doi:10.11959/j.issn.1000-436x.2017155

基于空域图像变换参数扰动的隐写术

孙曦[1,2]，张卫明[1,2]，俞能海[1,2]，魏尧[1,2]
（1. 中国科学技术大学信息科学技术学院，安徽 合肥 230001；2. 中国科学院电磁空间信息重点实验室，安徽 合肥 230001）

摘 要：在经典图像隐写算法研究中，使用的载体图像库多为实验室环境下的自然图像库，而在现实环境下，随着各种图像处理软件的流行，用户也越来越多地使用经图像处理后的图片。然而，如何利用图像处理来更好地指导隐写还未被系统地研究过。对此，以空域图像变换为例，提出参数扰动模型，将隐写带来的图像噪声隐藏在因参数微扰而带来的图像像素波动中，主动引入了图像集的失配因素。实验结果表明，在相同的隐写算法下，相较于直接使用原始图像库，隐写分析者在应对参数扰动的情况下的检错率有明显提高，从而显著增强了隐写算法的安全性，也使隐写算法更适用于现实环境。

关键词：隐写；隐写分析；图像处理；失配；空域图像变换
中图分类号：TP309 **文献标识码**：A

Steganography based on parameters' disturbance of spatial image transform

SUN Xi[1,2], ZHANG Wei-ming[1,2], YU Neng-hai[1,2], WEI Yao[1,2]
（1. School of Information Science and Technology, University of Science and Technology of China, Hefei 230001, China;
2. Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230001, China）

Abstract: In the research of state-of-the-art steganography algorithms, most of image sources were natural images in laboratory environment. However, with the rapid development of image process tools and applications, images after image processing were widely used in real world. How to use image process to improve steganography has not been systematically studied. Taking spatial image transform for consideration, a parameters' disturbance model was presented, which could hide the noise taken by steganography in the pixel fluctuation due to the disturbance. Meanwhile, it would introduce cover source mismatch for a steganalyzer. The experimental results show that, compared with using traditional image database, it can significantly enhance the security of steganography algorithms and accommodative to the real world situation.

Key words: steganography, steganalysis, image process, mismatch, spatial image transform

1 引言

隐写术是信息隐藏的一个分支[1]，可以将隐私数据嵌入到载体中，为安全通信提供了强有力的技术支持，对国家、集体、个人的信息安全以及隐私保护具有重要的意义。数字图像隐写是隐秘通信的经典模式之一，发送方通过一种嵌入算法将隐秘消息嵌入到载体图像中得到载密图，通常来说，载密图只是对载体图做了极其轻微的修改从而在传输过程中难以被检测到，接收方通过与发送方共享的密钥，使用提取算法从载密图中提取出隐秘消息。通过这种模式将整个通信过程

收稿日期：2016-12-30；**修回日期**：2017-05-08
通信作者：俞能海，ynh@ustc.edu.cn
基金项目：国家自然科学基金资助项目（No.U1636201, No.61572452, No.61502007）；中国博士后科学基金资助项目(No.2015M582015)
Foundation Items: The National Natural Science Foundation of China (No.U1636201, No.61572452, No.61502007), China Postdoctoral Science Foundation (No.2015M582015)

伪装成普通的图像传输过程，从而达到安全隐蔽的效果。

早期图像隐写假设修改不同位置载体元素造成的失真是相同的[2]，对于图像内容是“非自适应”的，但是快速发展的隐写分析技术[3~5]使隐写术从“非自适应”向“自适应”阶段发展，即优先修改失真小（难检测）的区域。针对这种需求，Filler等[6]提出了实用的最小化失真隐写编码，称为 STC 编码。STC 编码出现以后，隐写术研究就集中在了如何设计合理的失真函数。在空域隐写算法中，HUGO 算法[7]通过计算载体图像修改后的特征向量的差值来定义失真；WOW 算法[8]使用方向滤波，对于复杂区域的像素赋予高的修改失真、平滑区的像素赋予低的修改失真；SUNIWARD 算法[9]在 WOW 的基础上做了进一步的改进；李斌等[10]和 Denemark 等[11]独立提出“方向一致性原则”，即相邻像素的修改方向尽量保持一致可以减小修改代价。在变换域，JPEG 图像是最流行的图像格式。为与 JPEG 标准兼容，JPEG 隐写通常以量化 DCT 系数为载体。目前，广泛研究的 JPEG 隐写算法也是采用“最小化失真模型”，对 DCT 系数定义失真函数并利用 STC 嵌入消息。如 UED 算法[12]用 DCT 系数绝对值和块内、块间相关性定义失真；JUNIWARD 算法[9]采用修改 DCT 系数对空域像素相关性的影响定义失真。

然而，一方面，传统隐写载体多为实验室环境下的自然图像，如 BOSSbase 1.01[13]，而在现实环境下的社交平台中，随着 Photoshop、美图秀秀等图像处理软件的流行，人们也越来越多地使用经处理后的图像，然而对于图像处理的载体集隐写尚没有系统的研究；另一方面，Fridrich 等[14]指出传统隐写分析中训练集与测试集之间存在的各种失配，如训练集和测试集统计特征不一致导致的失配、嵌入率未知导致的失配、算法未知导致的失配等，会使传统隐写分析的错误率大幅提升。因此，可以试图在隐写过程中，把因隐写造成的图像改变淹没在图像处理的扰动中，主动引入失配因素从而提高隐写安全性能，隐写分析者或许可以检测出一张图片是否经过了图像变换处理，但更难检测出是否经过了隐写处理。

本文以图像作为隐写、隐写分析的研究对象，以隐写算法结合图像处理为切入点，定性定量地研究了因空域图像变换的参数扰动带来的图像失配对隐写及隐写分析的影响。

2　最小化失真隐写模型

目前的自适应隐写算法都是基于最小化失真模型，设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 为未嵌入消息的载体图像，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)$ 为嵌入消息后的载密图像，则隐写过程图像失真函数的定义为

$$D(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{n}\rho_i(x_i,y_i) \tag{1}$$

