A Robust Watermarking Algorithm for the Encryption of Medical Big Data

Lian XU, Jing-bing LI*, Meng-xing HUANG, Yu-cong DUAN

College of Information Science & Technology, Hainan University, Haikou, Hainan, China, 570228

***Abstract* — Medical big volume data should be encrypted for preserving personal privacy when it is stored in the cloud or transferred in a channel. In order to facilitate the third party to manipulate the encrypted medical volume data directly and eliminate the tedious process of deciphering, we propose in this paper a robust watermarking algorithm of medical volume data in the encrypted domain. In this algorithm, there are several steps. Firstly, the medical volume data is encrypted using Chen hyper-chaos in the 3D-DCT domain; and then the feature vector of encrypted medical volume data is extracted in the 3D-DFT domain; finally, the feature vector is associated with the watermark to implement the embedding and extraction of the watermark. The watermark embedding doesnot change the image data of the encrypted volume data, which is zero-watermarking technique. The difficulty of implementing the algorithm is to obtain the feature vector with good robustness in the encrypted domain of the medical volume data. The experimental results show that the watermark can be extracted correctly without deciphering the encrypted medical volume data, and it can resist some normal attacks and geometric attacks.**

Keywords - Medical volume data; feature vector; chaotic encryption; zero watermarking; robustness.

I. INTRODUCTION

Medical image watermarking technology has been used mainly to protect patients' personal information. The patients' electronic medical records and other information were hidden in the medical images as watermarks. The process of embedding and extraction in most of present watermarking technologies is implemented on the plain images. Nowadays personal privacy is attracting more and more attention. More people are aware that when the medical images are stored in the server, if the server is untrustworthy or insecure, the medical images will cause privacy leakage. Therefore, the medical images need to be encrypted when being stored and the watermarking technology based on the encrypted domain is regarded as a more secure watermarking technology.

In recent years, there are some people who research on the image processing in the encrypted domain. Bianchi [1] proposed the implementation of DFT and FFT in the encrypted domain. Zheng[2] proposed the implementation of DWT in the encrypted domain and the method to reduce the data expansion after encryption. Hsu[3] used the scale-invariant feature transform(SIFT) to extract the image feature in the encrypted domain. Zheng[4] implemented Walsh-Hadamrd transform in the homomorphic encrypted domain and applied it in the image watermarking, which realized an application that the third party implemented the embedding of the watermark without knowing anything about the plain image. Zhao[5] proposed an effective watermarking scheme in the encrypted domain for seller-buyer watermarking protocol. Subramanyam[6] proposed a robust watermarking algorithm of compressed and encrypted JPEG2000 images. Those algorithms utilize the concept of homomorphic encryption to realize secure image processing, while the process of computation is complex and the robustness needs to be improved. In order to solve the problem of data expansion in the process of encryption, a method of composite signal representation was proposed in [7]. Barni used branching programs and neural networks to implement secure classification of electrocardiogram signals presented in [8]. A problem of FingerCode-based identity matching in the encrypted domain was addressed in [9]. Lu. proposed a method of content-based retrieval over encrypted multimedia database[10]. Some approaches of commutive encryption and watermarking can be found in [11]-[12].

There is few researches on the watermarking algorithm for encrypted medical images. And there is almost no watermarking algorithm for the encrypted 3D medical volume data which can resist geometric attacks. While there is large amount of volume data among the medical images, i.e.: CT, MRI images are volume data composed by slices. The research on how to embed the watermark in the encrypted volume data is significant. Medical volume data cannot be changed for its special purpose, which makes the watermark embedding more difficult. In this paper, we proposed a novel watermarking algorithm for medical volume data in the encrypted domain. The rest of this paper is organized as follows. Section II gives the fundamental theory used in our proposed algorithm. Section III introduces the process of the proposed algorithm in details. Section IV reports the experimental results, and the conclusion is given in Section V.

