A Geometric-Rotation-Attack-Resistant Zero-Watermark Algorithm (Postprint)

Authors: Liu Wanjun, Sun Siyu, Qu Haicheng, Feng Lin, He Muze

Date: 2018-06-19T00:00:00+00:00

Abstract

To address the issue of weak robustness against geometric attacks in zero-watermarking algorithms, a zero-watermarking algorithm resistant to geometric rotation attacks is proposed. First, based on the pixel distortion of images after Scale-Invariant Feature Transform (SIFT) rotation correction, an approximately lossless safe region is determined in the central area. Second, this region undergoes a two-level redundant discrete wavelet transform to extract the low-frequency region, which is then partitioned into blocks; the maximum singular value is extracted from each block and utilized to construct a transition matrix. Third, a feature matrix is constructed by comparing each element value of the transition matrix with its mean. Finally, the encrypted watermark image and the feature matrix are combined to construct the zero watermark. Experimental results indicate that, compared with the algorithm using only SIFT correction, the robustness against rotation attacks is improved by an average of 13.26%; compared with GH rotation moments and pseudo-Zernike orthogonal moments algorithms, the robustness against rotation attacks is improved by 1.1% and 0.94%, respectively. The proposed algorithm also demonstrates strong robustness against conventional attacks, scaling attacks, cyclic translation, and small-range cropping attacks.

Full Text

Preamble

An Anti-Geometric Rotation Attack Zero-Watermarking Algorithm

Liu Wanjun, Sun Siyu†, Qu Haicheng, Feng Lin, He Muze

College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract: To address the weakness of zero-watermarking algorithms against geometric attacks, this paper proposes a zero-watermarking algorithm resistant to geometric rotation attacks. First, according to the pixel distortion characteristics of images after Scale-Invariant Feature Transform (SIFT) rotation correction, a secure region with approximately lossless pixels is identified in the central area. Second, this region undergoes two-level Redundant Discrete Wavelet Transform (RDWT) to extract the low-frequency component, which is then divided into blocks. The maximum singular value of each block is extracted to construct a transition matrix. Next, a characteristic matrix is generated by comparing each element value in the transition matrix with its mean value. Finally, the encrypted watermark image and the characteristic matrix are combined to construct the zero-watermark. Experimental results demonstrate that compared with SIFT correction alone, the robustness against rotation attacks improves by an average of 13.26%. Compared with GH rotation moment and pseudo-Zernike orthogonal moment algorithms, the anti-rotation attack robustness increases by 1.1% and 0.94%, respectively. The algorithm also exhibits strong robustness against conventional attacks, scaling attacks, cyclic translation, and small-scale shear attacks.

Keywords: Scale-Invariant Feature Transform; inscribed square region; redundant discrete wavelet; Arnold; characteristic matrix; singular value decomposition; anti-geometric attack

0 Introduction

Traditional digital watermarking faces a fundamental conflict between robustness and imperceptibility. Zero-watermarking overcomes this by utilizing internal features of the original carrier image to construct the watermark, thereby preserving the integrity of the carrier image. Geometric attacks disrupt the synchronization between the carrier image and watermark, making correct watermark extraction impossible. Compared to traditional digital watermarking, zero-watermarking can better resist geometric attacks. To effectively improve the anti-geometric attack capability of image watermarking algorithms, second-generation zero-watermarking based on image feature points has gained widespread attention.

Previous works have employed Scale-Invariant Feature Transform (SIFT) to match feature points before and after attacks as templates for correcting geometric distortion, thereby achieving rotation resistance. However, these methods fail to consider the pixel loss problem caused by rotation attacks. Jia et al. proposed an improved SIFT-based watermarking algorithm against geometric attacks, which fused rotation-invariant texture features into traditional SIFT feature vectors to improve matching accuracy. Zhang et al. used Speeded-Up Robust Features (SURF) and RANSAC algorithms to improve feature point quality, followed by affine transformation for higher correction precision. While

these approaches optimize feature detection, they do not address the issue of watermark loss caused by missing pixel values after rotation correction.

Although feature point matching algorithms can effectively resist rotation attacks, their robustness deteriorates as the rotation angle increases. The primary reason is that rotation attacks not only change element positions but also cause pixel value loss, resulting in missing pixels in the rotation-corrected image. To address this phenomenon, traditional watermarking algorithms construct invariant domains or invariant moments. Jia et al. proposed an anti-geometric attack digital watermarking algorithm based on local image normalization, which selects local regions unaffected by rotation correction for watermark embedding and extraction. However, this approach reduces algorithm capacity as the embedding region shrinks. Zhu et al. embedded watermark information into the pseudo-Zernike moment amplitudes of the carrier image' s low-frequency components, improving watermark robustness. Chen et al. used the Harris algorithm to extract carrier image feature points, constructed local feature regions, and embedded watermark information by quantizing and modulating pseudo-Zernike moment amplitudes, while using phase information for geometric correction against rotation attacks.

