SIFT Feature-Based Watermarking Method Aimed at Achieving IHC Ver.5

Masaki Kawamura(✉) and Kouta Uchida

Graduate School of Sciences and Technology for Innovation,
Yamaguchi University, Yamaguchi-shi 753-8512, Japan
m.kawamura@m.ieice.org

Abstract. We propose a watermarking method using scale-invariant feature transform (SIFT) features that have both scale and rotation invariance, and evaluate our method in accordance with the information hiding criteria (IHC) ver. 5. It is defined as evaluation criteria against several possible attacks; these attacks are JPEG compression and geometric attacks, e.g., scaling, rotation, and clipping. In our method, we use local feature regions located around the SIFT features that are robust against scaling and rotation. The regions are normalized in size and selected as marked regions. Watermarks are embedded in the marked regions. We also introduce two error-correction techniques: weighted majority voting (WMV) and low-density parity-check (LDPC) code. When a stego-image is attacked by scaling or rotation, the image is spatially distorted. WMV and LDPC code can correct errors of extracted watermarks in the distorted stego-image. On the other hand, it is not easy to detect rotated marked regions. Therefore, the correct orientation is searched for by brute force. We evaluated the proposed method in accordance with IHC ver. 5. Our method can achieve robustness against scaling and rotation attacks in the highest tolerance category.

Keywords: Feature-based watermarking · SIFT · IHC · Geometric attack

1 Introduction

Usage of digital watermarking techniques include protecting digital content (e.g., image, video, or audio data) from being modified either legally or illegally. Even if the digital content were attacked, the watermark could be extracted from the attacked content. Therefore, a robust watermarking method should be developed. By focusing on watermarking for still images, we want to extract watermarks from a processed image. Image processing, e.g., lossy compression, clipping, scaling, or rotation, is usually applied to an image. Therefore, the image processing is regarded as an *attack* on the image.

Evaluation standards are required to evaluate robustness and image quality for the watermarking techniques. Therefore, the Institute of Electronics, Information and Communication Engineers (IEICE) proposed information hiding criteria (IHC) [1]. In the IHC ver. 5, for example, the stego-images are attacked

J.-S. Pan et al. (eds.), *Advances in Intelligent Information Hiding and Multimedia Signal Processing*, Smart Innovation, Systems and Technologies 81,
DOI 10.1007/978-3-319-63856-0_46

by JPEG compression, clipping, scaling, rotation, and combinations of these attacks. The watermarks should be extracted from attacked images with almost no errors. The extraction should be performed blind, that is, the decoder cannot use any information about both the original images and the attacks. IHC ver. 5 [1] is summarized as follows. There are three roles in the model: watermarker, attacker, and detector. It is supposed that the six original IHC standard images are provided with a 4608×3456 pixel size. Ten messages are generated by using an M-sequence. The message length is 200 bits. The watermarker encodes the messages for the purpose of correcting errors. An encoded message is embedded into the original image. Since the JPEG format is popular, the image is compressed with JPEG to be distributed (1st compression). The file size should be less than 1/15 of the original size. The compressed image is called a stego-image in this model. The image quality of the stego-image is measured by peak signal-to-noise ratio (PSNR) and mean structural similarity (MSSIM) [2]. The PSNR of the stego-image should be more than 30 dB.

The attacker performs a geometrical attack, such as scaling, rotation, and combinations of these two attacks on the stego-image. After the geometrical attack, the image is clipped to an HDTV-size area (1920×1080) at four specified coordinates. The clipped images are compressed with JPEG again to be saved in the JPEG format (2nd compression). The obtained image is called an attacked image. The detector extractes the message (200 bits) from a given attacked image without referring to the original image, attack parameters, and any related information. The watermarks should be decoded from the attacked image with an almost zero bit error rate (BER). The accuracy of the decoded message is measured by the BER [1].

