基于图像高阶 MARKOV 链模型的扩频隐写分析

张 湛[1,2],刘光杰[1],王俊文[1],戴跃伟[1],王执铨[1]
(1.南京理工大学自动化学院,江苏南京 210094;2.重庆电子工程职业学院计算机应用系,重庆 401331)

摘 要: 扩频隐写分析是信息隐藏研究领域的一个重要方面.文章提出基于高阶 Markov 链的数字图像统计分布模型,在对常用图像扫描方法构成高阶 Markov 链的效果进行比较后,采用 Hilbert 扫描方式构建数字图像 n 阶 Markov 链模型,进而提出度量数字图像隐写统计安全性的 n 阶 Markov 链测度,并证明其有界.最后文章通过研究扩频隐写对高阶 Markov 链模型经验矩阵的影响,利用该模型提取图像统计特征,并使用支持向量机对几种常用图像扩频隐写方法进行分析.实验说明文章所提方法对扩频隐写分析效果良好,且随着模型阶数提高,分析准确率也随之提高.

关键词: 信息隐藏; 隐写分析; 高阶马尔科夫链; 扩频图像隐写
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112 (2010) 11-2578-07

Steganalysis of Spread Spectrum Image Steganography Based on High-Order Markov Chain Model

ZHANG Zhan[1,2], LIU Guang-jie[1], WANG Jun-wen[1], DAI Yue-wei[1], WANG Zhi-quan[1]
(1. *School of Automation, Nanjing University of Science and Technology. Nanjing, Jiangsu* 210094, *China*;
2. *Department of Computer Application, Chongqing College of Electronic Engineering, Chongqing* 401331, *China*)

Abstract: Steganalysis of spread spectrum image steganography is an important aspect in the domain of information hiding research. An image statistical distribution model based on high-order Markov chain was proposed. By a comparison on the effect of n-order Markov chain achieved by several general image scanning methods, Hilbert scanning method was used to construct the proposed model. After that, a n-order statistic security measure for digital image steganography based on the model was proposed, and that it is bounded was proved. Finally, after the effect of spread spectrum steganography on empirical matrix of the proposed model was studied, image statistical features which based on the proposed model were extracted, and several common spread spectrum steganography algorithms were analyzed using support vector machine. Experiment shows the active effects of steganalysis of spread spectrum image steganography based on the proposed model is obvious, furthermore, the steganalysis accuracy is improved as the model order increased.

Key words: information hiding; steganalysis; high-order Markov chain; spread spectrum image steganography

1 引言

近年来出现了许多新的数字图像隐写算法[1,2].在各种隐写算法中,扩频调制是一种重要的隐写方式,Cox 等[3]首先提出载体频域信息隐藏扩频调制方法,Marvel 等[4]则提出图像扩频隐写(Spread Spectrum Image Steganography, SSIS)方法,随后 Fridrich 等[5]提出的随机调制隐写算法大大增加了 SSIS 的嵌入容量,使其应用更为广泛,反之对其的隐写分析也更加重要.

针对 SSIS 的分析不少研究者做出了重要工作,Harmsen 等[6]分析了加性隐写的噪声模型,对加性 SSIS 进行了分析;Wang 等[7]利用高斯平稳过程描述图像模型,对加性 SSIS 进行了分析;Chandramouli 等[8]提出一种针对 SSIS 的主动分析方案;Ji 等[9]利用块间分布差异对分块 DCT 域 SSIS 进行了分析.

Sullivan 等[10,11]则考虑到自然图像像素间存在较强相关性,以及计算复杂度等原因,没有使用独立同分布(Independent and Identically Distributed, i.i.d.)模型描述自然图像,而对更符合自然图像相关性统计分布的 Markov 随机场模型进行简化,提出数字图像 Markov 链(Markov Chain, MC)模型,并基于该模型对空域和频域的加性与乘性 SSIS 进行分析,取得了较好效果.图像 MC 模型将自然图像简化为无后效应的 MC, MC 中任一位均只受前一位的影响, MC 模型虽包含了一定载体相关性信息,但由于通常自然图像每一像素点均至少与其 8 临域像素点有较强相关性,因此使用 MC 模型仍将丢失

稿日期:2009-09-10;修回日期:2010-02-08
基金项目:江苏省自然科学基金(No.BK201022321);江苏省自然科学基金(No.BK2008403);中国博士后基金(No.20070421017);教育部博士点基金(No.20093219120037);南京理工大学自主科研计划(No.2010ZYTS048)

载体大量相关性信息,虽利用 MC 模型提取载体相关性信息相比 Markov 随机场模型,计算复杂度大幅降低,但当分析者计算能力增强时,使用 MC 模型进行分析不能进一步提高分析的准确率.

本文以 n 阶 Markov 链(n-MC)[12]作为数字图像统计分布模型,在研究了几种常用图像扫描方法所构成 n-MC,对相邻像素相关性信息量的包含程度后,采用 Hilbert 扫描方式构建图像 n-MC 模型,利用统计分布散度距离提出图像隐写统计安全性的 n-MC 测度,并证明其有界及在特定条件下与 ε-secure 安全性指标[13]和图像 MC 模型统计分布测度[11]等价.最后本文通过研究加性和乘性 SSIS 对载体 n-MC 经验矩阵的影响,利用 n-MC 模型提取图像特征,对空域和 DCT 频域的加性和乘性 SSIS 进行分析.实验说明基于载体 n-MC 模型对 SSIS 的分析效果良好,且模型阶数提高时,分析准确率也随之提高,虚警率则随之降低.由于图像隐写 n-MC 模型计算复杂度可调,因此基于该模型对 SSIS 的分析更适用于计算能力不同的分析者.

