

基于辨识性统计特征的 PQ 隐密图像识别算法

卢记仓^{1,3}, 刘粉林^{2,3}, 罗向阳^{2,3}, 张轶^{2,4}

(1. 解放军信息工程大学 导航与空天目标工程学院, 河南 郑州 450001;

2. 解放军信息工程大学 网络空间安全学院, 河南 郑州 450001;

3. 数学工程与先进计算国家重点实验室, 河南 郑州 450001;

4. 中国科学院 信息工程研究所 信息安全国家重点实验室, 北京 100093)

摘要: 提出一种基于辨识性统计特征的 PQ (perturbed quantization) 隐密图像识别算法。该算法根据经典 PQ 隐写对图像数据的更改方式, 提取可有效区分该类隐密图像与其他类隐密图像的辨识性统计特征, 并运用 SVM (support vector machines) 分类器进行分类识别。实验结果表明, 本算法能够可靠地将 PQ 隐密图像从 5 类典型 JPEG 隐写 PQ、F5、nsF5、MB1 和 MOD 的隐密图像中识别出来; 即使 F5、nsF5、MB1 和 MOD 的隐密图像不参与分类器的训练, 本算法仍能有效识别 PQ 隐密图像。

关键词: 隐写分析; 隐密图像识别; PQ 隐写; 辨识性统计特征

中图分类号: TP391.4

文献标识码: A

Recognition of PQ stego images based on identifiable statistical feature

LU Ji-cang^{1,3}, LIU Fen-lin^{2,3}, LUO Xiang-yang^{2,3}, ZHANG Yi^{2,4}

(1. School of Navigation and Aerospace Engineering, Zhengzhou Information Science and Technology Institute, Zhengzhou 450001, China;

2. School of Cyberspace Security, Zhengzhou Information Science and Technology Institute, Zhengzhou 450001, China;

3. State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China;

4. State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China)

Abstract: A PQ (perturbed quantization) stego images recognition algorithm is proposed based on identifiable statistical feature. According to the specific changing ways of PQ steganography to image data, the proposed algorithm extracts the identifiable statistical feature that can distinguish PQ stego images from other types of stego images. Then, the SVM (support vector machines) classifier is trained to recognize PQ stego images. Experimental results show that, the proposed algorithm can reliably recognize PQ stego images from multi-class stego images generated by five types of well-known JPEG steganography (PQ, F5ns, F5, MB1 and MOD). Even though the stego images generated by F5, nsF5, MB1 and MOD are not used for training classifier, the proposed algorithm can still effectively recognize PQ stego images.

Key words: steganalysis; stego image recognition; PQ steganography; identifiable statistical feature

1 引言

隐写检测技术在信息安全领域所处的特殊地

位, 以及现实赋予的迫切需求, 已成为研究者对其开展深入研究的不懈动力^[1]。现有隐写检测研究主要集中在特定隐写检测和通用盲检测 2 个方面^[2]:

收稿日期: 2013-11-14; 修回日期: 2014-01-20

基金项目: 国家自然科学基金资助项目 (61272489, 61379151, 61302159); 中国博士后科学基金资助项目 (20110491838, 2012T50842); 信息保障技术重点实验室开放基金资助项目 (KJ-14-108); 河南省科技创新杰出青年基金资助项目 (14410051001); 解放军信息工程大学优博基金资助项目 (BSLWCX201203)

Foundation Items: The National Natural Science Foundation of China (61272489, 61379151, 61302159); China Postdoctoral Science Foundation (20110491838, 2012T50842); Foundation of Science and Technology on Information Assurance Laboratory (KJ-14-108); The Excellent Youth Foundation of Henan Province of China (14410051001); The Doctoral Dissertation Innovation Fund of Zhengzhou Information Science and Technology Institute (BSLWCX201203)

前者是在已知所使用隐写算法的情况下,判断待检测对象是否含有秘密信息,甚至能够估计信息长度和提取秘密信息等;后者通常是在可能使用的隐写算法集合已知的情况下,对待检测对象是否含有秘密信息进行判决。由此可见,如何准确识别和判定隐密图像所使用的隐写算法,对进一步估计信息长度、搜索隐写密钥甚至提取秘密信息等具有重要的意义,是进一步进行隐写分析和取证的关键^[3]。将从多类隐密图像集中特定隐密图像识别的角度出发,对上述问题展开研究。

现有多类隐密图像中特定隐密图像识别的研究主要集中在多类隐密图像的分类方面,这些方法通常是在隐密图像所使用隐写算法集合完全已知的情况下,基于两类分类的思想且仅使用一种盲检测特征对不同类隐密图像进行分类。如针对 JPEG 多类隐密图像,文献[4]最早基于系数直方图和共生矩阵等特征进行分类,文献[5]通过对直方图特征进行扩展,并加入 Markov 转移概率矩阵特征,改进了该分类算法;文献[6]基于小波系数的概率密度函数矩特征,并结合特征选择和降维,给出了一种分类识别算法;针对空域多类隐密图像,文献[7]基于像素差值矩阵的游程长度直方图特征,给出了一种空域常见隐写的分类算法。已有的多类隐密图像分类方法所使用的统计特征均为盲检测中的常见特征,当前盲检测算法所使用的其他类型的特征,如文献[8~10]等所描述的特征,可能也能够用于多类隐密图像的分类,但从现有研究来看,这些方法主要存在以下三方面不足:1)由于已有方法所提取特征的目的在于通用盲检测,因此所提取的特征往往无法反映特定隐写所引起的图像数据统计特性的特殊变化;2)已有方法必须知道隐密图像中的隐写算法集合,难以适用于存在未知类别隐密图像的情况;3)由于已有方法所使用的分类器通常基于已知类别的隐密图像进行训练,因此当加入新的类别的隐密图像时,已有方法需要训练新的分类器。

