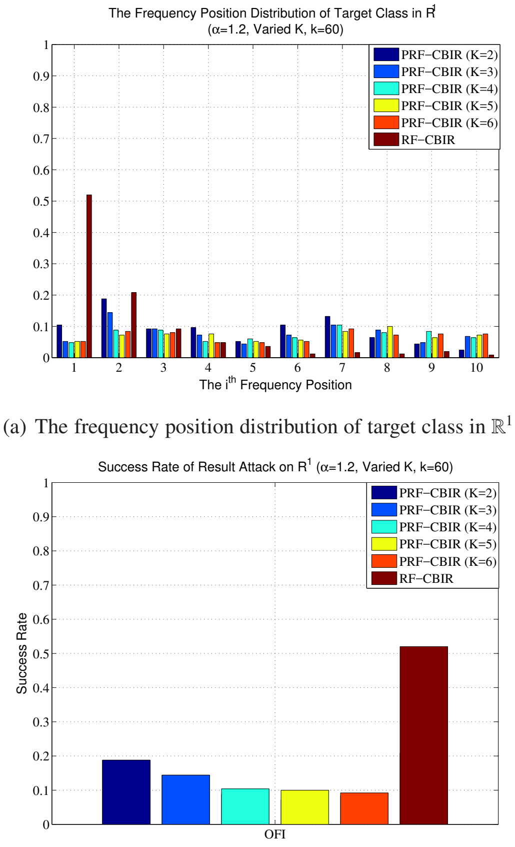
These experiments were performed to evaluate the result attack on $\mathbb{R}^1$. The private feedback procedure depicted in Section 6.2.2 is performed, and the top 300 images are returned as result $\mathbb{R}^1$. Fig. 10 shows that the attack success rate of PRF-CBIR is significantly lower than that of RF-CBIR. With the increase of K, the attack success rate in PRF-CBIR decreases. This observation is consistent with feedback attack. The reason is that $\mathbb{R}^1$ is retrieved with feedback image set, and has the similar image distribution as the feedback image set.

6.2.4 Attack Comparison

In this section, we compare different attacks empirically. Besides the result attack and feedback attack, we take into account *Synthetic Attack*.

Synthetic Attack: For this attack, we consider that the service provider can aggregate the exchanged information, including $\mathbb{R}^0$, $\mathbb{F}_p$ and $\mathbb{R}^1$, to infer the user's search intention.

The experiment setting here is the same with that in Section 6.2.2, and K is set to 5. The success rate of different attacks in PRF-CBIR are shown in Fig.11 comparatively. We see that feedback attack has the highest success rate. The synthetic attack and result attack on $\mathbb{R}^0$ has lower success rate compared with feedback attack. Result attack on $\mathbb{R}^1$ achieves the lowest success rate. This experimental results demonstrate that feedback attack can achieve the highest success rate, and how to deal with it is the key to PRF-CBIR.



(a) The frequency position distribution of target class in $\mathbb{R}^1$.

(b) Success rate of result attack on $\mathbb{R}^1$.

Fig. 10. Analysis of result attack on $\mathbb{R}^1$.

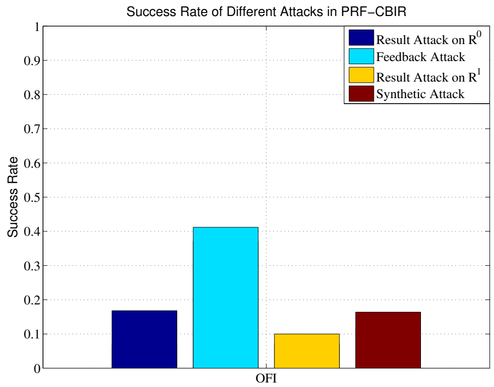


Fig. 11. Comparison of attacks.

6.3 Evaluation of Retrieval Performance

In this section, we evaluate the retrieval performance of the proposed PRF-CBIR. The retrieval precision of PRF-CBIR is measured on the result of local retrieval after private feedback. The private query and traditional RF-CBIR are implemented as references.

The retrieval performance of three competitors is presented in Fig.12. By comparison with private query, the retrieval precision is improved significantly in both RF-CBIR and PRF-CBIR. For example, the average precision of top 20 is 11%, 46% and 42% in private query, RF-CBIR and PRF-CBIR (K=5) respectively. The retrieval performance of PRF-CBIR is slightly lower compared to RF-CBIR. The retrieval performance of PRF-CBIR with different K is shown in Fig.12. We can see, with the increase of K, the retrieval precision drops a little. The reason is that more confusing classes are presented in the procedure of private feedback and local retrieval. From the figure, we observe that the precision of top results are less affected. For example, when K changes from 2 to 6, the precision of top 10 drops slightly from 64% to 62%, while the precision of top 100 decreases from 18% to 12%. Since the users pay more attention to the top ranked images, the retrieval performance sacrifice is acceptable.

