A Novel Fragile Watermark Algorithm for Detecting Speech Tampering

Xiaoting Zhao[1, 2, a], Wenhuan Lu[1, 2, b,*], Jianguo Wei[1, 2, c]

[1]School of Computer Software, Tianjin University, Tianjin, 300072, China

[2]Tianjin Key Lab Cognitive Computing and Application, Tianjin University, Tianjin, 300072, China

[a]email: zhaoxiaoting@tju.edu.cn, [b]email: wenhuan@tju.edu.cn, [c]email: jianguo@tju.edu.cn

Keywords: Fragile Watermark; Speech Authentication; Alteration Detection

Abstract. In this paper, we propose a fragile watermark scheme for detecting the tampering or hostile attack of speech signal. Watermark block is generated by Gaussian Random operation on a null matrix, then the watermark is embedded into the original signal which is processed by DCT (discrete cosine transform). During the signal detection processing, matrix to vector method is adopted, in virtue of the differential matrix row and column, a threshold value can be calculated. With the comparison between the watermarked signal and threshold, the tampering location can be detected and the watermark scheme can achieve valid identification on tampering, with the accuracy rate reached up to 95%.

Introduction

With the advent of information era, computer network gets a rapid development, the utilization of multimedia data has become gradually common in our daily life. While multimedia data can bring convenience, the new challenges about the field of information security have appeared. Since multimedia data can be tampered easily, it may be distorted during the network transmission so the integrity authentication of multimedia data becomes an urgent issue. The appearance of digital watermark brings the hope to solve the problem, it provides a kind of powerful tool about data authentication and verification. The fragile watermark can localize the tampering area, and distinguish behaviors that are hostile attacks. Varies of watermarking schemes which hammer at detecting multimedia tampering have been developed.

Many related researches have paid attention to the protection of multimedia data, especially in image protection, about different coefficients which are adopted in representing. Wang S employed discrete wavelet transform (DWT) [1], Jiri Fridrich employed discrete cosine transform (DCT) coefficients [2], and Eugene T. Lin adapted semi-fragile watermarks on image tampering detection [3], Zhang Xinpeng embedded watermarks in least significant bits combined with discrete cosine transform [4].

The researches on digital watermark in speech signal is less than image domain, as human auditory system is more sensitive than visual system, so after the embedding of watermark, the signal requires higher audibility. As speech signal security is very important, an air traffic control system of speech watermarking is proposed in [5], in which the pilot messages embedded flight information. Some robust watermark algorithms have been used in the field of copyright protection in speech signal [6-9]. Also watermark embedded into least significant bits are used which estimate the origin signal by solving a linear equation with least square QR-factorization method. Previous work often embed watermark information in different area of the signal, when parts of the signals tampered, the rest information of watermark can localize the damaged part [10]. The above researches devote to tampering detection and signal recovery, while in this study, we focus on the detection and the valid identification of compression on speech signal, for sometimes we don't treat compression as a hostile attack.

In this paper, we extend a watermark scheme to speech signal domain. We adopt a null matrix as our original material, the scheme avoids the watermark data waste problem. The original material is processed by Gaussian Random to generate the watermark, then we take advantage of human auditory system characteristic, sensitive to low frequency. We embed watermark into the

corresponding low frequency regions. When the watermarked signal is tampered or compressed, we can compare the coefficient we defined with the help of differential matrix ranks. The scheme can detect the tampering region accurately.

Watermark Embedding

The watermark algorithm consists of two parts, the watermark generation and the watermark embedding, in this study, we employ embedding method on the DCT domain of the watermark and the original speech signal. We partition the speech signal into non-overlap frames, and each frame is the fundamental unit during the experiment.

Assume a frame composed of n sampling sites, as the watermark embedding will be proposed on DCT domain, the watermark is constructed in the DCT domain can generate a smooth watermark. So we set n equal to 169, which is equal to 13*13, so each frame can be represented by a 13*13 matrix. The discrete cosine transform will assemble the high frequency part to the top left corner while the low frequency to the bottom right corner, we use Zig-Zag algorithm, which is showed in Fig.1 to handle the 169 sampling sites. The Zig-Zag algorithm convert one dimension speech signal to two dimension matrix, then the zero-mean, unit variance Gaussian Random will be proposed on the top left corner 91 sampling sites as the Fig.2 shows. After the random process, inverse discrete cosine transform (IDCT) will be performed to produce a spatial domain watermark W, which will be embedded into the original signal later.