2　数字图像 n-MC 模型

利用隐写对自然图像像素间相关性的影响对载密图像进行分析,需使用适当模型反映像素间的相关性情况.使用的模型应包含足够多的像素相关性信息,且计算复杂度低,以满足分析者实际计算能力和对大量载体隐写分析的要求.由于采用 n-MC 作为数字图像统计分布模型具有可根据计算复杂度和包含载体相关性信息程度的具体要求,通过改变 n 的大小来调节模型复杂度的优点,因此该模型更适合不同计算能力的分析者根据自身需求和实际情况采用.

2.1　数字图像 n-MC 及其经验矩阵的构建

以灰度图像 $\boldsymbol{B}$ 为例,对 $\boldsymbol{B}$ 按某种方式扫描得到 n 阶 Markov 链 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_t,\cdots,x_L\}$,$x_t\in\boldsymbol{G}$,$L$ 为链长,$\boldsymbol{G}$ 为 x_t 的可能取值集合(如 8 位灰度图像为 0 ~ 255),$\boldsymbol{X}$ 中元素 x_t 满足 $P(x_t\mid x_{t-1},x_{t-2},\cdots,x_1)=P(x_t\mid x_{t-1},x_{t-2},\cdots,x_{t-n})$,即 x_t 只与 x_{t-1}、x_{t-2}、$\cdots$、x_{t-n} 有关.

定义 $\eta_{i_1,i_2,\cdots,i_{n+1}}(\boldsymbol{X})$ 为 $\boldsymbol{X}$ 中数据从 i_1 经 i_2、i_3 等状态到达 i_{n+1} 的变换次数,即 $x_t=i_{n+1},x_{t-1}=i_n,\cdots,x_{t-n}=i_1$ 的出现次数;则 $\boldsymbol{X}$ 的经验矩阵 $\boldsymbol{M}^{\boldsymbol{X}}\triangleq\left\{m_{i_1,i_2,\cdots,i_{n+1}}=\dfrac{\eta_{i_1,i_2,\cdots,i_{n+1}}(\boldsymbol{X})}{L-n},i_k\in\boldsymbol{G}\right\}$,其中 $\dfrac{\eta_{i_1,i_2,\cdots,i_{n+1}}(\boldsymbol{X})}{L-n}$ 表明 $\boldsymbol{X}$ 中像素值从 i_1 经 i_2、i_3 等状态最终到达 i_{n+1} 的变换在总的像素变换中所占比例,即 $\boldsymbol{M}^{\boldsymbol{X}}$ 提供了对联合分布概率 $P(x_t=i_{n+1},x_{t-1}=i_n,\cdots,x_{t-n}=i_1)$ 的估计.

图 1 以二值图像的 2 阶 MC 为例说明图像 n 阶 MC 及其经验矩阵的构建.

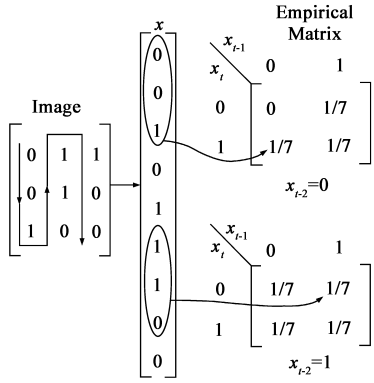

图1　2-MC及其经验矩阵构成的简单例举

对该二值图像按列扫描(此处为说明简便对图像按列扫描)得到数据链 $\boldsymbol{X}=001011100$.

若设定数据链 $\boldsymbol{X}$ 为 2-MC(即 x_t 由 x_{t-1}和 x_{t-2}唯一决定),则 $\boldsymbol{X}$ 的 2-MC 经验矩阵为 $\boldsymbol{M}^{\boldsymbol{X}}_{2\times2\times2}$,$\boldsymbol{M}^{\boldsymbol{X}}_{2\times2\times2}$中元素 $m^{\boldsymbol{X}}_{i_1,i_2,i_3}=\dfrac{\eta_{i_1,i_2,i_3}}{9-2}=\dfrac{\eta_{i_1,i_2,i_3}}{7}$.

2.2　构建图像 n-MC 的扫描方式选择

通常自然图像像素点间存在较强相关性,且像素点越临近,相关性一般越强.由于构建图像 n-MC 的目的是利用像素点间相关性分析隐写对图像高阶统计分布的影响,而图像 n-MC 定义为当前像素点只与数据链中前 n 个像素点相关,因此将图像扫描成 n-MC 时,应选择能在链中尽量将相邻像素排列在一起的扫描方式.

目前较常用的图像扫描方式主要有列(行)扫描、zig-zag 扫描及 Hilbert 扫描等.设 $x_{j,k}$为图像像素点,假设该点只与周围 8 个像素点以及这 8 个点外围的 16 个像素点相关,如表 1 所示.

表 1　与像素点 $x_{j,k}$相关的像素点

$x_{j-2,k-2}$	$x_{j-2,k-1}$	$x_{j-2,k}$	$x_{j-2,k+1}$	$x_{j-2,k+2}$
$x_{j-1,k-2}$	$x_{j-1,k-1}$	$x_{j-1,k}$	$x_{j-1,k+1}$	$x_{j-1,k+2}$
$x_{j,k-2}$	$x_{j,k-1}$	$x_{j,k}$	$x_{j,k+1}$	$x_{j,k+2}$
$x_{j+1,k-2}$	$x_{j+1,k-1}$	$x_{j+1,k}$	$x_{j+1,k+1}$	$x_{j+1,k+2}$
$x_{j+2,k-2}$	$x_{j+2,k-1}$	$x_{j+2,k}$	$x_{j+2,k+1}$	$x_{j+2,k+2}$

