Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.532 – 539

RESEARCH ARTICLE

Multicarrier Iterative Generalized Least Squares Data Extraction in Digital Images

Srinivas M[1], Dhana Lakshmi[2]

[1]M.Tech Student , Department of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, India

Mukthasrinivas1989@gmail.com

[2]Assistant Professor, Department of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, India

ABSTRACT: Information hiding techniques are become important in a number of application areas. In the field of Data Communication security of information is the major problem. The transmission of information via the Internet may expose it to detect and theft. So, data embedding technologies are developed to provide personal privacy, commercial and national security interests. In this work we consider the problem of extracting blindly data embedded over a wide band in a spectrum (transform) domain of a digital medium (image, audio, video).We develop a novel multicarrier/signature iterative generalized least-squares (M-IGLS) core procedure to seek unknown data hidden in hosts via multicarrier spread-spectrum embedding. Here the original host and the embedding carriers both are assumed as not available. Experimental results shows that the proposed algorithm can achieve recovery probability of error close to what may be attained with known embedding carriers and host autocorrelation matrix.

I. INTRODUCTION

In the field of Data Communication, security-issues have the major problem. The transmission of information via the Internet may uncover it to detect and theft. In the field of information technology Digital data embedding in digital media is rapidly growing commercial as well as national security interest. The main Applications of data hiding are annotation, copyright-marking, and watermarking, single-stream media merging (text, audio, image) and Steganography [1]. In annotation the secondary data are embedded into digital multimedia to provide a way to deliver side information for various purposes such as copyright-marking it shows the ownership by act as permanent "iron branding"; fragile watermarking may be intended to detect future tampering; hidden low-probability-to-detect (LPD) watermarking mainly used to identify the confidential data validation and digital fingerprinting for tracing purposes [2]–[4]. Steganography or Covert communication, steganography is a Greek Latin word which means "covered writing" , it is the process of hiding the data under a cover medium or host such as image, audio, or video to establish secret communication between trusting parties and hide the existence of embedded data [5]–[9].

This paper mainly focus on the blind recovery of secret data hidden in medium hosts via spread-spectrum (DS-SS) transform domain multicarrier/signature direct-sequence Embedding [13]–[20]. The original host and the embedding carriers (signatures or spreading sequences) both are assumed to be not known (fully blind data extraction). This blind hidden data extraction problem has also been referred to as "Watermarked content Only Attack" (WOA) in the watermarking security context [21]–[24].. In blind extraction of SS embedded data, the unknown host acts as a source of interference/disturbance to the data to be recovered and, in a way, the problem parallels blind signal separation (BSS) applications as they arise in the fields of array processing, biomedical signal processing, and code-division multiple-access (CDMA) communication systems. Under the assumption that the embedded secret messages are independent identically distributed (i.i.d.) random sequences and independent to the cover host, independent component analysis (ICA) may be utilized to pursue hidden data extraction [24], [25]. However, ICA-based BSS algorithms are not effective in the presence of correlated signal interference as is the case in SS multimedia embedding and degrade rapidly as the dimension of the carrier (signature) decreases relative to the message size. In [19], an iterative generalized least squares (IGLS) procedure was developed to blindly recover unknown messages hidden in image hosts via SS embedding. This algorithm has low complexity and strong recovery performance. But the scheme is designed solely for single-carrier SS embedding where messages are hidden with one signature only and is not generalizable to the multicarrier case. Realistically, an embedded would favor multicarrier SS transform-domain embedding to increase security and/or payload rate. In this paper, we develop a novel multicarrier iterative generalized least squares (M-IGLS) algorithm for SS hidden data extraction for improved recovery performance, in particular for small hidden messages that pose the greatest challenge, experimental studies indicate that a few independent M-IGLS reinitializations and executions on the host can lead to hidden data recovery with probability of error close to what may be attained with known embedding carriers and known original host autocorrelation matrix. The proposed algorithm can be treated as a tool to test security robustness of SS data hiding schemes.

