IRIW: Image Retrieval based Image Watermarking for Large-Scale Image Databases

Jong Yun Jun[1], Kunho Kim[1], Jae-Pil Heo[1], and Sung-eui Yoon[1,2]

[1] Dept. of Computer Science, KAIST, South Korea
[2] Div. of Web Sci. and Tech., KAIST, South Korea

Abstract. We present a novel, Image Retrieval based Image Watermark (IRIW) framework to identify copyright-violated images in both efficient and accurate manner for large-scale image databases. We first perform SIFT-based image retrieval to identify similar images given a query image and store them as an output list. Then we extract watermark patterns and check watermark similarity only for images stored in the list. As a final step, we re-rank images by considering various information available between each image in the list and the query image and by utilizing information even among images in the list. Also, in order to reduce any negative impacts on image retrieval by embedding watermark patterns on images, we propose to use a SIFT-aware image watermark detection method. Compared with the exhaustive method that checks all the images stored in an image database that consists of 10 K images, our method achieves more than two orders of magnitude performance improvement. More importantly, by identifying similar images given a query image and focusing on checking watermark similarities among those similar images, we are able to reduce false positive and false negative cases by a factor of up to two over the exhaustive method.

1 Introduction

Thanks to rapid advances of digital camera and various image processing tools, we can easily create new pictures and images for various purposes. This in turn results in a huge amount of images in the internet and even in personal computers. For example, flickr, an image hosting website, contains more than five billion images and flickr members update more than three thousand image every minute [3].

These huge image databases pose numerous technical challenges in terms of image processing, searching, storing, etc. One of the many challenging problems caused by easy image processing and modification technologies is the security problem. By the nature of digital data, it is very easy to copy, modify, redistribute the original image data. In order to address the security problem related to images, image watermark techniques have been studied actively in the last decade [21].

The main concept of image watermarking is to embed visually imperceptible patterns on images so that a copyright holder of images can claim his or

[3] http://blog.flickr.net/en/2010/09/19/5000000000/

her ownership by extracting those patterns. Therefore, most image watermark techniques focus on extracting the embedded watermark patterns in a highly accurate manner against many different image attack scenarios (e.g., geometric transformation, cropping, and noise addition).

Even with drastic advances on image watermarking, the state-of-the-art image watermark techniques have certain false negative and false positive probabilities. As a result, a high number of false negative and false positive cases can occur, if we attempt to identify copyright-violated images solely based on image watermark techniques for web-scale image databases such as flickr. Furthermore, extracting watermark patterns and matching those patterns against the watermark pattern of the input query image can take prohibitive time for a large-scale image database consisting of millions of images or more.

Main contributions: In order to efficiently and accurately identify images that are modified or are the exactly same images from a query image in large-scale image databases, we present a novel, Image Retrieval based Image Watermark (IRIW) framework. Instead of exhaustively scanning and extracting watermark patterns from all the images in the image database, we first identify similar images given a query image by using a SIFT-based image retrieval method (Sec. 4.1). Then we extract watermark patterns only from those similar images and measure watermark pattern similarities against the query image. Finally, we re-rank images by considering both image and watermark pattern similarities against the query image, in order to place images that are more likely to be copyright-violated in higher ranks in a final image list (Sec. 4.2). We propose to use a SIFT-aware image watermark method (Sec. 4.3) that does not embed watermark patterns on image regions where we get SIFT features, in order to minimize negative effects on our SIFT-based image retrieval method.

In order to verify the benefits of our method, we test our method in an image database that consists of 10 K images (Sec. 5). We found that our method improves the performance of searching copyright-violated images given a query image by more than two orders of magnitude over the exhaustive method that searches those copyright-violated images by accessing all the images in the database. More importantly, our method improves the accuracy of search results by reducing ratios of false negative and false positive cases up to two times over the exhaustive method. The performance and accuracy improvements of our method is mainly caused by identifying similar images based on image retrieval and by checking watermark similarities only for those images.

2 Related Work

In this section we review prior work on content-based image retrieval (CBIR), image watermarking, and their combinations.

2.1 Image Retrieval

CBIR has been drawing significant attention in recent years, and an excellent survey [7] is available. One of the most successful classes of CBIR techniques is based on using Scale Invariant Feature Transform (SIFT) [11] and the concept of visual words [17]. A visual word is a clustered set of similar SIFT features.

