Available online at www.elixirpublishers.com (Elixir International Journal)

Electrical Engineering

Elixir Elec. Engg. 97 (2016) 41829-41832

ISSN: 2229-712X

Digital Media Based Spread Spectrum Hidden Data

M.H.H.Sastry, Jagadeesh Thati and Alluri Srinivasa Rao
Department of ECE Tirumala Engineering College, NRT, Andhra Pradesh, India.

ARTICLE INFO

Article history:
Received: 10 June 2016;
Received in revised form:
20 July 2016;
Accepted: 15 July 2016;

Keywords
Spread Spectrum,
M-IGLS,
Multicarrier SS Embedding.

ABSTRACT
Hiding the information is a vital issue in the 21st century in the field of Data Communication security .It is an important issue because the virtual and digital information transmission faces critical setbacks due to hacking and hackers threats. 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.

1.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. 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 . 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.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, $H_1, H_2, \ldots, H_M$, 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 $T(H_m) \in R^{\frac{N_1 N_2}{M}}$ $m = 1,2,\ldots,M$. 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 $X(m) \in R^L, m = 1,2,\ldots,M$.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

$\{b_k(1), b_k(2), \ldots, b_k(M)\}, k = 1,2,\ldots,K, b_k(m) \in \{\pm 1\}, m = 1,\ldots,M$

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, $b_1(m), \ldots, b_k(m)$, is Simultaneously hidden in the m^{th} transform-domain host vector $X(m)$ via additive SS embedding by means of K spreading sequences (carriers)

$s_k \in R^L, \| s_k \| = 1, k = 1,2,\ldots,K$

Tele:
E-mail address: sastry_mh@yahoo.co.in

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

With corresponding amplitudes.

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

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

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

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

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

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

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 $y(m), m = 1,2, \dots, M$ 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]. 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

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

Where z(m) is modeled as a sequence of zero-mean (without loss of generality) vectors with autocovariancematrix $R_z = E\{zz^T\} = R_x + \sigma_n^2 I$. Let $V_k \triangleq A_k S_k \in R^L, k = 1, \dots, K$ be the amplitude-including embedding carriers. Then, we can further reformulate SS embedding as

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

$$= Vb(m) + z(m), m = 1, \dots, M, \quad (7)$$

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

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

Where multiplication by $R_Z^{-\frac{1}{2}}$ can be interpreted as pre-whitening 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 $\hat{B}$ in (9) can be computed independently of $\hat{V}$ by exhaustive search over all possible choices under the criterion function

$$\hat{B} = \arg \min_{B \in \{\pm 1\}^{K \times M}} \frac{\| R_Z^{-\frac{1}{2}} YP \perp B \|_F^2}{\| R_Z^{-\frac{1}{2}} YP \perp B \|_F^2} \quad (8)$$

1) $\mathbf{d} := \mathbf{0}$; Initialize $\hat{B}^{(0)} \in \{\pm 1\}^{K \times M}$ arbitrarily.

2):=d+1;

V^((d)) :=Y⟦(B^((d-1)))⟧^T [(B^((d-1)))⟦(B^((d-1)))⟧^T)]^(-1);

B^((d)):=sgn{((V^((d)))^T R^_y^(-1) (V^((d))))^(-1) (V^((d)))^T R^_y^(-1) Y}

3)Repeat Step 2 until $\hat{B}^{(d)} = \hat{B}^{(d-1)}$

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

TABLE I

IV. 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

$$\hat{V}_{GLS} = \arg \min_{V \in R^{L \times K}} \| R_Z^{-\frac{1}{2}} (Y - VB \|_F^2$$

$$= YB^T (BB^T)^{-1} \quad (9)$$

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

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

$$= (V^T R_Z^{-1} V)^{-1} V^T R_Z^{-1} Y \quad (10)$$

Observing that

$$(V^T R_Z^{-1} V)^{-1} V^T R_Z^{-1} = \left(V^T R_y^{-1} V\right)^{-1} V^T R_y^{-1} \quad (11)$$

