doi:10.3969/j.issn.1000-436x.2013.01.006

采用 Hilbert 扫描序列短重码统计的盲隐写检测方法

钟尚平，徐巧芬，郭文忠，廖彬
(福州大学 数学与计算机科学学院，福建 福州 350108)

摘 要：在 LSB 行扫描序列中，基于短重码间距统计的隐写检测方法对 LSB 匹配等隐写技术具有良好的盲检测性能。然而此方法只能利用相邻码元的相关性，影响检测性能。理论证明了短重码间距统计量的检测能力与重码累积成功概率、短重码维数有关，并提出采用 Hilbert 扫描序列以提高重码累积成功概率比率来提升检测性能的盲隐写检测方法。该方法在图像 LSB Hilbert 扫描序列中，基于码元相同短重码统计量的分布特征，通过 Poisson 分布显著性检验检测隐写信息，可充分利用 Hilbert 曲线良好的局部相关保持特性，不仅利用了相邻码元的相关性，还利用了局部区域码元的相关性。理论分析和实验结果表明了本文方法在有效控制虚警率的前提下，具有较好的隐写检测性能。

关键词：盲隐写检测；码元相同短重码统计；Hilbert 曲线扫描；相邻码元相关性；局部区域码元相关性
中图分类号：TP911.73 **文献标识码：**A **文章编号：**1000-436X(2013)01-0051-10

Blind detection for image steganography using short du ate codes statistical model for Hilbert scanning sequences

ZHONG Shang-ping, XU Qiao-fen, GUO Wen-zhong, LIAO Bin
(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract: By analyzing and proving the correlation between the detection capability of a short duplicate code statistical feature and the probability of cumulating short duplicate codes、the dimension of short duplicate codes, a method to improving the detection capability of a short duplicate code statistical feature was found. Then, a blind detection method for image steganography using short duplicate codes statistical model for Hilbert scanning sequences was proposed. The proposed method used Poisson distribution test to detect the stego-message based on the statistical feature of short duplicate code with same elements in LSB Hilbert scanning sequences. So, the proposed method could make full use of Hilbert curve to maintain the good properties of local correlation, and could not only use the correlation of adjacent elements, but also use the correlation of elements in local regions. Theoretical lysis and experiments show that the proposed method can effectively improve the detection rate under the condition of effectively controlling the false alarm rate.

Key words: blind detection for image steganography; statistics of short duplicate codes with same elements; Hilbert curve scanning; correlation of adjacent elements; correlation of elements in local regions

1 引言

目前，对隐写图像进行分析检测主要有两条途径。一是针对某一种具体的隐写方法设计出检测方法，称为专用隐写检测；二是盲（通用）隐写检测，是在隐写方法未知而只拥有检测对象的情况下，对其是否含有隐写信息作出判断，目前主要有序列随机性检测法和通过提取变换域特征并通过学习训练建立模式分类模型的方法。近年来，针对隐写图像的盲隐写检测技术受到业界越来越多的关注[1,2]。

收稿日期：2011-09-05；修回日期：2012-07-20
基金项目：国家自然科学基金资助项目(61103175)；福建省自然科学基金资助项目(2010J01331)；教育部科学技术研究重点基金资助项目(212086)
Foundation Items: The National Natural Science Foundation of China (61103175)；The Natural Science Foundation of Fujian Province (2010J01331); The Key Project of Chinese Ministry of Education (212086)

图像最不重要比特(LSB, least significant bit) 隐写是最早出现的时/空域隐写方法之一。目前 LSB 隐写主要有基于替换的方法[3]和基于匹配的方法[4]。对基于替换的 LSB 隐写，已经提出了不少有效的专用隐写检测方法，如，卡方统计检测法[5]、RS 分析法[6]、基于样值对分析的方法[7]和基于差分直方图的检测方法[8]等；而对基于匹配的 LSB 隐写，已有的专用隐写检测方法效果一般，特别是对于低嵌入率隐写图像易产生误判，且对载体图像的依赖性较强，如针对彩色图像的检测方法[9]、基于直方图特征函数的检测方法[10]和基于区域相关性的 LSB 匹配隐写分析方法[11]等。目前，对 LSB 隐写的盲隐写检测主要采用序列随机性检测方法[12]，如频率检测、串行检测、Poker 检测和游程检测等，普遍存在过高的虚警概率。另外，王国新等依据自然图像作为一个局部区域平稳的 Markov 信源的基本结论，根据 LSB 行扫描序列中短重码的分布特征提出了一种基于相邻行码元相关性的盲隐写检测方法[13]，该方法通过定义短重码间距统计概念，推导分析了间距为图像宽度(或图像宽度较小的整数倍)统计变量的分布特征，并采用 Poisson 分布显著性检验检测隐秘信息。此法在有效控制虚警率的前提下，对 LSB 匹配、LSB 替换、直方图补偿等隐写有良好的检测性能；该方法不仅适用于 LSB 位平面，也可以针对其他位平面；该方法在检测过程中不针对具体的隐写算法，具有较强的通用性。

Hilbert 空间填充曲线由德国数学家 David Hilbert 在 1891 年首先提出。利用 Hilbert 曲线构造图像像素空间连续的扫描方式，相比其他如行(列)扫描、zig-zag 扫描，由于 Hilbert 曲线具有更好的空间连续性，这种好的空间连续性能够更准确地描述图像局部像素的空间相关性[14]。目前，Hilbert 曲线良好的局部相关保持特性，在图像隐写和隐写分析[14,15]、网络信任管理[16]和数据挖掘[17]等领域都得到了有效应用。

本文采用文献[13]中基于 LSB 序列短重码间距统计模型进行盲隐写检测的思想，分析证明了短重码间距统计量的检测能力与重码累积成功概率、短重码维数之间的关系规律，提出了一种采用 Hilbert 扫描序列短重码统计的盲隐写检测方法。有别于文献[13]采用的 LSB 行扫描序列短重码间距统计模型只能利用相邻码元相关性的特点，本文设计的 Hilbert 扫描序列短重码统计模型不仅利用了图像相邻码元的相关性，还能够充分利用局部区域码元间的相关性，从而具有更大的重码累积成功概率比率。因此，Hilbert 扫描序列短重码统计量比文献[13]短重码间距统计量具有更强的检测能力。