An input query image is decomposed into a number (e.g., a few thousands) of SIFT features. Then, each SIFT feature of the query image is assigned to one or multiple visual words, which are precomputed with images stored in a database. Once we represent the query image with a set of visual words, then we find similar images from the database; the image similarity is defined in terms of the associated visual words for each image.

Since it can take a huge amount of time to identify similar images among a large number of images, hierarchical computation for visual words [13] or approximate computation [15] for similar images have been proposed. Our technique is based on one [13] of recent techniques that shows high runtime query performance and accuracy. However, our approach can be integrated with other SIFT-based image retrieval techniques.

A few CBIR techniques have been used to identify copyright-violated images by relying only on image features [4, 8]. These techniques can be classified as CBIR methods designed for near-duplicate (or near-identical) image detection [5, 20]. Since they do not use any watermarking techniques, it is unclear how robustly they can handle differently attacked images. Moreover, even though we identify copyright-violated images based on these near-duplicate image detection methods, these results provide limited legal claims over identifying copyright-violated images based on watermark techniques.

2.2 Image Watermarking

Image watermark algorithms have been extensively studied, and major image watermark techniques are well explained in a recent survey [21]. Most image watermark techniques are classified as spatial and transform/spectral domain techniques. Spatial domain techniques are easier to implement, but transform/spectral domain techniques [6] have been proven to be more robust for various image editing attacks.

In recent years, research on image watermark algorithms targets on achieving a higher robustness against to various geometrical distortions including RST (Rotation, Scaling, and Translation) attacks. Different approaches [21, p.26] have been proposed for these RST attacks. One class of techniques that are robust for RST attacks relies on using salient image features such as corners and edges of images. Utilizing such image features is useful, since the problem of geometric synchronization necessary for watermark extraction can be addressed by aligning those image features that are invariant to such geometric transformations.

Among image watermark techniques utilizing image features, Bas et al. [2] proposed a content-based synchronization algorithm by using image corners. They perform Delaunay triangulation [3] with the computed image corner points. Watermark patterns are embedded into each triangle of the constructed Delaunay triangulation. Also, Tang and Hang [18] use feature points computed by the Mexican Hat wavelet scale interaction that considers the intensity changes in images. Lee et al. [10] utilize SIFT features, the well-known image feature for image retrieval, for image watermarking. We propose to use this kind of techniques within our IRIW framework, in order to minimize any negative effects on the accuracy of our image retrieval component.

2.3 Image Retrieval with Watermarking

CBIR and image watermark techniques have been developed in separate fields. Recently there have been a few approaches that combine these two techniques.

Lu et al. [12] introduced a multipurpose watermarking scheme that embeds robust and fragile watermarks simultaneously in images. They also use image features that can be used for image retrieval as watermark patterns for images. Xu et al. [19] proposed an image retrieval technique that utilizes watermark patterns as features for image retrieval, and showed its retrieval performance in a small number of image data consisting of only eight different images. This method can allow users to identify images that have the exactly same watermark patterns. However, if watermark patterns of images are broken, this technique cannot identify similar images, since the method relies solely on watermark patterns for image retrieval. Furthermore, these two prior methods do not use image retrieval to improve the performance and accuracy of image watermark methods. In other words, results computed only based on watermark patterns may not include severely attacked images if their watermark patterns are broken. Also, this approach may report completely different images especially in large-scale image databases, because of certain false positive ratios of any watermarking techniques.

Unlike prior approaches that use image features or watermark patterns either for image retrieval or for image watermarking, we propose a novel, holistic framework that combines image watermark and retrieval techniques together such that it can improve both the performance and accuracy of image watermarking for large-scale image databases.

3 Overview

In this section we summarize issues with large-scale image databases and present the overview of our approach.

3.1 Issues with Large-Scale Image Databases

Suppose that a copyright holder wants to identify illegal usages (e.g., using the exact or modified images) of his/her images among images available on the internet. Even though addressing this kind of scenario is necessary because of the rapid advances of the internet, effective ways of handling large-scale image databases have not been actively studied in the context of image watermarking [21].

The most simple, but general approach for dealing with large-scale image databases is to exhaustively scan and extract watermark patterns from all the images in the database. More specifically, for each image on the internet, we can attempt to extract a watermark pattern and perform a *watermark similarity test* that measures a watermark similarity value by comparing the extracted pattern against the watermark pattern of the copyright holder. Then the exhaustive method reports a list of k images that have top k watermark similarity values in the image database.

