A New Image Watermarking Technique based on Random Forests

Li san-ping[1.a]

[1] Fuzhou University of International Studies and Trade, Fuzhou China,

[a]sandeli1273@163.com

Keywords: image watermark; wavelet transform; random forests

Abstract. “Random Forests” is an algorithm developed by Breiman and Cutler in 2001[1]. It runs by constructing multiple decidion trees while training and outputting the classes that is the mode of the classes output by individual trees. It has improved performance over single decision trees, and it is much more efficient than traditional machine learning techniques, e.g. artificial neural networks and support vector machine. In this paper, a new image watermarking technique based on Random Forests （RF） is proposed. First, the image is decomposed through discrete wavelet transform. Then we use the relationship between the selected coefficient and its neighboring coefficients to train RF. Thanks to the good learning ability of RF, the watermark is adaptively embedded in wavelet domain and also can be extracted by the well trained RF. Experimental results show that the algorithm is robust against high intensity JPEG, JPEG2000, geometric distortion.

1. Introduction

Watermarking techniques have been studied extensively in the past. By taking account of the domain in which the watermark is embedded, watermarking schemes can be classified into spatial-domain and frequency-domain techniques. The advantages of spatial domain techniques are nice transparency and efficiency, while the disadvantage is fragile to image processing. Since the watermark embedded into the spatial domain can be easily destroyed by image processing, the watermarking research in the transform domain is prevalent. The frequency domain watermarking techniques embed the watermark by modulating the magnitude of coefficients in a transform domain, such as DCT[2], DFT[3], DWT[4][5]. In contrast to the spatial domain technique, the frequency domain technique is more robust to image processing. In this paper, a new image watermarking technique based on Random Forests （RF） is proposed. First, the image is decomposed through discrete wavelet transform. Then we use the relationship between the selected coefficient and its neighboring coefficients to train RF. Thanks to the good learning ability of RF, the watermark is adaptively embedded in wavelet domain and also can be extracted by the well trained RF. Experimental results show that the algorithm is robust against high intensity JPEG, JPEG2000, geometric distortion.

2. Random Forests（RF）

Breiman (2001) proposed random forests, which add an additional layer of randomness to bagging[1]. The algorithm of random forests is as follows:

Step1. Draw n_{tree} bootstrap samples from the original data. **Step2.** For each of the bootstrap samples, grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m_{try} of the predictors and choose the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained when $m_{try} = p$, the number of predictors.)**Step3.** Predict new data by aggregating the predictions of the n_{tree} trees.(i.e., majority votes for classification, average for regression), this paper use the majority votes for classification. An estimate of the error rate can be obtained, based on the training data, by the following:

Step1. At each bootstrap iteration, predict the data not in the bootstrap sample (what Breiman calls "out-of-bag", or OOB, data) using the tree grown with the bootstrap sample. **Step2.** Aggregate the OOB predictions. (On the average, each data point would be out-of-bag around 36% of the times, so aggregate these predictions.) Calcuate the error rate, and call it the OOB estimate of error rate.

3. Watermarking Strategies

- Watermark sequence composition

The watermark is composed of two binary sequence information: $W = t_0 t_1 ... t_{N-1} s_0 s_1 ... s_{M-1}$. The first binary sequence is pseudorandom number $T = t_0 t_1 ... t_{N-1}$ ($N = 128$) which generated by Blum-Blum-Shub with the key k_1, and used to train RF model. The second binary sequence is $S = s_0 s_1 ... s_{M-1}$ which is watermark sequence information. Because of the advantages of DWT, this paper selects the information watermark into the low frequency domain of the DWT according to the idea of the document[6].