其中，$\rho_i(x_i,y_i)$ 为将第 i 个载体像素 x_i 修改为 y_i 的失真代价，定义载密图像素值的概率分布为 $\pi(\boldsymbol{y})=P(\boldsymbol{Y}=\boldsymbol{y}|\boldsymbol{x})$，则在一次隐写通信过程中可发送的信息总量为

$$H(\pi)=-\sum_{\boldsymbol{y}\in\boldsymbol{Y}}\pi(\boldsymbol{y})\mathrm{lb}\pi(\boldsymbol{y}) \tag{2}$$

失真函数的期望值为

$$\mathrm{E}_{\pi}[D]=\sum_{\boldsymbol{y}\in\boldsymbol{Y}}\pi(\boldsymbol{y})D(\boldsymbol{y}) \tag{3}$$

那么最小化失真嵌入即为以下 2 个优化问题。

1) 嵌入率受约束的条件下，使失真函数最小。

$$\min \mathrm{E}_{\pi}[D],\text{s.t. } H(\pi)=m \tag{4}$$

2) 失真期望值受约束的条件下，使嵌入率最大。

$$\max H(\pi),\text{s.t. } E_{\pi}[D]=D \tag{5}$$

3　图像变换参数扰动的隐写模型

目前，常用的空域隐写算法的载体集多为实验室环境下的自然图像，如 BOSSbase 1.01[13]，并且达到了相对理想的安全性能，如 S-UNIWARD 算法。然而在现实环境下，尤其在网络社交平台中，人们使用图像处理的行为越来越普遍，经处理后的图像也广泛存在；因此，在这部分提出以图像处理操作过程为掩盖来提升隐写安全性，隐写分析者也许可以检测到一张图是否经过了图像处理，但这不能成为其怀疑隐写存在的证据。

3.1　载体载密图像检测模型

与经典图像模型相同[15]，假设忽略空域滤波和马赛克处理，对于原始载体图像库的每张图像，其像素值为独立同分布于高斯模型的 n 维向量

$\boldsymbol{z}=(z_1,z_2,\cdots,z_n)$，即 $Z_n\sim N(\mu_n,\omega_n^2)$，其中，$n=1,2,\cdots,N$，$\mu_n$ 为无噪声的图像内容，ω_n^2 为高斯噪声方差。设 $\hat{\mu}_n\in Z$ 为第 n 个像素的平均估计值，则像素噪声残差 $x_n=z_n-\hat{\mu}_n$ 服从高斯分布 $X_n\sim N(0,\sigma_n^2)$，其中，$\sigma_n^2>\omega_n^2$，这是考虑到还有模型的误差。此时，像素噪声残差 x_n 的概率密度函数 P_{σ_n} 为

$$p_{\sigma_n}=P(x_n=k)\propto\frac{1}{\sigma_n\sqrt{2\pi}}\exp\left(-\frac{k^2}{2\sigma_n^2}\right) \tag{6}$$

给定载体图像，以 $\boldsymbol{x}=(x_1,x_2,\cdots,x_N)$ 表示，相应的载密图像表示为 $\boldsymbol{y}=(y_1,y_2,\cdots,y_N)$，对于±1的三元消息嵌入模式下符合如下概率规则。

$$\begin{cases}\mathbb{P}(y_n=x_n+1)=\beta_n\\ \mathbb{P}(y_n=x_n-1)=\beta_n\\ \mathbb{P}(y_n=x_n)=1-2\beta_n\end{cases} \tag{7}$$

其中，β_n 为特定像素点的±1修改概率，满足 $0\leqslant\beta_n\leqslant\frac{1}{3}$，在最小化失真隐写模型中，$\beta_n=\frac{\mathrm{e}^{-\lambda\rho_n}}{1+2\mathrm{e}^{-\lambda\rho_n}}$，$\rho_n$ 为失真值，λ 是与嵌入率和隐写算法相关的参数。则 y_n 的概率密度函数 Q_{σ_n,β_n} 为

$$\begin{aligned}q_{\sigma_n,\beta_n}(k)=\mathbb{P}(y_n=k)=(1-2\beta_n)p_{\sigma_n}(k)+\\ \beta_n p_{\sigma_n}(k+1)+\beta_n p_{\sigma_n}(k-1)\end{aligned} \tag{8}$$

设 $\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_N)$ 为隐写者的修改概率，$\boldsymbol{\gamma}=(\gamma_1,\gamma_2,\cdots,\gamma_N)$ 为隐写分析者对 β 的相应估计值，则隐写分析者对于图像 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 的检验假设为

$$\begin{cases}H_0=\{x_n\sim P_{\sigma_n},\forall\sigma_n>0\}\\ H_1=\{x_n\sim Q_{\sigma_n,\beta_n},\forall\sigma_n>0\}\end{cases} \tag{9}$$

由似然比检验理论得到判决门限函数为

$$\Lambda(x,\sigma)=\prod_{n=1}^{N}\Lambda_n=\prod_{n=1}^{N}\left(\frac{q_{\sigma_n,\gamma_n}(x_n)}{p_{\sigma_n}(x_n)}\right)\tau \tag{10}$$

由中心极限定理，当 $N\to\infty$ 时，有

$$\begin{aligned}\mathrm{lb}\Lambda^*(\boldsymbol{x},\sigma)&=\frac{\sum_{n=1}^{N}(\mathrm{lb}\Lambda_n-\mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N}Var_{H_0}[\mathrm{lb}\Lambda_n]}}\\ &\to\begin{cases}N(0,1),\ H_0\text{条件下}\\ N(\theta,1),\ H_1\text{条件下}\end{cases}\end{aligned} \tag{11}$$