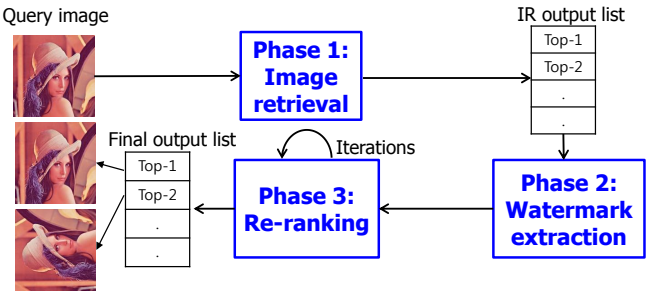

Fig. 1. This figure shows an overview of our IRIW framework.

In the list, however, we may fail to include copyright-violated images (e.g., the exact or modified images) given the query image or may incorrectly include irrelevant images, given an image watermark method, since any image watermark method has certain probabilities for false negative and false positive. Moreover, it is prohibitively expensive to search copyright-violated images by exhaustively scanning images in the image database and performing the watermark similarity tests.

One may want to accelerate the performance of identifying images that have top k watermark similarity values in the database, by transforming the problem of identifying such images into the problem of finding k nearest neighbors [1]. Then we can borrow well-established acceleration techniques for the nearest neighbor problem. One of the main acceleration techniques is to use a hierarchy (e.g., kd-trees) computed from image watermark patterns that are pre-extracted from images of the database, and to perform the nearest neighbor search given the watermark pattern of the query image.

This hierarchical approach, however, has a major limitation that makes the approach impractical. Since most image watermark techniques require a private key of the copyright holder to extract watermark patterns from images [21], it is impossible to even pre-extract watermark patterns until query images are available.

In this paper we aim to improve both the performance and accuracy of the exhaustive by adopting an image-retrieval technique as a culling step that does not need to pre-extract watermark patterns and still handles large-scale image databases.

3.2 Overview of Our Approach and Expected Benefits

As a pre-computation step of our IRIW approach, we construct a vocabulary tree with image features (e.g., SIFTs) of images in the database. Then, we perform our runtime algorithm that consists of three phases (Fig. 1): 1) image retrieval, 2) on-demand watermark extraction, and 3) re-ranking phases. Given a query image, we first identify similar images by performing our SIFT-based image retrieval method and store them in an output list, called *IR output list*. Then we extract a watermark pattern on demand for each image in the IR output list,

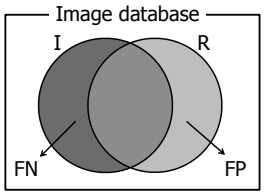

Fig. 2. The ground-truth images, I, that are modified from a query image, and a result set, R, computed by an image watermark method.

followed by performing the watermark similarity tests between images in the output list and the query image. As a final step, we re-rank images in the output list by considering the computed similarity values and other additional information (e.g., similarity values among images in the output list), and provide our final output list to users. Also, we use a SIFT-aware image watermark technique that does not interfere with our SIFT-based image retrieval with watermarked images.

Our proposed method has the following benefits:

- **Higher performance:** By identifying similar images given a query image and then performing the watermark similarity tests only against those similar images, we can drastically reduce the number of images that we need to consider for image watermarking, leading to fast runtime performance for large-scale image databases. Note that the image retrieval component serves as a culling step for an image watermark method employed in our IRIW framework.
- **Higher accuracy:** By excluding dissimilar images based on our image retrieval component from the IR output list and by measuring watermark similarities against images stored only in the list, we can reduce the number of false positive cases in the final output list. This is because that it is likely that strongly dissimilar images are not modified from the query image and thus they are not copyright-violated with respect to the query image. Moreover, we also reduce the number of false negative cases by identifying similar images and placing them in the final output list, even though they may have low watermark similarity values caused by severe image editing attacks.

4 Our Approach

In this section we describe different steps of our approach in a detailed manner.

4.1 Image Retrieval Phase

As the first step of our method, we perform our image retrieval method to identify images that are similar to the given query image.

Suppose that given a query image, I_q, we have a set, I, of images modified from the query image I_q in an image database (Fig. 2); I serves as ground-truth results that are modified from the query image. Any image watermark

methods aim to produce a set, R, of result images that contains all of those modified images. However, because of inaccuracy of image watermark methods, we may get a set, FP, of false positive images, which are irrelevant images (i.e. $FP \cap I = \phi$) given the query image, but are included in R. Moreover, we may fail to identify a subset, FN, of those modified images as false negative images; therefore, $FN \subseteq I$, but $FN \cap R = \phi$.