Once an image has been geometrically attacked (i.e. rotation, scaling, and clipping have been performed on the image), the positions of the pixels in the image would be moved, and then the marked regions would also be moved. To extract watermarks from the marked regions, the original marked positions need to be detected. This is called *synchronization* and is used to find the marked positions. Watermarking methods using a marker or synchronization code have been proposed [3,4]. In these methods, a watermark consists of an encoded message and a marker. The marker is a specific bit sequence, and is effective for a clipping attack to find marked positions. When the image is subjected to geometric attacks, the synchronization is performed by searching for markers in the attacked image. However, it is hard to find the marked positions due to distortion of the image. Since the attack parameters of the scaling ratio and rotation angles are unknown, the marked positions can only be searched for by brute force. Therefore, we focus on a synchronization technique using rotation- and scale-invariant features.

We propose a watermarking method using scale-invariant feature transform (SIFT) features for synchronization in accordance with IHC ver. 5. A message is encoded by using low-density parity-check (LDPC) code [5] as a watermark. Marked regions are selected around the SIFT feature points. The size of the regions is normalized for robustness against scaling. The watermark is embedded into the discrete cosine transform (DCT) domain of each region for robustness against JPEG compression. In the detecting process, the extracted regions are rotated due to synchronization. The correct orientation is searched for by brute force.

2 Related Works

Scale-invariant feature transform (SIFT) [6] is a promising feature detector. Even if rotation and scaling attacks are applied, the same feature points can be extracted by the SIFT detector, and the scale parameters, which are proportional to the scaling ratio, can also be extracted. Therefore, the SIFT features are represented by circular patches or disks whose centers are feature points, and their radii are equal to the scale parameters. Since the circular patches are rotation- and scale-invariant, we can obtain the same marked regions. There are two major embedding domains for invariant feature-based methods: pixel-value domain and frequency domain. For pixel-value domain watermarking, the circular patches are extracted by SIFT or other invariant feature detectors. The bit sequence of watermarks are converted to polar coordinates, and then they are embedded into the patches directly [7–11]. The image quality of stego-images created by these methods is not good due to directly changing of the pixel values. Also, the watermarks in the stego-image are vulnerable to compression. Moreover, the conversion to polar coordinates involves loss of the watermarks.

For frequency domain watermarking, watermarks are embedded into some frequency domains in marked regions around the feature points. Since the discrete Fourier transform (DFT) has invariance for scaling, the methods involving embedding in the DFT domain are robust against scaling attacks [12–15]. However, they are sensitive to compression. Embedding in the DCT domain is effective for robustness against compression [16,17], but existing methods are non-blind ones. The method of Pham *et al.* [16] needs a database to store the SIFT features. Wei and Yamaguchi's method [17] needs the original watermark to find the watermarks.

Since a rotation attack changes the marked regions, robustness against rotation should be considered. The Fourier-Mellin transform domain is effective for rotation, scaling, and translation (RST) [18]. The Fourier-Mellin transform is equivalent to computing the Fourier transform of a log-polar mapping. Tone and Hamada's method [12] uses a Harris-Affine detector and a log-polar mapping as the invariant feature detector. It can extract scale- and rotation-invariant features. However, the log-polar mapping distorts the watermarks, resulting in the number of errors of the watermarks becoming high. Therefore, these methods could not achieve the IHC.

Methods without the log-polar mapping seem to be better for loss-less embedding. The orientation of the marked regions is worthy of consideration. In one study, the dominant orientation, which is the peak of the histogram of the gradient direction of the pixels, was used [19]. In another study, the dominant orientation was obtained by SIFT detector to align the direction of the features [14].

In two studies, the characteristic orientation based on the moments of a local region was used [10,20]. These methods detect the significant orientation and rotate the local regions. Therefore, the region has robustness against rotation. Since the process rotating the region involves distortion of the image, watermarks are also damaged in the embedding process. Moreover, since the orientation is low precision, the probability of failing to detect the orientation is higher in the detecting process.