II. FUNDAMENTAL THEORY

A. 3D Discrete Cosine Transform(3D-DCT)

The formula of 3D(three-dimensional) DCT is as follows:

$$F(u,v,w)=\mathrm{c}(u)\,\mathrm{c}(v)\,\mathrm{c}(w)[\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\sum_{p=0}^{P-1} f(x,y,z)\cdot \cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)v\pi}{2N}\cos\frac{(2z+1)w\pi}{2P}] \quad (1)$$

$$u=0,1,\cdots,M-1; v=0,1,\cdots,N-1; w=0,1,\cdots,P-1;$$

Where

$$c(u)=\begin{cases}\sqrt{1/M} & u=0\\ \sqrt{1/M} & u=1,2,\cdots,M-1\end{cases}$$

$$c(v)=\begin{cases}\sqrt{1/N} & v=0\\ \sqrt{2/N} & v=1,2,\cdots,N-1\end{cases}$$

$$c(w)=\begin{cases}\sqrt{1/P} & w=0\\ \sqrt{2/P} & w=1,2,\cdots,P-1\end{cases}$$

The formula of 3D IDCT is as follows:

$$f(x,y,z)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\sum_{p=0}^{P-1}[c(u)c(v)c(w)F(u,v,w)\cdot \cos\frac{(2x+1)u\pi}{2M}\cos\frac{(2y+1)v\pi}{2N}\cos\frac{(2z+1)w\pi}{2P}] \quad (2)$$

$$x=0,1,\cdots,M-1; y=0,1,\cdots,N-1; z=0,1,\cdots,P-1;$$

Where (x,y,z) is the sampling value in the spatial domain and (u,v,w) is the sampling value in the frequency domain.

B. 3D Discrete Fourier Transform(3D-DFT)

We suppose that f(x,y,z) is a 3D function in a discrete space, so the formula of 3D discrete fourier transform and the inverse 3D discrete fourier transform are expressed as follows:

The direct formula of 3D discrete Fourier transform (3D-DFT) :

$$F(u,v,w)=\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\sum_{z=0}^{P-1} f(x,y,z)\cdot e^{-j2\pi xu/M}e^{-j2\pi yv/N}e^{-j2\pi zw/P} \quad (3)$$

$$u=0,1,\cdots,M-1; v=0,1,\cdots,N-1; w=0,1,\cdots,P-1;$$

The inverse formula of 3D discrete fourier transform (3D IDFT) :

$$f(x,y,z)=\frac{1}{MNP}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\sum_{w=0}^{P-1} F(u,v,w)\cdot e^{j2\pi xu/M}e^{j2\pi yv/N}e^{j2\pi zw/P} \quad (4)$$

$$x=0,1,\cdots,M-1; y=0,1,\cdots,N-1; z=0,1,\cdots,P-1;$$

In the direct and inverse formulas of discrete Fourier transform, f(x,y,z) is a function in the 3D spatial domain, F(u,v,w) is the corresponding function in the 3D frequency domain.

C. Chen's hyper-chaotic system

Chen's hyper-chaotic system defines as follows:

$$\begin{cases}\dot{x}=a(y-x)+q\\ \dot{y}=dx-xz+cy\\ \dot{z}=xy-bz\\ \dot{q}=yz+kq\end{cases} \quad (5)$$

Where $\dot{x},\dot{y},\dot{z},\dot{q}$ are state parameters, which are the differentials of time t, a, b, c, d, k are system parameters. When a=35, b=3, c=12, d=7 and $0.085 \le k \le 0.798$, Chen's hyper-chaotic system is in a chaotic state and it can produce four chaotic sequences. We suppose that the initial value of Chen's hyper-chaotic system is (x_0, y_0, z_0, q_0) and the system parameters are a, b, c, d, k, and the step length is Δt. The method of Runge-Kutta is used to solve the function and four sequences of x,y,z,q are obtained. In order to get better results, the integer part of the hyper-chaotic sequence is deleted and the decimal part of that is retained. In this way, four sequences with good randomness are obtained.

The algorithm process

The algorithm process is divided into three parts: the encryption of medical volume data, the embedding and the extraction of the watermark, the process is expressed as Fig. 1.

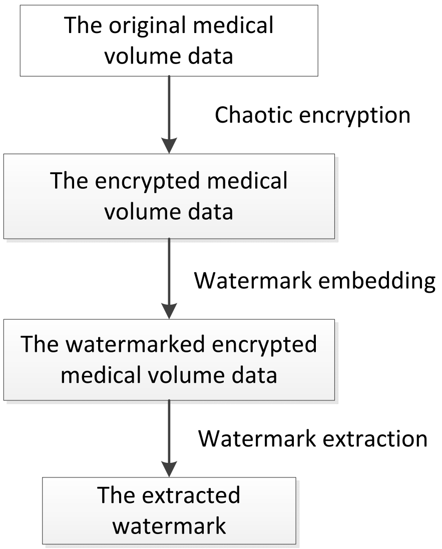

The original medical volume data
Chaotic encryption
The encrypted medical volume data
Watermark embedding
The watermarked encrypted medical volume data
Watermark extraction
The extracted watermark

Figure.1 The process of the algorithm.