The goal of our image retrieval phase is to compute an image output list such that the list reduces the cardinalities of two sets FP and FN. To achieve our goal, we propose to use a SIFT-based image retrieval method, since it has been studied extensively recently and reported to perform well in terms of identifying images that have similar image features [7]. By performing our SIFT-based image retrieval method, we compute an output list, called *IR output list*, of images sorted in terms of *image similarity*, which will be explained later.

Note that we identify images that are similar to the query image and report them in the IR output list. Dissimilar images cannot be in the IR output list and thus will be excluded in the final output list (Fig. 1), even when some of dissimilar images happen to have relatively high watermark similarity values against the query image. As a result, we can reduce false positive cases. Also, severely modified or attacked images may have low watermark similarity values against the query image. It is possible that they may not be included in the final output list, if the list is computed from the exhaustive method that reports images sorted only in terms of watermark similarity values. Nonetheless, those severely attacked images may still have similar image features and thus can be included in the IR output list computed from our SIFT-based image retrieval method. Since our final output list contains all the images of the IR output list with different ranks in the list, those severely attacked image can be included in the final output list.

Pre-computation: We perform our retrieval method based on the concept of visual words [17]. For all the images in the image database, we extract SIFT features and cluster them into visual words. In order to accelerate the clustering process, we adopt a hierarchical clustering method [13]. Starting from the root cluster that contains all the SIFT features, we recursively partition it into t different child clusters. We stop the recursive process if the depth of a cluster reaches a pre-defined threshold. Then we make those clusters leaf clusters that serve as visual words. For each leaf cluster, we compute a representative SIFT feature by averaging SIFT features assigned to the cluster and record images that are related to those contained SIFT features. This hierarchical construction method creates a t-ary tree that serves as a vocabulary tree.

Runtime process: Once a user provides a query image at runtime, we extract SIFT features from the query image. Then for each SIFT feature, we traverse the vocabulary tree and find a leaf cluster whose representative SIFT feature is closest to the SIFT feature. We also add the images associated with the leaf cluster into a *similar image list*. Once we represent the query image with a set of visual words, then we compute the *image similarity value* based on the visual words of the query image and those of images stored in the similar image list [13]. As the final step, we sort images in the similar image list based on the computed image similarity values and store top r images in the IR output list, which is fed to the next phase.

4.2 Watermark Detection and Re-Ranking Phases

After computing the IR output list, we measure watermark similarity values between the query image and images in the list; we will explain our image watermark method in the later section. Then, we re-rank images in the list by considering both image and watermark similarity values and store them in the final output list.

One can return the final output list, whose images are sorted only by the watermark similarity values. Note that it is highly likely that we get a very low watermark similarity values for severely modified or attacked images, even though our image retrieval method identifies them in the IR output list. As a result, these images are likely to be located near the bottom of the list and thus it hinders users to identify those modified images in an efficient manner. It is desirable to locate them higher in the list, even though they have low watermark similarity values.

In order to address this problem, we propose to re-rank images by utilizing information among the images stored in the list. Moreover, we re-rank images based on a weighted sum of image and watermark similarity values, instead of reporting images only according to the watermark similarity values for the final output list.

As an initial step, we associate a score value with each image in the IR output list, where the score value is initialized with the sum of image and watermark similarity values computed against the query image. According to the current score values, we sort images and store them in the list.

Then we perform our re-ranking by utilizing information available among images in the list. In each iteration of our re-ranking phase, we compute image similarity values between the first-ranked image and other images in the list. We accumulate the similarity value computed with each image in the list to the score associated with the image. As the final step of the iteration, we sort images in the list according to the current scores of those images. We iterate this process again with the next ranked image in the list. We found that running two iterations works well in our experiments.

4.3 SIFT-Aware Image Watermarking

Our image retrieval phase works by considering SIFT image features. If the image regions that contain SIFT image features are affected by embedded watermark patterns, results of image retrieval with watermarked images would be different from those before embedding watermark patterns on images. At the worst case, certain image features may not be extracted from the watermarked images. As a result, image retrieval may fail to identify similar images. This can deteriorate the accuracy of our framework, since our method performs image watermarking only with the IR output list computed from the image retrieval phase.

In order to prevent this problem, we propose to use an image watermark technique that takes advantage of SIFT features of images, inspired by image watermark techniques that utilize invariant image features [10]. We generate a donut-shaped watermark pattern (Fig. 3) and identify SIFT image features for each image. Then we embed the donut-shaped watermark pattern whose position is at the center of each extracted SIFT image feature.