3 Proposed Feature-Based Watermarking Method

3.1 Watermarking Process

In the proposed method, a message, $\boldsymbol{m}$, of length $N_m = 200$ is encoded to a codeword, $\boldsymbol{c}$ of length $N_c = 300$ by using LDPC code [5,21]. Since difficult attacks will be performed on the stego-images in accordance with IHC ver. 5, a lot of errors will appear in the extracted codewords. To measure the error rate, a check bit, $\boldsymbol{s}$, of length $N_s = 87$ is introduced, where it is given by $\boldsymbol{s} = (1, 1, 1, \cdots, 1)$. Therefore, the watermark, $\boldsymbol{w}$, consists of the codeword $\boldsymbol{c}$ and check bits $\boldsymbol{s}$, and is given by

$$w_i = \begin{cases} c_i, 1 \le i \le N_c \\ 1, \;\; N_c < i \le N_c + N_s \end{cases} . \tag{1}$$

The marked regions are selected around the SIFT feature points. The feature points (x_i, y_i) and the scale parameters σ_i are obtained by using the SIFT algorithm [6], and then the circular patches are constructed. However, circular forms should be avoided due to distortion of the watermarks. Therefore, the bounding squares of the circular patches are selected as marked regions in our method. The bounding squares are squares of side $2d\sigma_i$ pixels, where d is the radius magnification. In this paper, we set $d = 7$ in view of the length of the watermark. The scale parameters in the range of $\sigma_L < \sigma_i < \sigma_U$ are selected, since large bounding squares will overlap each other, and small ones will vanish as a result of shrinking [10]. We set $\sigma_L = 4$ and $\sigma_U = 10$. Even if the range of the scale parameters are limited, some of the squares may be overlaped each other. In this case, the feature point which has the largest value of difference of Gaussian (DoG) filter is selected. Each marked region is normalized to a square of side $h = 96$ pixels in preparation for the scaling attack.

The watermarks $\boldsymbol{w}$ are embedded in the DCT domain of the normalized regions. In the method of Hirata and Kawamura [3,4], the normalized region is divided into 8 × 8 pixel blocks. One bit of the watermark is embedded in one block. In our method, the region is divided into 32 × 32 pixel blocks, since the 8 × 8 pixel region was too small to embed a watermark in it. Each block is transformed by using the 2D DCT. A more than one bit ($N_B > 1$) watermark is embedded in the DCT coefficients of the block by using quantization index modulation (QIM) [22], where N_B is given by

$$N_B = \left\lceil \frac{32 \times 32}{h \times h} (N_c + N_s) \right\rceil , \tag{2}$$

where $\lceil x \rceil$ stands for the ceiling function. In the case of $h = 96, N_c = 300$, and $N_s = 87$, the number of the bits is $N_B = 43$. Since a stego-image will be clipped, the same watermarks are repeatedly embedded throughout an image. After embedding the watermarks, the stego-image is compressed by JPEG compression to be less than 1/15 of the original size.

3.2 Extraction and Decoding Process

The stego-image is attacked by JPEG compression, clipping, scaling, rotation, and combinations of these attacks in accordance with IHC ver. 5. Watermarks must be extracted from the attacked image. The synchronization is performed by SIFT. The process to obtain the marked regions is the same as the watermarking process except for the range of the scale parameters. In extraction, the scale parameters in the range of $0.8\sigma_L < \sigma_i < 1.2\sigma_U$ are selected, since the stego-image is magnified or shrunk by the scaling attack. We assume P marked regions are extracted from the attacked image and P candidates for watermarks are obtained by QIM [22].

Since the marked region may be rotated by the rotation attack, the region is rotated to find the correct marked position in the extraction process. Let a candidate for the p-th watermark rotated by a θ-degree angle be $\hat{\boldsymbol{w}}^p(\theta) = (\hat{\boldsymbol{c}}^p(\theta), \hat{\boldsymbol{s}}^p(\theta))$. Note that the first N_c bits correspond to the codeword bits, and the residual N_s bits correspond to the check bits. Now, the matching ratio for the check bits $\hat{\boldsymbol{s}}^p(\theta)$ is defined by

$$R^p(\theta) = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{s}_i^p(\theta) \,. \tag{3}$$

Since all of the check bits are 1, the estimated degree of the angle, $\hat{\theta}^p$, can be calculated by

$$\hat{\theta}^p = \arg \max_{0 \le \theta \le 90} \{R^p(\theta)\}. \tag{4}$$

By using the SIFT features and searching for the rotation angle, the synchronization for clipping, rotation, and scaling can be performed.