按列(行)扫描和 zig-zag 扫描图像构成数据链中位于元素 x_t(除非 x_t 位于边角部分)以前且与 x_t 紧邻的几个元素,只包含 x_t 在原图位置中与其紧邻的 8 个点中一个和外围 16 个点中一个.由于如图 2 所示 Hilbert 空间填充曲线扫描路径的特殊性,从直观可知,对采用 Hilbert 扫描图像所构成数据链中元素 x_t,位于 x_t 以前且与 x_t 紧邻的几个元素,包含 x_t 在原图位置中与其紧邻的 8 个点以及外围 16 个点中更多的像素点,即采用 Hilbert 扫描图像相较列(行)扫描和 zig-zag 扫描,n-MC 包含更多图像相关性信息,且在一定范围内 n-MC 的阶

数越高优势越显著,文献[14]使用该扫描方式设计隐写算法取得了较好效果.

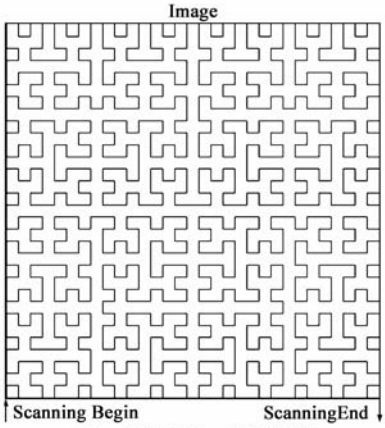

图2 图像Hilbert扫描路径

为简便,不妨设 $x_{j,k}$ 周围 8 个像素点与 $x_{j,k}$ 的相关度相同,外围 16 个像素点与 $x_{j,k}$ 的相关度相同但只为 $x_{j,k}$ 周围 8 个点与 $x_{j,k}$ 相关度的一半.由于 $x_{j,k}$ 只与这 24 个点相关,不妨设 $x_{j,k}$ 周围 8 个点与 $x_{j,k}$ 相关度的权重为 0.0625,而外围 16 个点与 $x_{j,k}$ 相关度的权重为 0.03125,其余像素点与 $x_{j,k}$ 相关度的权重为 0.

分别按列扫描、zig-zag 扫描和 Hilbert 扫描将 512×512 的图像扫描成数据链 $\boldsymbol{X}$,以 2-MC 研究扫描方式对包含图像相关性的影响,对 $\boldsymbol{X}$ 中每一点 x_t,观察 x_{t-1} 和 x_{t-2} 在原图中相对 x_t 的位置,并将所有点的相关度权重相加,所得之和 Σ 即代表了 $\boldsymbol{X}$ 包含的图像相关度信息的大小.2-MC 各扫描方式的 Σ 如下:

按列扫描:$\Sigma = 2.4608\mathrm{e}+004$;

按 zig-zag 扫描:$\Sigma = 2.4608\mathrm{e}+004$;

按 Hilbert 扫描:$\Sigma = 3.1129\mathrm{e}+004$;

对于 2 阶 MC 按列扫描与按 zig-zag 扫描所得 2-MC 包含的图像相关性信息量相同,而按 Hilbert 扫描所得 2-MC 包含的图像相关性信息量远多于其余两种扫描方式.对于 n 阶 MC,随着 n 的增加,按 Hilbert 扫描所得 n-MC 包含的图像相关性信息也会相应增加,但从 Hilbert 扫描路径可知,当 $n>4$ 时,n 的增加对 n-MC 所含图像相关性信息影响不大.因此采用 n-MC 模型分析隐写对图像高阶统计分布的影响时,应采用 Hilbert 扫描方式,考虑到 n 的增加对 n-MC 所含图像相关性信息量的增加以及对计算复杂度的影响,一般应采用 4 阶以下的 n-MC 模型.

2.3 数字图像 n-MC 模型的统计测度

由于隐写导致图像 n-MC 模型经验矩阵统计分布发生不同程度的改变,因此可采用该模型度量隐写对载体高阶统计分布的影响程度.

设 $\boldsymbol{X}$ 和 $\boldsymbol{S}$ 分别为载体和载密图像的 n-MC,$\boldsymbol{M}^{\boldsymbol{X}}_{i_1,i_2,\cdots,i_{n+1}}$ 和 $\boldsymbol{M}^{\boldsymbol{S}}_{i_1,i_2,\cdots,i_{n+1}}$ 分别为 $\boldsymbol{X}$ 和 $\boldsymbol{S}$ 的经验矩阵,$m^{\boldsymbol{X}}_{i_1,i_2,\cdots,i_{n+1}}$ 和 $m^{\boldsymbol{S}}_{i_1,i_2,\cdots,i_{n+1}}$ 为矩阵元素;$i_1,i_2,\cdots,i_{n+1}$ 为图像像素值,G 为 $i_1,i_2,\cdots,i_{n+1}$ 所有可能取值集合 $i_1,i_2,\cdots,i_{n+1}\in G$.

利用载体和载密图像 n-MC 经验矩阵的散度距离可反映隐写前后载体高阶统计分布的改变,考虑到 n-MC 模型在 n 为 1 时以及假设图像像素分布为 i.i.d.时分别与 MC 安全性指标[11]和 ε-secure 指标[13]相符,因此利用散度距离建立数字图像隐写 n-MC 统计测度 $D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})$.根据相关熵定义,此处规定 $0\times\log 0^{-1}=0$.

$$D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}}) \triangleq \sum_{i_1,i_2,\cdots,i_{n+1}\in G}\left\{m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}\cdot\log\left[\frac{m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}}{\sum\limits_{i_1}m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}}\left(\frac{m^{\boldsymbol{S}}_{i_1i_2\cdots i_{n+1}}}{\sum\limits_{i_1}m^{\boldsymbol{S}}_{i_1i_2\cdots i_{n+1}}}\right)^{-1}\right]\right\} \tag{1}$$

图像隐写 n-MC 统计测度 $D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})$ 提供了隐写对载体高阶统计分布的改变程度,即反映了载密图像本身固有的被分析者检测出的可能性.在实际运用中,隐写者可选择或设计使 $D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})$ 最小的隐写方案,也可在隐写方案一定的情况下选择 $D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})$ 最小的隐写载体;分析者则可根据隐写对载体 n-MC 模型经验矩阵的影响设计合适的分析方案.

