SECURE AND ROBUST SIFT WITH RESISTANCE TO CHOSEN-PLAINTEXT ATTACK

Chao-Yung Hsu,[1,2] *Chun-Shien Lu,*[2,*] *and Soo-Chang Pei*[1]

[1]Graduate Institute of Communication Eng., National Taiwan University, Taipei, Taiwan, ROC
[2]Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC

ABSTRACT

Scale-invariant feature transform (SIFT) is a powerful tool extensively used in the community of pattern recognition and computer vision. The security issue of SIFT, however, is relatively unexplored. We point out the potential weakness of SIFT, meaning that the SIFT features can be deleted or destroyed while maintaining acceptable visual qualities. To properly achieve the tradeoff between security and robustness of SIFT, we present a cube-based secure transformation mechanism to enable the SIFT method to resist up to the chosen plaintext attack while robustness against geometric attacks can still be maintained. Security analysis and robustness verification are provided to demonstrate the effectiveness of the proposed (and modified) SIFT method.

Keywords: Attack, Image hashing, Robustness, Security

1. INTRODUCTION

Scale-invariant feature transform (SIFT) conducted in the difference-of-Gaussian (DoG) scale space domain [2] has been widely used due to its powerful attack-resilient keypoint detection mechanism. For the applications of SIFT in multimedia security, Roy and Sun [4] proposed to generate hash sequences from thresholding SIFT feature vectors. An intuitive way to defeat Roy and Sun's method is to remove or insert the feature points in an image while keeping certain visual quality. While this is regarded impossible before, we first addressed the security issue of SIFT-based methods in [1]. In our previous work, we present two anti-SIFT attacks, which are validated via studying the relationship between image quality (in terms of PSNR) and keypoint removal rate. Then, in view of the fact that SIFT is indeed a powerful method in representing keypoints in an image, a secret key-based random process is introduced in an image such that the resultant perturbed keypoints in the transformed domain are hard to be deleted. Our results show that under the same PSNR the ratio of removed keypoints in the proposed secure SIFT method is significantly lowered than that in conventional SIFT.

Although the idea of detecting SIFT features in a secure transformed domain [1] is promising, its security could be further enhanced. In this paper, we propose a secure SIFT method with resistance up to chosen plaintext attack while preserving robustness against geometric attacks (e.g., Stirmark [3]). The idea comes from the observation that the notion of local extreme conventionally employed in SIFT is not the unique structure for robust keypoint extraction, in particular, when security is required to be taken into consideration. In view of this, we properly modify the original SIFT method in a way that a feature point is identified if the energy (DoG magnitudes) of a local area (i.e., a cube) centered at it is sufficiently large. This structure defined based on DoG energy is found to be very robust even under geometric attacks, which meets the wide applications of multimedia security, including media hashing and copy/duplicate detection. For security, the cubes of high energies are clustered to find the representative feature cubes (usually of size $3 \times 3 \times 3$) in the DoG domain. These representative feature cubes can then be mutually permutated to perturb SIFT feature detection in a rather secure way. As we have analyzed later, the SIFT features cannot be attacked because the secret key used for perturbation is very hard to be guessed or derived. In particular, security analysis and robustness verification demonstrate the effectiveness of the proposed image hashing method.

2. PROPOSED METHOD

In this section, we first discuss the insecurity of SIFT and then propose a secure and robust SIFT method.

2.1. Anti-SIFT Attack based on Unique Detection Structure of SIFT

In SIFT detection, a pixel is decided as a keypoint if and only if it is a local extremum in the scale space defined by difference-of-Gaussian (DoG) functions. A local extremum at a pixel is found if its DoG magnitude is larger than those of its neighbors. In [1], an anti-SIFT attack based on the unique detection structure of SIFT via enforcing duplicate extrema for restraining SIFT detection is proposed. Specifically, we aim to remove a keypoint by modifying intended pixels to yield more than one local extremum in a detection

[*]Corresponding Author: Dr. C. S. Lu (lcs@iis.sinica.edu.tw)

region. The idea behind our method is based on the observation that an original keypoint will not be detected by SIFT if another extremum is maliciously generated nearby. That is, there are two equal extrema in a detection region such that the duplicate extremum is enforced to be at one of the eight neighbors in the scale space to evade keypoint detection.

The anti-SIFT attacks can be successful because the inherent "structure" corresponding to a keypoint is "dominant" and can be exploited for anti-detection purpose. Obviously, such a dominant keypoint is visible mathematically (with a sharp bell in the DoG domain) or visually (via collage attack), and enables to be removed. Please refer to [1] for more details about anti-SIFT attack.