We segment the original speech signal into the same size frame which also contains 169 sampling sites, so the vector can be converted into a 13*13 matrix by Zig-Zag algorithm. Assume a frame of speech signal can be represented X which has been convert into a 13*13 matrix, and the embedded frame is Y:

$$Y = X + \delta W \tag{1}$$

Where σ represents the strength of the watermark

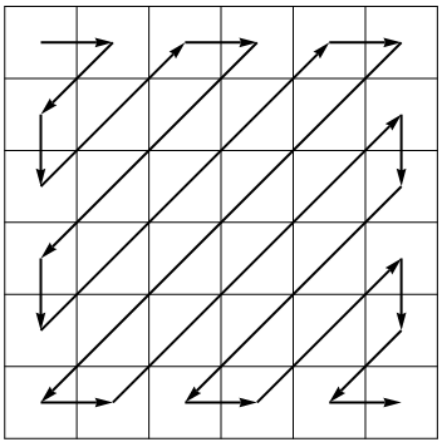


Fig.1. Zig-Zag transform

1	2	6	7	15	16	28	29	45	46	66	67	91
3	5	8	14	17	27	30	44	47	65	68	90	92
4	9	13	18	26	31	43	48	64	69	89	93	114
10	12	19	25	32	42	49	63	70	88	94	113	115
11	20	24	33	41	50	62	71	87	95	112	116	133
21	23	34	40	51	61	72	86	96	111	117	132	134
22	35	39	52	60	73	85	97	110	118	131	135	148
36	38	53	59	74	84	98	109	119	130	136	147	149
37	54	58	75	83	99	108	120	129	137	146	150	159
55	57	76	82	100	107	121	128	138	145	151	158	160
56	77	81	101	106	122	127	139	144	152	157	161	166
78	80	102	105	123	126	140	143	153	156	162	165	167
79	103	104	124	125	141	142	154	155	163	164	168	169

Fig.2. Watermark embedded location

Watermark Detection

The process of detection is performed on a frame by frame scheme to localize the parts that are likely altered, the detector is based on the difference of adjacent sampling sites value in the spatial domain. We introduce two matrix operations now, the differential matrix row and the differential column, we define operator ΔROW and operator ΔCOL respectively represent the above two operations.

The ΔROW and ΔCOL can be calculated with the following formula:

$$\Delta ROW(B) = \begin{cases} B(x,y) - B(x+1,y) & \text{if } x \in \{1,2,3,...,12\} \\ 0 & \text{if } x = 13 \end{cases} \tag{2}$$

$$\Delta COL(B) = \begin{cases} B(x,y) - B(x,y+1) & \text{if } y \in \{1,2,3,...,12\} \\ 0 & \text{if } y = 13 \end{cases} \tag{3}$$

Where the B represents an arbitrary frame that has been converted into a 13*13 matrix.

Set T_b as the frame of the speech signal which is going to be tested, and W_b denotes the corresponding watermark block. And let T_b^* be the reconstructed matrix after doing the differential matrix row and column operations, and let W_b^* the corresponding watermark block, the calculate process will be shown in the equation (4) and (5) below:

$$T_b^* = \{\ \Delta COL(T(x,y)) \mid \Delta ROW(T(x,y))\ \} \tag{4}$$

$$W_b^* = \{\ \Delta COL(W(x,y)) \mid \Delta ROW(W(x,y))\ \} \tag{5}$$

As T_b^* and W_b^* respectively involves $2n*(n-1)$ points, and the spatial correlation of the frame only has n*n points, so the detection of the frame is then the correlation ρ calculated in the formula (6):

$$\rho = \frac{T_b^* \cdot W_b^*}{\sqrt{(W_b^* \cdot W_b^*)(T_b^* \cdot T_b^*)}} \tag{6}$$

As ρ has been obtained for each frame (formula 6 above), we can use it to compare with the threshold value T, which should be determined by experiment. Usually, the selection of the threshold should satisfy that more than 98% of the speech frame value should larger than T after embedding the watermark, and the standard of the outcome is chosen as the follow:

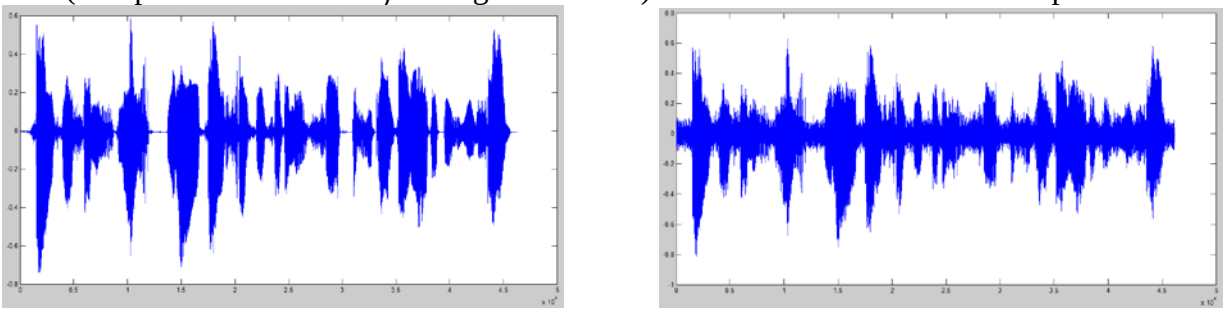
$\rho \leqslant T$: Frame is authentic.

