Information Hiding Algorithm Based on Predictive Coding

Junling Ren
School of Information Management
Beijing Information Science & Technology University
Beijing, China
e-mail: junlingren@hotmail.com

Yangbo Xia
School of Information Engineering
University of Science and Technology Beijing
Beijing, China
e-mail: renjunling@bistu.edu.cn

Zhifeng Ma
School of Information Science and Technology
Beijing Institution of Technology
Beijing, China
e-mail: mazhifeng@bit.edu.cn

***Abstract*—In view of the information security questions, the information hiding technology already becomes the hot spot in the research field. On the basis of the predictive coding, an algorithm using the prediction error to carry on the information hiding is proposed in this paper. In order to restrain the error diffusion which possibly appears during the anti-predictive coding in the information hiding process, an improved predictive coding algorithm is put forward. Through the experiments, the performances of the basic algorithm and improved algorithm are tested, resulting in the proof of the thread correctness. At the same time, the improved algorithm achieves the ultra large information capacity of 0.953 bits/Byte and the PSNR of 49.184dB so as to verify the validity of the improved algorithm.**

Keywords-information hiding; predictive coding; stochastic permutation algorithm; information hiding capacity;robustness

I. INTRODUCTION

Along with the development of the multimedia technology and the widespread application of network communication, the information security questions, such as information authentication, copyright protection and secret transmission, have become an increasing concern. Therefore, the information hiding technology arises and develops rapidly. Many scholars have made a lot of research on the information hiding technology. For instance, the improvement on the classical LSB algorithm [1], the lossless information hiding method [2, 3] based on the space domain and transform domain of the image proposed by Fridrich, the lossless data hiding method [4] based on prediction errors' expansion proposed by Thodi, the further study on the linear prediction [5, 6] carried on by Yuming Xie and Hongji Piao and so on so forth.

This article takes the images as carriers to show an information hiding method via predictive coding. First of all, the concept of predictive coding is interpreted, and then the fundamental algorithm of information embedding and extracting is given out, which is based on the predictive coding. In order to prevent the error from diffusing, another improved algorithm about information embedding is introduced. Finally, an analysis is carried on through the experiments to verify the validity of the algorithm above.

II. PREDICTIVE CODING

Predictive coding is a kind of information source coding based on data correlation, and the basic thought is reducing data dependency on the time domain and space domain. That is, the values of the adjacent samples are used to predict the current pixel value. Afterwards the difference (Prediction error) between the predicted value and real sample value is coded and transmitted.

Speaking of the image, there are two approaches usually used to acquire the predictive values. One is predicting the current pixel value from the adjacent pixel value. It is often happens that the previous pixel value is used to predict the current pixel value. The other is predicting the current pixel value from the adjacent image block. For example, the current pixel value is predicted from the weighted average of the several adjacent pixels.

III. INFORMATION HIDING ALGORITHM BASED ON PREDICTIVE CODING

A. Fundamental Information Hiding Algorithm

The main idea of the information hiding algorithm based on the predictive coding is to utilize the prediction errors for the transmitting of the secret information. In this paper, the bits of the corresponding difference signal (prediction error) obtained from the presumptive secret key table is compared with the secret information bits which are going to be embedded to carry on the information hiding. Here, the presumptive secret key table is established in advance. It assigns a bit for each possible difference.

Now we take the grey scale images for an example, and Fig. 1 illustrates the basic process of the algorithm.

It is embodied as below:

978-0-7695-3688-0/09 $25.00 © 2009 IEEE
DOI 10.1109/ITCS.2009.141

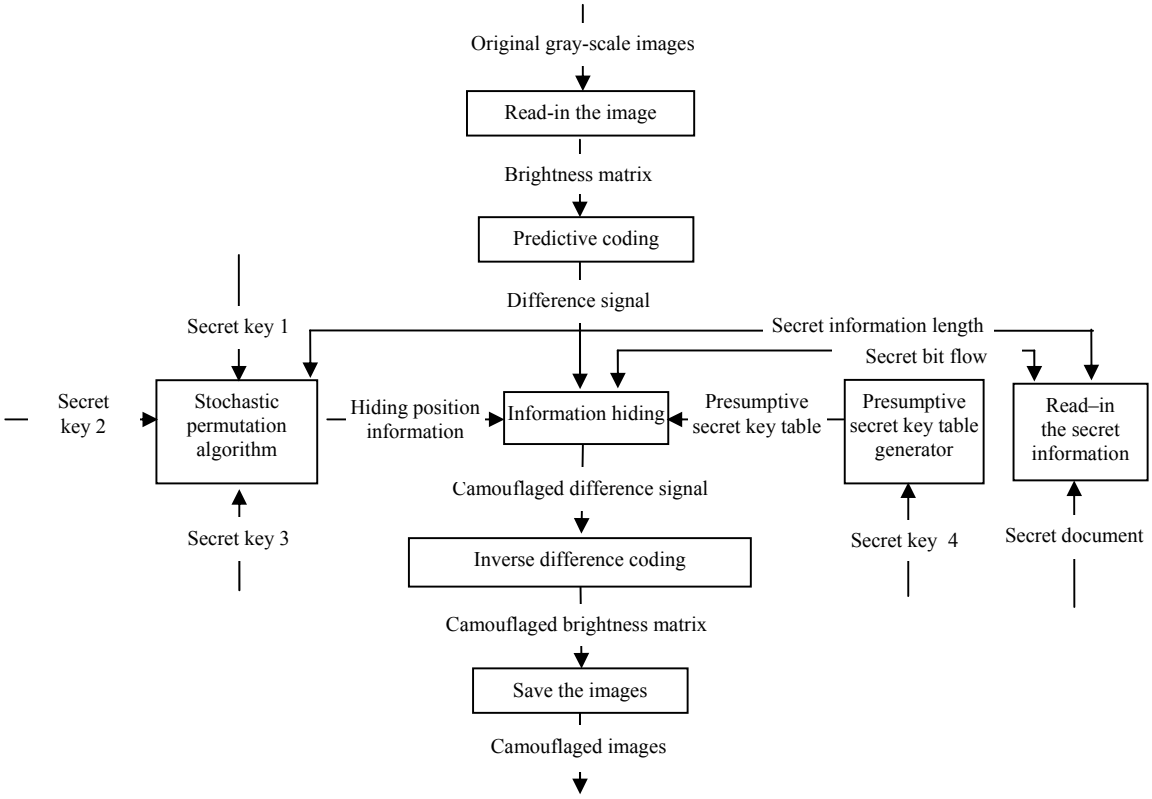

Figure 1. Embedded algorithm flow diagram

*1) Step 1:*Read the carrier image to acquire the brightness matrix.

*2) Step 2:*According to the brightness matrix, carry on the predictive coding to gain the difference signal.

In this article, the previous pixel value is used to predict the current pixel value, thus the signal value in the previous position subtracted from the signal value in the current position gives the difference signal. Here, the matrix made of the difference signals is called as difference matrix.

*3) Step 3:*Determine the secret information embedding position.

It is implemented by the stochastic permutation algorithm based on the MD5 algorithm. The main idea of the algorithm is: assume each input to be i, and i is an integer less than the sum of the units embedded into the carrier. We can get a number of J_i to express the embedding position of No.i bit in the secret information. The step of generating J_i is:

$$v = [i/X]; \quad (1)$$

$$u = i \bmod X; \quad (2)$$

$$v = (v+\mathrm{MD5}(u,k_1)) \bmod Y; \quad (3)$$

$$u = (u+\mathrm{MD5}(v,k_2)) \bmod X; \quad (4)$$

$$v = (v+\mathrm{MD5}(u,k_3)) \bmod Y; \quad (5)$$

$$J_i = vX+u. \quad (6)$$

Here, X and Y represent the row and column respectively. k_1, k_2 and k_3 are three secret keys.

*4) Step 4:*Input the secret key of k_4, and get the presumptive secret keys table by the stochastic number generator. The length of the presumptive secret keys table is determined by the range of the difference signals.

5) Step 5: The binary bits of secret information are read in, and form the secret bits flow waiting for being hidden.

6) Step 6: Embed the secret information.

For the No.i secret bit, find the corresponding $\triangle_i$ from the embedding position determined in the *Step3*. Compare the $\triangle_i$ with the corresponding bit in the presumptive secret keys table. If it is the same as the secret bit, does not do any revises. And if it differs from the secret bit, find $\triangle_i'$ which is nearest to $\triangle_i$ and the same as the secret bit in the presumptive secret keys table.

Now we illustrate the embedding process through an example. It is supposed that the embedded 10th secret bit is 0. In the hiding position information, we find its position is row 8 and column 10. And the difference in this position is 25. The corresponding presumptive secret keys table is shown in Table I. The bit according to 25 is 1 in the table. It is different from the secret bit. Find the difference which is not only the nearest to 25 but also equal to the bit of 0. The result is 24. So the difference in Row 8 and Column 10 is modified as 24 to hide the 10th secret bit. Repeat the same

TABLE I. EXAMPLE OF THE PRESUMPTIVE SECRET KEYS TABLE

$\triangle_i$	20	21	22	23	24	25	26	27	28
M_i	1	1	0	1	0	1	1	0	1

embedding process for each secret bit to complete the information hiding.