II. EMBEDDING AND EXTRACTION OF MULTICARRIER SS

Problem Formulation:

Let consider a hosts image $H \in \mathcal{M}^{N_1 \times N_2}$ where $\mathcal{M}$ is the finite image alphabet and $N_1 \times N_2$ is the image size in pixels. The image is partitioned into $\mathcal{M}$ local no overlapping blocks of size $\frac{N_1 N_2}{M}$ but without loss of generality. Each block, $H_1, H_2, \dots, H_{M,}$ is to carry hidden information bits (kM bits total image payload). Here embedding is performed in a 2-D transform domain (such as the Discrete Cosine Transform (DCT) and Wavelet Transform (WT), etc.). After transform calculation and vectorization (for example by conventional zig-zag scanning), we obtain $T(H_m) \in R^{\frac{N_1 N_2}{M}}$ $m = 1,2, \dots, M$. From the transform domain vectors $T(H_m)$
we choose a fixed subset of $L \leq \frac{N_1 N_2}{M}$ coefficients (bins) to form the final host vectors $X(m) \epsilon R^L, m = 1,2, \dots, M$.It is common and appropriate to avoid the dc coefficient (if applicable) due to high perceptual sensitivity in changes of the dc value. For our developments the autocorrelation matrix of the host data X is an important statistical quantity it defined as

$$R_x \triangleq E\{XX^T\} = \frac{1}{M}\sum_{m=1}^{M} X(m)X(m)X(m)^T$$

Generally it is easily verified that R_x is not constant- value diagonal or "white" in field language.

Multicarrier SS Embedding:

We consider K distinct message bit sequences
$\{b_k(1), b_k(2), \dots, b_k(M)\}, k = 1,2, \dots, K, b_k(m) \in \{\pm 1\}, m = 1, \dots, M$ and each of length M bits. The K message sequences may be to be delivered to Kdistinct corresponding recipients or they are just K portions of one large message sequence to be transmitted to one recipient. In particular, the m^{th} bit from each of the K sequences,$b_1(m), \dots, b_k(m)$,is simultaneously hidden in the m^{th} transform-domain host vector $X(m)$ via additive SS embedding by means of K spreading sequences (carriers) $s_k \in R^L, \| s_k \| = 1, k = 1,2, \dots, K$,

$$y(m) = \sum_{k=1}^{k} A_k b_k(m) s_k + X(m) + n(m), m = 1,2, \dots, M, \qquad (1)$$

With corresponding amplitudes. For the sake of generality, n(m) represents potential external white Gaussian noise1 of mean 0 and autocorrelation matrix $\sigma_n^2 I_L$, $\sigma_n^2 > 0$ It is assumed that $b_k(m)$ behave as equi-probable binary random variables that are independent in m message bit sequence) and k (across messages). The contribution of each individual embedded message bit b_k to the composite signal is $A_k b_k s_k$ and the block mean-squared distortion to the original host data x due to the embedded k message alone is

$$D_k = E\{\| A_k b_k s_k \|^2\} = A_k^2, k = 1,2, \ldots . K. \quad (2)$$

Under statistical independence of messages, the block mean squared distortion of the original image due to the total, multimessage, insertion of data is

$D = \sum_{K=1}^{k} A_k^2$

The intended recipient of the th message with knowledge of the k^{th}carrier s_k can perform embedded bit recovery by looking at the sign of the output of the minimum-mean-square-error(MMSE)filter

$$W_{MMSE,k} = R_y^{-1} s_k$$

$$\hat{b_k}(m) = \text{sgn}\{W_{MMSE,k}^T \, y(m)\} = \text{sgn}\{s_k^T R_y^{-1} y(m)\} \quad (3)$$

Where R_y is the autocorrelation matrix of the host-plus-data plus- noise vectors

$$R_y \triangleq E\{yy^T\} = R_x + \sum_{k=1}^{K} A_k^2 s_k s_k^T + \sigma_n^2 I_L \quad (4)$$

The autocorrelation matrix R_y can be estimated by sample averaging over the set of M received vectors

$$\{y(m)\}M_{m=1,} R_y = \frac{1}{M} \sum_{m=1}^{M} y(m) y(m)^T$$

.