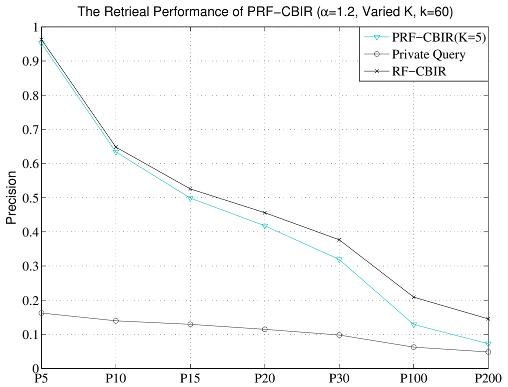


Fig. 12. Retrieval performance of PRF-CBIR.

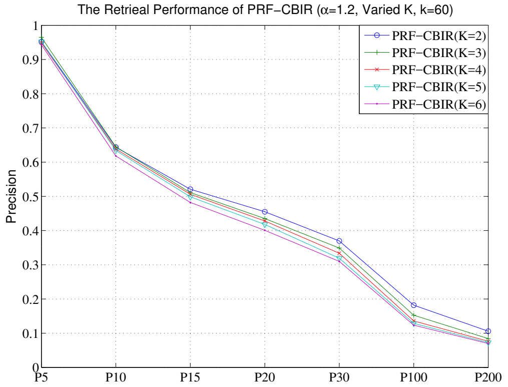


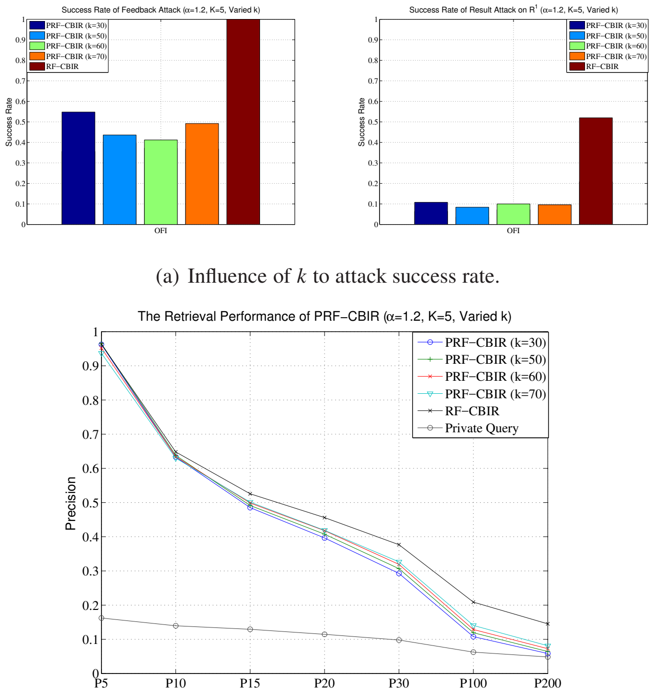
Fig. 13. Retrieval performance of PRF-CBIR with different K.

6.4 Influence of Parameters

In these experiments, we investigated the influence of two parameters in private feedback, the cluster number of k-means for confusing discovery— k and the adjustment factor for confusing selection — α. K is set to 5 for the K-anonymity feedback principle. The other experiments setting remain the same as those in Section 6.2.2.

6.4.1 Influence of Cluster Number

Fig.14 shows the attack success rate and retrieval precision with respect to the clustering number $k = 30, 50, 60, 70$. We find that the parameter k has slightly influence on privacy preserving and image retrieval. Generally, with a smaller k (e.g., $k = 30$) , the attack success rate is slightly higher and the retrieval precision



(a) Influence of k to attack success rate.

(b) Influence of k to retrieval performance.

Fig. 14. Influence of cluster number k.