A. The Encryption Of The Medical Volume Data

Step1: Use Chen's hyper-chaotic system to generate chaotic sequences;

Set the initial value (x_0, y_0, z_0, q_0) and the system parameters a, b, c, d, k and the step length Δt .Use Chen's hyper-chaotic system to generate four chaotic sequences x, y, z, q, and then the four chaotic sequences alternately form a chaotic sequence X.

Step 2: Obtain a symbol matrix;

Define a threshold function $Sign_1$. Use $Sign_1$ and the chaotic sequence X to obtain a symbol sequence. The symbol sequence transform to a symbol matrix S(i,j,k) according to the size of the medical volume data, where $1 \le n \le M \times N \times P$, $1 \le i \le M$, $1 \le j \le N$, $1 \le k \le P$.

$$Sign_1(x) = \begin{cases} 1, & x \ge 0.5 \\ -1, & x < 0.5 \end{cases} \tag{6}$$

$$S(i, j, k) = Sign_1(X(n)) \tag{7}$$

Step3: Perform 3D DCT on the medical volume data $F(i, j, k)$ and obtain a coefficient matrix $FD(i, j, k)$;

$$FD(i, j, k) = DCT3(F(i, j, k)) \tag{8}$$

Step4: The point multiplication is applied between the coefficient matrix and the symbol matrix, and the encrypted coefficient matrix $I(i, j, k)$ is obtained;

$$I(i, j, k) = FD(i, j, k) \cdot S(i, j, k) \tag{9}$$

Step5: Perform 3D IDCT on the encrypted coefficient matrix $I(i, j, k)$ and obtain the encrypted medical volume data $EF(i, j, k)$.

$$EF(i, j, k) = IDCT3(I(i, j, k)) \tag{10}$$

B. *Watermark Embedding*

Step1: Extract the feature vector of the encrypted medical volume data $V(j)$;

Perform 3D-DFT on the encrypted medical volume data $EF(i, j, k)$, and the coefficient matrix $EFD(i, j, k)$ is obtained. Select the first L/2 coefficients and define a threshold function $Sign_2$. Use $Sign_2$ to transform the real part and the imaginary part of the L/2 coefficients into a binary sequence of L bits. This binary sequence is the feature vector of the encrypted medical volume data. The process is shown in details as follows:

$$EFD(i, j, k) = DFT3(EF(i, j, k)) \tag{11}$$

$$Sign_2(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0 \end{cases} \tag{12}$$

$$\begin{aligned} V(2j-1) &= Sign_2(real(EFD(i, j, k))) \\ V(2j) &= Sign_2(imag(EFD(i, j, k))) \end{aligned} \tag{13}$$

Where $n = 1, 2, \cdots, L/2$.

Step2: A binary logical sequence $Key(i, j)$ is generated according to the binary watermark $BW(i, j)$ and the feature vector of the encrypted medical volume data $V(j)$;

$$Key(i, j) = V(j) \oplus BW(i, j) \tag{14}$$

Save $Key(i, j)$, and it will be used in the process of watermark extraction.

C. *Watermark Extraction*

Step1: Extract the feature vector of the encrypted medical volume data to be tested $V'(j)$.

Supposed the medical volume data to be tested is $EF'(i, j, k)$. Perform 3D-DFT on the $EF'(i, j, k)$ and obtain the coefficient matrix $EFD'(i, j, k)$, use the method of Step1 in section B to obtain the feature vector of encrypted medical volume data to be tested $V'(j)$;

$$EFD'(i, j, k) = DFT3(EF'(i, j, k)) \tag{15}$$

$$V'(2j-1) = Sign_2(real(EFD'(i, j, k))) \tag{16}$$

$$V'(2j) = Sign_2(imag(EFD'(i, j, k))) \tag{17}$$

Step2: Extract the watermark $BW'(i, j)$ in the encrypted medical volume data to be tested.