2.2. Secure and Robust SIFT

In our prior work, we present a secret key-based transformation process, which is performed on images before SIFT feature detection, such that the dominant features become recessive. This implies that the detection of SIFT features will be conducted in the transformed domain instead of the original spatial domain, and the goal of secure SIFT can be achieved. The previously proposed strategy is simple but needs a more sophisticated design to enhance security if known-plaintext attack or chosen-plaintext attack is considered. In this paper, we will address this issue.

As discussed in Sec. 2.1, there are basically two states in the original SIFT method [2] and our secure SIFT method [1]: a pixel's DoG magnitude is a local extreme which is a SIFT feature, and otherwise. Even the two states are proposed to be shuffled in Hsu *et al.*'s work, the adversary with stronger capability can still break it. This is because the security of Hsu *et al.*'s work is obtained from the XOR-based binary encryption. Thus, the secret key can be estimated from ciphertext-plaintext pairs when known-plaintext attack or chosen-plaintext attack is adopted for cryptanalysis. In this paper, we propose to properly modify the original SIFT detection by replacing the unique local extreme structure with the local cubes with large DoG energy. Specifically, we seek to find those local cubes (say $3 \times 3 \times 3$) with energy (also defined in the DoG domain) large enough as the regions, where SIFT features reside. In order to preserve security, these candidate cubes are further clustered into C clusters, where each of them defines a kind of feature points. Note that one cluster is composed of more than one candidate cubes, and the candidate cubes in a cluster can be interchanged to disturb and, thus, achieve secure SIFT extraction. The details will be elaborated as follows.

Let $G(x, y, s)$ be a DoG value defined at position (x, y) and scale s and let $G(x+i, y+j, s+k)$ be a DoG value neighboring to $G(x, y, s)$, where $-1 \leq i, j, k \leq 1$. That is, we consider $3 \times 3 \times 3$-dimensional cube in the DoG domain for SIFT detection. We also let $S(\cdot)$ be a decreasing sorting list of energies in a cube centered at $G(x, y, s)$. The top-T elements of $S(\cdot)$ are determines as:

$$T = \arg\min_{T} |\sum_{g=1}^{T} S(g) - \alpha \sum_{g=1}^{27} S(g)|, 0 < \alpha \leq 1, \quad (1)$$

where α is a user defined parameter that is used to sieve out the robust features. With the obtained T, the binarized cube centered at $G(x, y, s)$ is defined as:

$$B(G(x+i, y+j, s+k)) = \begin{cases} 1, \text{ if } G(x+i,y+j,s+k) \leq S(T) \\ 0, \text{ otherwise} \end{cases}. \quad (2)$$

The binarized cubes instead of the corresponding cubes with DoG magnitudes can facilitate clustering discussed in the following. Finally, rotation-invariant clustering is performed on the binarized cubes, generated from Eq. (2). As an illustrative example shown in Fig. 1, Fig. 1(a) shows a binarized traditional SIFT feature structure, Fig. 1(b) shows one example of our defined robust feature structure, and Figs. 1(c)-(f) show the four rotational versions of (b). To achieve rotation-invariance, Figs. 1(b)-(f) are clustered as the same class in this paper. In other words, the fact that any one of them is found will define the same SIFT feature.

After the feature structures are clustered, the elements of each class with DoG values ($i.e.$, $G()$) instead of binary values ($i.e.$, $B(G())$) are averaged to calculate the centroid of a class, which represents a feature cube in the DoG domain. In the next subsection, we propose secure linear transformations between feature cubes to satisfy secure SIFT.

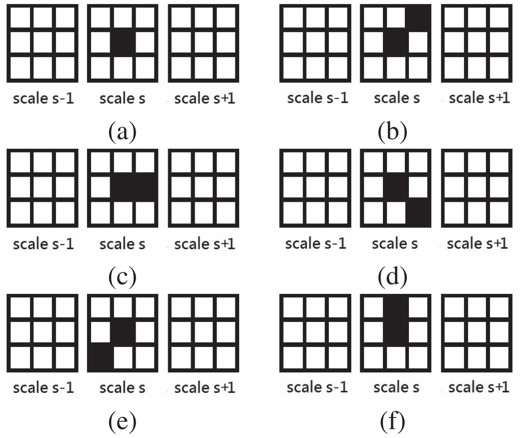

Fig. 1. Examples of rotation-invariant clustering for binary cubes: (a) illustrates the binarized traditional SIFT feature structure; (b) represents a feature structure defined in this paper; (c)-(f) show four rotational versions of (b). In our rotation-invariant clustering, (b)-(f) are clustered as the same group.