7) Step 7: Inverse predictive coding.

Finished hiding the secret information, the camouflaged difference matrix with the secret information is obtained, and the camouflaged brightness matrix can be acquired by carrying on the inverse predictive coding.

8) Step 8: Save it as the target camouflaged image. Save the camouflaged brightness information as the target camouflaged image. So far the process of the secret information hiding is completed.

B. Improved Information Hiding Algorithm

After hiding the information, the camouflaged brightness matrix is formed according to the modified difference signals. Since the secret information is concealed in the difference matrix, the error is diffused. For instance, the first six of some column among the brightness matrix is {15 11 17 18 8 10}. And it turns into {15 -3 6 1 -10 2} after predictive coding. Suppose that the second integer represents the position where secret information is hidden and it turns from -3 into -4 after hiding. Then the sequence becomes into {15 -4 6 1 -10 2}. After inverse predictive coding, the data becomes into {15 11 17 18 8 10}. Due to hiding such a message, each brightness value which is in the same row and behind the secret information hiding position changes, resulting in a diffusion of error. This will seriously affect the invisibility of information hiding.

Because the second integer indicates the hiding secret information position, as a result of the same secret key, the position is still where the secret information is hidden when the information is extracted. Only the difference signal according to the position impacts on the extracting of the secret information. And other signals belonging to the positions without the secret information will not affect it. Therefore, the brightness value of the position without secret information can maintain the original value, and there is no need to carry on predictive and inverse-predictive coding. This will not only have no effect on the extracting of secret information, but also greatly enhance invisibility of the information hiding. When concealing the information of the same capacity, the degree of the image modification gets smaller, and more information can be hidden. So the upper limit of the hidden information's capacity is increased. After hiding the secret information through this method and restoring brightness matrix, the series of the example above turns into {15 11 18 19 9 11}.

C. Algorithm of Information Extraction

The basic process of the extraction algorithm:

1) Step 1: Read the camouflaged image in, and obtain the camouflaged brightness matrix.

2) Step 2: Carry on the predictive coding for the camouflaged brightness matrix, obtain the camouflaged difference signal.

3) Step 3: Use the same method of information hiding to determine the position of the secret information embedded in.

4) Step 4: Generate the presumptive secret key table. The method is the same as that of the generation of the presumptive secret key table when hiding the information.

5) Step 5: Extract the secret information. For the No.*i* bit, examine the information (*m*,*n*) in the camouflaged position, and obtain the corresponding difference $\triangle_i$ from the camouflaged difference signal. According to presumptive secret key table, obtain the secret information bit.

6) Step 6: Save the secret information as a secret document to complete the extraction.

IV. EXPERIMENTS

To test the validity of the two algorithms in this paper, we do the experiments on the algorithms above. Here the basic hiding algorithm is called as Meth.1 and the improved hiding algorithm is called as Meth.2. Then we will carry on an analysis of their invisibility, information hiding capacity and robustness. Four gray-scale images are used for the carrier images. The size of these images is given as follows: carrier1:64×64, carrier2:128×128, carrier3:192×192, carrier4: 256×256.

A. Invisibility

In order to test the invisibilities of the two algorithms in the paper, we use separately Meth.1 and Meth.2 to hide the information of different content, such as 10Byte,50Byte and 100Byte. Then we will analyze the PSNR values(units:dB). The experimental results are shown in Table II. From Table II, we can see PSNR value of the Meth.2 is bigger than that of the Meth.1. This shows the invisibility of Meth.2 is better than that of the Meth.1.

B. Information Hiding Capacity

For the carrier images, there is an upper limit for the information hiding capacity. It is restrained by the size of carrier, the invisibility of the algorithm and the robustness. Table III shows the maximum hiding rate (MHR) and the corresponding PSNR of the two algorithms proposed in this article in different carrier images.

In Table IV, we compare the MHR and PSNR of the Meth.2 with those of other information hiding algorithms based on predictive coding. We can see that both the MHR and PSNR of the algorithm proposed in this paper have a great enhancement, thus the validity of the algorithm is proved.