重点对 JPEG 图像典型隐写 PQ^[11]所生成隐密图像的识别问题展开研究。在 PQ 隐写的检测方面,现有的研究主要基于盲检测的方法对载体图像和隐密图像进行分类,如文献[5,9,12,13]等所提出的盲检测方法,在前期研究中,基于 PQ 隐写原理和更改机制的分析,也提出了一种针对 PQ 系列隐写(PQt, texture_adaptive PQ 和 PQe, energy_adaptive PQ)^[14]的检测方法^[15]。但这些方法也都仅致力于判断待检

测图像中是否有隐藏信息,而对于如何从多类隐密图像中识别出 PQ 隐密图像的研究尚未见报道。

基于 PQ 隐写辨识性统计特征的分析 and 提取,并根据隐写算法集合是否存在已知算法的情况,分别给出了 2 种情况下的 PQ 隐密图像识别算法。首先,分析 PQ 隐写在信息嵌入过程中对图像数据的具体更改方式;然后,基于相邻 DCT 系数的统计相关性差异提取特征,将其作为 PQ 隐写区别于其他类隐写的辨识性统计特征;最后,基于所提取的统计特征,从多类隐密图像集中识别出 PQ 隐密图像。基于 5 类典型 JPEG 隐写 PQ、F5^[16]、nsF5^[14]、MB1^[17]和 MOD^[18],构建隐密图像集并进行了测试,实验结果表明,本算法能够可靠地将 PQ 隐密图像从上述多类隐密图像中识别出来,即使假设 F5、nsF5、MB1 和 MOD 对检测者来说是未知的(不参与分类器的训练),仅基于 PQ 隐密图像和载体图像,提取 184 维的 PQ 隐写辨识性统计特征并训练分类器,本算法也能有效地将 PQ 隐密图像从多类隐密图像中识别出来,平均正确率可达 91.7%。本算法尽管特征维数较低,但与已有的典型检测算法:文献[5]基于 548 维 CCPEV 特征的算法和文献[9]基于 22510 维 ccJRM 特征的算法等相比,本算法具有更优的检测性能。

2 PQ 隐密图像识别

2.1 PQ 隐写原理

PQ 隐写是一种典型的 JPEG 图像失真最小化隐写算法,是目前较为安全的隐写算法之一。该隐写在进行信息嵌入时,根据 DCT 系数在量化取整前的取值,选择更改后对图像失真较小的系数进行信息的嵌入。文献[11]给出了一种具体的信息嵌入方式。

首先,基于隐写前后 2 次量化步长之间的关系,确定可嵌入系数(可以用来嵌入信息的系数)在 8×8 分块中的位置。令 $Q^{(1)}$ 和 $Q^{(2)}$ 分别表示 PQ 隐写前后 JPEG 图像的 2 次量化矩阵, $q_{u,v}^{(1)}$ 和 $q_{u,v}^{(2)}$ 分别表示对应量化矩阵中 (u,v) 位置的量化步长, $1 \leq u,v \leq 8$, 则可嵌入系数及其位置应满足如下条件:当第一次量化取整后的系数值为 α 时,存在整数 β , 使式(1)成立。

$$\alpha q_{u,v}^{(1)} = \beta q_{u,v}^{(2)} + \frac{q_{u,v}^{(2)}}{2} \quad (1)$$

然后,基于失真最小的原则选择可嵌入系数来

隐藏信息。对可嵌入位置中2次量化后未取整的系数 d_i ，根据信息嵌入的长度，通过设定因子 σ 的取值，选择满足条件 $d_i - \lfloor d_i \rfloor \in [0.5 - \sigma, 0.5 + \sigma]$ 的可嵌入系数来隐藏信息，使隐写对图像可能造成的失真最小化，其中 $\lfloor d_i \rfloor$ 表示小于 d_i 的最大整数， σ 通常表示小于0.5的较小正实数。

式(1)中， $q_{u,v}^{(1)}$ 和 $q_{u,v}^{(2)}$ 、 α 和 β 均为整数，因此，第二次量化矩阵 $\mathbf{Q}^{(2)}$ 中满足式(1)关系的量化步长必为偶数，而 $\mathbf{Q}^{(2)}$ 中量化步长为奇数的位置上的系数不可能用于嵌入信息，即在信息嵌入过程中，PQ隐写不会对第二次量化步长为奇数时所对应位置上的系数进行改动，所有的数据更改均分布在偶数步长所对应位置上的系数中。关于PQ隐写原理的更详细分析可参考文献[11,15]。