根据 n-MC 经验矩阵的构成可知 $m_{i_1,i_2,\cdots,i_{n+1}}$ 为 n-MC 中像素值从 i_1 经 i_2、i_3 等状态最终到达 i_{n+1} 的联合概率分布,因此可得定理如下.

定理 1:令 $i_1,i_2,\cdots,i_{n+1}\in G$,$m^{\boldsymbol{X}}_{i_1,i_2,\cdots,i_{n+1}}$ 和 $m^{\boldsymbol{S}}_{i_1,i_2,\cdots,i_{n+1}}$ 分别为载体与载密图像 n-MC 经验矩阵 $\boldsymbol{M}^{\boldsymbol{X}}$ 和 $\boldsymbol{M}^{\boldsymbol{S}}$ 中元素.则

$$D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})\geqslant 0$$

且等号成立的充要条件为 $m^{\boldsymbol{X}}_{i_1,i_2,\cdots,i_{n+1}}$ 与 $m^{\boldsymbol{S}}_{i_1,i_2,\cdots,i_{n+1}}$ 处处相等.

证明:令 $A=\{(i_1,i_2,\cdots,i_{n+1})\mid m_{i_1,i_2,\cdots,i_{n+1}}>0\}$ 为 $m_{i_1,i_2,\cdots,i_{n+1}}$ 的支撑集.则根据式(1)有

$$-D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})=\sum_{i_1,i_2,\cdots,i_{n+1}\in A}\left\{m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}\cdot\log\left[\frac{m^{\boldsymbol{S}}_{i_1i_2\cdots i_{n+1}}}{m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}}\frac{\sum\limits_{i_1}m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}}{\sum\limits_{i_1}m^{\boldsymbol{S}}_{i_1i_2\cdots i_{n+1}}}\right]\right\}$$

根据 Jensen 不等式有

$$-D_n(\boldsymbol{M}^{\boldsymbol{X}},\boldsymbol{M}^{\boldsymbol{S}})\leqslant\log\left\{\sum_{i_1,i_2,\cdots,i_{n+1}\in A}\left[m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}\frac{m^{\boldsymbol{S}}_{i_1i_2\cdots i_{n+1}}}{m^{\boldsymbol{X}}_{i_1i_2\cdots i_{n+1}}}\right.\right.$$

$$\cdot \frac{\sum_{i_1} m^{X}_{i_1 i_2 \cdots i_{n+1}}}{\sum_{i_1} m^{S}_{i_1 i_2 \cdots i_{n+1}}}\Bigg]\Bigg\}$$

$$= \log\Bigg\{\sum_{i_1,i_2,\cdots,i_{n+1}\in A}\Bigg[m^{S}_{i_1 i_2 \cdots i_{n+1}} \cdot \frac{\sum_{i_1} m^{X}_{i_1 i_2 \cdots i_{n+1}}}{\sum_{i_1} m^{S}_{i_1 i_2 \cdots i_{n+1}}}\Bigg]\Bigg\}$$

$$= \log\Bigg\{\sum_{i_2,i_3,\cdots,i_{n+1}\in A}\Bigg[\frac{\sum_{i_1} m^{X}_{i_1 i_2 \cdots i_{n+1}}}{\sum_{i_1} m^{S}_{i_1 i_2 \cdots i_{n+1}}} \cdot \sum_{i_1} m^{S}_{i_1 i_2 \cdots i_{n+1}}\Bigg]\Bigg\}$$

$$= \log\Bigg\{\sum_{i_1,i_2,\cdots,i_{n+1}\in A} m^{X}_{i_1 i_2 \cdots i_{n+1}}\Bigg\}$$

$$\leqslant \log\Bigg\{\sum_{i_1,i_2,\cdots,i_{n+1}\in G} m^{X}_{i_1 i_2 \cdots i_{n+1}}\Bigg\}$$

$$= \log 1 = 0$$

由于 $\log t$ 为严格凹函数,因此等号成立的充要条件为 $\frac{m^{S}_{i_1 i_2 \cdots i_{n+1}}}{m^{X}_{i_1 i_2 \cdots i_{n+1}}} \cdot \frac{\sum_{i_1} m^{X}_{i_1 i_2 \cdots i_{n+1}}}{\sum_{i_1} m^{S}_{i_1 i_2 \cdots i_{n+1}}}$ 处处为 1,即 $m^{X}_{i_1,i_2,\cdots,i_{n+1}}$ 与 $m^{S}_{i_1,i_2,\cdots,i_{n+1}}$ 处处相等.

利用 $D_n(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 优化隐写算法以提高安全性时,通过调节 n 可根据隐写分析的实际限制情况控制优化隐写算法的计算复杂度.当 $n=1$ 时即为图像 MC 模型,$D_1(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 即为图像隐写 MC 模型的安全性检测指标 $D(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$[11].

若设图像像素点服从 i.i.d.,则 $D_n(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 与 ε-secure指标 $D(P_X \mid P_S)$[13] 等价(可由与文献[11]类似的方法证明).