Drawing from these works, we observe that embedding watermarks in regions of the carrier image that do not suffer pixel distortion from rotation attacks can effectively improve the robustness of SIFT-based geometric correction watermarking algorithms. Through extensive experiments, selecting an inscribed square region near the center of the carrier image with minimal pixel loss for zero-watermark construction and extraction can significantly reduce pixel deficiency phenomena. To compensate for capacity reduction caused by decreasing the carrier image area, this paper employs Redundant Discrete Wavelet Transform (RDWT) to extract stable low-frequency regions without reducing subband sizes, extracts maximum singular values from image blocks to construct a characteristic matrix, and combines this with encrypted watermark information to build the zero-watermark.

1.1 Scale-Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) is an algorithm for detecting local image features. It extracts scale, position, and rotation-invariant feature points in multi-scale space. Because SIFT feature operators maintain scale invariance and can handle image rotation and brightness changes, SIFT essentially constructs scale space and searches for highly robust salient feature points that persist despite illumination changes or occlusion. The process involves detecting extrema in scale space, filtering these extrema, extracting local characteristics around each stable feature point to form local descriptors, and performing matching via Euclidean distance to obtain scale, position, orientation, and descriptors for geometric correction.

1.1.1 Scale Space Construction

Scale space simulates multi-scale characteristics of image data. Gaussian convolution is the only linear kernel for scale transformation. The scale space $L(x, y, \sigma)$ of an image is the result of convolving the image $I(x, y)$ with a variable-scale Gaussian function $G(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

where $*$ denotes convolution, (x, y) represents spatial coordinates, and σ is the scale factor proportional to the smoothing degree. Greater smoothing produces blurrier images. The Gaussian function is defined as:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The Difference-of-Gaussian (DoG) scale space is obtained by subtracting adjacent Gaussian scale images in each octave of the Gaussian pyramid:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

1.1.2 Keypoint Orientation and Descriptor Determination

Keypoint orientation determination utilizes edge magnitude M and orientation θ, calculated as:

$$M(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}$$

where L represents the scale at the keypoint location. A histogram of gradient orientations in the neighborhood is used to determine the dominant orientation, ensuring descriptor rotation invariance. The feature vector generation process is illustrated in [Figure 1: see original paper]. Black dots represent feature keypoints, each small square represents a neighboring pixel, arrow directions indicate gradient orientation, and arrow lengths represent gradient magnitude.

1.1.3 SIFT Feature Point Matching and Correction

SIFT feature matching relies on similarity metrics between feature points. The Euclidean distance between SIFT descriptors of two images is computed, and a priority K-D tree searches for approximate nearest neighbors for each feature point. A match is accepted if the ratio of the shortest distance to the second-shortest distance is below a threshold. Adjusting this threshold controls the number of matches; a higher threshold yields more matches but increases false positives, while a lower threshold improves stability but reduces matches. Experiments show optimal matching occurs with threshold values in $[0.4, 0.6]$.

SIFT is a spatial-based local feature descriptor invariant to rotation, scaling, and translation. If two points A and B exist in the central region, rotating the image by angle α rotates both points equally. Therefore, rotation correction can be achieved by calculating the angle between lines connecting two pairs of feature points (where oo' connects the centers of both images). Extracting and matching SIFT feature points before and after rotation attack enables geometric correction.

Let two feature points in the original carrier image be (x_i, y_i) and (x_j, y_j), with corresponding points after rotation attack being (x'_i, y'_i) and (x'_j, y'_j). The rotation angle is illustrated in [Figure 2: see original paper].

a) The horizontal angle between lines connecting two feature point pairs corrects the image. The rotation angle for each pair is calculated using formula (5).

b) All rotation angles are divided into n intervals. A histogram determines the frequency m and corresponding rotation angle p for each interval. The maximum frequency m_{max} is found, and the rotation angle p_m with relatively large frequency is identified. Rotation angles within $[p_m - 1, p_m + 1]$ are filtered, and their average yields the final rotation correction angle, as shown in equations (6) and (7):

$$\alpha = \frac{1}{N} \sum_{i=1}^{num} \alpha_i$$

where num represents rotation angles in the $[p_m - 1, p_m + 1]$ interval with frequency $m(i) > 0.85 \times m_{max}$, and N is the number of filtered rotation angles.