2.2 基于PQ隐写原理的辨识性特征分析和提取

隐写检测中，载体统计特征的选取是影响检测性能的关键，特征在隐写前后的变化越大，其对载体和隐密图像的可分性就越好^[19]。同样，若要从多类隐密图像中识别PQ隐密图像，提取对PQ隐写较敏感、同时对其他类隐写不敏感或敏感度较低的统计特征是进行准确识别的关键。若要提取具有上述特性的统计特征，则需要根据PQ隐写不同与其他类隐写特有的信息嵌入和更改方式，构建合适的特征提取源，从而提取捕捉其特有更改方式的敏感特征。事实上，不同隐写对图像数据的更改方式不同，对图像中不同特征提取源数据统计特性的影响也不同。因此，针对某特定隐写，通过构建合适的特征提取源，就有可能提取具有上述特性的敏感特征，这在当前一些典型的特定隐写检测方法^[20-22]中已有所体现，如文献[22]根据DCT域隐写破坏系数直方图分布模型的特点，通过系数直方图的建模和分析，给出了一种针对DCT域隐写的检测方法。

对JPEG自然图像来说，相邻位置上的系数通常具有相似的分布特性。由2.1节的分析和介绍可知，PQ隐写根据图像量化步长的奇偶性以及第二次量化中的取整失真来选择可嵌入信息的系数，信息嵌入时将使偶数量化步长所对应位置上系数的分布发生变化，而奇数量化步长对应位置上系数的分布保持不变，致使两类位置上系数分布之间的相似性遭到破坏。而对于其他类DCT域隐写，如F5、nsF5、MB1和MOD等，均未根据量化步长来确定可嵌入信息的系数，即在信息嵌入过程中对偶数量

化步长和奇数量化步长同等对待，从而使隐写后两类位置上系数分布之间的关系不同于PQ隐写后的系数分布，甚至仍保持与自然图像系数分布相似的统计关系。上述特点正是PQ隐写不同与其他类隐写的独特之处。

根据系数所处的不同类别位置（是否可用于嵌入信息）分别构建特征提取源，该构建方式不同于以往的通用盲检测（未根据系数位置的不同类别分别构建特征提取源）。在特征提取时，从不同类别位置的系数中分别提取常见的直方图分布和共生矩阵特征，通过计算它们的直方图分布差异和共生矩阵差异，提取PQ隐写区别与其他类隐写的辨识性统计特征。

首先，提取直方图差异特征。JPEG自然图像中，DCT系数直方图通常关于0对称，因此，在提取特征前，将所有系数取绝对值。令 $h_{k,u,v}$ 表示 8×8 分块中 (u,v) 位置系数取值为 k 的直方图，则其与水平、垂直和对角方向相邻位置的直方图差异特征 $F_{h1,PQ}$ 、 $F_{h2,PQ}$ 和 $F_{h3,PQ}$ 分别为

$$F_{h1,PQ} = \{h_{k,u,v} - h_{k,u,v+1} \mid 0 \leq k \leq T, 0 \leq u \leq 8, 0 \leq v \leq 7\} \quad (2)$$

$$F_{h2,PQ} = \{h_{k,u,v} - h_{k,u+1,v} \mid 0 \leq k \leq T, 0 \leq u \leq 7, 0 \leq v \leq 8\} \quad (3)$$

$$F_{h3,PQ} = \{h_{k,u,v} - h_{k,u+1,v+1} \mid 0 \leq k \leq T, 0 \leq u \leq 7, 0 \leq v \leq 7\} \quad (4)$$

其中， T 为阈值，用以控制特征统计范围和维数。

然后，针对待检测图像奇数量化步长和偶数量化步长对应位置上的系数，分别统计2种位置上分块间相邻系数的共生矩阵，将其差异作为特征。令 \mathbf{J} 表示DCT系数矩阵，其水平和垂直方向分块数量分别为 X 和 Y ， Ω_1 表示奇数量化步长在量化矩阵中的位置集合， Ω_2 表示偶数量化步长对应的位置集合，并令 $F_{C1,PQ}$ 和 $F_{C2,PQ}$ 分别表示2种量化步长所对应的共生矩阵，其计算方式如下

$$F_{C1,PQ}(a,b) = \frac{\sum_{\xi=1}^{|\Omega_1|} (L(\Omega_1, \xi, 0, 1) + L(\Omega_1, \xi, 1, 0) + L(\Omega_1, \xi, 1, 1))}{|\Omega_1| \times (X(Y-1) + (X-1)Y + (X-1)(Y-1))} \quad (5)$$

$$F_{C2,PQ}(a,b) = \frac{\sum_{\omega=1}^{|\Omega_2|} (L(\Omega_2, \omega, 0, 1) + L(\Omega_2, \omega, 1, 0) + L(\Omega_2, \omega, 1, 1))}{|\Omega_2| \times (X(Y-1) + (X-1)Y + (X-1)(Y-1))} \quad (6)$$