2 基于 LSB 行扫描序列短重码间距统计的隐写检测方法概述

2.1 定义及符号说明

按文献[13]介绍要用到的术语和定义。

定义 1[13] 把图像的最低有效位平面的每一行串接起来，可以构成一个比特序列，称之为 LSB 行扫描序列。

设灰度 BMP 格式图像的宽和高分别为 W 和 H，其 LSB 行扫描序列可以看作一个长度为 $N=W\times H$ 的序列 V。

定义 2 如果序列 V 中存在维数相同的子序列 v_i 和 v_j，$i\neq j$，若 $v_i=v_j$，且子序列的维数 l 为一个较小的值时(如 $l\leq 8$)，则称子序列 v_i 和 v_j 为一对短重码；v_i, v_j $(i\neq j)$ 在 V 中的位置差称为短重码间距，记为 d，$d\in\{1,2,\mathrm{L},N-l\}$；序列 V 中相距为 d 的各种状态短重码个数的累积称为短重码数量，记为 $k, k\in\{0,1,\mathrm{L},N-d-l+1\}$。

定义 3 短重码间距统计。对待检测序列进行短重码统计，统计短重码数量和间距之间的对应关系，并忽略重码本身的状态和具体位置。统计结果为重码的各个状态在给定间距下的统计累积。

2.2 短重码间距统计模型

自然图像作为一个局部区域平稳的 Markov 信源，其中的某一局部区域内的像素值之间有很强的相关性。因此，对 LSB 行扫描序列，当重码间距接近图像宽度 W 或图像宽度较小的整数倍时，短重码相等的概率较大。而自然图像嵌入秘密信息(一般都经过加密处理)后，LSB 序列呈现出秘密信息特征，短重码间距统计没有间距为图像宽度时的统计优势。

图 1 为维数 $l=4$ 短序列在间距 $d=W$ 时统计短重码的直观表示。图中 v_i^h 与 v_i^{h+1} 为间距为 W (即相邻行，第 h 行与第 $h+1$ 行)的 2 个维数 $l=4$ 的短序列。

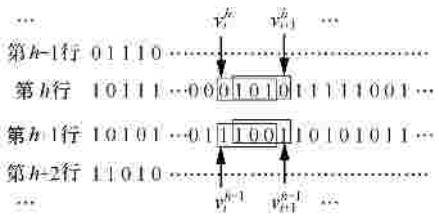

图 1 在间距 $d=W$ 时统计短重码

2.3　短重码间距统计特征

分析上述短重码间距统计模型，可得如下统计特征[13]。

定理 1　在自然图像 LSB 行扫描序列中，当 2 个维数为 l 的序列 v_i，v_j $(i \neq j)$ 处在相邻行(v_i 与 v_j 的位置关系如图 1 中 v_i^h 与 v_i^{h+1} 的位置关系)，也就是当两者的各个码元的位置相邻时，因码元间具有一定的相关性，v_i 与 v_j 相等的概率为 $q^l, 1/2 \le q \le 1$；而 2 个维数为 l，相互独立的二值完全随机白噪声序列(加密序列) v_i'，v_j' $(i \neq j)$ 相等的概率为 $1/2^l$。

证明　第 2 部分结论的证明请参见文献[13]。下面证明第 1 部分结论。

设第 1 个序列 v_i 在状态 i 时为 $b_1^i b_2^i \perp b_l^i$，第 2 个序列 v_j 在状态 j 时为 $b_1^j b_2^j \perp b_l^j$，当序列 v_i 与 v_j 两者相同且状态 k 时为 $b_1^k b_2^k \perp b_l^k$，其中，码元 b_1、b_2、…、$b_l \in \{0,1\}$。设事件 $B_1^i : b_1^i = b_1^k$；事件 $B_2^i : b_2^i = b_2^k$；…；事件 $B_l^i : b_l^i = b_l^k$；事件 $B_1^j : b_1^j = b_1^k$；事件 $B_2^j : b_2^j = b_2^k$；…；事件 $B_l^j : b_l^j = b_l^k$。设序列 v_i 与 v_j 两者相同的事件为 A。事件 A 的状态空间中的状态个数为 2^l。有

$$\begin{aligned} P\{A\} = P\{i = j\} &= \sum_{k=0}^{2^l-1} P\{i = k, j = k\} \\ &= \sum_{k=0}^{2^l-1} P\{i = k\} P\{j = k \mid i = k\} \\ &= \sum_{k=0}^{2^l-1} \frac{1}{2^l} \times P\{j = k \mid i = k\} \end{aligned}$$

而

$$\begin{aligned} &P\{j = k \mid i = k\} = P\{B_1^j B_2^j \perp B_l^j \mid B_1^i B_2^i \perp B_l^i\} \\ &= \frac{P\{B_1^j B_2^j \perp B_l^j B_1^i B_2^i \perp B_l^i\}}{P\{B_1^i B_2^i \perp B_l^i\}} \\ &= \frac{P\{B_1^j B_1^i\} P\{B_2^j B_2^i\} \perp P\{B_l^j B_l^i\}}{P\{B_1^i\} P\{B_2^i\} \perp P\{B_l^i\}} \\ &= P\{B_1^j \mid B_1^i\} P\{B_2^j \mid B_2^i\} \perp P\{B_l^j \mid B_l^i\} \end{aligned}$$

设 $P\{B_1^j \mid B_1^i\} = P\{B_2^j \mid B_2^i\} = \perp = P\{B_l^j \mid B_l^i\} = q$，$1/2 \le q \le 1$。故 $P\{A\} = q^l$。不妨称 q 为相邻码元相关性系数。定理获证。

定理 2　对于间距 $d = W$，自然图像 LSB 行扫描序列(长度为 N)中，维数为 l 短重码间距统计量 $X_l(W)$ 服从参数为 $N - d - l + 1, q^l$ 的二项分布，近似服从 Poisson 分布，即：$P(X_l(W) = k) \approx \dfrac{l^k e^{-l}}{k!}$，其中，$l = (N - d - l + 1) q^l$，$\dfrac{1}{2} \le q \le 1$，$e$ 为自然数；而对于间距 $d = W$，二值完全随机白噪声序列(长度为 N)的维数为 l 短重码间距统计量 $X_l(W)$ 服从参数为 $N - d - l + 1, (1/2)^l$ 的二项分布，近似服从 Poisson 分布，即：$P(X_l(W) = k) \approx \dfrac{l^k e^{-l}}{k!}$，其中，$l = (N - d - l + 1)\left(\dfrac{1}{2}\right)^l$，$e$ 为自然数。

