Security Analysis of ISS Watermarking Using Natural Scene Statistics

Dong Zhang[1], **Jiangqun Ni[1]**, Qiping Zeng[1], Dah-Jye Lee[2], and Jiwu Huang[1]

[1] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China, 510275

[2] Department of Electrical & Computer Engineering Brigham Young University, Provo, UT, U.S.A. 84602

IH2010 ·Calgary, Canada· June 28-30, 2010

Outline

- Introduction
- The GSM Model and Its Performance
- Methodology
- Security Analysis of ISS Watermarking Based on GSM
- Simulation and Discussion
- Conclusions

Robust vs. Security

- Robustness deals with blind attacks
- The security deals with intentional attacks and is more critical for watermarking
- The objective of security attack is gaining the knowledge about the secret keys of the system, therefore offering complete break

Previous Works (1)

- With Fisher information, Cayre *et al* quantified for the first time the security of Add-SS watermarking schemes

- Comesaña *et al* employed another measure for information leakage, i.e., the mutual information between the observed watermarked signals and the secret carriers

Previous Works (2)

- Ni *et al.* took advantage of natural scene statistics and presented a theoretical analysis of the Add-SS based watermarking algorithm with Shannon's mutual information

- Pérez-Freire *et al*. discussed the security of Improved Spread-Spectrum (ISS) watermarking under the assumption of Gaussian host

The Signal Model

- The information-theoretic evaluation and practical attacking algorithm for watermarking security require the statistical modeling of the signals involved in the problem: the host signal and the spreading carriers

- Usually Gaussian distributions are assumed for the involved signals in the interest of mathematical tractability

The Gaussian Scale Mixture Model – GSM

- The same issues are further investigated based on the NSS model (natural scene statistics)
- Both the models of generalized Gaussian (GG) and mixture of Gaussian (MoG) are used to characterize the behavior of natural images in transformed domain
- The security analysis with the above models suffers from the difficulty of mathematical manipulation

The Gaussian Scale Mixture Model - GSM

- Recently, the Gaussian scale mixture (GSM) has been proposed to model the natural image in wavelet domain

- M. J. Wainwright, E. P. Simoncelli, “Scale Mixtures of Gaussians and the Statistics of Natural Images,” Advances in Neural Information Processing Systems (NIPS*99), vol. 12, pp. 855-861, MIT Press, 2000.
- J. Portilla, V. Strela, M. J. Wainwright, E. P. Simocelli. Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Transactions on Image Processing*. 2003, 12 (11): 1338-1351.

The GSM Model and Performance

- The wavelet coefficient of natural image in one subband is modeled as the GSM RF (Random Field) , i.e.,

$$\mathbf{X} = \mathbf{S} \cdot \mathbf{U} = \{S_i \cdot U_i : i \in \mathbf{I}\}$$

- $\mathbf{U} = \{U_i : i \in \mathbf{I}\}$ is a global Gaussian scalar RF with mean zero and variance σ_U^2
- $\mathbf{S} = \{S_i : i \in \mathbf{I}\}$ is a RF of positive multipliers that control the variances of local coefficients