Fig. 3. This figure shows (a) the original Lena image with its SIFT features shown as circles, (b) watermark patterns that will be embedded around the SIFT features, and (c) watermarked image and its extracted SIFT features shown as rectangles with SIFT features of the original image shown in circles. We show only five SIFT features that have the top-five highest strength values.

Since a SIFT image feature is extracted from a 16 by 16 image region, the inner circle of each donut-shaped watermark pattern is computed to have a radius such that the inner circle can contain its associated 16 by 16 image region. For each image, about one thousand SIFT features are extracted. A *strength* variable for each SIFT feature is defined as the difference of Gaussians in two varying resolutions that contain the scale of the feature. Note that as a SIFT feature has a higher strength value, it is more likely that the SIFT feature survives with various attacks. As a result, we choose SIFT features that have high strength values and embed the donut-shaped watermark pattern at those SIFT features. More specifically, we choose SIFT features in the order of decreasing strength values, while avoiding any overlaps among the patterns associated with the SIFT features that are considered currently and were chosen previously. Also, we found that in this configuration, the chosen SIFT features are well distributed across the image and thus our technique can be robust for attacks such as image cropping.

Since we embed the donut-shaped watermark pattern on the SIFT image features, local gradient values around the center point of each SIFT image feature is not changed. As a result, even after embedding watermark patterns, we can extract most of the same SIFT image features and thus achieve a similar result with image retrieval even after embedding watermark patterns.

5 Results and Discussions

We have implemented our IRIW method and performed various tests with a 32 bit machine that consists of 2 GB memory and 3 GHz CPU.

Image benchmark: In order to test our method, we prepare an image benchmark that consists of 10 K images. The benchmark includes the well known images (e.g., Lena, Mandrill, and Goldhill) and images from the Cal-Tech 101 and UKBench image datasets. In our image benchmark, 100 different categories (e.g., airplanes, cups, cars, etc.) are defined. Also, each category has ten different, but similar images. In each category, we select two images among

ten similar images and embed two different watermarks into them. We leave the original un-watermarked images in our image benchmark. Since these original un-watermarked images do not have any watermark patterns, they can serve as images that could have been generated with ideal attacks, when we use watermarked query images. Also, to represent various attack scenarios, we attack each of watermarked images in eight different ways; we use the standard image attack generation tool, called Stirmark [14]. More precisely, these different attack scenarios include addictive noise (2% of the average pixel value), median filtering (3×3 box filter), center-cropping (75%), JPEG compression (lossy 70%), scaling (75%), rotation (45° and 90°), and shearing (1% extension along X and Y directions). Note that both JPEG compression and median filtering cause blurring that can affect SIFT features of images.

Vocabulary tree construction: Our image retrieval method is based on SIFT image features and uses the concept of vocabulary trees [13]. We perform the hierarchical k-means construction with SIFT features in order to construct a vocabulary tree. Our vocabulary tree has a depth of four with ten branches for each intermediate node; therefore, the tree has 10 K leaf nodes. From our image benchmark, we extract 4.5 million SIFT features, and it takes about 56 min to construct the vocabulary tree for the benchmark.

Comparison setting: In order to show the benefits of our method, we compare the runtime performance and accuracy of our method against those of the exhaustive method that checks all the images in the image database. In both methods, we set them to report 30 different images as their results given a query image. The image retrieval component of our method also computes the IR output list that contains 30 different images. In all the tests, we perform 100 different search queries to identify copyright-violated images, and compare the average performance and accuracy between these two different methods.

5.1 Runtime Performance

Achieving a higher runtime performance for identifying copyright-violated images is very important to support search queries in large-scale image databases for a more number of users. Therefore, we compare the runtime performance of our method against the exhaustive method.

The exhaustive method computes the watermark similarity value for each image in our database, and spends about 19 min. to compute top-30 images sorted according to only watermark similarity values. The exhaustive method spends most of its running time of extracting and comparing watermark patterns. On the other hand, our method spends 5.9 sec. to compute the top-30 images according to the sum of image and watermark similarity values. Since our image retrieval component identifies a small subset (e.g., 30 images) of images that serve as candidates for potentially modified images from the query image and we perform our image watermark extraction for only those images, our method achieves a much higher runtime performance, more than two orders of magnitude performance improvement, over the exhaustive method.

Within the average running time, 5.9 sec. of our method, our method spends 0.34 sec. and 0.71 sec. to extract SIFT features from the query image and identify top-30 similar images. The rest of the running time, 4.9 sec., is spent on performing watermark extraction and watermark similarity tests.