- training sequence embedded

Step1. Firstly, the original images I decomposed two levels through wavelet transform. The first low frequency subband is A_1 in which we select a random position $P_i = (x_i, y_i)_{\mathrm{i=1,2,...,128}}$ which generated by Blum-Blum-Shub with the key k_2, then the training sequence embedded by formula (1) and (2) , The α is Watermark embedding strength.

$$\begin{cases} A_1(P_i) = A_1(P_i) + \alpha(2t_i - 1) \\ A_1(P_i^{'}) = A_1(P_i^{'}) - \alpha(2t_i - 1) \end{cases} \tag{1}$$

Step2. After the training sequence is embedded, RF model is trained by formula (2).

$$F = \{TF_i = (V_i, d_i), i = 0, ..., N-1\} \tag{2}$$

V_i is the feature， d_i is label(" 0 " or " 1 ") that defined by the formula (3).

$$V_i = (\delta_{P_i}^1, \delta_{P_i}^2, \delta_{P_i}^3, \delta_{P_i}^4) \tag{3}$$

$\delta_{P_i}^1, \delta_{P_i}^2, \delta_{P_i}^3, \delta_{P_i}^4$ are the mean value between the coefficient of central point and its adjacent coefficients, Obtained by formula (4)~(7) respectively.

$$\delta_{P_i}^1 = | A_1(P_i) | - \frac{1}{7}\{| A_1(x_i - 1, y_i) | + | A_1(x_i + 1, y_i) | + | A_1(x_i, y_i - 1) | + A_1(x_i, y_i + 1) | + | A_2(x_i/2, y_i/2) | + | D_1(x_i, y_i) | + | D_2(x_i/2, y_i/2) | \tag{4}$$

$$\delta_{P_i}^2 = | A_1(P_i) | - \frac{1}{8}\{| A_1(x_i - 1, y_i - 1) | + | A_1(x_i - 1, y_i) | + | A_1(x_i - 1, y_i + 1) | + | A_1(x_i, y_i - 1) | + | A_1(x_i, y_i + 1) | + | A_1(x_i + 1, y_i - 1) | + | A_1(x_i + 1, y_i) | + | A_1(x_i + 1, y_i + 1) | \tag{5}$$

$$\delta_{P_i}^3 = | A_1(P_i) | - \frac{1}{8}\{| A_1(x_i, y_i - 2) | + | A_1(x_i, y_i - 1) | + | A_1(x_i, y_i + 1) | + | A_1(x_i, y_i + 2) | + | A_1(x_i - 2, y_i) | + | A_1(x_i - 1, y_i) | + | A_1(x_i + 1, y_i) | + | A_1(x_i + 2, y_i) | \tag{6}$$

$$\delta_{P_i}^4 = | A_1(P_i) | - \frac{1}{8}\{| A_1(x_i - 2, y_i - 2) | + | A_1(x_i - 2, y_i + 2) | + | A_1(x_i - 1, y_i - 1) | + | A_1(x_i - 1, y_i + 1) | + | A_1(x_i - 1, y_i - 1) | + | A_1(x_i + 1, y_i + 1) | + | A_1(x_i - 2, y_i - 2) | + | A_1(x_i + 2, y_i + 2) | \tag{7}$$

- Watermark sequence embedded

Step1. select a random position $P_i = (x_i, y_i)_{\mathrm{i=1,2,...,M}}$ which generated by Blum-Blum-Shub with the key k_3 and extract features in A_1 by formula (3) .

Step2. Use the RF model trained by formula (2) to classify the features V_i ,we can get $s_i^{'} = d(V_i)$.

$$s_i^{'} = \begin{cases} 0, ...if..d(V_{N+i}) = -1 \\ 1, ...if..d(V_{N+i}) = 1 \end{cases} \tag{8}$$

Step3. If $s_i = s_i^{'}$, modify the value of the location P_i according to formula (9).

$$A_{P_i} = A_{P_i} + \alpha(2t_i - 1) \tag{9}$$

If $s_i \neq s_i^{'}$, modify the value of the location P_i and its adjacent location point $P_i^{'}$ by formula (10).

$$\begin{cases} A_1(P_i) = A_1(P_i) + \alpha(2t_i - 1) \\ A_1(P_i^{'}) = A_1(P_i^{'}) - \alpha(2t_i - 1) \end{cases} \tag{10}$$

After the coefficient is modified, the features are reextracted and classified. If the result still $s_i \neq s_i^{'}$, revise coefficient again, untill $s_i = s_i^{'}$.