The GSM Model and Performance

- With the GSM, when conditioned on S_i, X_i is Gaussian distributed, i.e.,

$$p_{X_i|S_i}\left(x_i \mid s_i\right) \sim N\left(0, s_i^2 \sigma_U^2\right)$$

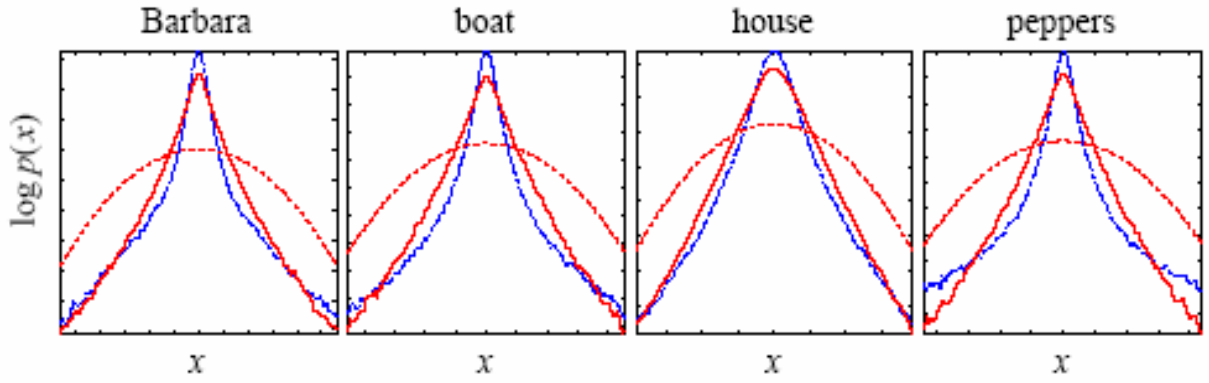


Fig. 2. Empirical marginal log distributions of coefficients from a multi-scale decomposition of photographic images (blue dot-dashed line), synthesized FoGSM samples from the same subband (red solid line), and a Gaussian with the same standard deviation (red dashed line).

The Evaluation of Watermarking Security

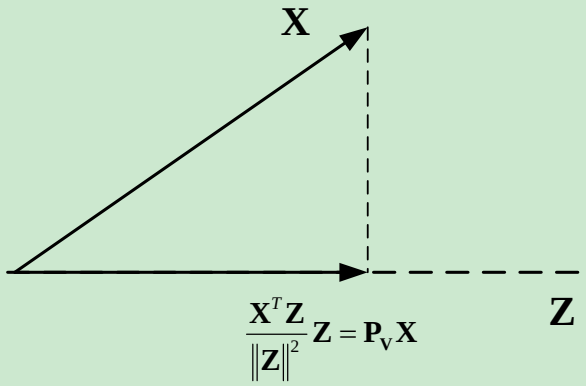
- The watermarking algorithm
  - QIM
  - Add-SS and ISS
- The security is quantified by the mutual information between observed watermarked signal $\mathbf{Y}^{N_0}$ and secret carrier $\mathbf{Z}$ under WOA/ KMA/KOA:

$$I\left(\mathbf{Z};\mathbf{Y}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\left|\mathbf{Z}\right.\right)$$

Improved Spread Spectrum Watermarking

$$\mathbf{Y}_j = \mathbf{X}_j + \mu\left(\mathbf{X}_j, M_j\right)\mathbf{Z}$$

$$\mathbf{Y}_j = \mathbf{X}_j + (-1)^{M_j} v\mathbf{Z} - \lambda \frac{\mathbf{X}_j^T \mathbf{Z}}{\|\mathbf{Z}\|^2}\mathbf{Z}$$



Notations

$\mathbf{X}_j$ $\mathbf{Y}_j$	the host and the watermarked signal of the j[th] observation
$\mathbf{Z}$	the secret carrier
M_j	the embedded message in the j[th] observation
$\mu(\bullet)$	A linear function for embedding
v	The parameter to control the embedding distortion
$\lambda\ (0 \le \lambda \le 1)$	The host-rejection parameter

Security Analysis on ISS watermarking for KMA

Residual entropy of secret carrier:

$$h\left(\mathbf{Z}\middle|\mathbf{Y}^{N_o},\mathbf{S}^{N_o},M^{N_o}\right)=h(\mathbf{Z})-I\left(\mathbf{Z};\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)$$

$$h(\mathbf{Z})=\frac{N_v}{2}\log\left(2\pi e\sigma_z^2\right)$$

Mutual information between observations and secret carrier:

$$I\left(\mathbf{Z};\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},M^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{Z},\mathbf{S}^{N_o},M^{N_o}\right)$$

N_o Number of observations

Security Analysis on ISS watermarking for KMA

Case $N_o = 1$

$$I\left(\mathbf{Y};\mathbf{Z}\middle|\mathbf{S},M\right)$$

$$= \frac{1}{2}\sum_{i=1}^{N_v}\log\left[s_i^2\sigma_u^2 + \frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left(-2\lambda s_i^2\left(N_v+2\right) + 2\lambda^2 s_i^2 + \lambda^2\sum_{l=1}^{N_v} s_l^2 \right) + v^2\sigma_Z^2 \right]$$

$$-\frac{1}{2}\log\left(\left(1-\lambda\right)^2 \sigma_u^{2N_v} \prod_{i=1}^{N_v} s_i^2 \right)$$