After the synchronization, a message will be estimated from the candidates $\hat{\boldsymbol{c}}^p(\hat{\theta}^p)$. The candidates include bit sequences extracted from incorrect feature points, in which the watermarks are not embedded. Even if candidates are the bit sequences in which the watermarks are embedded, they may be distorted by the attacks. Therefore, the candidates contain a lot of errors. To remove incorrect candidates, we introduce a weighted majority voting (WMV) algorithm [3,4]. The estimated codeword $\hat{\boldsymbol{c}} = (\hat{c}_1, \hat{c}_2, \cdots, \hat{c}_{N_c})$ can be calculated from the weighted candidates by the WMV, that is,

$$\hat{c}_i = \Theta\left(\sum_{p=1}^{\hat{P}} \alpha\left(R^p(\hat{\theta}^p)\right)\left\{\hat{w}_i^p(\hat{\theta}^p) - 0.5\right\}\right), 1 \le i \le N_c, \tag{5}$$

where $\alpha(x)$ is the weight function of the ratio $R^p(\hat{\theta}^p)$ and is defined by

$$\alpha(x) = \begin{cases} \tanh\left(\beta\left(x - T\right)\right), T \leq x \\ 0, \qquad\qquad\qquad x < T \end{cases}, \tag{6}$$

where T is the threshold and β is the weight coefficient. The function $\Theta\left(x\right)$ is the step function defined by

$$\Theta(x) = \begin{cases} 1\,, x \geq 0 \\ 0\,, x < 0 \end{cases}. \tag{7}$$

After obtaining the estimated codeword $\hat{\boldsymbol{c}}$, the estimated message $\hat{\boldsymbol{m}} = (\hat{m}_1, \hat{m}_2, \cdots, \hat{m}_{N_m})$ can be calculated by the sum-product algorithm of the LDPC code [23]. The BER is defined by

$$\mathrm{BER} = \frac{1}{N_m}\sum_{i=1}^{N_m} m_i \oplus \hat{m}_i, \tag{8}$$

where ⊕ stands for exclusive OR (XOR).

4 Evaluation by Computer Simulations

We evaluate our method by computer simulation on the basis of the IHC ver. 5. The attack procedure is shown in Fig. 1. Before attacking the stego-image, the Q-value of the JPEG compression is computed in advance to be less than 1/25 of the original size. The stego-image is attacked by scaling, rotation, and combinations of these two attacks. The scale parameter is $s \in \{80, 90, 110, 120\%\}$, and the rotation angle is $\theta \in \{3, 5, 7, 10°\}$. The combinations of the scaling and the rotation attack are $(s, \theta) \in \{(80, 9), (90, 7), (110, 5), (120, 3)\}$. These parameters are defined in the IHC ver. 5 [1]. The term 'No attack' in Fig. 1 means that the stego-image is not attacked by any geometric attacks. The next stage is clipping; the image is clipped by an HDTV-size area at four specified coordinates. The center points of the four clipped areas are $(x_c \pm 700, y_c \pm 500)$, where the point (x_c, y_c) is the center point of the stego-image. Each clipped area is saved as a new image by using the Q-value, which is computed in advance. The compressed image is called the attacked image.

The step size of the QIM is $\Delta = 72$. The threshold T and the weight coefficient β in the weight function are $T = 0.7$ and $\beta = 0.7$, respectively. We use a $(3, 4)$-regular LDPC code generated by the progressive edge-growth (PEG) algorithm [4,21]. The channel error rate in the sum-product algorithm [23] is 0.05. Ten different messages are generated by the M-sequence, they are embedded in six IHC standard images, and then four areas are clipped. Therefore, 240 attacked images are generated.