定理 2 的证明见文献[13]。定理 2 将“一次重码累积”看成是一次实验，而短重码间距统计过程相当于做了 $N - d - l + 1$ 重 Bernoulli 实验。设事件 C 为重码累积成功(即重码的统计累积值加 1)；设事件 C_1 为对于自然图像的 LSB 行扫描序列，重码累积成功；设事件 C_2 为对于二值完全随机白噪声序列，重码累积成功。显然有：$P(C_1) = q^l$，$P(C_2) = (1/2)^l$。设重码累积成功概率的比率 $d = \dfrac{P(C_1)}{P(C_2)} = (2q)^l$。

文献[13]利用短重码间距统计量 $X_l(W)$ 的统计分布规律，采用 Poisson 分布显著性检验来检测隐秘信息，检测算法见文献[13]。

3　提高 $X_l(W)$ 检测性能的可能途径分析

本文采用经典的 Fisher 判别率(FDR, Fisher's discriminant ratio)[18]

$$FDR = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

来定量估计服从 Poisson 分布的统计变量 $X_l(W)$ 的检测分类性能，其中，m_1, s_1 表示自然图像统计特征变量 $X_l(W)$ 的均值和方差；而 m_2, s_2 表示二值完全随机白噪声序列统计特征变量 $X_l(W)$ 的均值和方差。一般地，FDR 越大，则 $X_l(W)$ 的可检测分类能力越大。

由于 Poisson 分布的均值和方差相等，且都等于参数 l，根据定理 2，可得 $X_l(W)$ 的 Fisher 判别率

$$FDR((2q)^l, l) = \frac{(N - d - l + 1)^2 \left(q^l - \left(\dfrac{1}{2}\right)^l\right)^2}{(N - d - l + 1)^2 \left(q^{2l} + \left(\dfrac{1}{2}\right)^{2l}\right)}$$

$$= \frac{q^{2l} + \left(\frac{1}{2}\right)^{2l} - 2q^{l}\left(\frac{1}{2}\right)^{l}}{q^{2l} + \left(\frac{1}{2}\right)^{2l}}$$

$$= \frac{(2q)^{2l} + 1 - 2(2q)^{l}}{(2q)^{2l} + 1}$$

$$= \frac{(d)^{2} + 1 - 2d}{(d)^{2} + 1} \tag{1}$$

由式(1)可知，$X_l(W)$ 的检测分类性能仅与重码累积成功概率的比率 $d = (2q)^l$ 和短重码维数 l 有关，而与参数 $N - d - l + 1$ 无关。为了寻求提高 $X_l(W)$ 检测性能的途径，本文首先讨论短重码间距统计量的检测性能与重码累积成功概率比率、短重码维数大小的关系。

3.1 $X_l(W)$的检测性能与重码累积成功概率比率的关系规律

下述定理 3 揭示了 $X_l(W)$的检测性能与重码累积成功概率比率的关系规律，即重码累积成功概率比率越大，则 $X_l(W)$的检测性能越强。

定理 3 若重码累积成功概率比率 $d_2 \geq d_1$，则有 $FDR(d_2, l) \geq FDR(d_1, l)$。

证明 由式(1)可得

$$FDR(d_1, l) = \frac{(d_1)^2 + 1 - 2d_1}{(d_1)^2 + 1}$$

$$FDR(d_2, l) = \frac{(d_2)^2 + 1 - 2d_2}{(d_2)^2 + 1}$$

要证 $FDR(d_2, l) \geq FDR(d_1, l)$，即证

$$\frac{FDR(d_2, l)}{FDR(d_1, l)} = \frac{(d_2)^2(d_1)^2 + (d_1)^2 + (d_2)^2 + 1 - 2d_2(d_1)^2 - 2d_2}{(d_2)^2(d_1)^2 + (d_1)^2 + (d_2)^2 + 1 - 2d_1(d_2)^2 - 2d_1} \geq 1$$

只要证：$2d_2(d_1)^2 + 2d_2 \leq 2d_1(d_2)^2 + 2d_1$。即证

$$\frac{2d_2(d_1)^2 + 2d_2}{2d_1(d_2)^2 + 2d_1} = \frac{d_2(d_1)^2 + d_2}{d_1(d_2)^2 + d_1} \leq 1 \tag{2}$$

由 $d_2 \geq d_1$，可设 $d_2 = Cd_1$，其中，$C = 1 + e$，$e \geq 0$。有

$$\frac{d_2(d_1)^2 + d_2}{d_1(d_2)^2 + d_1} = \frac{C(d_1)^2 + C}{C^2(d_1)^2 + 1} = \frac{C(d_1)^2 + 1 + e}{C(1+e)(d_1)^2 + 1}$$

$$= \frac{C(d_1)^2 + 1 + e}{C(d_1)^2 + 1 + Ce(d_1)^2}$$

因为 $d_1 \geq 1$，$C \geq 1$，所以 $Ce(d_1)^2 \geq e$，即式(2)成立。故定理获证。

3.2 $X_l(W)$的检测性能与短重码维数大小的关系规律

定理 4 揭示了 $X_l(W)$ 的检测性能与短重码维数大小的关系。

定理 4 当相邻码元相关性系数 q 一定时，有 $FDR((2q)^l, l) \leq FDR((2q)^{l+1}, l+1)$。

证明 由式(1)可得 $FDR((2q)^{l+1}, l+1)$

$$= \frac{q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2} - 2q^{l+1}\left(\frac{1}{2}\right)^{l+1}}{q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2}}$$

考虑 $\frac{FDR((2q)^{l+1}, l+1)}{FDR((2q)^{l}, l)}$ 的大小，若比值大于等于 1，则定理成立。由于

$$\frac{FDR((2q)^{l+1}, l+1)}{FDR((2q)^{l}, l)} = \frac{C - 2q^{l}\left(\frac{1}{2}\right)^{l}\left(\frac{1}{2}q^{2l+1} + q\left(\frac{1}{2}\right)^{2l+1}\right)}{C - 2q^{l}\left(\frac{1}{2}\right)^{l}\left(q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2}\right)}$$

其中，$C = q^{4l+2} + \left(\frac{1}{2}\right)^{4l+2} + \left(\frac{1}{2}\right)^{2l+2} q^{2l} + \left(\frac{1}{2}\right)^{2l} q^{2l+2}$。

显然，要证 $\frac{FDR((2q)^{l+1}, l+1)}{FDR((2q)^{l}, l)} \geq 1$，只需

$$2q^{l}\left(\frac{1}{2}\right)^{l}\left(\frac{1}{2}q^{2l+1} + q\left(\frac{1}{2}\right)^{2l+1}\right) \leq 2q^{l}\left(\frac{1}{2}\right)^{l}\left(q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2}\right)$$