Perform XOR operation on $Key(i, j)$ which is generated in the process of watermark embedding and the feature vector of the medical volume data $V'(j)$,and finally the extracted watermark $BW'(i, j)$ is obtained.

$$BW'(i, j) = Key(i, j) \oplus V'(j) \tag{18}$$

Judge the ownership of the encrypted medical volume data to be tested according to the degree of correlation between $BW'(i, j)$ and $BW(i, j)$.

III. EEXPERIMENTAL RESULTS

The simulation platform is Matlab2010a, the experimental object is the volume data MRI in the database of Matlab, the size of the experimental object is $128 \times 128 \times 27$, the 3D imaging of the original medical volume data is shown in Fig.2(a), the slice (the 10th slice is selected) is shown in Fig.2(b); the Chen's hyper-chaotic system is used to encrypt the medical volume data, the state initial values $x_0 = 4.5$, $y_0 = 4.6$, $z_0 = 4.3$, $q_0 = 4.1$, the system parameters a=35, b=3, c=12, d=7, k=0.36, the step length is set at 0.001. The encrypted medical volume data is shown in Fig.2(c). The encrypted slice is shown in Fig.2 (d). This paper uses two evaluation standards as follows:

1. Normalized Cross Correlation (NC) is used to evaluate the degree of correlation between two images, the formula is as follow:

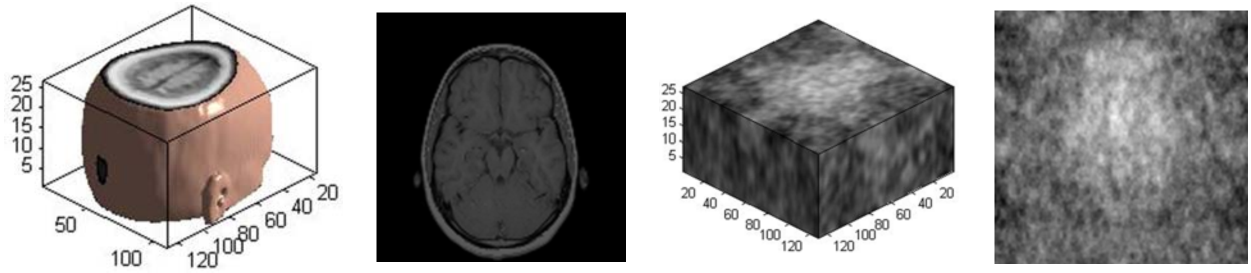
$$NC = \frac{\sum_i \sum_j W(i,j) W'(i,j)}{\sum_i \sum_j W^2(i,j)} \qquad (19)$$

Where $W(i,j)$ and $W'(i,j)$ are different images. The value of NC is larger, the degree of correlation between two images is higher.

2. Peak Signal to Noise Ratio (PSNR) is used to evaluate the image quality of the encrypted medical volume data to be tested. The formula is as follows:

$$PSNR = 10\lg\left[\frac{MNP \max_{i,j,k}(I_{(i,j,k)})^2}{\sum_i \sum_j \sum_k (I_{(i,j,k)} - I'_{(i,j,k)})^2}\right] \qquad (20)$$

Where I *(i,j,k)* the gray is value of the pixel at the coordinate *(i, j, k)* in the original encrypted medical volume data; I' *(i,j,k)* is the gray value of the pixel at the coordinate *(i, j, k)* in the encrypted medical volume data to be tested. M and are the numbers of pixels in a row and in a column respectively, P is the number of slices. The value of PSNR is larger, the image quality is better.

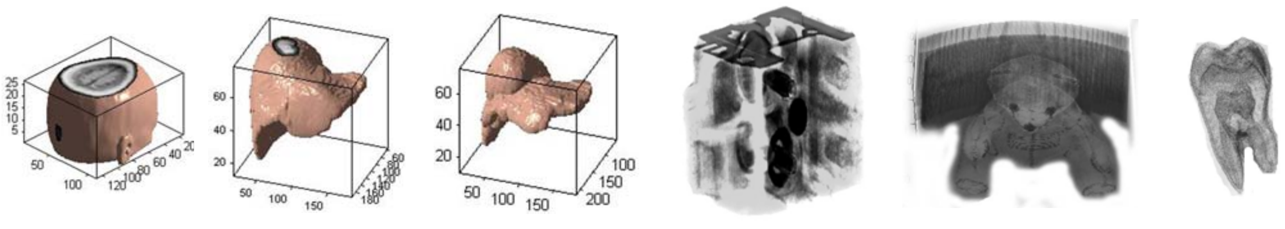


(a) (b) (c) (d)