There are two categories of evaluation criteria for comparing methods: "highest image quality" and "highest tolerance." Table 1 shows the average compression ratio, PSNR, and MSSIM. Since ten different messages were embedded, the

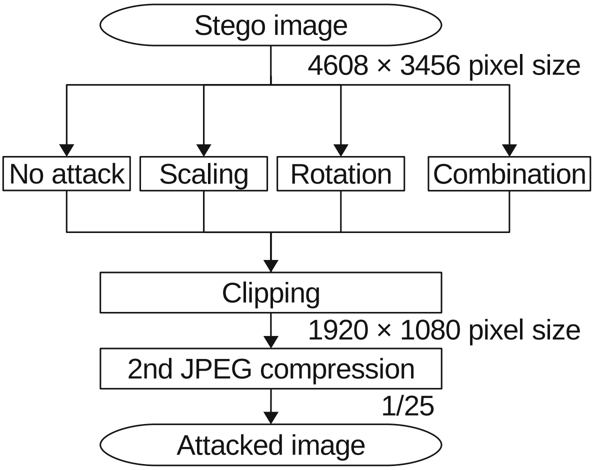

Fig. 1. Attack procedure

Table 1. Average compression ratio, PSNR, and MSSIM

	Compression ratio [%]	PSNR [dB]	MSSIM
Image 1	6.628	33.313	0.917
Image 2	6.588	33.800	0.917
Image 3	6.609	35.462	0.937
Image 4	6.621	36.851	0.949
Image 5	6.629	35.108	0.926
Image 6	6.656	33.718	0.919
Average	6.622	34.709	0.927

values are the average values for ten trials. The compression ratio was under $1/15 = 6.67\%$ for the first compression. The PSNRs and MSSIMs were calculated for the luminance signal of the generated stego-images before these images were attacked. All PSNRs were over 30 dB.

Table 2 shows the average error rate for the attacked images with additional attacks. In accordance with the highest tolerance category of the IHC [1], there are four clipped areas in each stego-image, and the BERs for three of the four areas must be zero. In other words, one area can be discounted. Therefore, the best three of the four BERs were used for the evaluation. The compression ratio in Table 2 was less than 1/25 of the original size for the second compression. As a result, the proposed method could achieve a BER of 0.0 for scaling or rotation attacks. However, the BERs for combined attacks were over 0.0 for many images. For the highest image-quality category of the IHC, the BER must be no more than 1% on average. In the worst case, the BER for the three clipped images must be less than 2%. The proposed method could not achieve this criterion.

Table 2. Average error rate for attacked images with additional attacks (%)

Image no.	No attack	Scaling (%)				Rotation (°)				Combination (s, θ)			
		80	90	110	120	3	5	7	10	$(80, 9)$	$(90, 7)$	$(110, 5)$	$(120, 3)$
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.003	0.004	0.001
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.003	0.001
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.001	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.002	0.002

However, the BERs for almost 80% of the attacked images were zero. In a few cases, the BERs were over 2%.

5 Conclusion

We proposed a SIFT feature-based watermarking method aimed at achieving the IHC ver. 5. To accomplish the criteria, the method has to be robust against compression, clipping, scaling, rotation, and combinations of these attacks. Since the IHC ver. 5 is the newest criteria, we demonstrated the ability of our method to operate in current conditions. We introduced SIFT against scaling and rotation attacks and WMV and the LDPC code against compression and clipping. As a result, our method does not have enough robustness against combined attacks of a scaling attack and a rotation attack but does have robustness against an individual attack.

Acknowledgments. This work was supported by JSPS KAKENHI Grant Number JP16K00156, and was performed in the Cooperative Research Project of the RIEC, Tohoku University. The computer simulations were carried out on PC clusters at Yamaguchi University.