故只要式(3)小于等于 1，则定理成立

$$\frac{2q^{l}\left(\frac{1}{2}\right)^{l}\left(\frac{1}{2}q^{2l+1} + q\left(\frac{1}{2}\right)^{2l+1}\right)}{2q^{l}\left(\frac{1}{2}\right)^{l}\left(q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2}\right)} = \frac{\frac{1}{2}q^{2l+1} + q\left(\frac{1}{2}\right)^{2l+1}}{q^{2l+2} + \left(\frac{1}{2}\right)^{2l+2}} \tag{3}$$

设 $q = \frac{1}{2} + e, 0 \leq e \leq \frac{1}{2}$，则式(3)为

$$\frac{\frac{1}{2}q^{2l+1} + e\left(\frac{1}{2}\right)^{2l+1} + \left(\frac{1}{2}\right)^{2l+2}}{\frac{1}{2}q^{2l+1} + eq^{2l+1} + \left(\frac{1}{2}\right)^{2l+2}} \tag{4}$$

因为 $\frac{1}{2} \le q \le 1$，故式(4)小于等于 1 成立。定理获证。

3.3 提高 $X_l(W)$检测性能的可能途径分析

在定理 2 中，统计量 $X_l(W)$近似服从 Poisson 分布的依据是下述著名的 Poisson 定理[19]。

定理 5 (Poisson 定理)若 $(N-d-l+1) \to \infty$，q^l 很小，$(N-d-l+1)\times q^l \to l$，则有：

$$C_{N-d-l+1}^{k}(q^l)^k(1-q^l)^{N-d-l-k+1} \approx \frac{l^k e^{-l}}{k!}$$

其中，$l=(N-d-l+1)q^l$，$\frac{1}{2} \le q \le 1$，e 为自然数。

由定理 5，若要采用 Poisson 分布显著性检验来检测隐秘信息，必须满足：$N-d-l+1$很大，q^l 很小，$(N-d-l+1)\times q^l$ 大小适中，否则将带来较大误差，影响短重码间距统计量 $X_l(W)$ 的检测性能。目前还没有确定 Poisson 定理中参数大小的具体标准[19]。当 $d=W$ 时，因为 $N=W\times H$，$N \gg d+l$，故 $N-d-l+1$很大；如选择 l 的范围为：$l=4,5,6,7,8$，可使 q^l 很小，且 $(N-d-l+1)\times q^l$ 大小适中。

另外，定理 4 表明：当 q 一定时，$FDR((2q)^l,l)$ 随 l 单调增加。但按文献[13]虚警概率也可能随 l 单调增加。实际应用统计量 $X_l(W)$ 检测隐写图像时，因为 $X_l(W)$ 仅仅是近似服从 Poisson 分布且检测需满足虚警率要求，不能仅用 $X_8(W)$ 作为每幅图像的检测统计量。需要根据待检测图像的质量和对检测率、虚警率的要求选择适当的短重码维数。

综上可知，提高 $X_l(W)$ 检测性能的可能途径是：增大重码累积成功概率的比率。

下一节，将建立 Hilbert 扫描序列短重码统计模型以增大重码累积成功概率的比率，提出一种检测性能更好的盲隐写检测方法。

4 采用 Hilbert 扫描序列短重码统计的盲隐写检测方法

4.1 Hilbert 扫描序列短重码统计及其统计特征

4.1.1 图像的 Hilbert 曲线扫描

二维 Hilbert 曲线不自交地通过格形方块每一个格点中心一次且仅一次，并且充满整个格形方块区域。相比其他如行(列)扫描、zig-zag 扫描，由于图像的 Hilbert 曲线扫描方式能够将相邻像素尽量排列在一起，具有更好的空间连续性，这种好的空间连续性能够更准确地描述图像的局部像素的空间相关性[14]。图 2 给出了 32×32 尺寸图像的 Hilbert 曲线扫描路径的例子。

4.1.2 定义及符号说明

定义 4 把图像的最低有效位平面按 Hilbert 扫描路径串接起来，可以构成一个比特序列，称之为 LSB Hilbert 扫描序列。

设灰度 BMP 格式图像的宽和高分别为 W 和 H，其 Hilbert 扫描序列可以看作一个长度为 $N=W\times H$ 的序列$\mathbb{V}$。

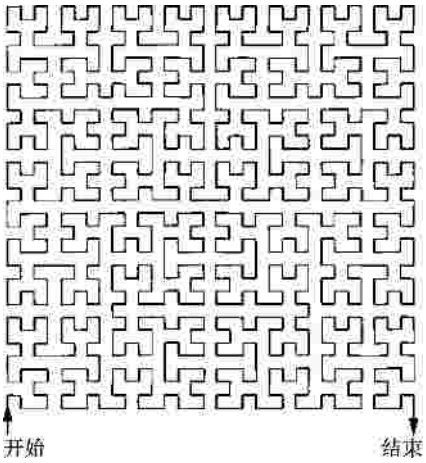

图 2　32×32 尺寸图像的 Hilbert 曲线扫描路径

定义 5 如果序列$\mathbb{V}$中存在维数为 l 的子序列 v_i，若 $v_i=b_1^1b_2^1\cdots b_l^1$ 或 $v_i=b_1^0b_2^0\cdots b_l^0$，其中，码元 b_1^1、b_2^1、…、b_l^1 为 1；码元 b_1^0、b_2^0、…、b_l^0 为 0，则称 v_i 为维数为 l 的码元相同子序列。

定义 6 如果序列$\mathbb{V}$中存在维数相同的码元相同子序列 v_i 和 v_j，$i\ne j$，且子序列的维数 l 为一个较小的值时(如：$l \le 8$)，则称子序列 v_i 和 v_j 为一对码元相同短重码；v_i，v_j $(i\ne j)$ 在$\mathbb{V}$中的位置差称为短重码间距，记为 d；序列$\mathbb{V}$中相距为 d (d 很小，本文仅考虑 $d=1$ 情形)的码元相同短重码个数的累积称为码元相同短重码数量，记为 k。

定义 7 码元相同短重码统计，对待检测序列进行码元相同短重码统计，统计短重码数量和间距 $(d=1)$之间的对应关系，并忽略重码本身的状态和具体位置。统计结果为 $d=1$ 时重码在$\mathbb{V}$中出现次数的统计累积。

4.1.3 Hilbert 扫描序列短重码统计模型

图 3 为维数 $l=4$ 短序列在间距 $d=1$ 时统计码元相同短重码的直观表示。图中 v_i 与 v_{i+1} 为 2 个维数 $l=4$，间距为 1 的短序列。重码统计过程的具体