$\rho > T$: Frame has been altered.

Evaluation of the watermark scheme

To evaluate the watermark scheme, we use the speech signal which is from Chinese High-tech R&D (CASIA-863) Program Speech Database. The speech signal we used consists of 273 frames which can obtain 273 ρ values, by statistical analysis, we can get that 272 values is less than 0.29 with strength of the watermark σ is set 0.5 , the percentage of the average threshold reaches up to 99.6%. So the 0.29 is selected as the threshold value in the experiment, and each frame's ρ will be compared with the threshold to conduct certification test.

Damage each frames from the start to the end, and 269 frames can be detected in the 272 valid frames (except the frame whose ρ is larger than 0.29). The detection rate can reach up to 98.9%.



(a) Original speech signal (b) Watermarked speech signal

Fig.3. Speech signal before and after embedding watermark

To check the perception of the watermarked speech signal, we introduce ABX auditory recognition experiment on 10 subjects from 20 to 24 years old. The ABX experiment means every subject knows there are two existing samples A and B, the original and the watermarked speech signal. Then an X unknown sample will be provided to the subject to recognize whether it's A or B. The experiment chooses 150 original speech signals and 150 watermarked speech signals from the former 150 samples. Every subject randomly selects 20 signals to do the experiment. 10 subjects have accepted total 200 experiments, and the accuracy reaches up to 55.1%. When the result is close to 50%, we consider the watermark which is embedded into the speech signal is no perceptible.

Conclusion

The proposed scheme of watermark achieve a high detection rate on tampering the watermarked

signal, which reaches more than 95%. The advantage of the scheme is the accurate positioning of the tampering region in the speech signal, as the frame is as small as the 169 sampling sites. The detector is based on correlation of spatial-domain difference takes advantage of the fact that most parts of the speech signal is relatively smooth. In the future work, semi-fragile watermark deserves to be paid more attention to research as the hostile attack should be distinguished with other kindly operations such as compression. Also the self-recovery of damaged signal will become a key research in the future.

Acknowledgement

In this paper, the research was supported in part by the grants from the National Natural Science Foundation of China (Program No. 61304250 and 61471259).

References

[1] Wang S, Zheng D, Zhao J, et al. Adaptive Watermarking and Tree Structure Based Image Quality Estimation [J]. IEEE Transactions on Multimedia, 2011, 16(2):74-77.

[2] Fridrich J, Goljan M. Images with self-correcting capabilities[C]. International Conference on Image Processing, 1999. ICIP 99. Proceedings. IEEE, 1999:792-796 vol.3.

[3] Lin E T, Delp E J. Detection of image alterations using semifragile watermarks [J]. Proceedings of SPIE - The International Society for Optical Engineering, 2000, 3971:152-163.

[4] Zhang X, Qian Z, Ren Y, et al. Watermarking With Flexible Self-Recovery Quality Based on Compressive Sensing and Compositive Reconstruction [J]. IEEE Transactions on Information Forensics & Security, 2011, 6(4):1223-1232.

[5] Hering H, Hagmuller M, Kubin G. Safety and security increase for air traffic management through unnoticeable watermark aircraft identification tag transmitted with the VHF voice communication [C], Digital Avionics Systems Conference, 2003. 2003:4.E.2 - 41-10.

[6] Cox I J, Miller M L. The First 50 Years of Electronic Watermarking [J]. Eurasip Journal on Applied Signal Processing, 2002, 2002(2):126-132.

[7] H. Ozer, B. Sank and N. Memon, An SVD-based audio watermarking technique, in Proceedings of the 7th ACM workshop multimedia security [C], pp. 51-56, 2005.

[8] El-Samie F E A. An efficient singular value decomposition algorithm for digital audio watermarking [J]. International Journal of Speech Technology, 2009, 12(1):27-45.

[9] Vivekananda B K, Sengupta I, Das A. An adaptive audio watermarking based on the singular value decomposition in the wavelet domain [J]. Digital Signal Processing, 2010, 20(6):1547-1558.

[10] Qiang S, Wang J, Zhang H. Tamper Detection and Self-Recovery of Image Based on Self-Embedding[C]// Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on. IEEE, 2009:76-79