We rewrite

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

where

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

The M-IGLS extraction algorithm is

$O(2K^3 + 2LMK + K^2(3L + M) + L^2 K$ 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 $s_{K=R^L - S_K, K=1, \dots, K}$. The sign-ambiguity problem can be overcome with a few known or guessed data symbols for supervised sign correction3 . 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 $\{(\hat{V}_1\hat{B}_1), \ldots (\hat{V}_p\hat{B}_1)$, has superior generalized- least-squares fit, we simply feed $(\hat{V}_i\hat{B}_i)$ to (9) (using R_y in place of R_z) and choose

$$\hat{V}_{final}, \hat{B}_{final} = \min_{arg\ (V,B)\in\{(\hat{V}_1\hat{B}_1),\ldots(\hat{V}_p\hat{B}_1)\}} \| R_Z^{-\frac{1}{2}}(Y - VB \|_F^2 \quad (12)$$

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

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

V.Simulation Results

Fig 5.1. 512*512 Plane image.

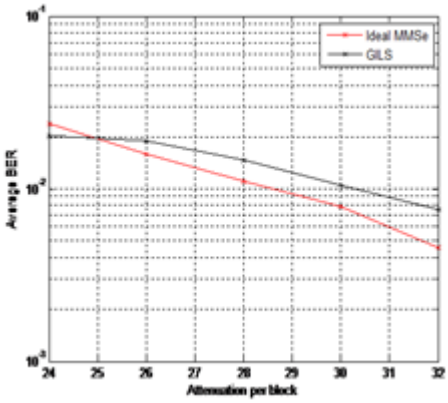
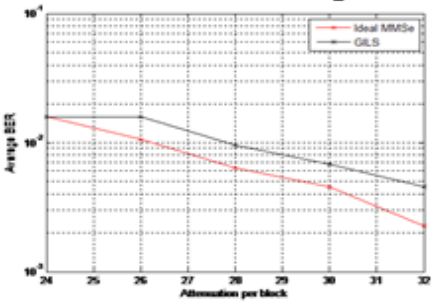


Fig 5.2 512*512 Plane image BER.

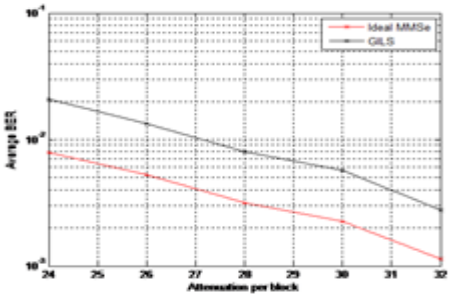
5.3 256*256 Plane image.



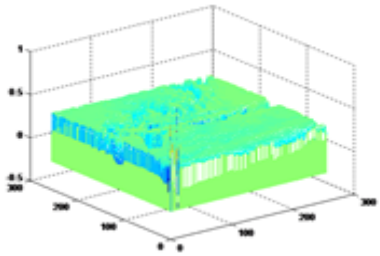
5.4 256*256 plane imane BER

5.5 512*512 boat image

5.4. 512*512 boat image BER.



5.5. 256*256 boat image.



5.6. Host data auto correlation matrix ,8*8 DCT.

VI.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.

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.

Author Profile

M .H. H. Sastry received his M.Tech degree from JNTUK, A.P, India in 2006. Presently he is working as an Associate professor, in Department of Electronics and Communication Engineering, Tirumala Engineering College, Guntur, A.P, India.His current areas of research interests include Communication Systems and Signal Processing.

Mr. Jagadeesh Thati is currently working as Associate Professor in Department of ECE at Tirumala Engineering College, Jonnalagadda, Narasaraopet, Guntur (dt). He has worked as dasa5 developer in Dasa Control systems AB Hammerdalsvgen 3, SE-352 46 Vxjo, Sweden. He did his MS from BTH, Sweden. He has published 23 international journals, 8 international conferences and two Books. He has appointed as reviewer for various journals. He has professional memberships in IETE and ISTE. He received best Teacher award from Tirumala Engineering College in 2012. His areas of Interests are Signal Processing, Digital Image Processing, Computer Vision, Neural Networks and Nano Technology.

Srinivasarao Alluri received his BE degree from Andhra University, visakapatnam, India in 2004 and M.Tech from JNTU Hyderabad, India in 2012. Presently he is working as a Associate Professor, in Department of Electronics and Communication Engineering, Tirumala Engineering College, Guntur, A.P, India. His current areas of research interests include VLSI, Analog IC Design and Communication systems.