步骤见算法 1。

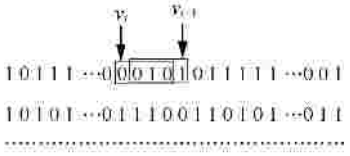

图 3 在间距 d=1 时统计短重码

算法 1 码元相同短重码统计过程。

输入：图像的 LSB Hilbert 扫描序列。

输出： $d=1$ 时，重码在 Hilbert 扫描序列中出现次数的统计累积。

算法步骤如下。

Step1 $i=1$。

Step2 统计 v_i 是否为维数 l 的码元相同短序列：若 v_i 是，则重码的统计累积值加 1，$i=i+1$；若 v_i 不是，则 $i=i+1$。

Step3 若 $i<N-l+1$，转 Step2；否则，转 Step4。

Step4 结束。

4.1.4 Hilbert 扫描序列短重码统计特征

分析上述 Hilbert 扫描序列短重码统计模型，可得如下统计特征。

定理 6 对自然图像 LSB Hilbert 扫描序列做码元相同短重码统计，维数为 l 的序列 v_i 为码元相同短序列的概率为 $p^{l-1}, 1/2\le p\le 1$，且 $p\ge q$，q 为定理 1 中定义的相邻码元相关性系数；而对二值完全随机白噪声序列(加密序列)做码元相同短重码统计，维数 l 的序列 v_i' 码元相同短序列的概率为 $\left(\frac{1}{2}\right)^{l-1}$。

证明 先证第 2 部分结论。

设 $v_i'=b_1'b_2'\cdots b_l'$ 为码元相同短序列的事件是 A'。事件 A' 的状态空间中的状态个数为 2。有：

$$P\{A'\}=P\{b_1'=1,b_2'=1,\cdots,b_l'=1\}+$$
$$P\{b_1'=0,b_2'=0,\cdots,b_l'=0\}=$$
$$P\{b_1'=1\}P\{b_2'=1\}\cdots P\{b_l'=1\}+$$
$$P\{b_1'=0\}P\{b_2'=0\}\cdots P\{b_l'=0\}=$$
$$\left(\frac{1}{2}\right)^l+\left(\frac{1}{2}\right)^l=\left(\frac{1}{2}\right)^{l-1}。$$

下面证第 1 部分结论。

设 $v_i=b_1b_2\cdots b_l$ 码元相同短序列的事件为 A。事件 A 的状态空间中的状态个数为 2。设事件 $B_1^1:b_1=1$；事件 $B_2^1:b_2=1$；…；事件 $B_l^1:b_l=1$；事件 $B_1^0:b_1=0$；事件 $B_2^0:b_2=0$；…；事件 $B_l^0:b_l=0$。有：$P\{A\}=P\{B_1^1B_2^1\cdots B_l^1\}+P\{B_1^0B_2^0\cdots B_l^0\}$，按乘法定理，有：$P\{A\}=P\{B_1^1\}P\{B_2^1\mid B_1^1\}\cdots P\{B_l^1\mid B_1^1B_2^1\cdots B_{l-1}^1\}+P\{B_1^0\}P\{B_2^0\mid B_1^0\}\cdots P\{B_l^0\mid B_1^0B_2^0\cdots B_{l-1}^0\}$

Hilbert 曲线扫描方式保证了：1) b_1 与 b_2、b_2 与 b_3、…、b_{l-1} 与 b_l 是相邻码元；2) 码元 b_1、b_2、…、b_l 同处在局部相关区域的概率很大。

因此，若设 $P\{B_2^1\mid B_1^1\}=q_2^1$，$P\{B_3^1\mid B_1^1B_2^1\}=q_3^1$，…，$P\{B_l^1\mid B_1^1B_2^1\cdots B_{l-1}^1\}=q_l^1$，$P\{B_2^0\mid B_1^0\}=q_2^0$，$P\{B_3^0\mid B_1^0B_2^0\}=q_3^0$，…，$P\{B_l^0\mid B_1^0B_2^0\cdots B_{l-1}^0\}=q_l^0$，由自然图像固有的局部区域相关性，有

$$q=q_2^1\le q_3^1\le\cdots\le q_l^1;q=q_2^0\le q_3^0\le\cdots\le q_l^0$$

不妨设 $q_k^1=q_k^0,k=2,3,\cdots,l$ 故：$P\{A\}=\frac{1}{2}\times q\times q_3^1\times\cdots\times q_l^1+\frac{1}{2}\times q\times q_3^0\times\cdots\times q_l^0=q\times q_3^1\times\cdots\times q_l^1@p^{l-1}$，显然，$\frac{1}{2}\le p\le 1$，且 $p\ge q$。定理获证。

定理 7 自然图像 LSB Hilbert 扫描序列(长度为 N)的维数为 l 码元相同短重码统计量 $X_l(1)$ 服从参数是 $N-l+1,p^{l-1}$ 的二项分布，近似服从 Poisson 分布，即 $P(X_l(1)=k)\approx\frac{\lambda^k e^{-\lambda}}{k!}$，其中，$\lambda=(N-l+1)p^{l-1}$，$\frac{1}{2}\le p\le 1$，$e$ 为自然数；而对于二值完全随机白噪声序列(长度为 N)的维数为 l 码元相同短重码统计量 $X_l(1)$ 服从参数是 $N-l+1,(1/2)^{l-1}$ 的二项分布，近似服从 Poisson 分布，即 $P(X_l(1)=k)\approx\frac{\lambda^k e^{-\lambda}}{k!}$，其中，$\lambda=(N-l+1)(\frac{1}{2})^{l-1}$，$e$ 为自然数。

证明 对长度为 N 的自然图像 LSB Hilbert 扫描序列做码元相同短重码统计，该统计过程满足如下内容。

1) 短序列是否为码元相同短序列，有且仅有是和否 2 种状态。

2) 短序列是否为码元相同短序列的概率与统计次数无关。

3) 在每 2 次比较短序列是否为码元相同短序列的过程中，2 次比较的结果互不影响。

将“一次重码累积”看成是一次实验，而 $d=1$ 码元相同短重码统计过程相当于做了 $N-l+1$ 重

Bernoulli 实验。因此，$d=1$ 码元相同短重码统计服从 Bernoulli 概型。结合定理 6，$X_l(1)$ 服从参数为 $N-l+1, p^{l-1}$ 的二项分布。又由定理 5，$X_l(1)$ 近似服从 Poisson 分布，即 $P(X_l(1)=k)\approx\frac{l^k e^{-l}}{k!}$，其中，$l=(N-l+1)p^{l-1}$，$\frac{1}{2}\le p\le 1$，$e$ 为自然数。