References

1. Information hiding and its criteria for evaluation, IEICE. http://www.ieice.org/iss/emm/ihc/. Accessed 10 Mar 2017
2. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
3. Hirata, N., Kawamura, M.: Watermarking method using concatenated code for scaling and rotation attacks. In: 14th International Workshop on Digital-Forensics and Watermarking (IWDW 2015) (2015)
4. Hirata, N., Nozaki, T., Kawamura, M.: Image watermarking method satisfying IHC by using PEG LDPC code. IEICE Trans. Inf. Syst. **E100–D**(1), 13–23 (2017)
5. Gallager, R.G.: Low-density parity-check codes. IRE Trans. Inf. Theory **IT–8**(1), 21–28 (1962)
6. Lowe, D., David, G.: Distinctive image features from scale-invariant keypoints. Inter. J. Comput. Vis. **60**(2), 91–110 (2004)

7. Lee, H.Y., Kim, H., Lee, H.K.: Robust image watermarking using local invariant features. Opt. Eng. **45**(3), 037002 (2006)
8. Verstrepen, L., Meesters, T., Dams, T., Dooms, A., Bardyn, D.: Circular spatial improved watermark embedding using a new global SIFT synchronization scheme. In: 16th International Conference on Digital Signal Processing, pp. 1–8 (2009)
9. Li, L.D., Guo, B.L.: Localized image watermarking in spatial domain resistant to geometric attacks. Int. J. Electron. Commun. **63**(2), 123–131 (2009)
10. Yu, Y., Ling, H., Zou, F., Lu, Z., Wang, L.: Robust localized image watermarking based on invariant regions. Digit. Signal Proc. **22**(1), 170–180 (2012)
11. Tsai, J.S., Huang, W.B., Kuo, Y.H., Horng, M.F.: Joint robustness and security enhancement for feature-based image watermarking using invariant feature regions. Sig. Process. **92**(6), 1431–1445 (2012)
12. Tone, M., Hamada, N.: Scale and rotation invariant digital image watermarking method. IEICE Trans. Inf. Syst. (Japanese Edition) **J88–D1**(12), 1750–1759 (2005)
13. Wang, X.Y., Hou, L.M., Wu, J.: A feature-based robust digital image watermarking against geometric attacks. Image Vis. Comput. **27**(7), 980–989 (2008)
14. Gao, X., Deng, C., Li, X., Tao, D.: Local feature based geometric-resistant image information hiding. Cogn. Comput. **2**(2), 68–77 (2010)
15. Yang, H., Xia, Z., Sun, X., Luo, H.: A robust image watermarking based on image restoration using SIFT. Radioengineering **20**(2), 525–532 (2011)
16. Pham, V.Q., Miyaki, T., Yamasaki, T., Aizawa, K.: Geometrically invariant object-based watermarking using SIFT feature. In: IEEE International Conference on Image Processing (ICIP), pp. 473–476 (2007)
17. Wei, N., Yamaguchi, K.: Image watermarking resistant to geometric attacks based on SIFT feature. IEICE Tech. Rep. EMM **111**(496), 43–48 (2012)
18. O'Ruanaidh, J., Pun, T.: Rotation, translation and scale invariant digital image watermarking. In: International Conference on Image Processing, vol. 1, pp. 536–536. IEEE Computer Society (1997)
19. Deng, C., Gao, X., Li, X., Tao, D.: A local Tchebichef moments-based robust image watermarking. Sig. Process. **89**(8), 1531–1539 (2009)
20. Nasir, I., Khelifi, F., Jiang, J., Ipson, S.: Robust image watermarking via geometrically invariant feature points and image normalisation. IET Image Proc. **6**(4), 354–363 (2012)
21. Hu, X.Y., Eleftheriou, E., Arnold, D.M.: Regular and irregular progressive edge-growth tanner graphs. IEEE Trans. Inform. Theory **51**(1), 386–398 (2005)
22. Chen, B., Wornell, G.W.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans. Inform. Theory **47**(4), 1423–1443 (2001)
23. Wadayama, T.: A coded modulation scheme based on low density parity check codes. IEICE Trans. Fundam. **E84–A**(10), 2523–2527 (2001)