N_v dimensions for host and observation

Security Analysis on ISS watermarking for KMA

Case $N_o > 1$

$$I\left(\mathbf{Z}, \mathbf{Y}^{N_o} \middle| \mathbf{S}^{N_o}, M^{N_o}\right)$$

$$\leq \frac{1}{2}\sum_{i=1}^{N_v} \log\left(\left(2\pi e\right)^{N_o} \left(1 + \sum_{j=1}^{N_o} \frac{v^2\sigma_Z^2}{s_{j,i}^2\sigma_u^2 + \frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left[-2\lambda s_{j,i}^2\left(N_v+2\right) + 2\lambda^2 s_{j,i}^2 + \lambda^2 \sum_{l=1}^{N_v} s_{j,l}^2 \right]} \right) \right.$$

$$\left. \cdot \prod_{j=1}^{N_o} \left(s_{j,i}^2\sigma_u^2 + \frac{\sigma_u^2}{N_v\left(N_v+2\right)}\left[-2\lambda s_{j,i}^2\left(N_v+2\right) + 2\lambda^2 s_{j,i}^2 + \lambda^2 \sum_{l=1}^{N_v} s_{j,l}^2 \right] \right) \right)$$

$$- \frac{1}{2}\sum_{j=1}^{N_o} \log\left[\left(2\pi e\right)^{N_v} \left(1-\lambda\right)^2 \sigma_u^{2N_v} \prod_{i=1}^{N_v} s_{j,i}^2 \right]$$

Security Analysis on ISS watermarking for WOA

$$h\left(\mathbf{Z}\middle|\mathbf{S}^{N_o},\mathbf{Y}^{N_o}\right)=h\left(\mathbf{Z}\middle|\mathbf{S}^{N_o}\right)-I\left(\mathbf{Z};\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)=h(\mathbf{Z})-I\left(\mathbf{Z};\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)$$

$$I\left(\mathbf{Z};\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)=h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o}\right)-h\left(\mathbf{Y}^{N_o}\middle|\mathbf{S}^{N_o},\mathbf{Z}\right)$$

$$I\left(\mathbf{Z};\mathbf{Y}\middle|\mathbf{S}\right)=h\left(\mathbf{Y}\middle|\mathbf{S}\right)-h\left(\mathbf{Y}\middle|\mathbf{S},\mathbf{Z}\right)$$

$$\geq h\left(\mathbf{Y}\middle|\mathbf{S},M=0\right)-h\left(\mathbf{Y}\middle|\mathbf{S},\mathbf{Z},M\right)-\log(2)$$

$$I\left(\mathbf{Z};\mathbf{Y}\middle|\mathbf{S},M\right)\geq I\left(\mathbf{Z};\mathbf{Y}\middle|\mathbf{S}\right)\geq I\left(\mathbf{Z};\mathbf{Y}\middle|\mathbf{S},M\right)-\log(2)$$

Simulation and Discussion (1)

- Set-up
  - An i.i.d. Gaussian vector was used as secret carrier
  - Natural images with different texture characteristics, including aerial, baboon, barb, boat, f16, lena, peppers and sailboat, were used as hosts
  - Bi-orthogonal 9/7 wavelet
  - Coefficients from HL2, LH2 and HH2 were randomly selected as hosts

Simulation and Discussion (2)