同理可证：对于二值完全随机白噪声序列(长度为 N)的维数为 l 码元相同短重码统计量 $X_l(1)$ 服从参数是 $N-l+1,(1/2)^{l-1}$ 的二项分布，近似服从 Poisson 分布，即：$P(X_l(1)=k)\approx\frac{l^k e^{-l}}{k!}$，其中，$l=(N-l+1)\left(\frac{1}{2}\right)^{l-1}$，$e$ 为自然数。

定理得证。

4.2 采用 Hilbert 扫描序列短重码统计的盲隐写检测算法

4.2.1 算法步骤

本文依据定理 7 揭示的维数为 l 码元相同短重码统计量 $X_l(1)$ 的统计分布规律，与文献[13]一样，采用 Poisson 分布显著性检验来检测隐秘信息。检测步骤见算法 2。

算法 2　采用 Hilbert 扫描序列短重码统计的盲隐写检测算法。

算法步骤如下。

Step1　取出待检测图像的 LSB Hilbert 扫描序列。

Step2　针对待检测 LSB Hilbert 扫描序列，取合适的维数 l 进行码元相同短重码统计(即执行算法 3)。

Step3　以加密序列的统计结果所服从的 Poisson 分布为标准，设定显著性系数 α，对待检测 LSB Hilbert 扫描序列的码元相同短重码统计结果进行 Poisson 显著性判决。

Step4　根据显著性判决结果，判断待检测图像中是否隐含秘密信息。

4.2.2 算法的检测性能分析

在算法 1 中，设事件 $\hat{C}$ 为重码累积成功(即重码的统计累积值加 1)；设事件 $\hat{C}_1$ 为对于自然图像 Hilbert 扫描序列，重码累积成功；设事件 $\hat{C}_2$ 为对于二值完全随机白噪声序列，重码累积成功。显然有：$P(\hat{C}_1)=p^{l-1}$，$P(\hat{C}_2)=(1/2)^{l-1}$。

由定理 1，文献[13]提出的盲隐写检测方法只利用了图像相邻码元的相关性。而由定理 6 及其证明过程可知，本文方法不仅利用了图像相邻码元的相关性，还利用了图像局部区域码元的相关性。因此，比较文献[13]方法的重码累积成功概率比率 $d=(2q)^l$，本文方法的重码累积成功概率比率 $\hat{d}=(2p)^{l-1}, p\ge q$。因为维数 l 可在一定范围选择(如 $l=4,5,\cdots,8$)，可取一个较大的维数 l，使得 $\hat{d}\ge d$，而由本文第 3 节的结论：提高 $X_l(W)$ 检测性能的可能途径是增大重码累积成功概率的比率。因此，本文方法可有效提高盲隐写检测性能。

5 实验结果与分析

5.1 实验建立

实验使用的图像库包含 5 000 幅 512×512 Bmp 格式灰度图像，这些图片取自 USC-SIPI 图像库[20]和自采集的图像库(这些图像来自不同的摄影者、不同的相机和不同的场景)。图片主题包括场景、人物、植物等。

采用 LSB 匹配隐写算法[11]把服从均匀分布的加密序列嵌入到图像的最低位平面中，且按 10 种嵌入率 100%，90%，80%，…，20%，10%等比例嵌入秘密信息。

对载体图像和嵌入不同比例秘密信息的载密图像，用本文提出的方法进行隐写检测，并与文献[13]方法、频率检测、串行检测、Poker 检测和游程检测[12]等方法进行比较分析。本实验中分场景、人物、植物等 3 种类型图像，并在图像库中各任选 1 000 幅图像进行实验。实验中设定的显著性判决系数同为 0.01。

5.2 实验结果分析

表 1~表 3 为不同检测方法针对场景图像、人物图像和植物图像 10 种嵌入方式的检出率。表中的“Cover”项为虚警率。

实验表明，频率检测、串行检测、Poker 检测和游程检测的虚警率都偏高，影响这些检测方法的实际应用。本文方法与文献[13]方法的虚警率相当，明显小于频率检测、串行检测、Poker 检测和游程检测，有效控制了虚警概率。针对 3 种不同类型的图像，本文方法与文献[13]方法一样，其虚警率有显著差别：对场景图像检测的虚警率最小，人物图像次之，对植物图像检测的虚警率最大。另外，3 种类型图像的隐写检测实验结果近似反映了定理 4 的结论：随着选择的短重码维数 l 的增大，检测率也增大。

表 1 **针对场景图像的不同检测方法检测结果**

序号	检测方法	嵌入率										Cover
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
1	**本文方法**(l=4)	7.1	8.9	13.1	16.2	34.6	56.3	71.0	84.4	93.1	97.1	4.1
2	**本文方法**(l=5)	6.6	7.9	13.3	18.0	32.3	55.2	69.1	85.2	93.1	92.9	3.3
3	**本文方法**(l=6)	6.6	7.9	11.5	21.1	38.4	54.5	68.7	87.2	90.7	93.3	1.2
4	**本文方法**(l=7)	8.1	8.1	12.6	21.1	38.5	57.6	72.3	83.3	91.2	94.0	1.2
5	**本文方法**(l=8)	7.3	9.4	13.2	21.3	40.6	57.4	78.8	87.9	94.5	98.3	1.5
6	**文献**[13]**方法**(l=4)	4.3	5.5	12.2	14.3	26.4	43.0	71.9	76.1	78.5	80.0	1.3
7	**文献**[13]**方法**(l=5)	4.8	7.3	10.3	11.1	26.1	42.9	68.3	76.5	78.9	80.2	1.6
8	**文献**[13]**方法**(l=6)	4.7	6.9	10.0	10.1	28.5	45.1	64.4	73.5	81.8	83.4	2.2
9	**文献**[13]**方法**(l=7)	4.9	6.9	9.8	8.9	25.6	47.1	67.3	76.7	83.3	84.6	4.5
10	**文献**[13]**方法**(l=8)	5.9	7.3	10.5	19.9	28.9	53.4	71.9	83.3	88.7	83.9	3.9
11	0，1 **频率检测方法**	74.1	76.0	75.8	78.4	82.0	82.1	88.9	93.1	95.2	100	72.4
12	**串行检测方法**	36.5	34.1	38.2	37.4	50.3	73.3	83.2	92.4	95.3	100	34.1
13	Poker **检测方法**	24.3	22.7	27.9	28.6	33.4	50.5	67.3	92.4	96.5	99.1	21.8
14	**游程检测方法**	20.8	19.6	21.1	21.9	26.7	48.5	67.7	98.1	100	99.3	19.5