2.3. Secure Linear Transformation

Assume we have M feature cubes available, which are represented as C_1, C_2, ..., and C_M. In addition, only one type of them is approved of containing the SIFT feature we would like to extract, and the remainder are mainly used to hide this fact from adversary. Recall that our secure SIFT scheme is proposed to be conducted in the domain controled by a secret key and only one type of feature cubes is assigned to define the existence of feature points. Therefore, a secret key is used to conduct a series of transformations among these feature cubes. To simplify analysis here, if M feature cubes are used, then there are in total $(M-1)!$ circular permutations. For example, $C_6- > C_3- > C_8- > C_1- > C_5- > C_7- > C2- > C_4$ defines a circular permutation for feature cubes.

Nevertheless, the transformation mentioned above can be broken when the adversary try to predict the rule of transformation based on some pairs of the plaintext (the image in the original DoG domain) and ciphertext (the image in the transformed DoG domain via cube permutation), which is known as the known plaintext attack (KPA). It should be noted that this attack indeed relies on the knowledge of ciphertext, which is, however, not directly accessible to the adversary in our secure SIFT scheme since the ciphertext, an intermediate, will not be finally produced for feature detection.

To further resist again other (advanced) attacks, the transformation conducted in a single path manner may be not suitable. As a matter of fact, even multiple transformations in a single path can be simplified as only one linear transformation. In the following, we consider the secure transformation conducted on more than one path to increase the randomness of transformed results.

Let's explain the idea using two-path transform, which is illustrated in Fig. 2 as an example. First, all the circular permutations in a pool are grouped in a pair-wise manner. Second, a secret key K is used to select a series of pair-wise circular permutations, i.e., an element of a secret key chooses a pair of circular permutations. Third, to further increase the randomness of feature cube exchange, a robust content (image)-dependent key, where each element depends on the underlying image is necessary. Our empirical results also reveal that the local maximum of the energy cube in the DoG domain are robust features and adapt to different image distortions. Such a characteristic is well suitable to define the image-dependent key. The energy cube in the DoG domain centered at position (x, y) and scale s is defined as:

$$EDoG(x, y, s) = \sqrt{\frac{1}{27}\sum_{i,j,k} G^2(x+i, y+j, s+k)}, \quad (3)$$

where i, j, and k are integer indices within $[-1, 1]$. A binomial random variable X, which is dependent on the values in the energy map, is expressed as:

$$\begin{aligned} P(X=1) &= P(EDoG(x, y, s)\text{is local maximum}) \\ &= 1 - P(X=0). \end{aligned} \quad (4)$$

Finally, the content-dependent key CK formed by the random variable X is incorporated with the secret key mentioned above to indicate $\frac{(M-1)!}{2}$ transforms for the case of two-path transformation.

Fig. 2 shows an example of a two-path transformation (like a binary tree). Given a feature cube, it can be finally transformed to different cubes according to two keys K and CK. As a result, the results of secure transform for the same feature cube using different secret keys and content-dependent keys become unpredictable. It is worth noting that for the same input cube (say C_1 in Fig. 2), the resultant output (i.e., permutated result) will be different. Based on the permutated results, we randomly select a type of feature cubes to define the SIFT features. The phenomena will hinder the adversary from knowing where the features should reside and launching the corresponding attacks. In the next section, we shall discuss the security of the proposed content-dependent multi-path secure transform against ciphertext only attack (COA) and chosen plaintext attack (CPA).

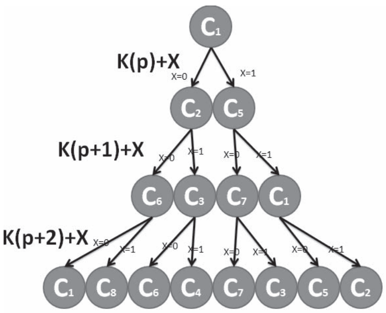

Fig. 2. An example of 3-layer secure transformation for a given feature cube.

3. SECURITY ANALYSIS

To verify the security of the proposed secure SIFT method, we analyze ciphertext only attack (COA) and chosen plaintext attack (CPA) in this section.