其中, $F_{C_1,PQ}(a,b)$ 和 $F_{C_2,PQ}(a,b)$ 分别表示 $F_{C_1,PQ}$ 和 $F_{C_2,PQ}$ 中与值对 (a,b) 应的矩阵元素; a 和 b 为共生矩阵统计值, 即待统计的系数值, 当阈值为 T 时, $0 \leq a, b \leq T$; $|\Omega_1|$ 和 $|\Omega_2|$ 分别表示相应集合 Ω_1 和 Ω_2 的元素个数; 函数 $L(\bullet)$ 计算方式如下

$$L(\Omega, \xi, \Delta_1, \Delta_2) = \sum_{x=1}^{X-\Delta_1} \sum_{y=1}^{Y-\Delta_2} \cdot$$

$$\delta(J_{\Omega(\xi)}(x, y) = a, J_{\Omega(\xi)}(x + \Delta_1, y + \Delta_2) = b) \quad (7)$$

其中, $\Omega(\xi)$ 表示第 ξ 个位置, $1 \leq \xi \leq |\Omega|$; Δ_1 和 Δ_2 用于控制相邻系数的方向, 如 $\Delta_1 = 0$, $\Delta_2 = 1$ 表示水平方向相邻系数; $\delta(\Delta_1, \Delta_2)$ 为判别函数, 当且仅当条件 Δ_1 和 Δ_2 都满足时, $\delta(\Delta_1, \Delta_2) = 1$, 否则 $\delta(\Delta_1, \Delta_2) = 0$ 。可得共生矩阵差异特征为

$$F_{C,PQ} = \{F_{C_2,PQ}(a,b) - F_{C_1,PQ}(a,b) \mid 0 \leq a, b \leq T\} \quad (8)$$

最后, 可得 PQ 隐写区别与其他类隐写的辨识性统计特征为

$$F_{PQ} = F_{h_1,PQ} \cup F_{h_2,PQ} \cup F_{h_3,PQ} \cup F_{C,PQ} \quad (9)$$

2.3 隐密图像识别算法

特定类隐密图像的辨识性特征仅在该特定类隐写嵌入信息后变化比较明显, 而对其他类隐写来说, 这类辨识性特征在隐写前后几乎保持不变或相对变化很小。因此, 若要从多类隐密图像中识别出特定类隐密图像, 在训练分类器时, 若多类隐密图像中仅部分隐写算法对检测者来说是已知的, 则仅使用部分其他类隐密图像和该特定类隐密图像作为训练图像集; 若隐密图像可能使用的所有隐写算法对检测者来说均未知, 则仅将载体图像和该特定类隐密图像作为训练图像集。然后, 在不同情况下, 运用训练好的分类器, 将该特定类隐密图像从多类隐密图像中可靠识别出来。

基于 2.2 节所分析和提取的 PQ 隐写辨识性统计特征, 将根据待检测隐密图像可能使用的隐写算法集合是否存在已知隐写的情况, 给出如下 2 种 PQ 隐密图像分类识别的具体算法。

1) 部分隐写算法已知

主要步骤如下。

①分类器训练。首先基于 PQ 隐写和已知的隐写算法集合 G 构建隐密图像训练集, 并采用 PQ 隐写的辨识性统计特征提取方法提取所有隐密图像

的统计特征; 然后, 基于所提取的特征, 选择合适的分类器训练相应的识别分类模型。

②待像特征提取检测图。根据 2.2 节的特征提取方法, 提取所有待检测图像集的特征。

③图像分类。运用训练好的识别分类模型, 将所有待检测隐密图像分为 PQ 隐密图像和非 PQ 隐密图像。

其算法结构如下。

算法 1 部分隐写算法已知时 PQ 隐密图像识别算法结构描述。

输入: 待检测隐密图像集 I , 已知的 n 种隐写算法集合 G , 载体图像集 C 。

输出: PQ 隐密图像集 I_{PQ} , 非 PQ 隐密图像集 I_{no-PQ} 。

```

 $S_{PQ} \leftarrow Embed(C, PQ);$  // 构建用于分类器训练的 PQ 隐密图像集,  $Embed()$  为隐写函数
for  $i \leftarrow 1$  to  $n$  do
     $S_{G(i)} \leftarrow Embed(C, G(i));$  // 构建用于分类器训练的第  $i$  种已知隐写  $G(i)$  的隐密图像集
end for
 $F_{Train} \leftarrow Fext(S_{PQ}, S_G);$  // 根据 2.2 节的特征提取方法, 提取所构建隐密图像集的特征,
//  $Fext()$  为特征提取函数
 $Cls \leftarrow Train(F_{Train});$  // 训练分类器,  $Train()$  为分类器训练函数
 $F_I \leftarrow Fext(I);$  // 根据 2.2 节的特征提取方法, 提取待检测隐密图像集  $I$  的特征,
 $(I_{PQ}, I_{no-PQ}) \leftarrow Classify(Cls, F_I);$  // 使用分类器进行分类,  $Classify()$  为分类函数
end

```

2) 所有隐写算法均未知

主要步骤如下。

①分类器训练。首先基于 PQ 隐写构建隐密图像训练集, 并采用 PQ 隐写的辨识性统计特征提取方法提取 PQ 隐密图像和载体图像的统计特征; 然后, 基于所提取的特征, 选择合适的分类器训练相应的识别分类模型;