Figure.1 The original medical volume data and its 10th slice and their encrypted version: (a)the medical volume data; (b)the 10th slice of '(a)'; (c) the encrypted medical volume data; (d)the 10th slice of '(c)'

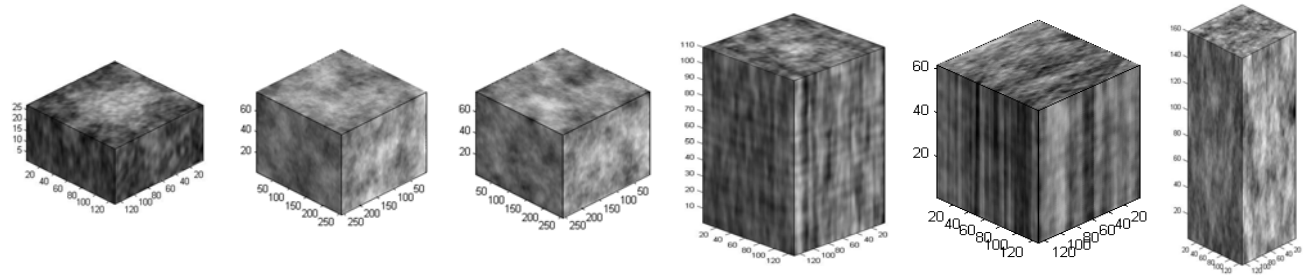
A. Collision Resistance Test

In order to testify that the feature vector extracted in the way mentioned above is an important feature of the encrypted medical volume data, this paper encrypted different volume data (shown in Fig.3). The encrypted images are shown in the Fig.4, the feature vectors of the encrypted images are extracted. The length of the feature vector is 64 bits. The NC between the feature vectors is shown in Table I, which shows that the NC between the encrypted volume data and itself is the largest, at 1.00. The NC between other encrypted volume data is much lower.



(a) (b) (c) (d) (e) (f)

Figure.2 Different 3D images: (a)head; (b)liver1; (c)liver2; (d)engine; (e)teddy bear; (f)tooth



(a) (b) (c) (d) (e) (f)

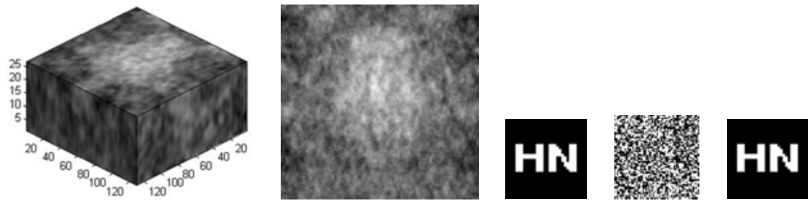
Figure.3 The corresponding encrypted images of different volume data: (a)the encrypted head; (b)the encrypted liver1; (c)the encrypted liver2; (d)the encrypted engine; (e)the encrypted teddy bear; (f)the encrypted tooth.

TABLE I. NC BETWEEN DIFFERENT ENCRYPTED 3D IMAGES

	Va	Vb	Vc	Vd	Ve	Vf
Va	1.00	-0.1409	-0.0148	0.0390	-0.0148	-0.2312
Vb	-0.1409	1.00	0.3695	0.2412	-0.1350	0.1431
Vc	-0.0148	0.3695	1.00	0.1772	-0.0089	0.2072
Vd	0.0390	0.2412	0.1772	1.00	-0.0148	0.2172
Ve	-0.0148	-0.1350	-0.0089	-0.0148	1.00	-0.0490
Vf	0.2312	0.1431	0.2072	0.2172	-0.0490	1.00

B. *Robustness Test*

The original watermark is a binary image of 64×64 bits, which is shown in Fig.5(c). In order to improve the security, the watermark is encrypted using Logistic Map. The length of the feature vector is set at 64 bits. The encrypted watermark is shown in Fig.5 (d). The original watermarked encrypted medical volume data is shown in Fig.5 (a). The 10th slice of the original watermarked encrypted medical volume data is shown in Fig.5 (b). The decryption of the extracted watermark without attacks is shown in Fig.5 (e). The NC between the decryption of the extracted watermark and the original watermark is 1.00.

