

A Digital Watermarking Algorithm Based on Image Similarity Calculation

J. Yang

Information Technology Department
Hubei University of Police
Wuhan, China

J. Qian, T.L. Ma, H.X. Zhang

School of Printing and Packing
Wuhan University
Wuhan, China

Abstract—Normalized correlation algorithms have been used in digital watermarking evaluation system at home and abroad which still exists some shortcomings. In this paper, several common image similarity algorithms in image retrieval technology are selected and applied to watermarking technology. This paper chooses an adaptive embedding and extraction algorithm based on discrete cosine transform (DCT) and uses some different methods to calculate the similarity between the extracted watermark images and the original watermark image. Before embedding, the watermark image has been scrambled. The entire process of image watermarking system is conducted by Matlab, which demonstrates that applying the similarity calculating methods used in image retrieval into digital watermarking technology can improve the accuracy when evaluate the robustness of a watermark system.

Keywords—digital watermarking; image retrieval; image feature; the similarity calculating

I. INTRODUCTION

In digital watermarking technology, how to calculate the similarity between the extracted watermark information with the original watermark information is a very important problem. Because it determines whether the watermark information can be detected properly or not, and whether the evaluation of the robustness of a watermarking algorithm is really reasonable. Presently normalized correlation has been widely used in digital watermarking evaluation system. Although traditional normalized correlation formula can show the similarity between original watermark and extracted watermark, it also exists some shortcomings. It is necessary to seek a more appropriate similarity calculating method, and apply it to the digital watermarking technology. The objective of this paper is to research the digital watermarking technology based on image retrieval similarity calculation.

This paper is organized as follows. The theories analysis are presented in Section 1. The experimental methods are illustrated in Section 2. The experiment results and discussion are shown in Section 3. Finally, the concluding remarks are given in Section 4.

II. THEORIES ANALYSIS

A. Analysis of the Common Watermark Similarity Algorithm

Assume that $\rho(w, \tilde{w})$ represents the similarity, w and \tilde{w} represent the original watermark of $M \times N$ and the extracted watermark.

The watermark similarity is calculated in Eq.(1) proposed by Lin et al[1] :

$$\rho_1(w, \tilde{w}) = \frac{w^T \tilde{w}}{\sqrt{w^T w} \sqrt{\tilde{w}^T \tilde{w}}} \quad (1)$$

The similarity proposed by [Luo et al][2] is defined as

$$\rho_2(w, \tilde{w}) = \frac{w^T \tilde{w}}{w^T w} \quad (2)$$

[Sun et al][3] proposes to determine the similarity with Eq.(3):

$$\rho_3(w, \tilde{w}) = \frac{w^T \tilde{w}}{w w} \quad (3)$$

If the watermark information is the length of R of binary sequences, the similarity algorithm defined by [Lv W Y et al][4] is described as

$$\rho_4(w, \tilde{w}) = \frac{\sum_{i=1}^I (w_i \otimes \tilde{w}_i)}{I} \quad (4)$$

Eq.(1) and Eq.(4) are generally known as the "normalized correlation". In the Eq.(1), the consideration of similarity calculation is the angle between the two-dimensional vectors. Due to the length of the vector does not affect the size of the angle between them, there will be as long as the angles between the extracted watermark and the original watermark are equal, their NC value is equal to 1. As is showed in Figure 1 left, The pixel values of the original watermark (a) is two times as much as extracted watermark (b). The similarity $NC = 1$ calculated by Eq.(1) represents the consistency of two images however they are visually very different.

As is showed in Figure 1 right, the similar problem caused by the Eq.(2) and the Eq.(3) is that the results of the calculation represent the consistency of two images however they are visually very different.

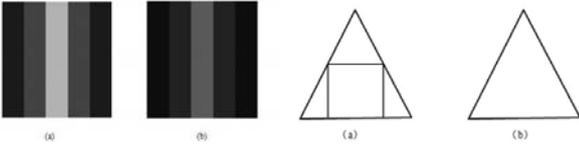


FIGURE 1. THE ORIGINAL WATERMARK (A) AND THE EXTRACTED WATERMARK (B)

B. Similarity Algorithms Used in Image Retrieval

To solve the above shortcomings of normalized correlation, some researches had been done in the field of image retrieval. In this field, image features are extracted based on the extraction of color feature, texture feature and shape feature, which can be used to calculate the similarity[5].

1) *Geometric distance*: At present there are many distance functions several commonly used in the calculation of image similarity, such as Euclidean distance, Manhattan distance and Chebyshev distance etc.