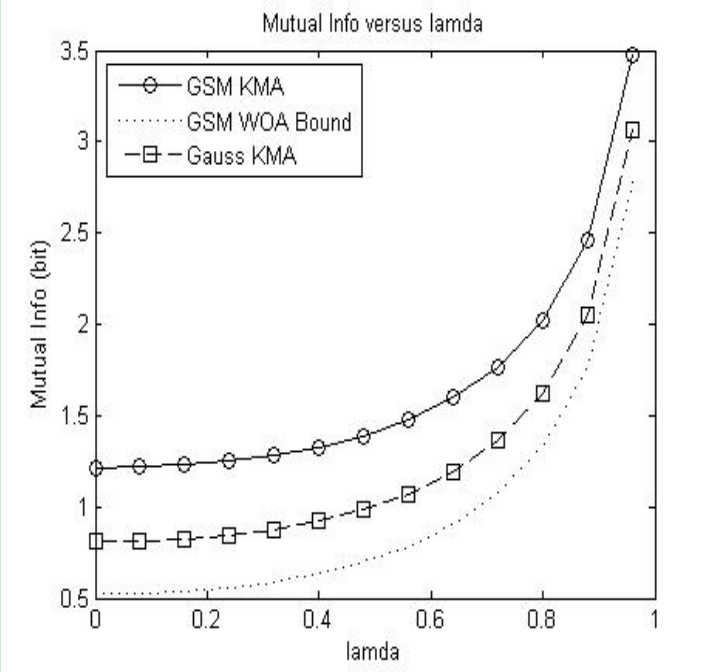


Fig.1 Relationship between the information leakage of secret carrier and the host-rejection parameter

$N_o = 1 \quad N_v = 512$ DWR=25dB

- With the increase of λ , information leakage increases for both Gaussian and GSM model
- ISS watermarking performs better in robustness and worse in security than Add-SS
- With GSM, there is more information leakage of secret carrier for ISS watermarking
- GSM characterizes natural image more accurately and enables the attacker to improve his estimations on secret carrier.

Simulation and Discussion (3)

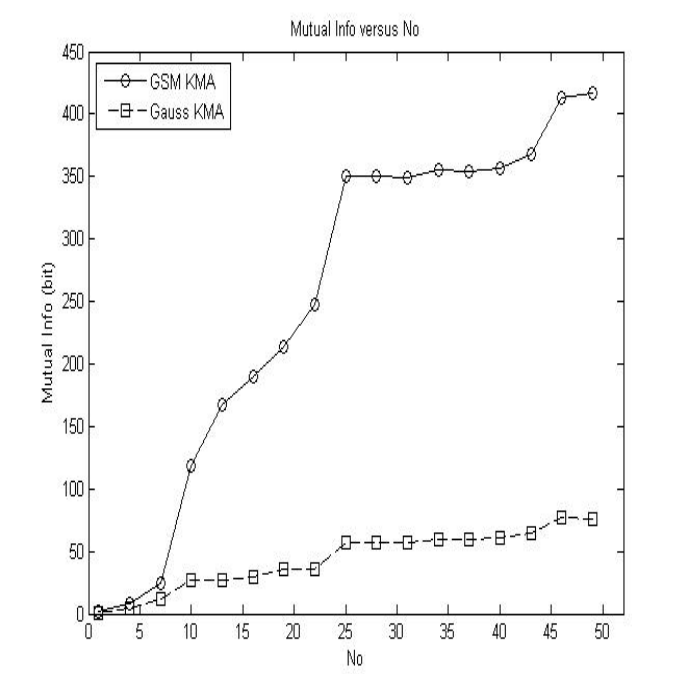


Fig. 2 Relationship between the information leakage of secret carrier and the number of observations

$N_v = 512 \quad \lambda = 0.5$ DWR=25dB

- The information leakage of the secret carrier increases with the accumulation of observations
- The attacker could improve his knowledge about the secret carrier by increasing the number of observations
- Due to the correlation among observations, the information leakage of secret carrier grows non-linearly against the number of observations for both GSM and Gaussian based approaches.

.