②待检测图像特征提取。根据 2.2 节的特征提取方法, 提取所有待检测图像集的特征。

③图像分类。运用训练好的识别分类模型, 将所有待检测隐密图像分为 PQ 隐密图像和非 PQ 隐密图像。

算法 2 所有算法均未知时 PQ 隐密图像识别

算法结构描述。

输入：待检测隐密图像集 I ，载体图像集 C 。

输出：PQ 隐密图像集 I_{PQ} ，非 PQ 隐密图像集

I_{no-PQ} 。

begin

$S_{PQ} \leftarrow Embed(C, PQ)$; // 构建用于分类器训练的 PQ 隐密图像集, $Embed()$ 为隐写函数

$F_{Train} \leftarrow Fext(S_{PQ}, C)$; // 根据 2.2 节的特征提取方法, 提取所构建隐密图像集和载体图像的特征,
// $Fext()$ 为特征提取函数

$Cls \leftarrow Train(F_{Train})$; // 训练分类器, $Train()$ 为分类器训练函数

$F_I \leftarrow Fext(I)$; // 根据 2.2 节的特征提取方法, 提取待检测隐密图像集 I 的特征,

$(I_{PQ}, I_{no-PQ}) \leftarrow Classify(Cls, F_I)$; // 使用分类器进行分类, $Classify()$ 为分类函数

end

3 实验结果及分析

基于主流的测试图像库对本算法进行实验验证和分析。原始图像是来自于 BossBase-1.01^[21] 的 5000 幅 PGM 格式的灰度图像, 尺寸为 512×512 。图像库构建时, 首先将图像转换为质量因子为 85 的 JPEG 图像; 然后, 基于 PQ 隐写构建信息嵌入比率分别为 0.05、0.10、0.15 和 0.20 bpnc (bit per non-zero AC DCT coefficient: 嵌入的信息长度占非 0 AC DCT 系数总量的比率) 的隐密图像, 第二次质量因子为 70, 同时生成相应的载体图像。基于 F5、nsF5、MB1 和 MOD 等 4 种 JPEG 隐写, 构建信息嵌入比率分别为 0.05、0.10、0.15 和 0.20 bpnc 的隐密图像。综上, 可得到包含 $5\,000 \times 4 \times 5 = 100\,000$ 幅隐密图像和 5 000 幅载体图像的图像库, 具体如表 1 所示。

表 1 实验图像库

图像类别	信息嵌入比率/bpnc	图像数量
载体图像	0	5 000
PQ 隐密图像	0.05、0.10、0.15、0.20	$5\,000 \times 4 = 20\,000$
F5 隐密图像	0.05、0.10、0.15、0.20	$5\,000 \times 4 = 20\,000$
nsF5 隐密图像	0.05、0.10、0.15、0.20	$5\,000 \times 4 = 20\,000$
MB1 隐密图像	0.05、0.10、0.15、0.20	$5\,000 \times 4 = 20\,000$
MOD 隐密图像	0.05、0.10、0.15、0.20	$5\,000 \times 4 = 20\,000$

注1: BOSS: Available: <http://exile.felk.cvut.cz/boss/BOSSFinal/2013>。

根据第 2 节 PQ 隐写的辨识性特征提取方法, 将特征提取阈值 T 设置为 7, 为了控制特征维数, 在直方图差异特征提取时, 每个方向仅提取低频 5 个相邻位置对的差异特征, 即式(2)~式(4)中 $(u, v) \in \{(1, 2), (1, 3), (2, 1), (2, 2), (3, 1)\}$, 则 3 个方向提取的直方图差异特征维数为 $8 \times 5 \times 3 = 120$, 同时根据式(8)可得共生矩阵差异特征维数为 $8 \times 8 = 64$, PQ 隐写辨识性特征的总维数为 184。

在分类器方法选择方面, 相对于较早使用的 Fisher 线性分类器和 BP (back propagation) 神经网络分类器等, 近年来, 基于低维统计特征的隐写检测算法往往使用基于高斯核的 SVM 分类器^[23], 该分类器训练方法通过将统计特征映射到更高维的特征空间, 进而寻找超平面以对样本进行分类, 通常具有更优的检测性能。因此, 在实验测试时, 本算法也将使用基于高斯核的 SVM 分类器, 每组实验首先重复实验 10 次; 然后, 将 10 次结果的平均值作为最终的结果。接下来的实验主要从 3 个方面进行, 分别如下。

1) PQ 隐写辨识性特征实验分析。该部分实验对所构建的多类隐密图像库, 其中的任两类图像均进行训练和分类, 分析基于 PQ 隐写原理分析提取的 PQ 隐写辨识性特征分别在 PQ 隐密图像与其他类隐密图像的分类识别、以及其他类隐密图像之间的分类识别的性能。

2) 部分隐写算法已知时的 PQ 隐密图像识别。该部分实验假设多类隐密图像集中仅有部分隐写算法对检测者来说是已知的, 即仅基于 PQ 隐密图像和部分已知隐写的隐密图像训练分类器, 其他类隐密图像不参与训练, 分析所提的 PQ 隐写辨识性特征在这种情况下的分类识别性能。

3) 所有隐写算法均未知时的 PQ 隐密图像识别。该部分实验假设多类隐密图像集中所有隐写算法对检测者来说均是未知的, 即仅基于 PQ 隐密图像和载体图像训练分类器, 其他类隐密图像均不参与训练, 进而分析所提的 PQ 隐写辨识性特征在这种情况下的分类识别性能。