C. Robustness

Robustness refers to the ability of the correct secret information extracting from the carrier under certain attacks. In the experiments, we focus on the robustness against JPEG attack and measure it in similarity ratio. Here, similarity ration is defined as the percentage of the bits of

TABLE II. PSNR VALUES OF THE CARRIER IMAGES HIDING THE SECRET INFORMATION OF DIFFERENT CONTENT

Carrier Image	10Byte		50 Byte		100 Byte	
	Meth.1	*Meth.2*	*Meth.1*	*Meth.2*	*Meth.1*	*Meth.2*
Carrier1	51.46	67.18	45.14	60.79	41.50	57.25
Carrier2	56.74	75.10	47.72	67.18	45.22	63.90
Carrier3	58.73	78.00	47.22	68.55	44.85	65.67
Carrier4	56.82	77.55	50.46	71.93	47.99	69.38

TABLE III. COMPARISON OF THE MHR AND PSNR

Carrier Image	Meth.1		Meth.2	
	MHR (bits/Byte)	*PSNR (dB)*	*MHR (bits/Byte)*	*PSNR (dB)*
Carrier1	0.040	48.177	0.961	49.342
Carrier2	0.031	46.614	0.976	49.002
Carrier3	0.028	42.794	0.896	48.987
Carrier4	0.032	42.576	0.977	49.403

TABLE IV. COMPARISON OF THE METH.2 AND OTHER ALGORITHMS

Information Hiding Algorithm	MHR (bits/Byte)	PSNR (dB)
Meth.2 in this paper	0.953	49.184
Thodi algorithm	0.182	40.754
Improved Thodi algorithm [6]	0.364	31.27

TABLE V. COMPARISON OF SIMILARITY RATIO

Carrier Image	Meth.1		Meth.2	
	Condition 1	*Condition 2*	*Condition 1*	*Condition 2*
Carrier1	93.75%	55%	97.625%	53.125%
Carrier2	98.375%	55.5%	99.875%	53.5%
Carrier3	99.875%	54.72%	100%	53.5%
Carrier4	100%	52.25%	100%	52.375%

extracted information which is the same as the primitive secret information to the bits of primitive secret information.

Table V shows the similarity ratio of the two algorithms except from the JPEG attack (Condition 1) and under the attack of JPEG (Condition 2) when 100 bytes are hidden in four carrier images.

From the experimental result, we can see that the robustness is weak under the JPEG attack no matter which method is used. The reason is that the essence of the information hiding algorithm based on predictive coding is hiding the information in the noise signals. So the robustness is poor.

V. CONCLUSIONS

In this paper, an information hiding method based on predictive coding is proposed. In order to restrain the error diffusion which possibly appears during the inverse predictive coding in the information hiding process, an improved predictive coding algorithm is put forward. The analysis of invisibility and hiding capacity verifies the validity of the algorithm. And it also confirms the improved predictive coding algorithm is superior to the basic predictive coding algorithm in the performance. Simultaneously, the comparison of the improved algorithm and other algorithms indicates that the improved algorithm proposed in this paper surpasses the other predictive coding algorithms in both the hiding capacity and the signal-to-noise ratio. So it is a high performance algorithm with the ultra large information capacity.

REFERENCES

[1] Bing Fu, "A method for raising the hiding of LSB information", Journal of Yangtze University (Natural Science Edition), ,vol.3, Dec. 2006, pp. 73-75.

[2] Fridrich J, Goldjan M, Du R., "Invertible authentication", *Proceedings of SPIE*. San Jos, California, vol. 4314, pp.197-208, January 2001.

[3] Goldjan M, Fridrich J, Du R., "Distortion-free data embed-ding", *Proceedings of the 4th Information Hiding Work-shop*, Pittsburg, PA,USA, pp.27-41, April 2001,

[4] Thodi D M, Rodrfguez J J., "Reversible watermarking by prediction-error expansion", *Proceedings of the 6th IEEE* Southwest Symposium on Image Analysis and Interpretation, Lake Tahoe, Nevada, vol.6, pp. 21-25, March 2004.

[5] Yuming Xie, Yimin Cheng, and Yixiao Wang, "Lossless Data Hiding Method in Image Based on Linear Prediction", Journal of Computer-Aided Design & Computer Graphics, China Computer Federation, Beijing, vol.18, Apr.2006, pp. 585-591.

[6] Hongji Piao, Pin Zheng, Xiong Tian, "Data Hiding Algorithm Based on Linear-Prediction and Bit-Operation", *Computer Technology and Development*, China Computer Federation Shanxi Province branch, Xi'an, Shanxi, vol.18, Jan.2008, pp.185-187.