表 2 **针对人物图像的不同检测方法检测结果**

序号	检测方法	嵌入率										Cover
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
1	**本文方法**(l=4)	10.2	18.0	31.3	34.1	46.2	67.2	72.1	83.6	88.6	93.6	9.1
2	**本文方法**(l=5)	11.2	18.4	29.1	35.4	45.3	63.8	64.5	84.5	92.4	90.8	7.4
3	**本文方法**(l=6)	10.6	17.5	27.6	36.6	44.2	66.1	71.3	78.4	88.4	94.3	8.4
4	**本文方法**(l=7)	12.3	19.3	25.6	34.8	43.4	68.2	71.9	80.4	97.3	91.5	8.5
5	**本文方法**(l=8)	14.8	23.6	28.7	36.9	46.2	67.7	75.3	82.9	91.8	94.9	9.0
6	**文献**[13]**方法**(l=4)	12.1	20.1	20.6	25.6	39.9	46.5	61.3	84.1	87.6	83.3	7.8
7	**文献**[13]**方法**(l=5)	12.2	17.3	25.3	31.5	38.1	44.3	58.3	80.3	86.3	83.4	9.1
8	**文献**[13]**方法**(l=6)	8.8	10.3	17.9	23.7	31.1	46.4	66.4	80.2	89.4	86.6	7.5
9	**文献**[13]**方法**(l=7)	10.3	10.5	19.4	21.8	30.4	45.6	67.6	81.9	88.6	91.1	7.5
10	**文献**[13]**方法**(l=8)	12.4	9.8	28.6	35.6	44.9	51.2	58.9	83.9	91.1	90.5	9.1
11	0，1 **频率检测方法**	73.4	76.9	77.5	77.4	79.6	80.0	81.2	88.0	86.5	100	73.4
12	**串行检测方法**	60.3	63.4	61.6	63.0	65.3	79.9	80.5	88.9	91.4	99.3	60.6
13	Poker **检测方法**	41.2	44.8	47.7	47.9	52.3	72.1	81.4	86.3	94.5	96.6	41.3
14	**游程检测方法**	34.6	35.5	42.1	40.9	48.5	75.4	85.2	100	100	100	34.7

表 3 **针对植物图像的不同检测方法检测结果**

序号	检测方法	嵌入率										Cover
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	
1	**本文方法**(l=4)	24.8	29.1	30.9	36.7	50.0	62.5	78.9	88.6	89.8	96.1	14.4
2	**本文方法**(l=5)	24.2	28.7	28.7	36.5	42.2	62.5	80.6	88.6	86.3	88.5	12.3
3	**本文方法**(l=6)	26.3	26.5	30.3	34.6	44.5	58.6	82.5	90.3	84.4	88.9	12.3
4	**本文方法**(l=7)	22.7	30.3	30.2	30.5	46.1	70.1	84.9	90.4	86.5	94.3	14.5
5	**本文方法**(l=8)	22.8	24.8	35.4	40.1	50.5	78.0	90.1	84.8	92.4	96.7	17.2
6	**文献**[13]**方法**(l=4)	18.5	26.3	28.4	34.1	36.6	52.4	64.3	84.9	78.1	88.3	14.5
7	**文献**[13]**方法**(l=5)	20.0	26.6	28.5	34.1	40.9	46.7	70.2	76.4	76.3	90.0	14.5
8	**文献**[13]**方法**(l=6)	20.5	24.0	30.6	34.8	42.3	46.6	70.3	72.2	78.5	92.2	14.6
9	**文献**[13]**方法**(l=7)	18.6	26.8	32.7	34.9	42.9	50.9	64.6	70.5	80.7	92.3	16.3
10	**文献**[13]**方法**(l=8)	18.9	22.2	34.8	32.7	46.4	56.8	60.7	78.1	78.9	88.4	16.6
11	0，1 **频率检测方法**	90.5	90.2	92.1	94.3	94.4	94.5	94.5	94.8	98.3	100	88.6
12	**串行检测方法**	68.4	68.4	76.2	70.4	76.8	82.3	86.5	94.2	96.2	100	64.3
13	Poker **检测方法**	42.1	42.2	44.6	46.7	50.2	70.2	84.2	96.0	98.5	94.1	36.2
14	**游程检测方法**	32.1	34.6	36.6	42.2	44.7	64.6	82.8	100	98.5	98.2	32.5

针对 3 种不同类型的图像，本文方法在不同嵌入率情形下都取得了较为明显的优于文献[13]方法的实验结果。直观的比较如图 4 所示，图中的检测率是对不同的 l 取得的最大检测率。