3.1 PQ 隐写辨识性特征实验分析

基于所构建的隐密图像库, 分别对任意两类隐密图像进行训练和分类, 以测试所提取的 PQ 隐写辨识性特征对不同类隐密图像对的分类性能, 如 PQ vs F5 表示将 PQ 和 F5 隐密图像作为一对。每种嵌入率下 3 000 幅隐密图像用于训练,

剩下的 2 000 幅用于测试, 则 4 种嵌入率 0.05、0.10、0.15 和 0.20 bpnc 下每对隐密图像参与训练和测试的数量分别为: $3\,000 \times 4 \times 2 = 24\,000$ 和 $2\,000 \times 4 \times 2 = 16\,000$ 。PQ 隐密图像分别与其他类隐密图像进行测试的平均正确率(正确分类的图像数量与测试图像总数的比率)如表 2 所示, 其他类隐密图像中的任两类进行测试的平均正确率如表 3 所示。

由表 2 和表 3 的对比结果可知, 基于所提取的 PQ 隐写辨识性统计特征对多类隐密图像中的任意两类隐密图像进行分类时, 若两类隐密图像中有一类是 PQ 隐密图像, 则所提取的特征能够很好地将两类隐密图像区分开, 当嵌入率大于 0.05 bpnc 时, 平均正确检测率均在 91.65%以上。而当两类隐密图像均不是 PQ 隐密图像时, 由表 3 结果可知, 基于 PQ 隐写的辨识性特征在进行分类时的正确率较低, 在对嵌入率为 0.05 bpnc 的隐密图像进行分类时几乎等于随机猜测, 在嵌入率为 0.20 bpnc 时的最高平均正确率也仅为 90.20%。

上述结果表明, 基于 PQ 隐写原理分析所提取的特征仅在 PQ 隐密图像与其他类隐密图像进行分类时具有较高的检测正确率, 而在除 PQ 隐密图像以外的其他类隐密图像之间进行分类时正确率较低, 从而验证了基于 PQ 隐写原理分析所提取的特征能够作为 PQ 隐写区别与其他类隐写的辨识性统计特征。

3.2 部分隐写算法已知时的 PQ 隐密图像识别

现有的隐密图像分类识别研究通常采用检测精度较高的通用盲检测特征, 文献[5]是一种较为典

型的多类隐密图像分类算法, 其基于图像校准所提取的 548 维 CCPEV 特征在隐密图像分类中具有较好的性能, 该算法所使用的分类器训练方法也是基于高斯核的 SVM 分类器; 此外, 文献[9]提取高维富模型特征集(特征维数为 22 510)作为隐写检测特征, 而对于 SVM 分类器来说, 高维富模型特征集往往会出现“维数灾难”问题, 并且空间和时间复杂度会随维数的增加而呈级数增长, 因此, 使用基于集成分类的方法来训练分类器, 进而对隐密图像进行检测, 不但降低了复杂度, 而且在现有的隐写检测方法中性能表现优越。本算法基于 PQ 隐写辨识性特征的识别结果将与上述文献[5]基于 CCPEV 特征的识别结果和文献[9]基于 ccJRM 特征的识别结果进行比较, 以验证本算法的有效性。为了客观反映文献[5]和文献[9]中检测方法的性能, 这 2 种检测算法仍采用原文中的分类器训练方法进行分类检测。

假设多类隐密图像集中仅有 F5 隐写对检测者来说是已知的, 即仅基于 PQ 隐密图像和 F5 隐密图像训练分类器, 其他类隐密图像(nsF5、MB1 和 MOD)不参与训练, 进而将多类隐密图像识别为 PQ 隐密图像和非 PQ 隐密图像。两类隐密图像中每种嵌入率下仍使用 3 000 幅用于训练, 剩下的 2 000 幅用于测试, 则 4 种嵌入率 0.05、0.10、0.15 和 0.20 bpnc 下参与训练的隐密图像数量为: $3\,000 \times 4 \times 2 = 24\,000$, 其他类隐密图像(nsF5、MB1 和 MOD)也使用相对应的 2 000 幅用于测试, 测试图像数量为: $2\,000 \times 4 \times 5 = 40\,000$ 。本算法和文献[5]及文献[9]算法的检测结果对比如表 4 所示。

表 2 PQ 隐密图像分别与其他类隐密图像进行训练和分类的平均正确率

嵌入率/bpnc	PQ vs F5	PQ vs nsF5	PQ vs MB1	PQ vs MOD
0.05	92.47	91.65	92.27	91.87
0.10	97.67	96.93	97.65	97.15
0.15	98.27	97.23	98.51	97.66
0.20	98.55	97.22	98.81	97.94

表 3 其他类隐密图像中的任两类进行训练和分类的平均正确率

嵌入率/bpnc	F5 vs nsF5	F5 vs MB1	F5 vs MOD	nsF5 vs MB1	nsF5 vs MOD	MB1 vs MOD
0.05	57.57	63.18	59.05	63.08	56.67	63.13
0.10	68.47	75.46	70.26	77.24	65.58	77.07
0.15	76.11	84.40	78.14	85.40	71.58	85.84
0.20	78.85	90.20	83.47	87.84	74.98	89.39