1.2.1 Pixel Loss Patterns After Rotation Correction

During 1° to 359° rotation, SIFT carrier images experience varying degrees of pixel loss at edges, which cannot be recovered after rotation correction. [Figure 3: see original paper] shows the original carrier image and images under different rotation attacks, while [Figure 4: see original paper] shows correction results

using SIFT alone. Table 1 details rotation correction results and loss conditions for various angles.

From [Figure 4: see original paper], we derive the following observations:

a) Regardless of rotation angle, the central region of the original carrier image always contains some pixel data without loss.

b) The pixel loss area is identical for rotation attacks of 15° and 75°.

c) When the carrier image is rotated by 45°, 135°, 225°, and 315°, the lost area reaches its maximum and remains identical.

Thus, under rotation attacks from 0° to 360°, the lost area and its position exhibit periodicity with a 45° cycle. Specifically, at 45°, 135°, 225°, and 315°, the lost area is maximal and positioned identically. Therefore, finding an inscribed square in the 45° rotation-corrected image allows similar squares to be determined for other angles.

1.2.2 Determination of Secure Region

From a human visual perspective, when rotation angles are 45°, 135°, 225°, and 315°, the selected region (marked with red boxes) experiences zero loss. For angles exceeding 45°, the selected region remains equally secure and unaffected by rotation attacks. This region is termed the **inscribed square secure region**. The principle and effect of region selection are shown in [Figure 5: see original paper].

To address image distortion from rotation attacks and pixel value loss after correction, previous work used pseudo-Zernike moments to reconstruct approximate circular regions, but pixel selection and block processing were cumbersome. This paper selects an inscribed square as the secure region. Regarding the square' s size: if too large, more pixel loss occurs; if too small, watermark capacity decreases. To ensure the extracted characteristic matrix suffers no pixel loss while maximizing watermark information, this paper determines the inscribed square dimensions in the rotation-corrected image with maximum loss area and maintains constant side length, ensuring pixels within this square remain unchanged across all rotation angles. Thus, constructing a square region with stable internal pixels after geometric correction is feasible.

To verify whether pixel loss occurs in the secure region under rotation attacks, [Figure 6: see original paper] shows the carrier image under various rotation angles (with normal inscribed square dimensions). The results confirm the selected region remains intact.

1.3 Redundant Discrete Wavelet Transform

Since Redundant Discrete Wavelet Transform (RDWT) produces subbands at each decomposition level with the same dimensions as the original image matrix, it can extract stable low-frequency regions while maintaining matrix size, thereby improving algorithm robustness and capacity. A two-level RDWT decomposition is illustrated in [Figure 7: see original paper].

Let image matrix A be $m \times m$. After one-level RDWT decomposition, low-frequency information LL_1 (approximation) and high-frequency information LH_1, HL_1, HH_1 (vertical, horizontal, and diagonal details) are obtained, each of size $m \times m$. [FIGURE:7(b) shows the second-level decomposition of LL_1, yielding another low-frequency component and three high-frequency components, all maintaining original dimensions. Higher decomposition levels concentrate most image information in deeper low-frequency subbands. Unlike high-frequency information vulnerable to noise, low-frequency subbands exhibit better stability and stronger resistance to external interference.

Compared with Discrete Wavelet Transform (DWT), where first-level subbands are 1/4 the size of the original image, RDWT maintains subband dimensions, offering greater capacity advantages for watermarking algorithms. Literature confirms RDWT provides stronger robustness than DWT in watermarking applications. Therefore, this paper selects low-frequency subband information from RDWT and employs SVD to further enhance algorithm stability and robustness.

---

2 Zero Watermark Construction and Extraction

This algorithm optimizes existing watermarking methods for rotation-corrected images. During the rotation correction phase, based on the SIFT method from literature [5], it additionally incorporates inscribed square secure region determination theory. The algorithm consists of two main stages: zero-watermark construction and watermark extraction. Before construction, the watermark undergoes Arnold scrambling encryption to improve security and robustness. The scrambled watermark effect is shown in [Figure 8: see original paper].