Step4. When the embedding of the training information and watermark information is completed, the watermark image I^* is obtained by the wavelet inverse transform.

- Watermark sequence extracted

The extraction of watermark information is the inverse process of watermark information embedding. So we finally need to extract the watermark sequence, the following steps are as follows:

Step1. the watermark image I^* decomposed two levels through wavelet transform. The first low frequency subband is A_1 in which we select position $P_i = (x_i, y_i)_{\mathrm{i}=1,2,...,128}$ with the same key k_2 by Blum-Blum-Shub, And extract the features set $F^{'}$ in A_1 by formula (3).

$$F^{'} = \{TF_i^{'} = (V_i^{'}, d_i), i = 0...N-1\} \tag{11}$$

All training features $TF_i^{'}$ are used to train new $RF^{'}$.

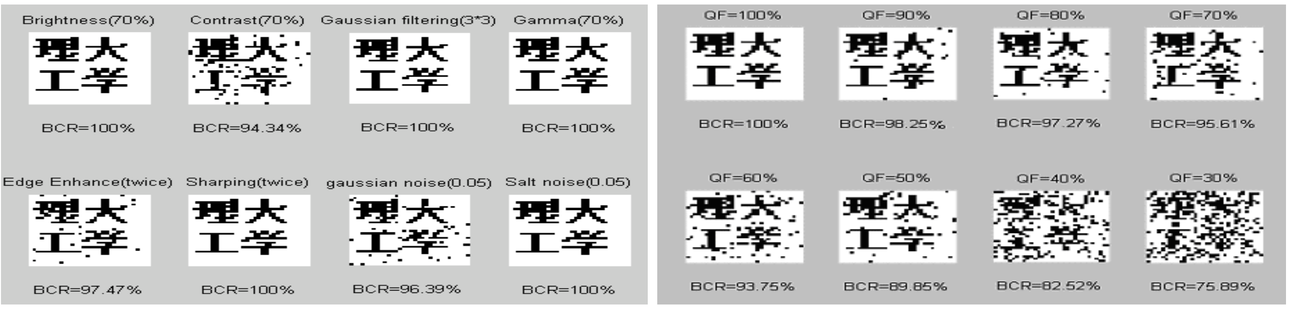
Step2. Select position $P_i = (x_i, y_i)_{\mathrm{i}=1,2,...,\mathrm{M}}$ with the same key k_3 by Blum-Blum-Shub and extract the features set $V_{N+i}^{'}$ in A_1 by formula (3). Then classify the $V_{N+i}^{'}$ according to the formula (8) with the new $RF^{'}$ model.

4. EXPERIMENTAL RESULTS

In our experiments, some necessary parameters used in the process of training RF model are mainly determined by experiments. The watermark intensity α is the parameter that regulates the invisibility and robustness of the watermark which is $\alpha = 1$ and the parameter of RF model is $\mathrm{n}_{\mathrm{TREE}} = 500$. The watermark information is converted to a one dimension sequence using the binary image of the size 32×32, so the watermark information length is $M = 1024$. The original image what we tested is 512*512 grey image(lena). The watermarked image with PSNR(Peak Signal-to-Noise Ratio) 44.582dB.

- JPEG compression and JPEG2000 compression

This paper test the robustness of the proposed method with several typical images attacked by JPEG and JPEG2000 compression. Fig.1 (a) and (b) shows the watermarks extracted from JPEG and JPEG2000 compressed versions of the watermarked image with various compression quality factors. we can see that the extracted watermark is still weakly recognizable when the compression quality factor reaches 30, which show the proposed method has good robustness to JPEG and JPEG2000.



（a）　　　　　　　　　　　　　　　　（b）

Fig. 1.(a) Extracted watermarks from the watermarked image with JPEG compression quality factors from 100 to 30.(b). Extracted watermarks from the watermarked image with JPEG2000 compression quality factors from 100 to 30.