Using $\hat{R_y}$ in (3) in place of R_y , we obtain what is known as the sample-matrix-inversion MMSE (SMI-MMSE) detector implementation [25].

Formulation of the Extraction Problem:

From a given host image to blindly extract spread-spectrum embedded data, first the analyst needs convert the host to observation vectors of the form of $y(m), m = 1,2, \ldots, M$ in (1). This requires knowledge of partition, transform domain, subset of coefficients, and number of carriers used by the embedder. The host image partition (and block size $\frac{N_1 N_2}{M}$ in our notation) may be estimated by neighboring- pixels difference techniques as in [27]. Regarding the subset of coefficients used in embedding, the conservative approach is to assume that all coefficients are used, except may be the dc value, and set accordingly. Finally, determination of the transform domain used in embedding seems to be a hurdle not yet tackled by current research. The natural approach would be to consider individually and exhaustively one transform at a time starting from the most common (for example, 2D-DCT, common wavelet transforms, and so on). This paper focus the technical presentation solely after the point that the analyst obtains transform-domain observations in the form of in y(m) (1), upon performing appropriate image parted3rd3dsw2sw2ition and transform calculation. We denote the combined "disturbance" to the hidden data (host plus noise) by partition and transform calculation. We denote the combined "disturbance" to the hidden data (host plus noise) by $z(m) \triangleq X(m) + n(m)$ and rewrite SS embedding by (1) as

$$y(m) = \sum_{k=1}^{K} A_k b_k(m) s_k + z(m), m = 1, \ldots . , M, \quad (5)$$

Where z(m) is modeled as a sequence of zero-mean (without loss of generality) vectors with autocovariancematrix

$$R_z = E\{zz^T\} = R_x + \sigma_n^2 I. \text{Let } V_k \triangleq A_k S_k \in R^L, k = 1, \ldots . , K$$

be the amplitude-including embedding carriers. Then, we can further reformulate SS embedding as

$$y(m) = \sum_{k=1}^{K} b_k(m)V_k + z(m) \quad (6)$$

$$= Vb(m) + z(m), m = 1, \ldots, M, \quad \textbf{(7)}$$

where
$V \triangleq \quad [V_{1,\ldots,} V_K] \in R^{L\times K}$ is the amplitude-including carrier matrix and

$b(m) \in \{\pm 1\}^{K\times u}$ is the vector of bits embedded in the m^{th} host block. For notational simplicity, we can write the whole observation data in the form of one matrix

$$Y = VB + Z \quad (8)$$

,
Where

$$Y \triangleq [y(1)\, y(2) \ldots y(M)] \in R^{L\times M}, B \triangleq [b(1)\, b(2) \ldots b(M)] \in \{\pm 1\}^{K\times M}$$

and

$$Z \triangleq [z(1)\, z(2) \ldots z(M)] \in R^{L\times M}$$

Our main objective is to blindly extract the unknown hidden data B from the observation matrix Y without prior knowledge of the embedding carriers s_k and amplitudes A_k $k = 1, \ldots, K$, in $V = [A_1 s_1, \ldots, A_K s_K]$ or the host medium itself $X(1), \ldots, X(M)$ in

$$Z = [X(1) + n(1), \ldots, X(M) + n(M)]$$

III. EXTRACTION OF HIDDEN DATA

If Z were to be modeled as Gaussian distributed, the joint maximum-likelihood (ML) estimator of and decoder of V and decoder of B would be

$$\widehat{V}, \widehat{B} = \arg \min_{\substack{B\in\{\pm 1\}^{K\times M} \\ V\in R^{L\times K}}} \| R_Z^{-\frac{1}{2}}(Y - VB) \|_F^2$$

Where multiplication by $R_Z^{-\frac{1}{2}}$ can be interpreted as prewhitening of the compound observation data. If Gaussianity of Z is not to be invoked, then (9) can be simply referred to as the joint generalized least-squares (GLS) solution2 of V and B .The global GLS-optimal message matrix $\widehat{B}$ in (9) can be computed independently of $\widehat{V}$ by exhaustive search over all possible choices under the criterion function

$$\| R_Z^{-\frac{1}{2}} YP \perp B \|_F^2$$

$$\widehat{B} = \arg\min_{B\in\{\pm 1\}^{K\times M}} \| R_Z^{-\frac{1}{2}} YP \perp B \|_F^2 \quad (10)$$