2.1 Zero Watermark Construction Process

a) Perform Arnold scrambling on the watermark image to obtain encrypted watermark W and save secret key K.

b) Based on Section 1.2, determine the inscribed square region M in the carrier image that remains pixel-loss-free under rotation attacks, where $\mathrm{mod}(m, n) = 0$. Save the top-left coordinates and side length of the inscribed square.

c) Apply two-level RDWT to M, extract its approximation subband, and divide it into blocks $B_{i,j}$, where $i, j = 1, 2, ..., m/n$.

d) Perform SVD on each image block and extract the maximum singular value from each block to construct transition matrix Y.

e) Generate characteristic matrix C based on the relationship between elements in Y and their mean value mean_Y, as shown in equation (8):

$$C(i,j) = \begin{cases} 1, & Y(i,j) \geq \text{mean}_Y \\ 0, & Y(i,j) < \text{mean}_Y \end{cases}$$

f) Perform XOR operation between characteristic matrix C and encrypted watermark W to obtain zero-watermark CW, which is registered with a trusted third-party authority (IPR).

2.2 Watermark Extraction

a) Detect and correct rotation in the potentially attacked image to obtain detection image P.

b) Follow steps b)-e) from the construction process to obtain characteristic matrix C^*.

c) Retrieve zero-watermark CW from the third-party authority, perform XOR operation with C^*, and apply inverse Arnold scrambling using secret key K to extract watermark W^*.

---

3 Experimental Results

Simulations were conducted on MATLAB 2014a using six grayscale images (512×512) with different textures and a 32×32 binary watermark image containing the "Liaoning Tech University" logo. The scrambling secret key K was set to 20. Bit Error Ratio (BER) and Normalized Cross-Correlation (NC) measured similarity between extracted and original watermarks. BER represents the proportion of erroneous bits, ranging from 0 to 1, where values closer to 0 indicate better robustness. NC ranges from 0 to 1, with values closer to 1 indicating better robustness.

BER is defined as:

$$\text{BER} = \frac{N_{\text{error}}}{N_{\text{bits}}} \times 100\%$$

where N_{error} is the number of erroneous bits and N_{bits} is the total number of watermark bits.

NC is defined as:

$$\mathrm{NC} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} w(i,j)\cdot w^{*}(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{M}[w(i,j)]^{2}}$$

where $w(i,j)$ is the original watermark and $w^{*}(i,j)$ is the extracted watermark, with M being the watermark dimension.

3.1 Rotation Attack Testing

This section tests rotation attacks centered on the original carrier image' s center point. For rotation-attacked images, SIFT feature point correction is first applied, followed by secure region selection for watermark extraction. Six different images undergo rotation attack testing.

[Figure 9: see original paper] shows test images under rotation attacks and extracted watermarks. Watermarks remain clearly visible and recognizable for copyright authentication across different rotation angles. NC values for six images resisting various rotation angles are presented in . Overall, extracted watermark NC values exceed 0.9962 for any rotation angle, reaching 1.0 at 90° rotation. This occurs because 90° rotation causes no pixel loss, and SVD possesses rotation invariance, making such attacks harmless to watermark extraction. Similarly, NC values equal 1.0 for 180° and 270° rotations.

3.3 Comparative Experiment Testing

Conventional attacks include geometric and non-geometric attacks. Since Section 3.1 addressed rotation resistance, this section focuses on non-geometric attacks using Lena, Barbara, and Baboon images. Results are shown in .

As attack intensity increases, extracted watermark NC values gradually decrease but remain above 0.95. For JPEG compression with quality factor 1 (100:100 compression ratio), NC values are 0.9716, 0.9723, and 0.9510, demonstrating strong resistance to high-intensity JPEG compression. Similar robustness is observed for other attack types, with NC values approaching 1.

Geometric attack results are shown in . SIFT-based geometric correction effectively resists scaling and row/column shifting attacks, with extracted watermark NC values near 1. For shear attacks, cropping 1/64 of the carrier image does not affect the feature point extraction region, yielding complete watermark extraction. However, larger shear areas remove portions of the feature extraction region, degrading watermark quality. Thus, the algorithm resists small-scale shear attacks.

To demonstrate superior rotation resistance, comparisons were made with literature [5, 8, 13]. Literature [5] uses only SIFT correction without addressing post-correction losses, while [8] and [13] select appropriate regions after rotation

correction. Literature [8] employs Gaussian-Hermite moments to determine local invariant regions, while [13] uses normalized pseudo-Zernike moments with quantization modulation. Literature [5] selects perceptually masked blocks and modifies relationships between first-column elements of left singular matrices from second-level DWT low-frequency components. Literature [13] uses normalized pseudo-Zernike moments, while our method employs non-embedding zero-watermarking, causing no visual impact and achieving stronger imperceptibility.