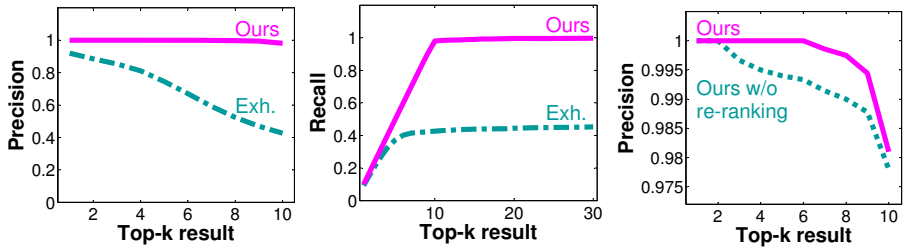

Fig. 4. The left and middle graphs show precision and recall curves of our method and the exhaustive method. The right graph shows precision curves w/ and w/o re-ranking images.

One may think that we can pre-compute watermark patterns for images stored in the database and construct a hierarchical acceleration structure to improve the performance of the exhaustive method. However, as highlighted in Sec. 3.1, it is impossible in practice to pre-extract watermark patterns from images because many watermark methods can be used and some of them can use private keys associated with query images that disallow the pre-extraction. Therefore, we decide to compare our method against the exhaustive method that does not have such problems and works in a wide variety of usage scenarios for detecting copyright-violated images.

5.2 Accuracy

We measure the accuracy of two methods in terms of ratios of false negative and false positive results given the ground-truth results of query images. Inspired by notions of *precision* and *recall* used for image retrieval, we also connect ratios of false positive and false negative results with precision and recall respectively for image watermark methods.

We define the ratio, FP_r, of false positive results to be a ratio of the number of irrelevant images that are not in the ground-truth result of the query image, but are in the final output list, to the size of the final output list; therefore, $1 - FP_r$ can be interpreted as precision. We also define the ratio, FN_r, of false negative results to be a ratio of the number of ground-truth images given a query image that are not in the final output list, to the size of the final output list. As a result, $1 - FN_r$ can be thought of as recall. Since the concepts of recall and precision are more intuitive, we represent the accuracy of different techniques in terms of those two concepts.

Fig. 4 shows the precision and recall curves of our and the exhaustive methods. Note that in our image benchmark there are ten ground-truth images (i.e. one original image, its watermarked image, and eight differently attacked images from the watermarked image) given a (watermarked) query image; ground-truth images for query images used in Fig. 5 are shown in Fig. 1 in the supplementary report, which is available at `http://sglab.kaist.ac.kr/IRIW`. As can be seen in the recall curve, our method achieves a near-linear recall curve up to the top-8 image in the final output list and reaches a recall value close to 1 around

the top-12 and the top-13 images in the list. On the other hand, the exhaustive method does not achieve a recall value of more than 0.5, even though we allow up to top-30 images in the list. This is because many irrelevant images have more higher watermark similarity values than those of ground-truth images in the exhaustive method. Similarly, our method achieves up to two times higher precision results over the exhaustive method as we vary the size of the final output list. Improvement achieved by our re-ranking method is shown in the right graph of Fig. 4. Results before and after re-ranking are available in Fig. 2 of the supplementary report.

Examples of our results given two different query images are shown in Fig. 5. The exhaustive method achieves comparable results over our method for the Mandrill image up to top seven images. However, its result deteriorates after the top-7 images, while our method achieves accurate results up to top ten images; see Fig. 3 of the supplementary report for top-6 to top-10 images. In the Mona Lisa image, the exhaustive method reports an irrelevant image (Fig. 5-(p)) at the top-5 place, while our method reports one of ground-truth images, the original image, at the top-5 place. Since the original image does not have any watermark in it, it serves as one of images attacked by ideal image editing scenarios and– thus is very hard to be identified by prior image watermark methods. This result supports that our approach can detect copyright violated images even if their watermark patterns has been removed.

5.3 Discussions

Our approach with other image watermark methods: To show benefits of our IRIW approach even with other image watermark methods, we combine our approach with a DCT-based image watermark method [16]. This DCT-based image watermark method works in the frequency domain, while our SIFT-aware image watermark method works in the spatial domain. Compared with the exhaustive method that uses the DCT-based image watermark method, our IRIW approach with the DCT-based method still achieves 233:1 performance improvement. Moreover, our IRIW approach with the DCT-based one achieves up to 2:1 accuracy (i.e. precision and recall) improvements in a similar manner shown in Fig. 4.