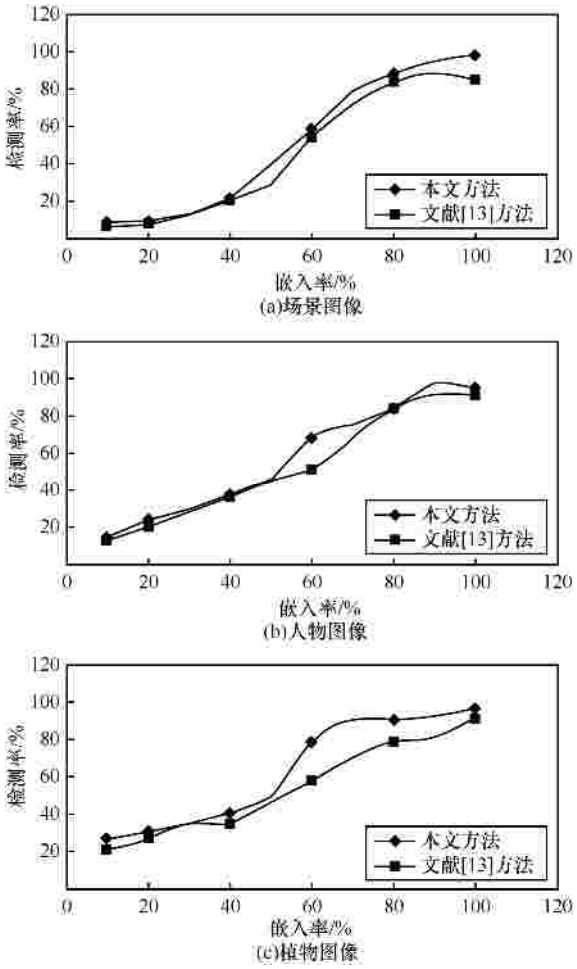

图 4　针对不同类型图像，本文方法和文献[13]方法(对不同的 l，取最大检测率)的检测率比较

另外，本文方法不只是对 LSB 匹配隐写具有有效的盲检测性能，本文方法保持了文献[13]方法一样的通用性。对其他 LSB 隐写的盲检测实验结果不做赘述。

6　结束语

本文提出了一种采用 Hilbert 扫描序列短重码统计的盲隐写检测方法。它有如下特点：充分利用了图像的 Hilbert 曲线扫描方式能够将相邻像素尽量排列在一起，可更为准确地描述图像局部像素空间相关性的特性，使基于码元相同短重码统计的本文方法不仅利用了图像相邻码元的相关性，还利用了图像局部区域码元的相关性，从而比较文献[13]方法，具有更大的重码累积成功概率比率，在有效控制虚警率的前提下，有效提高了隐写检测性能。

本文方法保持了文献[13]方法的通用性，也一样存在如下不足：若隐写算法不在载体图像位平面自相关性强的区域嵌入信息，则检测算法失效；另外，检测性能与自然图像 LSB 序列本身的分布有关，若自然图像位平面本身很近似服从均匀分布，则检测算法失效。实际上，如何降低载体图像内容本身对通用检测性能的影响一直是个公开课题[1]。也是笔者下一步努力的方向。

参考文献：

[1] LUO X Y, WANG D S, WANG P, LIU F L. A review on blind detection for image Steganography[J]. Signal Processing, 2008, 88: 2138-2157.

[2] 王朔中，张新鹏，张卫明. 以数字图像为载体的隐写分析研究进展[J].计算机学报，2009, 32(7): 1247-1263.
WANG S Z, ZHANG X P, ZHANG W M. Recent advances in image-based steganalysis research[J]. Chinese Journal of Computers, 2009, 32(7):1247-1263.

[3] 王朔中，张新鹏，张开文. 数字密写和密写分析[M]. 北京: 清华大学出版社, 2005.
WANG S Z, ZHANG X P, ZHANG K W. Steganography and Steganalysis[M]. Beijing: Tsinghua University Press, 2005.

[4] ANDREW D K. Improved detection of LSB steganography in grayscale images[A]. Proceedings of the 6th International W on Information Hiding[C]. Toronto, Canada: Springer-Verlag, 2004. 97-115.

[5] WESTFELD A, PFITZMANN A. Attacks on steganographic systems[A]. Proceedings of the 3rd International Workshop on Information Hiding[C]. Dresden, Germany: Springer-Verlag, 1999. 61-76.

[6] FRIDRICH J, GOLJAN M, DU R. Detecting LSB steganography in color and gray-scale images[J]. IEEE Multimedia, 2001, 8(4):22-28.

[7] DUMITRESCU S, WU X, WANG Z. Detection of LSB steganography via sample pair analysis[J]. IEEE Trans Signal Process, 2003, 51(7): 1995-2007.

[8] 张涛,平西建. 基于差分直方图实现 LSB 信息伪装的可靠检测[J]. 软件学报,2004,15(1): 151-158.
ZHANG T, PING X J. Reliable detection of spatial LSB steganography based on difference histogram[J]. Journal of Software,2004,15(1): 151-158.

[9] WESTFELD A. Detecting low embedding rates[A]. Proceedings of the 5th International Workshop on Information Hiding[C]. Noordwijkerhout, The Netherlands: Springer-Verlag, 2002. 324-339.

[10] KER A D. Steganalysis of LSB matching in grayscale images[J]. IEEE Signal Process Lett, 2005,12(6):441-444.

[11] 陈铭，张茹，刘凡凡等. 基于区域相关性的 LSB 匹配隐写分析[J]. 通信学报，2010,31(3): 1-11.
CHEN M, ZHANG R, LIU FF, *et al*. Steganalysis of LSB matching

based on regional correlation[J]. Journal on Communications, 2010, 31(3):1-11.

[12] MENEZES M, OORSCHOT P V, VANSTONE S. Handbook of Applied Cryptography[M]. CRC Press, 1996. 169-187.

[13] 王国新, 平西建, 许漫坤等.一种基于短重码间距统计的隐写分析方法[J].中国科学 F 辑:信息科学,2009, 39(4):416-421.
WANG G X, PING X J, XU M K, *et al.* A steganalysis method based on spacing statistics of short duplicate codes[J]. Science in China Series F: Information Sciences, 2009, 39(4): 416-421.

[14] WESTFELD A. Space filling curves in steganalysis[A]. Proceedings of SPIE: Security, Steganography, and Watermarking of Multimedia[C]. San Jose, California, USA:SPIE, 2005.28-37.

[15] 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 Y W, LIU G J, YE S G. Adaptive steganography based on Hilbert filling curve[J]. Acta Electronica Sinica, 2008, 36(12A):35-38.

[16] 高迎, 程涛远, 王珊.基于 Hilbert 曲线的许可证存储策略及查找算法[J]. 软件学报,2006,17(2): 305-314.
GAO Y, CHENG T Y, WANG S. Certificates storage strategy and search algorithm based on Hilbert curve[J]. Journal of Software, 2006,17(2):305-314.

[17] MOON B, JAGADISH H V, FALOUTSOS C, *et al.* Analysis of the clustering properties of the Hilbert space filling cur]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13 (1) :124-141.

[18] THEODORIDIS S, KOUTROUMBAS K. Pattern Recognition[M]. Third Edition. Singapore: Elsevier Pte Ltd, 2006.

[19] 周概容.概率论与数理统计[M].北京:高等教育出版社,2009.
ZHOU G R. Probability and Mathematical Statistics[M]. Beijing: Higher Education Press,2009.

[20] USC SIPI. The USC-SIPI image database[EB/OL]. http://sipi.usc.edu/services/ database/database.html,2012.1.6.

作者简介：

钟尚平（1969-），男，福建武平人，博士，福州大学副教授，主要研究方向为网络信息安全、模式识别等。

徐巧芬（1989-），女，福建德化人，福州大学硕士生，主要研究方向为多媒体安全、模式识别等。

郭文忠（1979-），男，福建泉港人，博士，福州大学副教授，主要研究方向为计算智能与计算机网络等。

廖彬（1984-），男，福建龙海人，福州大学讲师，主要研究方向为网络信息安全、模式识别等。