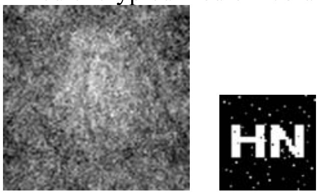


(a) (b) (c) (d) (e)

Figure.4 Experimental results: (a)The original watermarked encrypted medical volume data ; (b)The 10th slice of '(a)' image; (c)The original watermark; (d)The encrypted watermark; (e)The decryption of the extracted watermark under no attacks.

1) Gaussian Noise

Use the function imnoise () to add Gaussian noise to the original watermarked encrypted medical volume data.



(a) (b)

Figure.5 Experimental results: (a) The watermarked encrypted slice under Gaussian noise(variance=0.01); (b)The decryption of the extracted watermark (NC=0.97).

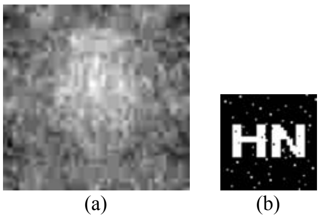
When the variance of Gaussian noise is 0.01, the decryption of the extracted watermark is shown in Fig.6 (b). From Table II, we can conclude that the proposed algorithm has good robustness against Gaussian noise attacks.

TABLE II. THE EXPERIMENTAL DATA OF ADDING GAUSSIAN NOISE

Variance (%)	1	4	8	10	15	20
PSNR(dB)	19.97	19.35	17.86	17.01	14.90	13.08
NC	0.97	0.94	0.90	0.88	0.86	0.82

2) JPEG compression

JPEG compression is applied on the watermarked encrypted medical volume data. When the compression quality is 20%, the decryption of the extracted watermark is shown in Fig.7 (b). From Table III, we can conclude that the proposed algorithm has strong robustness against JPEG compression.



(a) (b)

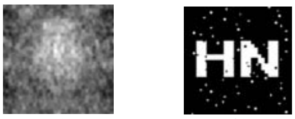
Figure.6 Experimental results: (a) The watermarked encrypted slice under JPEG compression (compression quality=20%); (b) The decryption of the extracted watermark (NC=0.97).

TABLE III. THE EXPERIMENTAL DATA OF JPEG COMPRESSION

Compression Quality(%)	**4**	**8**	**12**	**20**	**25**	**30**
PSNR(dB)	6.55	6.55	6.55	6.55	6.55	6.55
NC	0.85	0.91	0.92	0.97	0.98	1.00

3) Resizing

Resize the watermarked encrypted medical volume data from size 128×128×27 to 64×64×27. The decryption of the extracted watermark is shown in Fig.8 (b). From Table IV, we can conclude that the proposed algorithm has strong robustness against resizing attacks.



(a) (b)

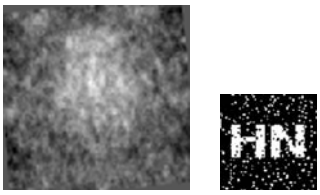
Figure.7 Experimental results: (a) The watermarked encrypted slice after resizing (compression quality=20%); (b) The decryption of the extracted watermark (NC=0.97).

TABLE IV. THE EXPERIMENTAL DATA OF RESIZING

Resizing Factor	**0.5**	**1.2**	**2**	**2.5**	**5**
NC	0.97	0.98	0.95	0.95	0.95

4) Rotation

When the watermarked encrypted medical volume data is rotated 5° counterclockwise. The decryption of the extracted watermark is shown in Fig.9 (b). From Table V, we can conclude that the proposed algorithm can resist rotation attacks within a small range.



(a) (b)

Figure.8 Experimental results: (a) The watermarked encrypted slice is rotated 5° counterclockwise; (b) The decryption of the extracted watermark (NC=0.84).

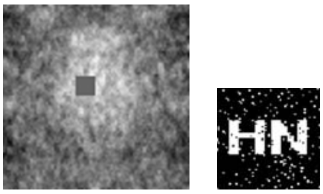
TABLE V. THE EXPERIMENTAL DATA OF ROTATION

Rotational Degree	**1°**	**3°**	**5°**	**8°**	**10°**
NC	0.97	0.87	0.84	0.74	0.68

Note: The minus sign denotes C.C.W., no sign denotes C.W.

5) Cropping

Random 1% of the volume of the watermarked encrypted medical volume data is cropped. The decryption of the extracted watermark is shown in Fig.10 (b). From Table VI, we can conclude that the proposed algorithm has good robustness against random cropping attacks.