Where $P \perp B \triangleq I - B^T(BB^T)^{-1}B$

Multicarrier Iterative Generalized Least-Squares

Data Extraction:

Unacceptable and attempt to reach a quality approximation of the solution of (10) (or (9), to that Respect) by alternating generalized least-squares estimates of V and B, iteratively, as described below. Pretend B is known.

The generalized least-squares estimate of V is

$$\widehat{V}_{GLS} = \arg \min_{V \in R^{L \times K}} \| R_Z^{-\frac{1}{2}} (Y - VB) \|_F^2$$
$$= YB^T (BB^T)^{-1} \quad (11)$$

Pretend, in turn, that is V is known. Then, the least-squares estimate of B over the real field is

$$\widehat{B}_{GLS}^{real} = \arg \min_{B \in R^{K \times M}} \| R_Z^{-\frac{1}{2}} (Y - VB) \|_F^2$$
$$= (V^T R_Z^{-1} V)^{-1} V^T R_Z^{-1} Y \quad (12)$$

Observing that

$$(V^T R_Z^{-1} V)^{-1} V^T R_Z^{-1} = \left(V^T R_y^{-1} V\right)^{-1} V^T R_y^{-1} \quad (13)$$

We rewrite

$$\widehat{B}_{GLS}^{real} = \left(V^T R_y^{-1} V\right)^{-1} V^T R_y^{-1} Y$$

and suggest the approximate binary message solution

$$\widehat{B}_{GLS}^{real} \widehat{V}_{GLS} = \arg \min_{B \in \{\pm\}^{K \times M}} \| R_Z^{-\frac{1}{2}} (Y - VB) \|_F^2$$
$$\simeq \operatorname{sgn}\left\{\left(V^T R_y^{-1} Y\right) V^T R_Z^{-1} Y\right\} \quad (15)$$

The multicarrier iterative generalized least-squares (M-IGLS) procedure suggested by the two equations (11) and (15) is now straightforward. Initialize B arbitrarily and alternate iteratively between (11) and (15) to obtain at each step conditionally generalized least squares estimates of one matrix parameter given the other. Stop when convergence is observed. Notice that (15) utilizes knowledge of the autocorrelation matrix R_y , which can be estimated by sample averaging over the received data observations

$$\widehat{R}_y = \frac{1}{M} \sum_{m=1}^{M} y(m) y(m)^T$$

The M-IGLS extraction algorithm Is $O(2K^3 + 2LMK + K^2(3L + M) + L^2K$ Summarized in Table I. Superscripts denote iteration index. The computational complexity of each iteration of the M-IGLS algorithms and, experimentally, the number of iterations executed is between20 and 50 in general For the sake of mathematical accuracy, we recall that in least squares there is always a symbol sign (phase in complex domains) ambiguity when joint data extraction and carrier estimations pursued (i.e., data bits on carrier $b_{K \in \{\pm\}^M}$ have the same least-squares error with data bits on carrier $s_{k=R^L - S_{K,K=1,\ldots,K}}$. The sign-ambiguity problem can be overcome with a few known or guessed data symbols for supervised sign correction3 . Moreover, in a multicarrier least-squares scenario as the one that we face herein, the index association remains unresolved (i.e., given a recovered (message, carrier) pair (b,s), the corresponding index $k \in \{1, \ldots, K\}$ in (1) cannot be obtained). To the extent that the application of the work presented in this paper is to simply extract blindly the embedded bits with the least possible errors, the carrier indexing problem is not dealt with any further. Returning to the proposed data extraction algorithm, we understand that with arbitrary initialization convergence of the M-IGLS procedure described in Table I to the optimal GLS solution of (9) is not guaranteed in general. Extensive experimentation with the algorithm in Table I indicates that, for sufficiently long messages hidden by each carrier (M=4 Kbits or more, for example), satisfactory quality message decisions B can be directly obtained. However, when the message size is small, M-IGLS may very well converge to inappropriate points/solutions. The quality (generalized-least-squares fit) of the end convergence point depends heavily on the initialization point and arbitrary initialization—which at first sight is unavoidable for blind data extraction—offers little assurance that the iterative scheme will lead us to appropriate, "reliable" (close to minimal generalized least-squares fit) solutions. To that respect, re initialization and re execution of the M-IGLS procedure, say P times, is always possible. To assess which of the returned solutions, say $\{\widehat{V}_1 \widehat{B}_1), \ldots . (\widehat{V}_p \widehat{B}_1)$, has superior generalized- least-squares fit, we simply feed $(\widehat{V}_i \widehat{B}_i)$ to (9) (using R_y in place of R_z) and choose