从隐写和分析两方面看,图像隐写 n-MC 统计测度 $D_n(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 对于研究隐写对有记忆效应载体所产生的影响,或利用隐写对该类型载体高阶统计分布的改变进行分析均是有益的.由于 $D_n(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 的阶数 n 可调,在考虑到分析者不同计算能力的情况下,对研究隐写算法不同程度的高阶统计安全性,或在计算复杂度一定的情况下研究或选择对载体统计分布改变最小的隐写算法,$D_n(\boldsymbol{M}^{X},\boldsymbol{M}^{S})$ 均可提供指导或参考.

图像 n-MC 模型通过对 n 的调节将载体像素相关性限制在 $n+1$ 个相邻像素间,在分析者所能运用的计算复杂度一定的情况下,n-MC 模型经验矩阵为分析者提供了充足的统计量,也为隐写者优化隐写方案提供了指导.许多隐写和分析方案均不同程度使用了 n-MC 模型,如文献[15,16]使用载体一维直方图特性分析载密图像,一维直方图即为 n-MC 模型经验矩阵的边缘概率,Sullivan 等[10,11] 利用图像载体 MC 模型统计分布的特性对 SSIS 进行分析,张湛等[17] 则利用载体和载密图像 MC 模型统计分布散度距离优化隐写算法,而 MC 模型即为 n-MC 模型 $n=1$ 的情况.

3　基于数字图像 n-MC 模型扩频隐写分析

SSIS 是指将与隐秘信息调制后的 0 均值单位方差高斯白噪声嵌入图像空域或变换域中的隐写算法.设 x_t 为载体空域或变换域信号,ξ_t 为与隐秘信息调制后的白噪声,α 为嵌入能量系数,常采用以下两种方式得到载密信号 s_t.

$$s_t = x_t + \alpha\xi_t \tag{2}$$

$$s_t = x_t(1+\alpha\xi_t) \tag{3}$$

本文称式(2)为加性嵌入,式(3)为乘性嵌入.

3.1　SSIS 对载体 n-MC 模型经验矩阵的影响

由于载体相邻像素间存在较强相关性,其 n-MC 模型经验矩阵的数值基本集中在主对角线附近.图 3 为 lena 图像 2-MC 经验矩阵 $\boldsymbol{M}^{X}_{i_1,i_2,i_3}$ 在 i_1i_2、i_1i_3、i_2i_3 平面的投影.

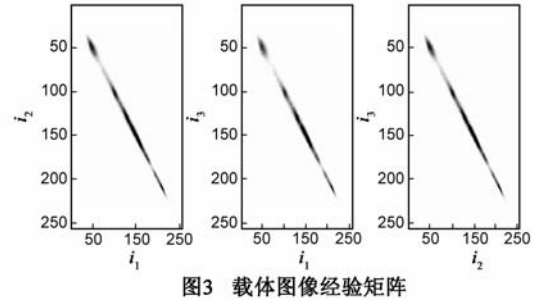

图3　载体图像经验矩阵

载体图像经 SSIS 后,由于隐写破坏了像素间相关性的高阶统计分布,使得载密图像相较原图像其 n-MC 经验矩阵向主对角线集中的程度有所降低.图 4 为 Lena 分别以加性和乘性SSIS嵌入隐秘信息后,其2-MC模型

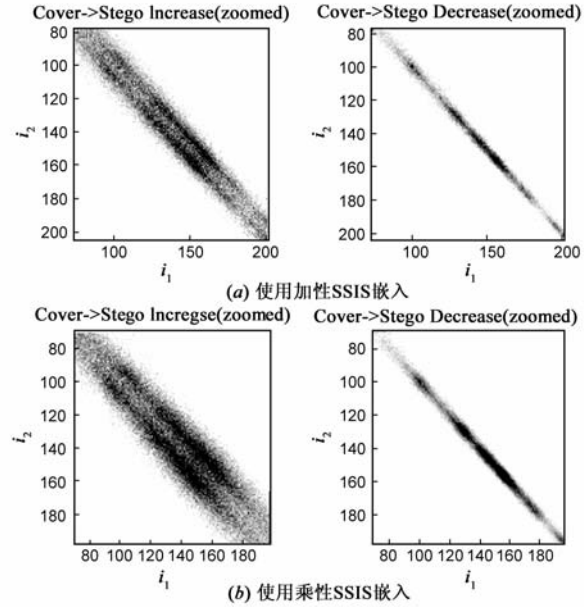

图4　载体与载密图像经验矩阵的改变情况

经验矩阵的变化情况(i_1i_3、i_2i_3 平面投影情况与 i_1i_2 平面类似).

载体嵌入隐秘信息后,因隐写破坏了相邻像素相关性,载密图像 2-MC 经验矩阵数值向主对角线集中的程度没有原图像强,随着嵌入量的增加,该现象愈加明显.此外,该现象随着 n 的增加也更为显著.

3.2 基于 n-MC 模型的扩频隐写分析

图像 n-MC 统计测度 $D_n(\boldsymbol{M}^X,\boldsymbol{M}^S)$反映了隐写对载体高阶统计分布的影响.$D_n(\boldsymbol{M}^X,\boldsymbol{M}^S)$的计算需原始图像,而分析者常不具有掌握原始图像的条件,因此采用 n-MC 模型进行隐写分析时,不能直接使用 $D_n(\boldsymbol{M}^X,\boldsymbol{M}^S)$.若使用 n-MC 经验矩阵的整体作为分类分析的特征,则特征向量维数为$(2^r)^{n+1}$(r 为图像位数),使得分析者计算复杂度过高.由于隐写后载体经验矩阵产生向主对角线周围扩散的现象,可利用该现象对载密图像 n-MC 经验矩阵的特征提取进行降维,以满足计算复杂度要求.本节以 8 位灰度图像 2-MC 模型为例说明基于 n-MC 模型对图像高阶特征提取,并用分类工具进行扩频隐写分析的方法.