表4 仅F5隐写已知时将多类隐密图像识别为PQ隐密图像的概率(%), 对PQ为正确率, 对非PQ为错误率

嵌入率/bpnc	检测算法	PQ	非PQ = (F5 + nsF5 + MB1 + MOD) / 4
0.05	文献[5]算法	90.77	20.87 = (11.31 + 25.01 + 23.31 + 23.85) / 4
	文献[9]算法	99.21	10.17 = (0.55 + 14.29 + 12.51 + 13.31) / 4
	本算法	90.03	6.59 = (5.09 + 7.44 + 6.68 + 7.16) / 4
0.10	文献[5]算法	95.34	18.42 = (4.87 + 25.15 + 23.14 + 20.54) / 4
	文献[9]算法	99.73	9.54 = (0.25 + 13.91 + 11.74 + 12.24) / 4
	本算法	98.52	5.52 = (3.19 + 7.18 + 5.64 + 6.08) / 4
0.15	文献[5]算法	95.46	17.03 = (2.14 + 24.46 + 23.85 + 17.66) / 4
	文献[9]算法	99.73	9.12 = (0.16 + 13.71 + 11.50 + 11.12) / 4
	本算法	98.52	5.52 = (3.19 + 7.18 + 5.64 + 6.08) / 4
0.20	文献[5]算法	95.84	16.54 = (0.96 + 24.06 + 25.97 + 15.58) / 4
	文献[9]算法	99.78	8.95 = (0.10 + 13.50 + 11.93 + 10.25) / 4
	本算法	98.68	4.26 = (1.58 + 6.26 + 4.77 + 4.43) / 4

由表4结果可知, 在仅有F5隐写已知(即nsF5、MB1和MOD隐密图像不参与分类器训练)的情况下, 基于所提出的PQ隐写辨识性特征, 能够较好地识别PQ隐密图像。与文献[5]和文献[9]算法相比, 尽管本算法对PQ隐密图像的识别率略低于文献[9]算法, 但文献[9]算法将更多的非PQ隐密图像误判为PQ隐密图像。对未参与训练的非PQ隐密图像, 本算法的误判率均在8%以下, 而文献[5]算法的误判率绝大多数在20%以上, 文献[9]算法的误判率均在10%以上。上述结果表明, 尽管多类隐密图像中的部分类别隐密图像不参与分类器的训练, 本算法仍能较为准确地识别PQ隐密图像和非PQ隐密图像, 即所提出的特征能够作为PQ隐写的辨识性特征。

此外, 经过实验表明, 当仅有一种隐写算法nsF5、MB1或MOD已知时, 3种检测算法在不同嵌入率下对不同类隐密图像的分类结果与表4结果相似, 仅通过不同隐写算法已知时PQ隐密图像识别结果的ROC(receiver operation characteristic)曲线图, 这是当前隐写检测研究中常用的性能比较方法, 比较不同检测算法的性能, 对比结果如图1所示(其中虚警率是非PQ隐密图像误判为PQ隐密图像的概率, 正确检测率是对PQ隐密图像正确识别的概率)。同时, 计算ROC曲线下的面积(面积越大, 检测算法性能越好), 其与每种情况下的平均正确率(所有类别隐密图像正确分类的数量与隐

密图像总数的比率)和平均时间耗费(分类用于训练和测试的总时间耗费)分别列于表5中。

图1中的ROC曲线图及表5中的平均正确率和ROC曲线下面积对比结果进一步反映了本文算法整体性能优于文献[5]和文献[9]算法, 验证了本算法在仅部分算法已知时能够更为可靠地从多类隐密图像中识别PQ隐密图像和非PQ隐密图像。同时, 本算法的时间耗费明显小于文献[5]和文献[9]算法, 具有更快的分类识别速度。

3.3 所有隐写算法均未知时的PQ隐密图像识别

假设多类隐密图像集中所有隐写算法对检测者来说均是未知的, 即仅基于PQ隐密图像和载体图像训练分类器, 其他类隐密图像(F5、nsF5、MB1和MOD)均不参与训练, 进而从多类隐密图像集中识别出PQ隐密图像。PQ隐密图像中每种嵌入率下仍使用3000幅用于训练, 并用对应的3000幅载体图像参与训练, 则参与训练的图像数量为: $3000 \times 4 + 3000 = 15000$, 剩下的2000幅用于测试, 其他类隐密图像也使用相对应的2000幅用于测试, 测试的隐密图像数量为: $2000 \times 4 \times 5 = 40000$ 。本算法和文献[5]及文献[9]算法的检测结果对比如表6所示。

由表6结果可知, 在所有隐写算法均未知的情况下(其他类隐密图像均不参与分类器训练), 基于所提出的PQ隐写辨识性特征, 仍能较好地识别PQ隐密图像。在嵌入率为0.05bpnc时, 本算法对PQ隐密图像的识别率以及对nsF5和