其中，

$$\begin{aligned}\theta&=\frac{\sum_{n=1}^{N}(\mathrm{E}_{H_1}[\mathrm{lb}\Lambda_n]-\mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N}Var_{H_0}[\mathrm{lb}\Lambda_n]}}\\ &=\frac{2\sum_{n=1}^{N}\sigma_n^{-4}\beta_n\gamma_n}{\sqrt{2\sum_{n=1}^{N}\sigma_n^{-4}\gamma_n^2}}\end{aligned} \tag{12}$$

设 $N(0,1)$ 的概率密度函数为 $f_0(x)$，$N(\theta,1)$ 的概率密度函数为 $f_1(x)$，则有 $f_1(x)=f_0(x-\theta)$，检测的虚警率为

$$P_{\mathrm{FA}}=\int_{\tau}^{+\infty}f_0(x)\mathrm{d}x \tag{13}$$

漏警率为

$$P_{\mathrm{MD}}=\int_{-\infty}^{\tau}f_1(x)\mathrm{d}x=\int_{-\infty}^{\tau-\theta}f_0(x)\mathrm{d}x \tag{14}$$

总体检错率为

$$P_{\mathrm{E}}=\frac{1}{2}(P_{\mathrm{FA}}+P_{\mathrm{MD}}) \tag{15}$$

3.2 空域图像变换

空域图像变换是对像素值进行变换，不失一般性，设变换函数为

$$z_n'=f(z_n,\xi_n) \tag{16}$$

其中，ξ_n 为变换参数。为便于统一地分析与讨论，令

$$z_n'=\alpha_n z_n \tag{17}$$

其中，α_n 是关于参数 ξ_n 的函数。

1) 在空域图像变换过程中存在大量复杂操作，其中，对于线性的变换本文以基础的灰度线性变换为例进行讨论，它主要应用于图像的对比度增强。

灰度线性变换是对图像像素值的简单线性变换处理，其变换过程如下。

$$z_n'=f(z_n,\xi_n)=\xi_n z_n \tag{18}$$

当 $\xi_n>1$ 时，图像变亮；当 $0<\xi_n<1$ 时，图像变暗；当 $\xi_n=1$ 时，图像不变；式(17)中系数 $\alpha_n=\xi_n$。

2) 空域图像变换还包括大量非线性变换，对此，本文以基础的 Gamma 变换为例进行讨论，它主要应用于相机图片的亮度矫正。

Gamma 变换是常用的矫正数字图像亮度的图像处理操作之一，设 $z_n\in[0,z_m]$ 为变换前图像像素

点的像素值，$z'_n \in [0, z_m]$ 为变换后相对应位置的像素点的像素值，其变换过程如下。

$$z'_n = f(z_n, \xi_n) = \left(\frac{z_n}{z_m}\right)^{\xi_n} z_m = \left(\frac{z_n}{z_m}\right)^{\xi_n - 1} z_n \quad (19)$$

当 $0 < \xi_n < 1$ 时，图像变亮；当 $\xi_n > 1$ 时，图像变暗；当 $\xi_n = 1$ 时，图像不变；式(17)中系数 $\alpha_n = \left(\frac{z_n}{z_m}\right)^{\xi_n - 1}$。

3.3　参数扰动模型

如图 1 所示，设原始图像库为 S_0，把 S_0 看作是参数 ξ_0=1 的灰度线性变换或 Gamma 变换，给定扰动参数 δ 下，对原始图像库的每张图片 I_i，隐写者随机选取变换参数 $\xi_n \in [\xi_0 - \delta, \xi_0 + \delta]$ 做空域图像变换，从而得到变换图像库 S'，并在此图像库上应用隐写算法。当 $\delta = 0$ 时，$\xi_n = \xi_0 = 1$，模型退化到无扰动模型。为了单独研究扰动参数对隐写算法和隐写分析的影响，本文假设隐写算法及嵌入率等信息均为匹配情况，对于隐写分析者，在无法预知或估计不准隐写者的扰动参数 δ 情况下，隐写分析者将基于原始图像库 S_0 来训练分类器。

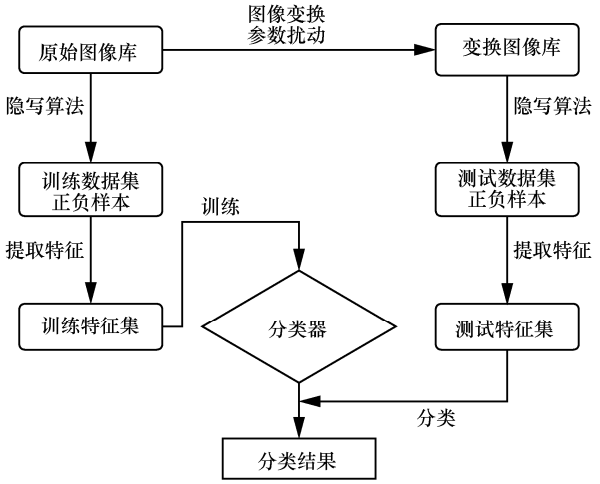


图 1　图像变换参数扰动的隐写模型

隐写分析者使用和隐写者相同的隐写算法，使用原始图像库 S_0 生成载密图像库，分别对 2 个正负数据集提取图像特征，在特征空间训练得到分类器，并用此分类器对隐写者的变换图像库的载体载密图像进行分类测试。

根据 $z'_n = \alpha_n z_n$，变换后的载体图像 x'_n 的概率密度函数为

$$p'_{\sigma_n}(k) = P(x'_n = k) \propto \frac{1}{\alpha_n \sigma_n \sqrt{2\pi}} \exp\left(-\frac{k^2}{2\alpha_n^2 \sigma_n^2}\right) \quad (20)$$