通过 3.1 的分析和大量实验证实,由于隐写破坏载体像素相关性,无论加性 SSIS 或乘性 SSIS,载密图像相较原图像,n-MC 经验矩阵均相对主对角线更为发散,即主对角线和主对角线附近数值减小,而更外侧数值增加.可利用该变化选取图像高阶统计特征进行扩频隐写分析.

待分析图像 $\boldsymbol{B}$ 经 Hilbert 扫描得到 2-MC 链 $\boldsymbol{X}$,其经验矩阵为 $256\times256\times256$ 的立方矩阵 $\boldsymbol{M}_{i_1,i_2,i_3}$.由于 $\boldsymbol{M}$ 中元素 m_{i_1,i_2,i_3}代表 $\boldsymbol{X}$ 中像素 i_1 经 i_2 到达 i_3 过程的概率,则 $\boldsymbol{M}$ 主对角线中数值最大的点 $m_{i,i,i}$中的 i($i\in[0,255]$)即为 $\boldsymbol{X}$ 中连续三个像素相等情况出现最多次数的像素值 i.

为充分利用载密图像 2-MC 经验矩阵 $\boldsymbol{M}$ 向主对角线周围扩散的现象,选取 $\boldsymbol{M}$ 主对角线元素 $m_{i,i,i}$中数值最大的 10 个点,且对选取的每一点 $m_{i,i,i}$分别在 i_1,i_2,i_3 三个方向取最临近的 10 个点,得到反映扩散情况的 310 个特征.

对于载密图像 2-MC 经验矩阵 $\boldsymbol{M}$ 主对角线元素数值减小的现象,在主对角线上从 $m_{1,1,1}$开始,每隔 4 点选取一个点作为特征,共 64 个特征反映主对角线元素的变化情况.

综上,对每幅待分析图像共选取其 2-MC 经验矩阵 $\boldsymbol{M}$ 的 374 个特征,组成一个 374 维特征向量来反映 $\boldsymbol{M}$ 主对角线及其附近区域的变化.

SVM 在隐写分析中已得到广泛应用[10,11]且取得了良好效果,因此本文对 SSIS 的隐写分析选择 SVM 作为分类工具.由于在大量实验中发现 SVM 不同核函数的选择对分类结果影响不大,因此仍使用线性核函数.

分析者对训练集中每幅原始图像和使用 SSIS 嵌入隐秘信息的载密图像提取其 2-MC 经验矩阵的 374 维特征向量,使用 SVM 进行训练,并使用完成训练的 SVM 进行 SSIS 隐写分析.

3.3 模型阶数对计算复杂度的影响

基于 n-MC 模型扩频隐写分析算法计算复杂度主要包含图像特征提取和使用分类器分类两方面.设待分析图像 $\boldsymbol{B}_{k\times k}$的像素位数为 r,采用 Hilbert 扫描生成长度为 $k\times k$ 的数据链 $\boldsymbol{X}$.

特征提取方面,在模型经验矩阵 $\boldsymbol{M}$ 生成阶段,由于 $\boldsymbol{M}$ 的生成是从数据链 $\boldsymbol{X}$ 的第 $n+1$ 位开始针对每一位及其前 n 位情况在 $\boldsymbol{M}$ 相应位置累加,因此 $\boldsymbol{M}$ 生成的计算复杂度主要与图像大小有关,为 $O(k\times k)$.在特征提取阶段,虽阶数 n 增加导致 $\boldsymbol{M}$ 的规模增大,但其主对角线元素个数仍为 2^r,在主对角线上每隔 4 点选取一个特征点的计算复杂度为 $O(2^r/4)$,对主对角线元素排序选取数值最大 10 个点的计算复杂度为 $O(2^r\log_2(2^r))$,因此在特征提取阶段计算复杂度为 $O\left(2^r\log_2(2^r)+\frac{2^r}{4}\right)$.综上,基于 n-MC 模型对待分析图像特征提取的计算复杂度,主要与图像规模 k 和图像位数 r 有关.实验环境为 4G 内存、intel core2 2GHz CPU、Matlab R2008a,采用 1 – MC 模型提取一幅 512×512 的 8 位灰度图像特征需时 5.26 秒,而采用 2-MC 模型提取同一图像特征需时 5.68 秒.

使用分类器分类方面,随着模型阶数 n 的增大,所提取的特征维数亦相应增大.当使用分类器进行分析时,随着 n 的增大所需时间亦相应增加.采用与图像特征提取相同的实验环境,使用台湾大学林智仁(Lin Chih-Jen)等开发的 LIBSVM 软件包对已提取的 1000 幅图像特征进行分类时,基于 1-MC 模型的分析时间为 4.61 秒,而 2-MC 模型分析时间为 11.17 秒.

随着模型阶数 n 的增大而引起分析所需存储空间增长方面,影响存储规模的主要因素为 $\boldsymbol{M}$ 的规模,而 $\boldsymbol{M}$ 的规模为$(2^r)^{n+1}$,即随 n 的增加其规模会急剧增大.通常自然图像的 $\boldsymbol{M}$ 为稀疏矩阵,因此在使用较高阶数模型进行分析时,可考虑对 $\boldsymbol{M}$ 的存储方式进行优化.

4 实验结果

此部分介绍基于图像 2-MC 模型对 8 位灰度图像空域及 DCT 域的加性和乘性 SSIS 进行隐写分析的实验情况,并与基于 MC 模型[10,11]的 SSIS 隐写分析情况比较.实验采用 UCID.V2 图像库[18]的 1338 幅未压缩图像以

及通过扫描得到的 662 幅未压缩图像作为训练和测试图像库.