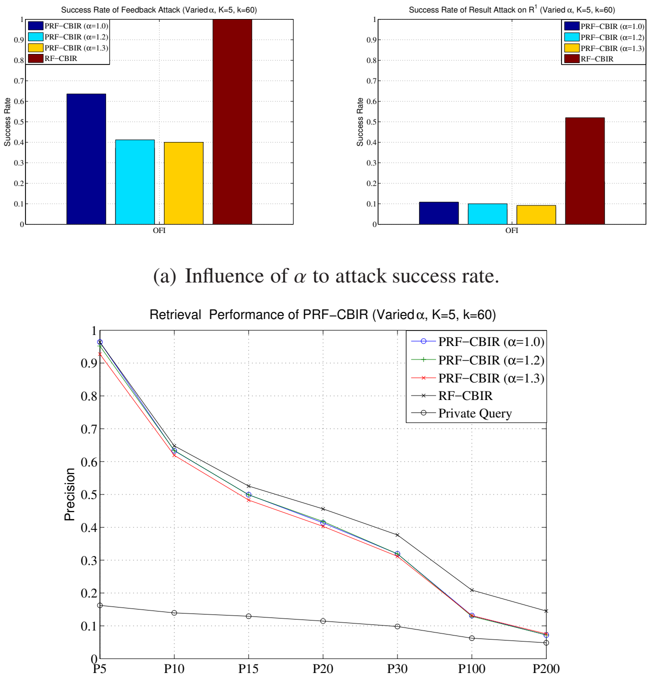
is little lower. The reason is that the purity of confusing classes candidates is lower than that with a much suitable k. With a larger k (e.g., $k = 70$), the attack success rate is slightly higher and the retrieval precision is little higher. The reason is that although the purity is enhanced, however, the cluster size of k-means become smaller as well, decreasing the number of valid confusing classes candidates. In this paper, we choose the k to 60 as an optimal value. It is should be noted that a k in $[30, 70]$ works well in our experiments.

6.4.2 Influence of Adjustment Factor

The influence of the adjustment factor α on PRF-CBIR is reported in Fig.15. The figures show that, with the increase of α, the attack success rate and the retrieval precision both decrease. The reason is that with bigger size of confusing classes, and the probability of the target class dispersing in the 1^{th} to the K^{th} frequency position increases.

7 Conclusion

This paper addressed a new problem of privacy protection in RF-CBIR. A new PRF-CBIR scheme is proposed to protect the user's search intention and leverage the performance gain of relevance feedback. PRF-CBIR consists of three stage: private query, private feedback and local retrieval. PRF-CBIR can deal with query attack, result attack and feedback attack existing in RF-CBIR. We provided a theoretical analysis on privacy protection of the new scheme. It shows the new scheme can effectively control privacy leakage and significantly reduce the attack success probability. Moreover, we carried out a large number of experiments on a real-world image collection. The results demonstrate that the privacy preserving performance of PRF-CBIR is significantly improved compared to RF-CBIR, while the retrieval performance sacrifice is acceptable. The success rate of feedback attack with OFI (Optimal Frequency Inferring) drops dramatically from 100% in RF-CBIR to 41% in PRF-CBIR (K=5), while the average precision of top 20 decreases slightly from 46% to 42%. We also investigated the parameters of the new scheme and give the suggestions.



(a) Influence of α to attack success rate.

(b) Influence of α to retrieval performance.

Fig. 15. Influence of adjustment factor α.

Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 61300077) and Beijing Natural Science Foundation (No. 4172054).

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.

[3] R. Datta, J. Li, and J. Z. Wang, "Content-based image retrieval: Approaches and trends of the new age," in *The 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*, Hilton, Singapore, 2005, pp. 253–262.

[4] W. Lu, A. Swaminathan, A. L. Varna, and M. Wu, "Enabling search over encrypted multimedia databases," in *Proc. SPIE*, vol. 7254, San Jose, CA, 2009.

[5] W. Lu, A. L. Varna, A. Swaminathan, and M. Wu, "Secure image retrieval through feature protection," in *The 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, 2009, pp. 1533–1536.

[6] W. Lu, A. L. Varna, and M. Wu, “Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance-preserving randomization,” *IEEE Access*, vol. 2, pp. 125–141, 2014.

[7] J. Yuan, S. Yu, and L. Guo, “SEISA: Secure and efficient encrypted image search with access control,” in *2015 IEEE Conference on Computer Communications (INFOCOM)*, Kowloon, 2015, pp. 2083–2091.

[8] 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*, no. 99, 2015.

[9] B. Ferreira, J. Rodrigues, J. Leitao, and H. Domingos, “Privacy-preserving content-based image retrieval in the cloud,” in *The 2015 IEEE 34th Symposium on Reliable Distributed Systems (SRDS),*, Montreal, QC, 2015, pp. 11–20.

[10] L. Zhang, T. Jung, P. Feng, K. Liu, X. Y. Li, and Y. Liu, “PIC: Enable large-scale privacy preserving content-based image search on cloud,” in *The 44th International Conference on Parallel Processing (ICPP)*, Beijing, 2015, pp. 949–958.

[11] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[12] J. Shashank, P. Kowshik, K. Srinathan, and C. V. Jawahar, “Private content based image retrieval,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, 2008, pp. 1–8.

[13] L. Weng, L. Amsaleg, A. Morton, and S. Marchand-Maillet, “A privacy-preserving framework for large-scale content-based information retrieval,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 152–167, 2015.

[14] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, “Relevance feedback: a power tool for interactive content-based image retrieval,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644–655, 1998.

[15] S. T. Zhou, Sean Xiangand Huang, “Relevance feedback in image retrieval: A comprehensive review,” *Multimedia Systems*, vol. 8, no. 6, pp. 536–544, 2003.