$$\widehat{V}_{final}, \widehat{B}_{final} = \min_{arg\ (V,B)\in\{\widehat{V}_1\widehat{B}_1),\ldots.(\widehat{V}_p\widehat{B}_1)\}} \| R_Z^{-\frac{1}{2}}(Y - VB \|_F^2 \quad (16)$$

The computational complexity of the P-times reinitialized M-IGLS is, of course

$$(PD(2K^3 + 2LMK + K^2(3L + M) + L^2K))$$

,

Where D represents the number of internal iterations in Table I.

IV. RESULTS

Fig: 512*512 Plane image

Fig: 512*512 Plane image BER

V. CONCLUSION

In this paper we considered the problem of blindly extracting unknown messages hidden in image hosts via multicarrier/signature spread-spectrum embedding In this neither the original host nor the embedding carriers are assumed available we developed a low complexity multicarrier iterative generalized least-squares (M-IGLS) core algorithm. Experimental results showed that M-IGLS can achieve probability of error rather close to what may be attained with known embedding signatures and known original host autocorrelation matrix and presents itself as an effective countermeasure to conventional SS data embedding/ hiding.

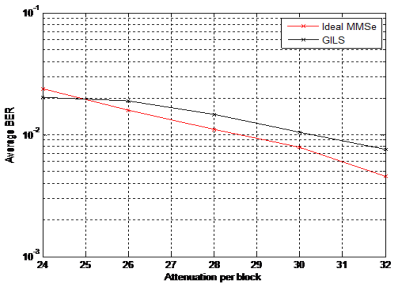

Fig: 256*256 Plane image

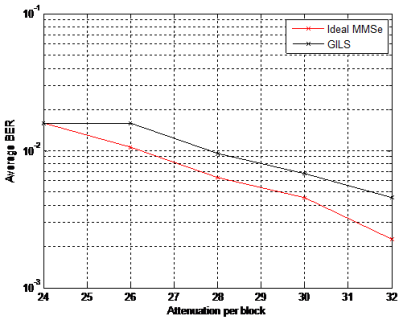

Fig: 256*256 plane image BER

REFERENCES

[1] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding: A survey," Proc. IEEE, Special Issue on Identification and Protection of Multimedia Information, vol. 87, no. 7, pp. 1062–1078, Jul. 1999.
[2] I. J. Cox, M. L. Miller, and J. A. Bloom, Digital Watermarking. San Francisco, CA, USA: Morgan-Kaufmann, 2002.
[3] F. Hartung and M. Kutter, "Multimedia watermarking techniques,"
Proc. IEEE, Special Issue on Identification and Protection of Multimedia Information, vol. 87, pp. 1079–1107, Jul. 1999.
[4] G. C.Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data: A state-of-the-art overview," IEEE Signal Process. Mag., vol. 17, no. 5, pp. 20–46, Sep. 2000.
[5] N. F. Johnson and S. Katzenbeisser, , S. Katzenbeisser and F. Petitcolas, Eds., "A survey of steganographic techniques," in Information Hiding. Norwood, MA, USA: Artech House, 2000, pp. 43–78.
[6] S. Wang and H. Wang, "Cyber warfare: Steganography vs. steganalysis," Commun. ACM, vol. 47, pp. 76–82, Oct. 2004.
[7] C. Cachin, "An information-theoretic model for steganography," in Proc. 2nd Int. Workshop on Information Hiding, Portland, OR, USA, Apr. 1998, pp. 306–318.
[8] G. J. Simmons, "The prisoner's problem and the subliminal channel," in Advances in Cryptology: Proc. CRYPTO'83, New York, NY, USA, 1984, pp. 51–67, Plenum.
[9] J. Fridrich, Steganography in Digital Media, Principles, Algorithms, and Applications. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[10] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," IEEE Trans. Inf. Theory, vol. 54, no. 6, pp. 2706–2722, Jun. 2008.