将图像库中图像转换为 8 位灰度 BMP 图像.在其中随机选取 1000 幅作为训练集,其余作为测试集.在训练集和测试集中分别随机选取 500 幅作载密图像,剩余作原始图像.各种 SSIS 除随机调制隐写[5]外,常需采用纠错编码机制,本文实验中在对 SSIS 进行分析时,将纠错编码后的待嵌码流均看作需嵌入的隐秘信息,仍采用 bpp(bits per pixel)衡量嵌入量.对训练集和测试集中载密图像使用加性 SSIS($\alpha = 0.375$)和乘性 SSIS($\alpha = 0.05$),以嵌入量 0.91bpp 在图像空域和 8×8 分块 DCT 系数域嵌入 0 – 1 均匀分布的隐秘信息,分别得到空域和 DCT 系数域的加性和乘性 SSIS 载密图像.

对训练和测试集中图像分别扫描构成 2-MC 和 MC,其中 2-MC 采用 Hilbert 扫描方式.由于 MC 链中元素 x_t 仅与前一元素 x_{t-1} 相关,无论采用列扫描、zig-zag 扫描还是 Hilbert 扫描,x_{t-1} 均为 x_t 在图像 8 领域中元素,扫描方式的改变对效果影响不大,因此实验中对于 MC 模型仍按文献[11]所述采用列扫描方式.对扫描所得图像 2-MC 按照第 3 节所述方法提取图像特征,对 MC 按照文献[11]的方法提取图像特征,使用所提特征对 SVM 进行训练和测试.实验采用台湾大学林智仁(Lin Chih-Jen)等开发的 LIBSVM 软件包和线性核函数 $K(\boldsymbol{v}_j, \boldsymbol{v}_k) = \boldsymbol{v}_j' \times \boldsymbol{v}_k$,其中 $\boldsymbol{v}$ 为特征向量.SVM 的惩罚因子 C 控制着对错分样本的惩罚程度,其选择是样本识别率和机器学习时间的折衷,C 的选择与训练样本有关,常通过实验选择对特定训练集较适用的值.本文实验中选取 $C = 10$.重复 20 次实验的平均结果如表 2 所示(其中 TP 为 True Positive;FP 为 False Positive;TN 为 True Negative;准确率为(TP + TN)/2),ROC 曲线如图 5 所示.

表 2　SSIS 分析实验结果(重复 20 次实验的平均结果)

嵌入域	嵌入方法	2-MC			MC		
		TP(%)	FP(%)	准确率	TP (%)	FP (%)	准确率
空域	加性	96.45	1.18	97.63	97.04	5.92	95.56
	乘性	85.80	16.57	84.62	85.21	23.67	80.77
DCT	加性	94.08	0.59	96.75	97.63	8.88	94.38
	乘性	85.21	16.77	84.32	76.92	21.30	77.81

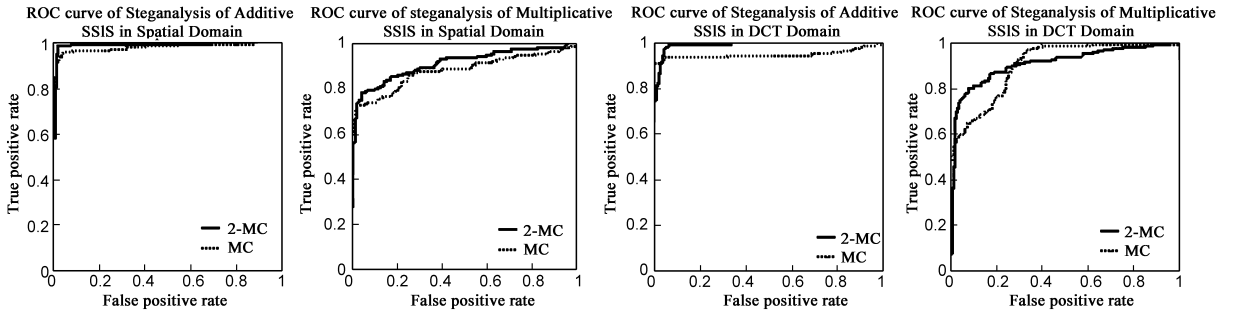

图5　扩频隐写分析实验ROC曲线

从表 2 和图 5 可知基于图像 n-MC 模型的隐写分析对图像空域和变换域加性 SSIS,随模型阶数增加,2-MC 模型相比 MC 模型的分析准确率提高了两个百分点左右,且虚警率(FP)大大减小.对空域和变换域的乘性 SSIS,随着模型阶数增加,2-MC 模型相比 MC 模型的分析准确率有较大提高,且虚警率也大为减小.

5　结论

本文以 n-MC 作为研究图像隐写的统计分布模型,通过对各扫描方法所构成 n-MC,包含相邻像素相关性信息量的研究,采用 Hilbert 扫描方式构建图像 n-MC 模型.利用散度距离提出图像隐写统计安全性的 n-MC 测度,并证明其有界.最后通过对加性和乘性 SSIS 载体 n-MC 经验矩阵变化情况的研究,得出由于 SSIS 破坏了载体相邻像素相关性,因此载密图像 n-MC 经验矩阵较原图像相对主对角线发散的结论并通过实验进行了直观说明,进而根据这一结论利用高阶 MC 模型提取图像特征,并使用 SVM 对空域和 DCT 域加性和乘性 SSIS 进行分析.实验说明基于图像 n-MC 模型对 SSIS 的分析具有较好效果,且随着模型阶数的提高,分析准确率也随之提高,虚警率则随之降低.由于图像 n-MC 模型计算复杂度可调,基于该模型对 SSIS 的分析适用于计算能力不同的分析者.