[16] X. Qian, X. Tan, Y. Zhang, R. Hong, and M. Wang, “Enhancing sketch-based image retrieval by re-ranking and relevance feedback,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 195–208, 2016.

[17] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[18] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[19] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[20] L. Sweeney, “K-anonymity: A model for protecting privacy,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[21] C. Dwork, “Differential privacy,” in *Proceedings of the 33rd International Conference on Automata, Languages and Programming*, Venice, Italy, 2006, pp. 1–12.

[22] A. K. Jain, “Data clustering: 50 years beyond k-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[23] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[24] C. E. Brodley and M. A. Friedl, “Identifying mislabeled training data,” *Journal of Artificial Intelligence Research*, vol. 11, pp. 131–167, 1999.

[25] J. Zhang and L. Ye, “Content based image retrieval using unclean positive examples,” *IEEE Transactions on Image Processing*, vol. 18, no. 10, pp. 2370–2375, 2009.

[26] G. Griffin, A. Holub, and P. Perona, “Caltech-256 object category dataset,” Technical Report, California Institute of Technology, http://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001 , accessed Aug. 1, 2016.

[27] J.-R. Ohm, L. Cieplinski, H. J. Kim, S. Krishnamachari, B. Manjunath, D. S. Messing, and A. Yamada, “The MPEG-7 color descriptors,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2001.

[28] S. A. Chatzichristofis and Y. S. Boutalis, “CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval,” in *International Conference on Computer Vision Systems*, Santorini, Greece, 2008, pp. 312–322.

[29] T. Sikora, “The MPEG-7 visual standard for content description-an overview,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 696–702, 2001.

[30] S. A. Chatzichristofis and Y. S. Boutalis, “FCTH: Fuzzy color and texture histogram - a low level feature for accurate image retrieval,” in *The 9th International Workshop on Image Analysis for Multimedia Interactive Services*, Klagenfurt, 2008, pp. 191–196.

[31] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

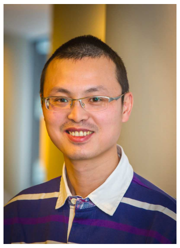
[32] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837–842, 1996.

[33] M. Lux and S. A. Chatzichristofis, “Lire: Lucene image retrieval: An extensible java CBIR library,” in *Proceedings of the 16th ACM International Conference on Multimedia*, Vancouver, British Columbia, Canada, 2008, pp. 1085–1088.

[34] C. Chang and C. Lin, “LIBSVM: a library for support vector machines,” National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin/libsvm/, accessed Aug. 15, 2016.

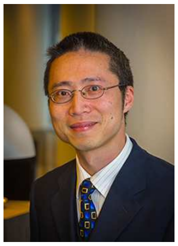
[35] C. W. Hsu, C. C. Chang, and C. J. Lin, “A practical guide to support vector classification,” National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, accessed Aug. 15, 2016.

Yonggang Huang (M'14) received his PhD from Beihang University, China, in 2012. He is now a Assistant Professor with School of Computer Science and Technology, Beijing Institute of Technology. His research interests include image retrieval, image privacy and security. He has published 10 more research papers in refereed international journals and conferences, such as IEEE Transactions on Information Forensics and Security, Multimedia Tools and Applications and IEICE Transactions on Information and Systems. He is a member of the IEEE.



Jun Zhang (M'12) received his PhD from University of Wollongong, Australia, in 2011. He is currently an Associate Professor with School of Software and Electrical Engineering, Swinburne University of Technology. His research interests include multimedia retrieval, multimedia privacy and network security. He has published 50 more research papers in refereed international journals and conferences, such as IEEE Transactions on Information Forensics and Security, IEEE Transactions on Dependable and Secure Computing and IEEE Transactions on Image Processing. He is a member of the IEEE.

Lei Pan (M'12) received his PhD degree in computer forensics from Deakin University, Australia, in 2008. He is now a Lecturer with School of Information Technology, Deakin University. His research interests are cyber security and privacy. He has published more than 30 research papers in refereed international journals and conferences, such as IEEE Security & Privacy, Journal of Multimedia, and Digital Investigation. He is a member of the IEEE.



Yang Xiang (A'08–M'09–SM'12) received his PhD degree in computer science from Deakin University, Australia. He is currently a Full Professor with School of Software and Electrical Engineering, Swinburne University of Technology. His research interests are cyber security and privacy. He has published more than 150 research papers in many international journals and conferences, such as IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Dependable and Secure Computing, and IEEE Journal on Selected Areas in Communications. He serves as the Associate Editor of IEEE Transactions on Computers, IEEE Transactions on Parallel and Distributed Systems, Security and Communication Networks, and the Editor of Journal of Network and Computer Applications. He is the Coordinator, Asia for IEEE Computer Society Technical Committee on Distributed Processing. He is a Senior Member of the IEEE.