在同样的隐写算法下，对应的载密图像 y'_n 的概率密度函数为

$$q'_{\sigma_n, \beta_n}(k) = P(y'_n = k) = (1 - 2\beta_n) p'_{\sigma_n}(k) + \beta_n p'_{\sigma_n}(k+1) + \beta_n p'_{\sigma_n}(k-1) \quad (21)$$

此时，式(9)检验假设的真实情况则为

$$\begin{cases} H'_0 = \{x'_n \sim P'_{\sigma_n}, \forall \sigma_n > 0\} \\ H'_1 = \{x'_n \sim Q'_{\sigma_n, \beta_n}, \forall \sigma_n > 0\} \end{cases} \quad (22)$$

判决门限的收敛函数(11)将改变为

$$\mathrm{lb}\Lambda^*(\boldsymbol{x}, \sigma) = \frac{\sum_{n=1}^{N} (\mathrm{lb}\Lambda_n - \mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N} Var_{H_0}[\mathrm{lb}\Lambda_n]}} \rightarrow \begin{cases} N(\eta, 1), & H'_0 \text{ 条件下} \\ N(\theta + \eta, 1), & H'_1 \text{ 条件下} \end{cases} \quad (23)$$

其中，

$$\eta = \frac{\sum_{n=1}^{N} \sigma_n^{-2} \gamma_n (\alpha_n^2 - 1)}{\sqrt{2 \sum_{n=1}^{N} \sigma_n^{-4} \gamma_n^2}} \quad (24)$$

式(23)的推导证明详见附录。也就是说，变换图像库的判决门限函数相比原始模型有了 η 的偏移量，考虑关于 η 的以下 2 种特别情况。

1) 当 $\alpha_n = 1$，即 $\delta = 0$，图像不发生变换操作时，$\eta = 0$，模型自然退化到原始图像库上的假设检验模型。

2) 当 $\alpha_n = \xi_n$，是在 $[\xi_0 - \delta, \xi_0 + \delta]$ 上的随机分布时，η 对于 α_n 的平均期望值为

$$\eta_0 = \mathrm{E}_\alpha(\eta) = \frac{\sum_{n=1}^{N} \sigma_n^{-2} \gamma_n}{3\sqrt{2 \sum_{n=1}^{N} \sigma_n^{-4} \gamma_n^2}} \delta^2 \quad (25)$$

说明偏移量 η 随着扰动参数 δ 的增大而增大。

设 $N(\eta,1)$ 的概率密度函数为 $f'_0(x)$，$N(\theta+\eta,1)$ 的概率密度函数为 $f'_1(x)$，则有 $f'_0(x) = f_0(x - \eta)$、$f'_1(x) = f_1(x - \eta)$。实际情况的虚警率、漏警率和总体检错率将改变为

$$P'_{\mathrm{FA}} = \int_\tau^{+\infty} f'_0(x)\mathrm{d}x = \int_{\tau - \eta}^{+\infty} f_0(x)\mathrm{d}x > P_{\mathrm{FA}} \quad (26)$$

$$P'_{\mathrm{MD}}=\int_{-\infty}^{\tau}f_1'(x)\mathrm{d}x=\int_{-\infty}^{\tau-\eta}f_1(x)\mathrm{d}x$$
$$=\int_{-\infty}^{\tau-\theta-\eta}f_0(x)\mathrm{d}x<P_{\mathrm{MD}} \tag{27}$$

$$P'_{\mathrm{E}}=\frac{1}{2}(P'_{\mathrm{FA}}+P'_{\mathrm{MD}}) \tag{28}$$

与原始假设检验模型相比，检错率的差值为

$$\Delta P_{\mathrm{E}}=P'_{\mathrm{E}}-P_{\mathrm{E}}=\int_{\tau-\eta}^{\tau}f_0(x)\mathrm{d}x+\int_{\tau-\theta-\eta}^{\tau-\theta}f_0(x)\mathrm{d}x$$
$$=(\phi(\tau-\theta)+\phi(\tau))-(\phi(\tau-\theta-\eta)+\phi(\tau-\eta))>0 \tag{29}$$

因此，在理论上，隐写分析者对于变换图像的实际分析情况为：虚警率将增高、漏警率将降低、总检错率仍然增高。根据文献[19]，在真实场景中，载体对象的数量通常远远大于载密对象。因此，检测虚警率的提高将意味着，大量被误判为载密的载体对象则会把隐写分析系统淹没，使隐写分析者的检测性能急剧下降。

4 实验方法

4.1 实验对象选择

BOSSbase 1.01 是目前隐写算法最常使用的图像库，它包含了 10 000 张大小为 512 pixel×512 pixel 的 RAW 格式灰度图片，所以，为了便于与经典的隐写算法进行比较，本文采用空域 PGM 格式的 BOSSbase 图像库（即 10 000 张大小为 8 bit 的 512 pixel×512 pixel 的 PGM 格式灰度图片）。为了控制整体图像库的变换幅度在“微扰”的条件下，本文分别选取微扰参数 $\delta\in\{0.05,0.1,0.2\}$ 进行实验。

4.2 隐写及隐写分析方法选择

本文采用的隐写算法为自适应的空域经典隐写算法 SUNIWARD，分别在嵌入率为 α=0.1 bit/pixel、α=0.2 bit/pixel、α=0.3 bit/pixel、α=0.4 bit/pixel 以及 α=0.5 bit/pixel 情况下隐写。