Euclidean distance is also called L2 distance, in n-dimensional space the shortest length of the line between two points is the Euclidean distance[6]. The Euclidean distance is described as:

$$\rho(x, y) = \left[\sum_{i=1}^n |x_i - y_i|^2 \right]^{\frac{1}{2}}, \quad i=1,2,\dots,v \quad (5)$$

Manhattan distance derives from the city block distance and is also referred to as blocks distance or L1 distance. It is the sum of the calculated absolute wheelbase two points on the coordinate system of the standard. The Manhattan distance formula is given by:

$$\rho(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (6)$$

Chebyshev distance is also called chessboard distance or L ∞ distance. The chebyshev distance between two points is equal to the maximum value of the coordinate difference[7]. The actual calculation formula is as follows:

$$\rho(x, y) = \max_{i=1,2,\dots,n} \{ |x_i - y_i| \} \quad (7)$$

2) *Mahalanobis distance*: In calculating the distance between image feature vector, if each characteristic component affects each other or the characteristic components play different roles in distinguishing images, a new method to calculate the metric between feature vectors differences can be used which named Mahalanobis distance [8]. Its actual calculation formula is as below:

$$D_s = \sum_{i=1}^N \frac{(X_i^{(1)} - X_i^{(2)})^2}{C_i} \quad (8)$$

3) *Histogram distance*: In measuring the diversities between different histograms, the most commonly used method is to calculate the histogram intersection. The same number of pixels in each cell between two histograms is the two histograms' histogram intersection[9]. If the two color histogram X and Y contains N cells, in order to make the range of histogram intersection fall between [0, 1], the formula can be standardized as:

$$\rho(x, y) = \sum_{i=1}^n \min\{x_i, y_i\} / \sum_{i=1}^n y_i \quad (9)$$

4) *Bhattacharyya distance*: In statistics, Bhattacharyya distance is mainly used to measure the two discrete probability distribution. In calculating the histogram similarity, Bhattacharyya has the best effect, but its calculation is the most complex. At the same domain Z, the Bhattacharyya distance formula of the probability distribution P and Q is as follows[10]:

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (10)$$

III. EXPERIMENTAL METHODS

A. Experimental Considerations

The experiment idea is as shown in Figure 2.

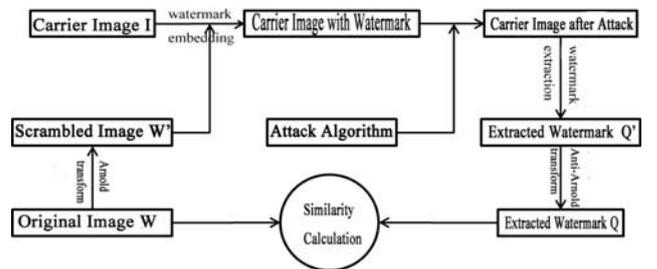


FIGURE II. EXPERIMENTAL CONSIDERATIONS

Vector image used in the experiment is 512 * 512 grayscale lena. Meanwhile the watermark image is 128*128 "CS" binary image whose format is Bitmap. Watermarking algorithm tested in the experiment is based on the discrete cosine (DCT).The operating system is windows 7. It is easy to observe and compare through a very intuitive GUI(Graphic User Interface) interface which is generated in matlab 7.0 simulation environment. Part of the attack algorithms are also carried out by Matlab, the final calculation of similarity between the original watermark information and extracted watermark information are completed by Matlab 7.0.

IV. DISCUSSION AND RESULTS

To verify the effects of diffirent methods used to calculate the similarity in digital watermarking technology, the experiment watermarked image has been tested with attacks such as image cropping, rotation, Gauss fuzzy, Guassian noise addition.

A. Image Cropping

The four corners of the watermarked image is cropped in experiment 1. As shown in Figure 4, followed by the sheared image, the extracted watermark image after shearing and anti scrambling the watermark image. In the case of cropping, although the extracted watermark here has been a lot of noise, it can still identify original watermark image information "CS".

B. Rotation

The watermark is rotate with 30° in experiment 2. As shown in Figure 5, the extracted watermark image after rotation and anti scrambling the watermark image are followed by the rotated image. As we can see, after rotation the extracted watermark image has been unable to identify the original watermark image information.