Effects on the accuracy of image retrieval: To further verify the amount of effects of our SIFT-aware image watermark method on the accuracy of our SIFT-based image retrieval, we measure the mean Average Precision (mAP) of our SIFT-based image retrieval method. Our image retrieval method shows 0.99 mAP with images that do not have any watermarks. We also measure the mAP after embedding watermarks on all the images and mAP is changed only a bit (e.g., less than 1% changes). This result verifies that our SIFT-based image watermark method does not have major effects on the image retrieval accuracy even after embedding watermark patterns on images.

Limitations: Our IRIW approach employs an image retrieval component to cull most of irrelevant images given a query image. If our image retrieval component fails to identify similar images that are copyright-violated, our method cannot report such images in the final output list. However, we found that our SIFT-based image retrieval method works quite well in our tested image benchmark. Also, there may be attack scenarios, where our IRIW method may not

work well. For example, one can apply severe blurring on images to affect most SIFT features and then deblur the blurred images based on recent advanced deblurring techniques. We expect that our method may not work well in such extreme cases, while the exhaustive method is also expected not to work well. Also, we can improve the accuracy of the exhaustive method by adopting simple geometric verifications [15] and culling irrelevant images based on simple image information (e.g., color histrom) given a query image. However, we can also adopt the same approach to our IRIW approach to further improve its accuracy.

6 Conclusion

We have presented a novel, Image Retrieval based Image Watermark (IRIW) approach that uses a SIFT-based image retrieval component to efficiently and accurately identify similar images from a query image. Our method extracts watermark patterns and measures watermark similarity values against images only in similar images identified from our image retrieval component. We have also proposed a re-ranking method to place severely attacked images even in higher positions in the final output list. Finally, we have proposed to use a SIFT-aware image watermark method that does not have negative effects on the image retrieval component. As a result, we were able to show more than two orders of magnitude performance improvement and up to two times accuracy improvement over the exhaustive method that scans all the images in an image database.

There are many interesting future research directions. In addition to addressing current limitations of our system, we would like to design an interactive IRIW system for web-scale image databases based on a recent large-scale image retrieval method [9]. It would require massive parallelization on all the components of our current system. Also, we would like to investigate efficient watermark extraction methods by utilizing GPUs. Also, we found that sometimes users can provide additional information about similarities among images. Therefore, we would like to design effective visualization and browsing tools for large-scale image databases. Finally, we would like to design a frequency-domain image watermark method that maintains SIFT image features even after embedding watermark patterns.

Acknowledgements

We would like to thank anonymous reviewers and members of SGLab for their constructive comments and feedbacks. The first author was supported in part by the Korea Student Aid Foundation (KOSAF) grant funded by the Korea government (MEST) (S2-2009-000-00971-1). Also, this research is supported in part by MKE/KEIT [KI001810035261], MSRA, MKE/MCST/IITA [2008-F-033-02], KRF-2008-313-D00922, BK, DAPA/ADD (UD080042AD), MEST/NRF/WCU (R31-2010-000-30007-0), KMCC, and MCST/KOCCA/CT/R&D 2011.