对于隐写分析，采用的图像提取特征为 34 671 维度的 SRM 特征[16]，选用的隐写分类器为目前隐写分析中常用的 ensemble 分类器[17]（版本为 2.0，使用默认设置）。SRM（spatial rich model）首先计算 22 个一阶及 22 个三阶残差矩阵、12 个二阶残差矩阵、2 个 SQUARE 残差矩阵、10 个 EDGE 3×3 及 10 个 EDGE 5×5 残差矩阵。共计 22+22+12+2+10+10=78 个残差矩阵。分别计算上述残差矩阵的四阶马尔可夫特征，范围参数 $T=2$，即每个残差矩阵有 $(2T+1)^4=625$ 维。利用符号对称性及方向对称性降低残差矩阵个数及特征维度。可将一阶及三阶残差矩阵降至 12 个，二阶残差矩阵降至 7 个，SQUARE 残差矩阵降至 2 个，EDGE 3×3 及 EDGE 5×5 残差矩阵降至 6 个。可将 12 个一阶特征降至 169 维，其他 33 个特征降至 325 维，共计 12×169+33×325=12 753 维。上述 12 753 维度的特征采用步长 $q=1$ 进行量化，若量化步长按照式(30)确定，则可以得到 2×(2×169+10×325)+3×(10×169+23×325)=34 671 维的 SRM 特征，其中，c 为残差矩阵阶数。

$$q\in\begin{cases}\{c,1.5c,2c\}, & c>1\\ \{1,2\}, & c=1\end{cases} \tag{30}$$

按照隐写分析中的一般性的安全性能衡量方法，P_{FA} 为检测结果的虚警概率，P_{MD} 为检测结果的漏警概率，使用二者均值表示隐写分析的检错概率 P_{E}。

5 实验分析

5.1 参数扰动对隐写图像特征的影响

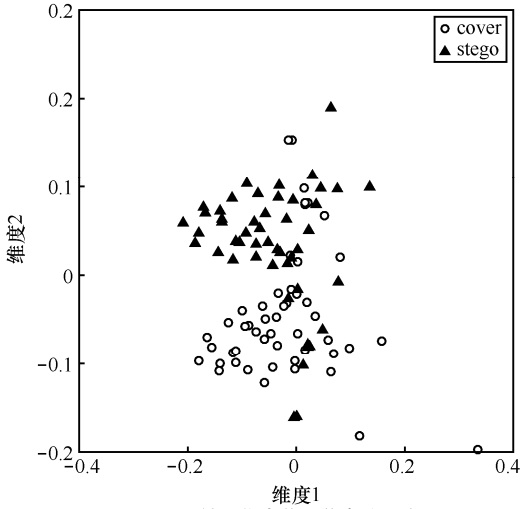
为了验证图像变换的参数扰动对隐写及隐写分析的影响，鉴于分类器的训练和测试都是基于图像特征，所以本文首先仍以 Gamma 变换为例，定性分析 Gamma 变换参数扰动对于图像集特征的影响。

1) 从原始图像库 BOSSbase1.01 中随机选取 50 张图片作为载体图像 cover，使用 SUNIWARD 算法在 0.5 bit/pixel 嵌入率下生成对应载密图像 stego，对载体载密图像分别提取 686 维的 SPAM 特征，然后分别应用 PCA 主成分降维，取前 2 维特征绘制特征空间点，如图 2(a)所示。

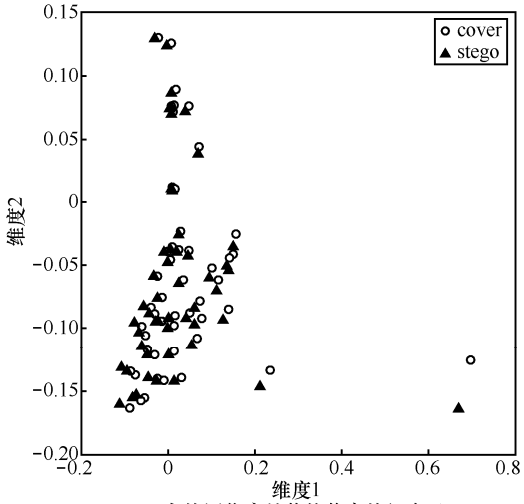
2) 对 1)中的 50 张图片随机作扰动参数 $\delta=0.2$ 的 Gamma 变换，即对 1)中的每张图片随机作参数 $\xi\in[0.8,1.2]$ 的 Gamma 变换，作为新的载体图像 cover，使用 SUNIWARD 算法在 0.5 bit/pixel 嵌入率下生成对应新的载密图像 stego，对载体载密图像分别提取 686 维的 SPAM 特征，然后分别应用 PCA 主成分降维，取前 2 维特征绘制特征空间点，如图 2(b)所示。

可以看出，1) 中载体 cover 和载密 stego 的空间特征点有较明显的分类平面，cover 特征集中于图像下方，stego 特征集中于图像上方。而 2) 中载体 cover 和载密 stego 的空间特征点的分布更具有空间随机性，二者之间也不再有较明显的分类平面，stego 特

征相对于 cover 而言有随机方向上不同程度的偏移，直观来看，参数扰动造成失配的特征对于分类器来说将更难以分辨，从而可以提高隐写的安全性能。



(a) 原始图像库载体载密特征表示



(b) 变换图像库的载体载密特征表示

图 2　载体载密特征表示

5.2　灰度线性变换的参数扰动

基于 5.1 节的定性分析，下面本文将分别针对隐写分析者在灰度线性变换和 Gamma 变换做定量的隐写分析实验。2 种情况下，隐写分析者都使用原始图像库 $S_0(\delta=0)$ 进行隐写得到分析者的 10 000 对载体载密图像集，提取 SRM 特征，训练使用 ensemble 分类器做隐写分析，测试隐写者的载密图像集，并与直接在原始图像库 $S_0(\delta=0.0)$ 隐写进行比较。