图像 n-MC 模型通过对 n 的调节可方便调节模型计算复杂度,因此在考虑到分析者不同计算能力的情况下,如何利用该模型及其经验矩阵所含图像高阶统计分布信息以及隐写对其经验矩阵分布情况的破坏,研究对其它隐写算法的分析方法;或者从隐写角度,如何利用 n-MC 高阶统计分布测度研究高阶统计安全的隐写算法,均可作为进一步研究的方向.

参考文献:

[1] 张新鹏,王朔中.基于稀疏表示的密写编码[J].电子学报,2007,35(10):1892 – 1896.
Zhang Xinpeng, Wang Shuozhong. Steganographic encoding based on sparse representation[J]. Acta Electronica Sinica, 2007,35(10):1892 – 1896.(in Chinese)

[2] 刘劲,康志伟,何怡刚.一种基于小波对比度和 LSB 的密写[J].电子学报,2007,35(7):1391 – 1393.
Liu Jin, Kang Zhiwei, He Yigang. A steganographic method based on wavelet contrast and LSB[J]. Acta Electronica Sinica, 2007, 35(7): 1391 – 1393. (in Chinese)

[3] Cox I, Kilian J, Leighton T, et al. Secure spread spec-trum watermarking for multimedia[J]. IEEE Transactions on Image Processing, 1997, 6(12): 1673 – 1687.

[4] Marvel L M, Boncelet C G, Retter C T. Spread spectrum image steganography[J]. IEEE Transactions on Image Processing, 1999, 8(8): 1075 – 1083.

[5] Fridrich J, Goljan M. Digital image steganography using stochastic modulation[A]. Proceedings of SPIE-The International Society for Optical Engineering[C]. Santa Clara: SPIE, 2003. 191 – 202.

[6] Harmsen J J, Pearlman W A. Steganalysis of additive noise modelable information hiding[A]. Proceedings of SPIE-The International Society for Optical Engineering[C]. Santa Clara: SPIE, 2003. 131 – 142.

[7] Wang Y, Moulin P. Steganalysis of block-structured stegotext [A]. Proceedings of the SPIE-The International Society for Optical Engineering[C]. San Jose: SPIE, 2004. 477 – 488.

[8] Chandramouli R. Subbalakshmi K P. Active steganalysis of spread spectrum image steganography[A]. Proceedings of the 2003 IEEE International Symposium on Circuits and Systems [C]. Bangkok: IEEE, 2003. 830 – 833.

[9] Ji R R, Yao H X, Liu S H, et al. A new steganalysis method for adaptive spread spectrum steganography[A]. Proceedings-2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing[C]. Pasadena: IEEE, 2006. 363 – 368.

[10] Sullivan K, Madhow U, Chandrasekaran S, et al. Steganalysis of spread spectrum data hiding exploiting cover memory[A]. Proceedings of SPIE-The International Society for Optical Engineering[C]. San Jose: SPIE, 2005. 38 – 46.

[11] Sullivan K, Madhow U, Chandrasekaran S, et al. Steganalysis for Markov cover data with applications to images[J]. IEEE Transactions on Information Forensics and Security, 2006, 1 (2): 275 – 287.

[12] Bishop Y M, Fienberg S E, Holland P W. Discrete Multivariate Analysis: Theory and Practice[M]. New York: Springer, 2007. 267 – 270.

[13] Cachin C. An information-theoretic model for steganography [J]. Information and Computation, 2004, 192(1): 41 – 56.

[14] 戴跃伟,刘光杰,叶曙光.基于 Hilbert 填充曲线的自适应隐写[J].电子学报,2008,36(12A):35 – 38.
Dai Yuewei, Liu Guangjie, Ye Shuguang. Adaptive steganography based on Hilbert filling curve[J]. Acta Electronica Sinica, 2008, 36(12A): 35 – 38. (in Chinese)

[15] Dabeer O, Sullivan K, Madhow U, et al. Detection of hiding in the least signicant bit[J]. IEEE Transactions on Signal Processing, Supplement on Secure Media I, 2004, 52(10): 3046 – 3058.

[16] Hogan M T, Hurley N J, Silvestre G C M, et al. ML detection of steganography[A]. Proceedings of the SPIE-The International Society for Optical Engineering[C]. San Jose: SPIE, 2005. 16 – 27.

[17] 张湛,刘光杰,王俊文等.基于 Markov 链安全性的量化隐写算法[J].光电子·激光,2009,20(7):944 – 949.
Zhang Zhan, Liu Guangjie, Wang Junwen, et al. A novel quantization-embedded steganographic algorithm based on Markov chain security[J]. Journal of Optoelectronics·Laser, 2009, 20(7): 944 – 949. (in Chinese)

[18] Schaefer G, Stich M. UCID-An Uncompressed Colour Image Database[DB/OL]. http://vision.cs.aston.ac.uk/datasets/UCID/ucid.html, 2004-12-30/2009-5-21.

作者简介:

张 湛 男,1974 年生于重庆万盛,南京理工大学自动化学院博士研究生,研究方向为信息隐藏、隐写与分析.
E-mail: blacksnown@126.com

刘光杰 男,1980 年生于江苏徐州,南京理工大学自动化学院副研究员.博士.研究方向为隐写分析、数字取证.

王俊文 男,1984 年生于安徽安庆,南京理工大学自动化学院博士研究生,研究方向为信息安全、数字取证.

戴跃伟 男,1962 年生于江苏镇江,南京理工大学自动化学院教授,博士生导师,主要研究领域为多媒体信息安全、数字水印.

王执铨 男,1939 年生于湖北武汉.南京理工大学自动化学院教授.博士生导师.主要研究领域为容错控制、信息安全.