References

1. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.: An optimal algorithm for approximate nearest neighbor searching. In: Symp. on Discrete Alg. pp. 573–582 (1994)
2. Bas, P., m. Chassery, J., Macq, B.: Geometrically invariant watermarking using feature points. IEEE Trans. on Image Processing 11, 1014–1028 (2002)
3. de Berg, M., Cheong, O., van Kreveld, M., Overmars, M.: Computational Geometry: Algorithms and Applications. Springer-Verlag TELOS, Santa Clara, CA (2008)
4. Berrani, S.A., Amsaleg, L., Gros, P.: Robust content-based image searches for copyright protection. In: Proceedings of the 1st ACM international workshop on Multimedia databases. pp. 70–77 (2003)
5. Chum, O., Philbin, J., Isard, M., Zisserman, A.: Scalable near identical image and shot detection. In: ACM international conference on Image and video retrieval. pp. 549–556 (2007)
6. Cox, I., Kilian, J., Leighton, F., Shamoon, T.: Secure spread spectrum watermarking for multimedia. Image Processing, IEEE Transactions on 6(12), 1673 –1687 (1997)
7. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Survey 40(2), 1–60 (2008)
8. Huston, L., Sukthankar, R., Ke, Y.: Evaluating keypoint methods for content-based copyright protection of digital images. In: Multimedia and Expo (ICME). IEEE International Conference on. p. 4 pp. (july 2005)
9. Jégou, H., Douze, M., Schmid, C., Pérez, P.: Aggregating local descriptors into a compact image representation. In: CVPR. pp. 3304 –3311 (2010)
10. Lee, H.Y., Kim, H.S., Lee, H.K.: Robust image watermarking using invariant features. Optical Engineering 45(3), 1–11 (2006)
11. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)
12. Lu, Z.M., Skibbe, H., Burkhardt, H.: Image retrieval based on a multipurpose watermarking scheme. In: Knowledge-Based Intelligent Information and Engineering Systems. pp. 573–579 (2005)
13. Nister, D., Stewenius, H.: Scalable recognition with a vocabulary tree. In: CVPR. pp. 2161–2168 (2006)
14. Petitcolas, F.: Watermarking schemes evaluation. IEEE Signal Processing Magazine 17(5), 58–64 (2000)
15. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: CVPR. pp. 1–8 (2007)
16. Piva, A., Barni, M., Bartolini, F., Cappellini, V.: Dct-based watermark recovering without resorting to the uncorrupted original image. In: ICIP. pp. 520– (1997)
17. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: ICCV. vol. 2, pp. 1470–1477 (2003)
18. Tang, C.W., Hang, H.M.: A feature-based robust digital image watermarking scheme. IEEE Trans. on Signal Processing 51(4), 950–959 (2003)
19. Xu, J., hua Qin, W., ying Ni, M.: A new scheme of image retrieval based upon digital watermarking. In: Int. Symp. on Computer Science and Computational Tech. pp. 617–620 (2008)
20. Zhao, W.L., Ngo, C.W.: Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. Image Processing, IEEE Transactions on 18(2), 412–423 (2009)
21. Zheng, D., Liu, Y., Zhao, J., Saddik, A.E.: A survey of rst invariant image watermarking algorithms. ACM Computing Survey 39(2), 5 (2007)

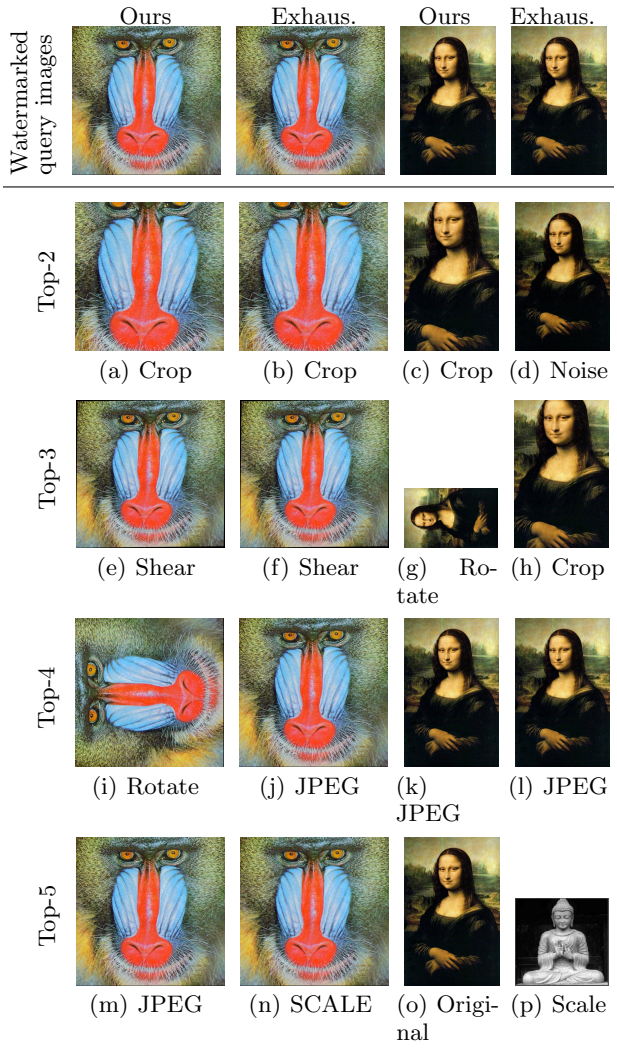

Fig. 5. This figure shows returned results in the top-5 images of our and exhaustive methods given the watermarked query images shown in the top row. We do not show the top-1 images since the query images are returned at the top-1 images in all the cases. Sub-captions from (a) to (p) represent image attack used to create the corresponding images. Top-6 to top-10 images are available at the supplementary report.