Simulation and Discussion (4)

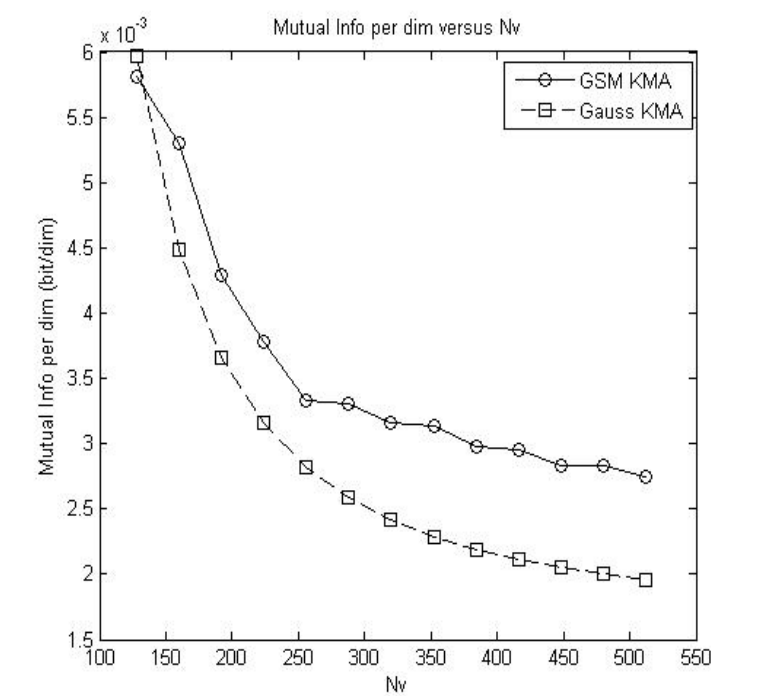


Fig.3 Relationship between the information leakage of secret carrier and the length of carrier

$N_0 = 1$ $\lambda = 0.5$ DWR=25dB

- Information leakage per-dimension decreases with the increase of the length of carrier
- Using longer secret carrier will provide higher security to ISS watermarking

Simulation and Discussion (5)

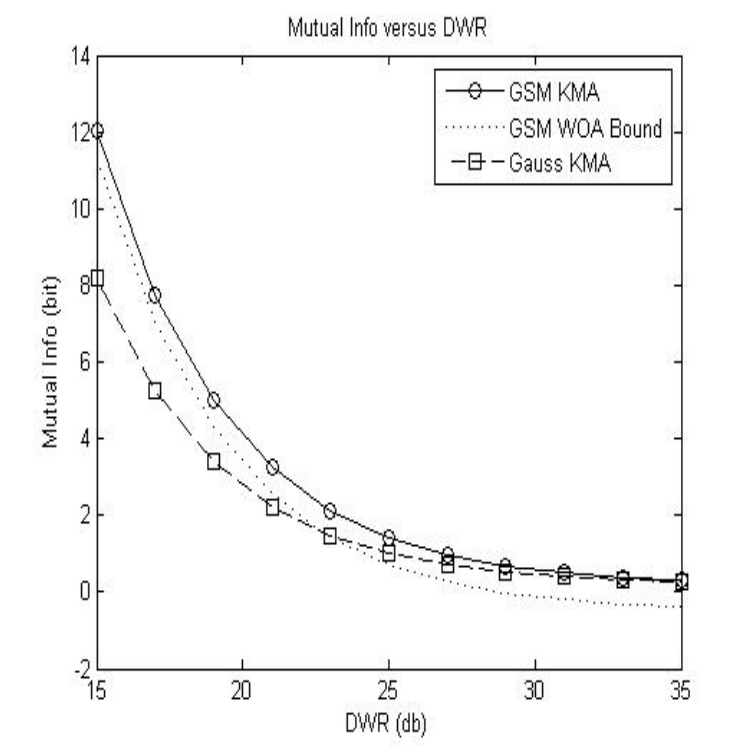


Fig.4 Relationship between the information leakage of secret carrier and DWR

$N_o = 1$ $N_v = 512$ $\lambda = 0.5$

- Information leakage of secret carrier decreases against the increasing of DWR
- Security performance of ISS watermarking is improved by increasing the value of DWR

Conclusions

- This paper presents a theoretical analysis of ISS watermarking security by incorporating NSS model
- The new security measures are derived based on the NSS model GSM
- The information leakage of the secret carries with GSM host is consistently greater than that with Gaussian host, which shows the over-estimated security level of ISS watermarking when host is characterized with Gaussian