(a) (b)

Figure.9 Experimental results: (a) The watermarked encrypted slice under random cropping (cropping percentage=1%); (b) The decryption of the extracted watermark (NC=0.91).

TABLE VI. THE EXPERIMENTAL DATA OF CROPPING

Cropping Percentage (%)	**1**	**2**	**3**	**4**	**5**
PSNR(dB)	25.85	19.85	17.41	16.58	16.18
NC	0.91	0.82	0.68	0.60	0.57

IV. CONCLUSION

The medical volume data is encrypted using Chen's hyper-chaotic system to attain large key space and efficient encryption. Through experimenting on different objects, it indicates that the feature vector extracted in the encrypted volume data in the proposed method has good uniqueness and it can represent the encrypted volume data. In addition, this algorithm realizes watermark embedding and watermark extraction in the encrypted domain. The experiment results show that the proposed watermarking algorithm can resist JPEG compression and resizing in a large degree. It can also resist Gaussian noise and cropping in a modest degree, alongside resisting rotation in a smaller degree. The proposed algorithm uses zero-watermarking technique, which doesn't change the image data of the medical volume data and well protects the medical volume data.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGMENT

This work is partly supported by the National Natural Science Foundation of China (NO:61263033 and NO:61363007), and by the International Science and Technology Cooperation Project of Hainan (NO:

KJHZ2015-04 and NO:KJHZ2014-16) and the Institutions of Higher Learning Scientific Research Special Project of Hainan(Hnkyzx2014-2).

REFERENCES

[1] T. Bianchi, A. Piva, M. Barni. "On the Implementation of the Discrete Fourier Transform in the Encrypted Domain", IEEE Transactions Information Forensics and Security, vol.4, no.1, pp.86-97, March 2009.

[2] P. Zheng, J. Huang, "Discrete Wavelet Transform and Data Expansion Reduction in Homomorphic Encrypted Domain", IEEE Transactions on Image Processing, vol.22, no.6, pp. 2455-2468, June 2013.

[3] Hsu, C. Lu, S. Pei, "Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT", IEEE Transactions on Image Processing, vol.21, no.11, pp. 4593-4607, November 2012.

[4] P. Zheng, J. Huang, "Walsh-Hadamard Transform in the Homomorphic Encrypted Domain and Its Application in Image Watermarking", Information Hiding, Springer, vol.7692, pp. 240-254, 2013.

[5] B. Zhao, W. Kou, H. Li, et al., "Effective watermarking scheme in the encrypted domain for buyer-seller watermarking protocol", Information Sciences, vol.180, pp. 4672-4684 , August 2010.

[6] Subramanyam, S. Emmanuel, "Robust Watermarking of Compressed and Encrypted JPEG2000 Image", IEEE Transactions on Multimedia, vol.14, no.3, pp.703-716, June 2012.

[7] T. Bianchi, A. Piva, M. Barni, "Composite Signal Representation for Fast and Storage-Efficient Processing of Encrypted Signals", IEEE Transactions Information Forensics and Security, vol.5, no.1, pp. 180-187, 2010.

[8] M. Barni, P. Failla, R. Lazzeretti, et al., "Privacy-Preserving ECG Classification with Branching Programs and Neural Networks", IEEE Transactions Information Forensics and Security, vol.6, no.2, pp. 452-468, 2011.

[9] T. Bianchi, S. Turchi, A. Piva, et al., "Implementing FingerCode-Based Identity Matching in the Encrypted Domain", IEEE Workshop Biometric Measurements and Systems for Security and Medical Applications, pp. 15-21 , 2010.

[10] W. Lu, A. Swaminathan, A. L. Varna, et al., "Enabling Search Over Encrypted Multimedia Databases", Proc. SPIE. vol.7254, pp. 1-11, January 2009.

[11] S. Lian, Z. Liu, R. Zhen, et al., "Commutative Encryption and Watermarking in Video Compression", IEEE Transactions Circuits and System for Video Technology, vol.17, no.6, pp.774-778, 2007.

[12] R. Schmitz, S. Li, D. Tao, "A New Approach to Commutative Watermarking-Encryption", Communications and Multimedia Security, Lecture Notes in Computer Science Springer, Berlin Heidelberg, vol.7394, pp. 117-130, May 2012.