对于隐写者，分别选取扰动参数 $\delta\in\{0.05,0.1,0.2\}$，对原始图像库 $S_0(\delta=0.0)$ 的每张图片作随机灰度线性变换，在 0.1~0.5 bit/pixel 的嵌入率下使用 SUNIWARD 算法生成 10 000 对载密图像集，再进行隐写分析实验得到实验数据，表 1 为检测虚警率，表 2 为检测漏警率，表 3 为整体检错率。

表 1　灰度线性变换参数扰动下的隐写分析虚警率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.417 9	0.329 3	0.272 9	0.218 2	0.179 8
δ=0.05	0.531 4	0.452 9	0.355 8	0.281 5	0.216 2
δ=0.10	0.596 4	0.538 9	0.436 1	0.353 0	0.272 1
δ=0.20	0.625 6	0.591 4	0.486 4	0.402 3	0.318 2

表 2　灰度线性变换参数扰动下的隐写分析漏警率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.388 6	0.307 0	0.237 1	0.189 4	0.144 9
δ=0.05	0.311 0	0.233 8	0.191 1	0.156 2	0.120 9
δ=0.10	0.271 8	0.194 0	0.163 6	0.131 9	0.106 3
δ=0.20	0.263 3	0.185 8	0.162 0	0.132 7	0.108 2

表 3　灰度线性变换参数扰动下的隐写分析检错率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.403 2	0.318 2	0.255 0	0.203 8	0.162 3
δ=0.05	0.421 2	0.343 3	0.273 5	0.218 8	0.168 5
δ=0.10	0.434 1	0.366 5	0.299 9	0.242 5	0.189 2
δ=0.20	0.444 4	0.388 6	0.324 2	0.267 5	0.213 2

将表 3 的检错率数据整理后绘制得到图 3。

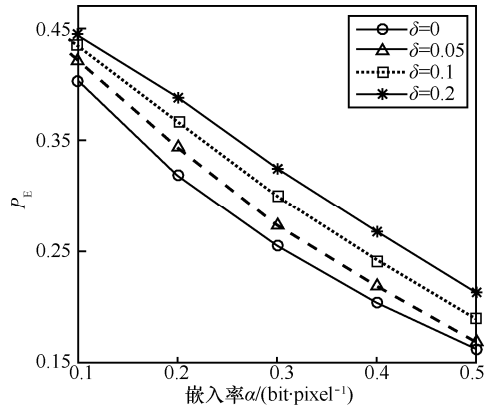


图 3　灰度线性变换参数扰动下的隐写分析

综合表 3 和图 3 可以看出，使用参数扰动后的灰度线性变换图像库作载体图像集进行隐写，安全性有了较大提高，且隐写分析者的检错率随着扰动参数的增大而增大，这是因为随着扰动参数的增大，图像集失配的程度也更加严重，隐写分析者的

门限函数的偏移值η_0也随之增大。

数据也表明，当对原始图像库进行扰动参数的灰度线性变换时，隐写分析者检测时的虚警率增高、漏警率降低、整体检错率提升，且在一定范围内，这种差值会随着扰动范围的增大而增大，这与 3.3 节的参数扰动模型相拟合。在$\delta=0.05$时检错率大约提高 0.6%~3%，在 δ=0.10 时检错率大约提高 3%~5%，在 δ=0.20 时检错率大约提高 4%~7%。

将 δ=0.20 时的微扰与经典隐写算法进行对比，如图 4 所示。

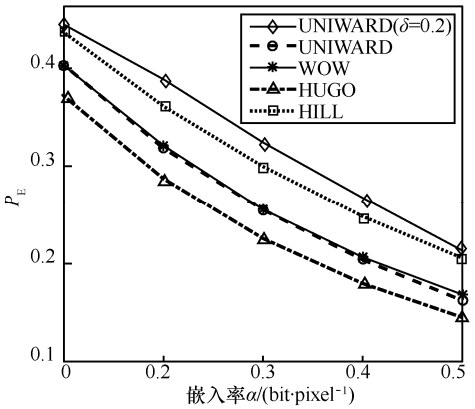


图 4　灰度线性变换微扰与经典隐写算法对比

可以看出，UNIWARD 与 WOW 有着相似的安全性能，它们都要优于 HUGO，HILL 在 4 种经典隐写算法中安全性最高，$\delta=0.2$时的线性变换微扰情况在同等嵌入率下检错率都高于 HILL 算法。

5.3　Gamma 变换的参数扰动

对于隐写者，分别选取扰动参数 $\delta\in\{0.05, 0.1, 0.2\}$，对原始图像库 $S_0(\delta=0)$的每张图片作随机 Gamma 变换，在 0.1~0.5 bit/pixel 的嵌入率下使用 SUNIWARD 算法生成 10 000 对载密图像集，再进行隐写分析实验得到实验数据，表 4 为检测虚警率，表 5 为检测漏警率，表 6 为整体检错率。

表 4　Gamma 变换参数扰动下的隐写分析虚警率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.417 9	0.329 3	0.272 9	0.218 2	0.179 8
δ=0.05	0.553 3	0.469 6	0.365 6	0.291 3	0.229 0
δ=0.10	0.621 0	0.555 0	0.433 2	0.353 0	0.274 5
δ=0.20	0.642 1	0.597 9	0.471 4	0.391 1	0.301 2

表 5　Gamma 变换参数扰动下的隐写分析漏警率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.388 6	0.307 0	0.237 1	0.189 4	0.144 9
δ=0.05	0.292 5	0.223 3	0.183 4	0.147 9	0.117 2
δ=0.10	0.250 0	0.182 8	0.157 2	0.125 3	0.099 2
δ=0.20	0.245 5	0.173 4	0.155 7	0.126 8	0.107 2

表 6　Gamma 变换参数扰动下的隐写分析检错率

扰动参数	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5
δ=0	0.403 2	0.318 2	0.255 0	0.203 8	0.162 3
δ=0.05	0.422 9	0.346 5	0.274 5	0.219 6	0.173 1
δ=0.10	0.435 5	0.368 9	0.295 2	0.239 2	0.186 9
δ=0.20	0.443 8	0.385 6	0.313 6	0.259 0	0.204 2