C. Gauss Fuzzy Experiment

Experiment 3 is a Gauss fuzzy operation on watermarked image, which is known as Gaussian low-pass filtering. As shown in Figure 6, the extracted watermark image after rotation and anti scrambling the watermark image are followed by the image after Gauss fuzzy. It is clear to see that after Gauss fuzzy operation the extracted watermark image has been unable to identify the original watermark image information.

D. Guassian Noise Addition

Experiment 4 adds Gauss white noise to embedded watermarking image in Matlab, the white Gauss noise density is 0.02. As shown in Figure 7, the extracted watermark image after adding Gaussian noise and anti scrambling the watermark image are followed by the image with additive Gaussian noise. The result is that after shear attack the extracted watermark although there has been a lot of noise, it can still identify original watermark image information "CS".



FIGURE III. IMAGE CROPPING EXPERIMENT



FIGURE IV. ROTATION ATTACK EXPERIMENT



FIGURE V. GAUSSIAN BLUR



FIGURE VI. ADD GAUSSIAN NOISE

The normalized correlation values of two images of NC1, NC2, HD correlation value histogram intersection, Euclidean distance OD and Bhattacharyya distance BD are calculated, the results are given in table 1.

TABLE I. THE CALCULATION OF WATERMARKING IMAGE SIMILARITY

Processing mode	Attack parameters	NC1	NC2	HD	OD	BD
Cropping	Four corners	0.859	0.741	0.945	0.077	0.950
	shear	6	2	1	7	0
rotation	10 deg	0.289	0.401	0.803	0.278	0.848
		7	4	2	3	4
Gauss fuzzy		0.252	0.362	0.771	0.323	0.825
		1	6	0	9	6
Gaussian noise	density 0.02	0.786	0.915	0.942	0.107	0.938
		7	6	1	4	0

The experimental results shows that the resistance of selected watermarking algorithm to five kinds of attacks was ordered as: Cut > Gaussian noise > Rotate > Gaussian Blur.

This result is consistent with the results of evaluation of OD and BD and more accurate than the results of NC1 and NC2.

V. CONCLUSION

Discussing the contrast problems of extracted watermark images and the original watermark images in the field of digital watermark, the paper proposes a method to calculate the similarity of extracted watermark images and original watermark images with similarity calculation algorithm in image retrieval, and emulates and simulates the total process of embedding, attacking and detection and extraction by means of GUI. The results verifies that applying the method of similarity calculation in image retrieval to the digital watermarking technology can improve the accuracy of the digital watermarking robustness evaluation.

ACKNOWLEDGMENT

The authors thank the financial support of the Technology R&D Program of Hubei Province: the project of digital copyright protection and trade service cloud platform (No. YJG0264) and the Hubei Provincial Service Industry Guidance of Capital Construction Program: the project of digital newspaper publishing and copyright trading technology. (Development and Reform Commission issued a notice to guide capital No.[2013]765, Special Account Funds Direct Financial Payment Application No.[2013]206)

REFERENCES

- [1] Lin C, Wu M, Bloom J A, et al. Rotation, scale, and translation resilient watermarking for images[J]. *Image Processing*, IEEE Transactions on, 2001,10(5):767-782.
- [2] Luo W, Heileman G L, Pizano C E. Fast and robust watermarking of JPEG files[C]// *In Proceedings of the Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*,. IEEE, 2002: 158-162.
- [3] Sun J F, Wen Q, Wang S X. Based on chaotic array image watermarking algorithm[J]. *ACTA ELECTRONICA SINICA*, 2003,32(1):149-153.
- [4] Li X Y, Zhuang Y T, Pan Y H. The technique and system of content-based image retrieval [J]. *Journal of Computer Research & Development*, 2001, 38(3):344-354.
- [5] Li M L. Research of image retrieval technology & design and implementation of system based on color and shape feature[D]. Xi'an: Northwest University, 2010.

- [6] Yang W X. Composite Model research based on shape for image matching[D]. Beijing:Signal and Information Processing North China Electric Power University, 2010.
- [7] Fu X J. Research on Algorithm of content-based image retrieval integrated based on color and texture features[D]. South China Normal University,2011.
- [8] Di Q, Feng X. Survey on content-based image retrieval techniques[J]. *Computer Engineering*, 2005,31(z1):223-225.
- [9] Yong H. Watermarking algorithm research and application based on transform domain[D]. Yingchuan: Ningxia University circuit and system, 2010.
- [10] Chen J B. Research and implementation of image features extraction and its similarity[D]. Xi'an: Software Engineering Xi'an Electronic Science & Technology University, 2012.