[11] Federal Plan for Cyber Security and Information Assurance Research and Development Interagency Working Group on Cyber Security and Information Assurance, Apr. 2006.
[12] R. Chandramouli, "A mathematical framework for active steganalysis," ACM Multimedia Syst., Special Issue on Multimedia Watermarking, vol. 9, pp. 303–311, Sep. 2003.
[13] H. S.Malvar and D. A. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," IEEE Trans. Signal Proc., vol. 51, no. 4, pp. 898–905, Apr. 2003.
[14] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shannon, "Secure spread spectrum watermarking for multimedia," IEEE Trans. Image Process., vol. 6, no. 12, pp. 1673–1687, Dec. 1997.
[15] J. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," IEEE Trans. Image Process., vol. 9, no. 1, pp. 55–68, Jan. 2000.
[16] C. Qiang and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," IEEE Trans. Multimedia, vol. 3, no. 3, pp. 273–284, Sep. 2001.
[17] C. Fei, D. Kundur, and R. H. Kwong, "Analysis and design of watermarking algorithms for improved resistance to compression," IEEE Trans. Image Process., vol. 13, no. 2, pp. 126–144, Feb. 2004.
[18] M. Gkizeli, D. A. Pados, and M. J. Medley, "SINR, bit error rate, and Shannon capacity optimized spread-spectrum steganography," in Proc. IEEE Int. Conf. Image Proce. (ICIP), Singapore, Oct. 2004, pp. 1561–1564.
[19] M. Gkizeli, D. A. Pados, S. N. Batalama, andM. J.Medley, "Blind iterative recovery of spread-spectrum steganographic messages," in Proc. IEEE Int. Conf. Image Proc. (ICIP), Genova, Italy, Sep. 2005, vol. 2, pp. 11–14.
[20] M. Gkizeli, D. A. Pados, and M. J. Medley, "Optimal signature design for spread-spectrum steganography," IEEE Trans. Image Process., vol. 16, no. 2, pp. 391–405, Feb. 2007.
[21] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: Theory and practice," IEEE Trans. Signal Process., vol. 53, no. 10, pt. 2, pp. 3976–3987, Oct. 2005.
[22] L. Pérez-Freire, P. Comesana, J. R. Troncoso-Pastoriza, and F. Pérez-González, "Watermarking security: A survey," LNCS Trans. Data Hiding Multimedia Security, vol. 4300, pp. 41–72, Oct. 2006.
[23] M. Barni, F. Bartolini, and T. Furon, "A general framework for robust watermarking security," ACM J. Signal Process., Special Section: Security of Data Hiding Technologies, vol. 83, pp. 2069–2084, Oct. 2003.
[24] L. Pérez-Freire and F. Pérez-González, "Spread-spectrum watermarking security," IEEE Trans. Inf. Forensics Security, vol. 4, no. 1, pp. 2–24, Mar. 2009.

BIOGRAPHIES

M Srinivas obtained B.Tech degree in Electronics and Communication Engineering from Nalanda Institute of Engg. Tech, Guntur. In 2011. He is M.Tech (ECE) student at Acharya Nagarjuna University College of Engineering and technology, Guntur, India. His areas of interest are Digital Image Processing and Digital Signal processing. Processing and Radar systems.

P. Dhana Lakshmi is currently working as Asst. Professor in Dept. of Electronics and Communication Engineering at University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, India.