将表 6 的数据整理后绘制得到图 5。

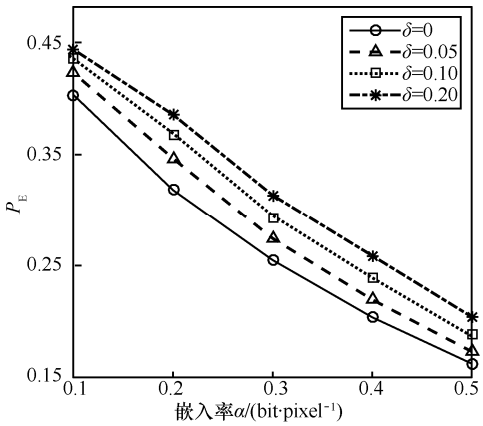


图 5　Gamma 变换参数扰动下的隐写分析

综合表 6 和图 5 可以看出，与 5.2 节实验结果类似，使用参数扰动后的 Gamma 变换图像库作载体图像集进行隐写，安全性有了较大提高，且隐写分析者的检错率随着扰动参数的增大而增大，这也表明随着扰动参数的增大，图像集失配的程度也更加严重，隐写分析者的门限函数的偏移值η_0也随之增大。

数据也表明，当对原始图像库进行扰动参数的 Gamma 变换时，隐写分析者检测时的虚警率增高，漏警率降低，整体检错率提升，且在一定范围内，这种差值整体会随着扰动范围的增大而增大，这与 3.3 节的参数扰动模型相拟合。在$\delta=0.05$时检错率大约提高 1%~3%，在$\delta=0.1$时检错率大约提高 2%~5%，在$\delta=0.2$时检错率大约提高 4%~7%。

将$\delta=0.2$时的 Gamma 变换微扰与经典隐写算法进行对比得到图 6。

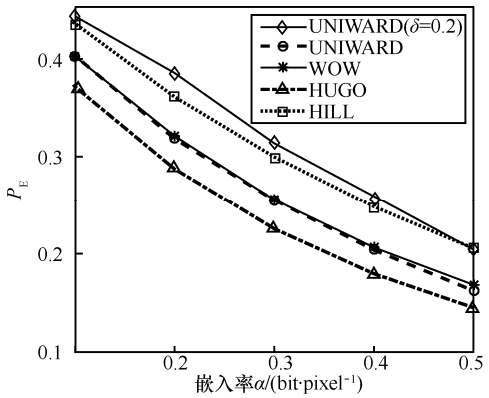


图 6　Gamma 变换微扰与经典隐写算法对比

与图 4 的结论类似，$\delta = 0.2$ 时的 Gamma 变换微扰情况在同等嵌入率下检错率都高于 HILL 等经典隐写算法。

6　结束语

本文以 BOSSbase1.01 图像库为基础，以图像处理中的灰度线性变换和 Gamma 变换为例，采用 SUNIWARD 算法，推导了图像变换对于隐写的参数扰动模型，详细探讨了参数扰动对于隐写及隐写分析的影响。经过参数扰动的图像变换，能使载体载密图像在特征空间上的分布更加随机均匀，把隐写带来的图像噪声隐藏在了因参数微扰而带来的图像像素波动中，引入了图像集的失配因素，提高了隐写分析者的检测难度，通过实验验证了其检测结果的虚警率将增高，漏警率将降低，总的检错率仍然增高。而且在一定范围内，隐写分析者的检错率将随着波动参数的增大而增大。

同时，灰度线性变换和 Gamma 变换也是最为基础的图像处理之一，因此，在接下来的工作中，本文将进一步探讨这种参数扰动带来的图像集失配性质是否也适用于对比度增强、风格变换等其他图像处理操作。

附录　式(23)和式(25)的证明

1) 式(23)的证明

由文献[15]，对于式(10)和式(11)

$$\mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n] = 0 \tag{31}$$

$$Var_{H_0}[\mathrm{lb}\Lambda_n] = \frac{2\gamma_n^2}{\sigma_n^4} \tag{32}$$

$$\mathrm{lb}\Lambda_n = \gamma_n\left(-\frac{1}{\sigma_n^2} + \frac{x_n^2}{\sigma_n^4}\right) \tag{33}$$

$$\theta = \frac{2\sum_{n=1}^{N}\sigma_n^{-4}\beta_n\gamma_n}{\sqrt{2\sum_{n=1}^{N}\sigma_n^{-4}\gamma_n^2}} \tag{34}$$

在 H_0' 假设下，$\dfrac{x_n^2}{\alpha_n\sigma_n} \sim N(0,1)$，所以 $\dfrac{x_n^2}{\alpha_n^2\sigma_n^2} \sim \chi_1^2$，根据 $\mathrm{E}[\chi_1^2] = 1$，有

$$\mathrm{E}_{H_0'}\left[\frac{x_n^2}{\sigma_n^4}\right] = \frac{\alpha_n^2}{\sigma_n^2} \tag{35}$$

所以，

$$\mathrm{E}_{H_0'}\left[\mathrm{lb}\Lambda_n\right] = \frac{\alpha_n^2 - 1}{\sigma_n^2} \tag{36}$$

$$\frac{\sum_{n=1}^{N}(\mathrm{E}_{H_0'}[\mathrm{lb}\Lambda_n] - \mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N}Var_{H_0}[\mathrm{lb}\Lambda_n]}} = \frac{\sum_{n=1}^{N}\sigma_n^{-2}\gamma_n(\alpha_n^2 - 1)}{\sqrt{2\sum_{n=1}^{N}\sigma_n^{-4}\gamma_n^2}} \triangleq \eta \tag{37}$$