COA is completely successful if the corresponding plaintext or secret key can be deduced from the given ciphertext. The probability that an attacker hits the secret key depends on the length L of the secret key and the number M of representative feature cubes (or clusters):

$$P(K_{COA} = K_{true}) = \frac{1}{[\frac{(M-1)!}{2}]^L}, \quad (5)$$

where K_{COA} represents the randomly guessed key by attacker and K_{true} is the secret key used in the proposed method. The lower bound of key length is $\log_2 M$ since the transform must be performed at least $\log_2 M$ times to guarantee that each incoming feature cube can be transformed to all M classes (see the leave nodes of Fig. 2). On the other hand, the security of our method won't be increased with the increase of the key's length if guessing the features directly is easier than key estimation. Therefore, we have

$$(\frac{(M-1)!}{2})^L \leq M^R, \tag{6}$$

where R is the number of feature cubes after clustering and M^R denotes the number of cases in guessing the feature cubes directly. According to Eq. (6), the upper bound of key length can be derived as:

$$L \leq R\frac{\log_2(M)}{\log_2(\frac{(M-1)!}{2})}. \tag{7}$$

In sum, the length of secret key is obtained as:

$$\log_2 M \leq L \leq R\frac{\log_2(M)}{\log_2(\frac{(M-1)!}{2})}, \tag{8}$$

which provides a guideline to set the length of a secret key.

For the case of CPA, the adversary can choose arbitrary plaintexts to be encrypted and obtain the corresponding ciphertexts when our algorithm is treated as a black box. The goal is to gain further information, which enable to reduce the degree of security for a method. To defeat our method, the best plaintext for testing is an image with a single feature, which can help the adversary easily understand the output of secure transform without needing the knowledge of the content-dependent key. Under this circumstance, M images with different single features make the probability of successful attack increase to

$$P(K_{CPA} = K_{true}) = \frac{M^M}{(\frac{(M-1)!}{2})^L}. \tag{9}$$

In Fig. 3, we show that under different parameters of M and L, $P(K_{CPA} = K_{true})$ is sufficiently low.

4. ROBUSTNESS VERIFICATION

We verify the robustness of the proposed secure SIFT method to see if robustness is sacrificed when security is enhanced. Please also refer to [1] for the efficiency of SIFT keypoint removal using the anti-SIFT attack.

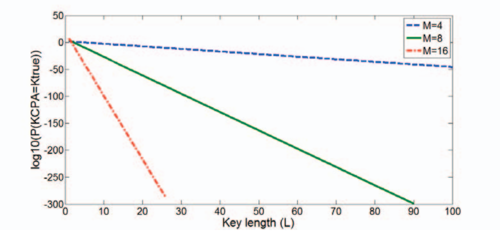

Fig. 3. Logarithmic probability (y-axis) of $P(K_{CPA} = K_{true})$ under different M's and L's.

Ten color images with different contents (I_1: Splash; I_2: Lenna; I_3: F-16; I_4: Tank; I_5: Bridge; I_6: Baboon; I_7: Goldhill; I_8: Clock; I_9: Sailboat; and I_{10}: Peppers) were used to verify the robustness of our scheme against miscellaneous attacks. The standard benchmark [3], Stirmark 3.1 and 4.0, was quite suitable for simulating various manipulations of the digital images. In this test, the original image was used as a query to find out how many modified versions could be successfully detected by comparing the detected SIFT feature vectors.

Similar to the results reported in [1], the secure SIFT method proposed here can resist geometric attacks defined in Stirmark very well except for the noise adding attacks, which can interfere the detection of either local extreme of DoG magnitudes or energy cube described in this paper. Our experimental results confirm that the introduced secure transformation mechanism only slightly affect robustness. Thus, the tradeoff between security and robustness can be maintained properly.

5. CONCLUSIONS

Scale-invariant feature transform (SIFT) is a robust keypoint detector and has been extensively used in the literature. We first present an anti-SIFT attack to combat the belief that the feature points representing the essence of an image cannot be destroyed without significantly degrading visual quality. We then propose a cube-based secure transformation mechanism to enable the SIFT method to resist against chosen plaintext attack while robustness can still be maintained. Secureity analysis and robustness verification are provided to demonstrate the security and robustness of the proposed method.

6. ACKNOWLEDGMENTS

This work was supported by National Science Council, Taiwan, ROC, under Grant NSC 98-2631-H-001-013.

7. REFERENCES

[1] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Secure and Robust SIFT," *Proc. ACM Multimedia Conference*, pp. 637-640, Beijing, China, 2009.

[2] D. Lowe, "Distinctive image features from scale invariant keypoints," *IJCV*, Vol. 60, pp. 91-110, 2004.

[3] F. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Attacks on Copyright Marking Systems," *Information Hiding Workshop*, LNCS 1575, 1998.

[4] S. Roy and Q. Sun, "Robust Hash for Detecting and Localizing Image Tampering," *IEEE ICIP*, 2007.