Rotation attack comparison data is presented in . The optimized rotation correction algorithm demonstrates significantly superior robustness compared to correction-only methods. Literature [8] achieves average NC of 0.9872, literature [13] achieves 0.9888, while our algorithm reaches 0.9981. Our approach resembles literature [8] in selecting local regions for embedding meaningful binary watermarks, but outperforms it by utilizing image matrix singular values, which resist perturbation and interpolation-based restoration interference. Although literature [13] improves rotation resistance, its watermark is a meaningless 128-bit binary sequence with limited capacity.

compares NC values under conventional attacks on Lena image. Our algorithm effectively resists JPEG compression, median filtering, Gaussian noise, salt-and-pepper noise, scaling, cyclic shifting, and shear attacks. Under JPEG compression, our NC values range from 0.9858 to 0.9987, while literature [5] and [13] range from 0.9685-1 and 0.9070-1, respectively, showing our superior JPEG compression resistance. For noise attacks at 0.02 Gaussian noise, our robustness exceeds literature [5] and [13] by 15.61% and 17.29%, respectively. At 0.04 salt-and-pepper noise, improvements are 9.95% and 39.10%. Median filtering resistance also surpasses both methods. Scaling and cyclic shifting performance is similar, though shear attack resistance is inferior because our spatial domain region selection may be removed by shearing. Nevertheless, small-area shear attacks are handled robustly.

[Figure 10: see original paper] visually compares rotation attack results across different methods.

---

4 Conclusion

To address incomplete watermark extraction caused by pixel information loss from geometric rotation attacks, this paper proposes an optimized zero-watermarking algorithm for rotation-corrected images. During carrier image preprocessing, after SIFT geometric correction, constructing an inscribed square secure region enables zero-watermark construction without pixel loss in the characteristic matrix, restoring watermark synchronization. In construction and extraction, RDWT preserves subband dimensions matching the original image, improving capacity, while low-frequency information and SVD stability enhance robustness.

Extensive experiments demonstrate effective resistance to rotation, scaling, cyclic translation, and shear attacks, showing excellent robustness. However, resistance to large-area shear attacks requires further improvement.

References

[1] Wen Q, Sun T F, Wang S X. Concept and application of zero-watermark [J]. Acta Electronica Sinica, 2003, 31 (2): 214-216.

[2] Liao Qinan. New watermarked image geometric correction algorithm based on SIFT feature points matching [J]. Application Research of Computers, 2011, 28 (6): 2247-2249.

[3] Li Hao, Li Hongchang. Geometric attack resisting double-watermarking algorithm based on CS-SIFT [J]. Computer Science, 2014, 41 (S2): 263-267.

[4] Lyu Jianping, Peng Shu. A DWT-domain watermarking algorithm against geometric attacks based on SIFT [J]. Journal of Xi' an University of Posts and Telecommunications, 2015, 20 (2): 88-92.

[5] Qi Xiangming, Gao Ting. Invisible and robust watermarking algorithm based on an image block [J]. Journal of Image and Graphics, 2017, 22 (6): 719-730.

[6] Jia Chao, Zhang Zhengbao. Resistance to geometric attacks watermarking algorithm based on improved SIFT [J]. Journal of Chinese Computer Systems, 2014, 35 (12): 2655-2658.

[7] Zhang W, Chen J, Wang R, et al. Affine correction based image watermarking robust to geometric attacks [C]// Proc of IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies. 2017.

[8] Jia X, Yang Z, Qi Y, et al. The anti-geometric attack digital watermarking algorithm based on image normalization of local information [C]// Proc of IEEE Information Technology, Networking, Electronic and Automation Control Conference. 2016: 1120-1124.

[9] Zhu Dandan, Lyu Lizhi. Anti-geometric-attack watermarking algorithm based on pseudo-zernike moments and contourlet transform [J]. Computer Science, 2016, 43 (6): 131-134.

[10] Chen Qing, Weng Xufeng. A robust image watermarking based on harris feature points and pseudo-zernike moments [J]. Electronic technology, 2016, 29 (3): 183-186.

[11] Rassem T H, Makbol N M, Khoo B E. Performance evaluation of RDWT-SVD and DWT-SVD watermarking schemes [C]// Proc of AIP Conference Proceedings. [S. l.]: AIP Publishing, 2016: 050021.

[12] Sun S, Yang S, Zhao L. Noncooperative bovine iris recognition via SIFT [J]. Neurocomputing, 2013, 120: 310-317.

[13] Chen Qing, Weng Xufeng. Novel blind image watermarking based on pseudo Zernike moments [J]. Application Research of Computers, 2016, 33 (9): 2810-2812.

Note: Figure translations are in progress. See original paper for figures.