故而在 H_0' 假设下，有

$$\begin{aligned}\mathrm{lb}\Lambda^*(\boldsymbol{x},\sigma) &= \frac{\sum_{n=1}^{N}(\mathrm{lb}\Lambda_n - \mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N}Var_{H_0}[\mathrm{lb}\Lambda_n]}} \\ &\to N(\eta,1)\end{aligned} \tag{38}$$

同理，在 H_1' 假设下，有

$$\begin{aligned}\mathrm{lb}\Lambda^*(\boldsymbol{x},\sigma) &= \frac{\sum_{n=1}^{N}(\mathrm{lb}\Lambda_n - \mathrm{E}_{H_0}[\mathrm{lb}\Lambda_n])}{\sqrt{\sum_{n=1}^{N}Var_{H_0}[\mathrm{lb}\Lambda_n]}} \\ &\to N(\theta+\eta,1)\end{aligned} \tag{39}$$

2) 式(25)的证明

当 $\alpha_n = \xi_n$，是在 $\xi_n \in [\xi_0 - \delta, \xi_0 + \delta]$ 上的随机分布且与 $\xi_0 = 1$ 时，ξ_n 的概率密度函数为 $\dfrac{1}{2\delta}$，对 $\mathrm{E}_{H_0'}\left[\mathrm{lb}\Lambda_n\right]$ 求变量 ξ_n 的期望值

$$\begin{aligned}\mathrm{E}_{\xi_N}\left[\mathrm{E}_{H_0'}\left[\mathrm{lb}\Lambda_n\right]\right] &= \frac{1}{2\delta}\int_{1-\delta}^{1+\delta}\frac{\xi_n^2 - 1}{\sigma_n^2}\mathrm{d}\xi_n \\ &= \frac{1}{3\sigma_n^2}\delta^2\end{aligned} \tag{40}$$

即，η 对于 α_n 的平均期望值为

$$\eta_0 = \mathrm{E}_{\alpha}(\eta) = \frac{\sum_{n=1}^{N}\sigma_n^{-2}\gamma_n}{3\sqrt{2\sum_{n=1}^{N}\sigma_n^{-4}\gamma_n^2}}\delta^2 \tag{41}$$

参考文献：

[1] FRIDRICH J. Steganography in digital media: principles, algorithms, and applications[M]//Steganography in Digital Media : Principles, Algorithms, and applications. Cambridge University Press, 2010:1-20.

[2] CRANDALL R. Some notes on steganography[J]. Posted on Stegano-Graphy Mailing List, 1998.

[3] PEVNÝ T, BAS P, FRIDRICH J. Steganalysis by subtractive pixel adjacency matrix[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 215-224.

[4] FRIDRICH J, KODOVSKÝ J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.

[5] SHI Y Q, SUTTHIWAN P, CHEN L. Textural features for steganalysis[C]//Information Hiding. 2012: 63-77.

[6] FILLER T, JUDAS J, FRIDRICH J. Minimizing additive distortion in steganography using syndrome-trellis codes[J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 920-935.

[7] PEVNÝ T, FILLER T, BAS P. Using high-dimensional image models to perform highly undetectable steganography[C]//Information Hiding. Springer Berlin Heidelberg. 2010: 161-177.

[8] HOLUB V, FRIDRICH J. Designing steganographic distortion using directional filters[C]//IEEE International Workshop on Information 2013: 234-239.

[9] HOLUB V, FRIDRICH J. Digital image steganography using universal distortion[C]//The First ACM Workshop on Information Hiding and Multimedia Security. 2013: 59-68.

[10] LI B, WANG M, LI X, et al. A strategy of clustering modification directions in spatial image steganography[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(9): 1905-1917.

[11] DENEMARK T, FRIDRICH J. Improving steganographic security by synchronizing the selection channel[C]//The 3rd ACM Workshop on Information Hiding and Multimedia Security. 2015: 5-14.

[12] GUO L, NI J, SHI Y Q. Uniform embedding for efficient JPEG steganography[J]. IEEE Transactions on Information Forensics and Security, 2014, 9(5): 814-825.

[13] BAS P, FILLER T, PEVNÝ T. Break our steganographic system: the Ins and outs of organizing BOSS[M]// Information Hiding. 2011: 59-70.

[14] FRIDRICH J. Study of cover source mismatch in steganalysis and ways to mitigate its impact[J]. SPIE-International Society for Optical Engineering, 2014, 9028(2):96-101.

[15] SEDIGHI V, COGRANNE R, FRIDRICH J. Content-adaptive steganography by minimizing statistical detectability[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(2): 221-234.

[16] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics & Security, 2012, 7(3): 868-882.

[17] KODOVSKY J, FRIDRICH J, HOLUB V. Ensemble classifiers for steganalysis of digital media[J]. IEEE Transactions on Information Forensics & Security, 2012, 7(2):432-444.

[18] SEDIGHI V, FRIDRICH J, COGRANNE R. Toss that BOSSbase, Alice![J]. Electronic Imaging, 2016.

[19] PEVNÝ T, KER A D. Towards dependable steganalysis[J]. SPIE-International Society for Optical Engineering, 2015, 9409: 94090I- 94090I-14.

作者简介：

孙曦（1994-），男，安徽阜阳人，中国科学技术大学硕士生，主要研究方向为信息隐藏、计算机视觉等。

张卫明（1976-），男，河北保定人，中国科学技术大学教授，主要研究方向为信息隐藏、数字内容安全。

俞能海（1964-），男，安徽无为人，中国科学技术大学教授，主要研究方向为多媒体数据处理与分析、数字内容安全。

魏尧（1992-），男，陕西宝鸡人，中国科学技术大学硕士生，主要研究方向为网络安全、信息隐藏等。