- Geometric deformation attack

Fig.2 is a typical geometric distortion attack (including shear, rotation and scaling) to the watermark image. From fig.5, when the watermarked image by 1/16 and 1/4 shear attack, the watermark detection accuracy rate were 99.52% and 94.18%; when the watermarked image is rotated 5 degrees and 10 degrees of the attack, the watermark detection is correct the rate is 100% and 99.83% respectively; When the watermark image is reduced by 0.9 and the amplification factor is 1.1, the correct rate of watermark detection is 100%. When the watermark image is reduced by 0.8 and the amplification factor is 1.2, the correct detection rate of watermark is 98.93% and 99.01%, respectively. It can be seen that the method also has good robustness for geometric distortion attack.

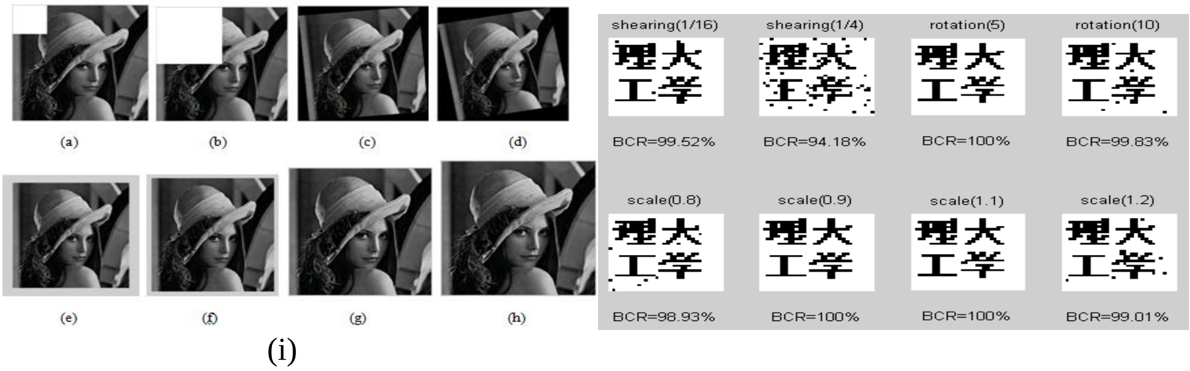


Fig.2. The watermark image is subjected to a typical geometric distortion attack: (a) shearing (1/16);(b) shearing (1/4); (c) rotating(5o);(d) rotating(10o); (e) reducing(0.8);(f) reducing(0.9);(g) enlarging (1.1);(h) enlarging (1.2);(i) The corresponding watermark extracted from the watermark images after various geometric attacks.

5. conclusions

This paper presents an image watermarking algorithm in wavelet domain based on RF model. First, the image is decomposed through wavelet transform. Then we use the relationship between the selected coefficient and its neighboring to train RF. Thanks to the good learning ability of RF, the watermark is adaptively embedded in wavelet domain and also can be extracted by the well trained RF. the watermark information is embedded into the good imperceptibility, and extract the watermark without the original image. Experimental results show that the algorithm is robust against high intensity JPEG compression attacks, JPEG2000 compression attacks, geometric distortion attacks.

References

[1]L. Breiman. Random forests. Machine Learning, 45(1):5–32, 2001.

[2]Wang, Y.W., Doherty, J.F., Robert, E.V., A wavelet-based watermarking algorithm for ownership verification of digital images. IEEE Transactions on Image Processing, 11 (2), 77–88, 2002.

[3]Cox, I.J., Kilian, J., Leighton, T., Shamoon, T., Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing , 6 (12), 1673–1687, 1997.

[4]Shelby, P., Thierry, P., Robust template matching for affine resistant image watermarks. IEEE Transaction on Image Processing, 9 (6), 1123 –1129, 2000.

[5]Barkat B, Sattar F. Time-Frequency and Time-Scale-Based Fragile Watermarking Methods for Image Authentication[J]. Eurasip Journal on Advances in Signal Processing, 2010(1):1-14, 2010

[6]R. Buccigrossi and E. Simoncelli, “Image compression via joint statistical characterization in the wavelet domain,” IEEE Transactions on Image Processing, vol. 8, no. 12, pp. 1688–1701,1999.