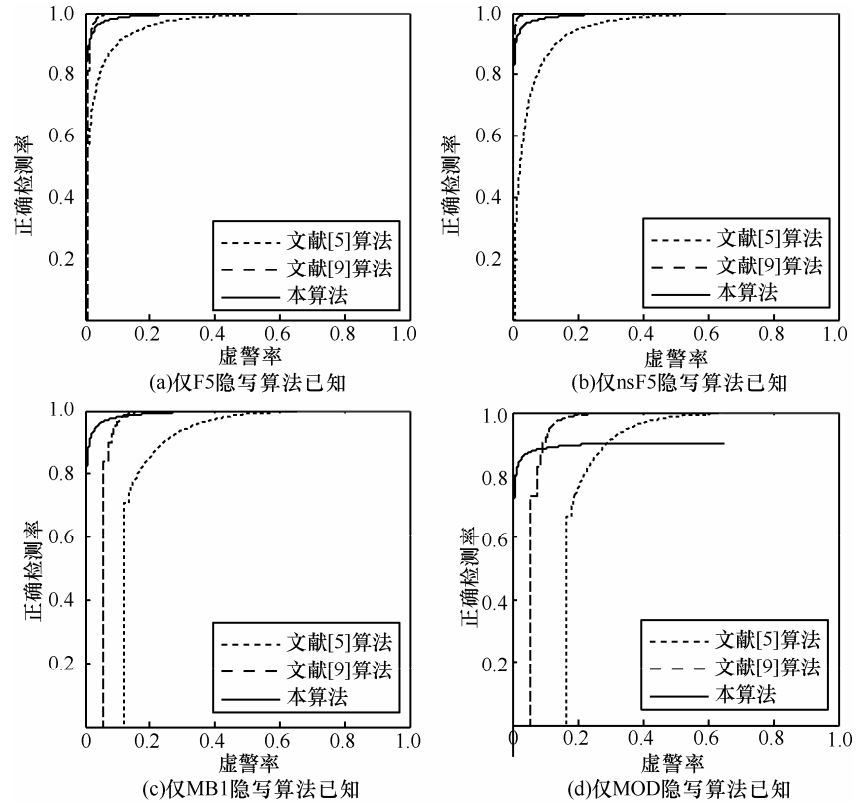


图 1 部分隐写算法已知时的 PQ 隐密图像识别结果 ROC 曲线对比

表 5 部分隐写算法已知时的 PQ 隐密图像识别平均正确率、平均时间耗费(min:sec)及 ROC 曲线下面积对比

已知隐写算法	平均正确率			平均时间耗费			ROC 曲线下面积		
	文献[5]算法	文献[9]算法	本算法	文献[5]算法	文献[9]算法	本算法	文献[5]算法	文献[9]算法	本算法
F5	84.28	92.37	95.08	38:42	31:23	09:49	0.962 2	0.988 5	0.991 6
nsF5	84.50	97.92	96.06	58:38	37:16	11:33	0.944 7	0.996 0	0.991 7
MB1	62.47	84.19	94.46	18:04	33:50	09:59	0.845 9	0.939 2	0.990 1
MOD	65.44	82.43	95.54	29:15	33:39	09:34	0.797 9	0.931 4	0.991 8

表 6 所有隐写算法均未知时将多类隐密图像识别为 PQ 隐密图像的概率(%), 对 PQ 为正确率, 对非 PQ 为错误率

嵌入率/bpnc	检测算法	PQ	非 PQ = (F5 + nsF5 + MB1 + MOD) / 4
0.05	文献[5]算法	95.31	18.87 = (29.70 + 11.21 + 28.98 + 5.59) / 4
	文献[9]算法	98.66	6.88 = (13.62 + 3.21 + 8.93 + 1.75) / 4
	本算法	93.09	8.96 = (8.29 + 9.02 + 9.66 + 8.85) / 4
0.10	文献[5]算法	98.68	45.08 = (81.91 + 26.09 + 66.25 + 6.05) / 4
	文献[9]算法	99.67	25.98 = (62.70 + 7.27 + 31.98 + 1.95) / 4
	本算法	99.19	9.33 = (8.09 + 9.30 + 11.13 + 8.78) / 4
0.15	文献[5]算法	98.87	60.88 = (97.65 + 50.73 + 88.36 + 6.78) / 4
	文献[9]算法	99.60	43.90 = (94.14 + 18.96 + 60.16 + 2.35) / 4
	本算法	99.29	9.99 = (8.42 + 9.66 + 13.01 + 8.88) / 4
0.20	文献[5]算法	99.06	69.55 = (99.57 + 74.52 + 95.89 + 8.20) / 4
	文献[9]算法	99.71	55.34 = (99.44 + 41.17 + 77.69 + 3.05) / 4
	本算法	99.26	10.83 = (8.95 + 10.04 + 15.55 + 8.79) / 4

MOD 隐密图像的误识率低于文献[9]算法，这是因为相对于其他隐写，nsF5 和 MOD 隐写在低嵌入率对图像造成的失真更小，隐写前后的特征变化很微小，其特征更接近于载体图像的特征，而这种情况下的分类器是基于 PQ 隐密图像和载体图像进行训练的，因此 2 种算法会更多地将其与载体图像分为一类，即较好地将其识别为非 PQ 隐密图像。本算法在嵌入率为 0.05 bpnc 时对 PQ 隐密图像的识别率低于文献[9]算法，这也是文献[9]算法的优势之处，对参与训练的类别的隐密图像具有较好的性能。但是，文献[5]和文献[9]算法会将更多的 F5 和 MB1 隐密图像误判为 PQ 隐密图像。随着嵌入率的升高，如 0.15 bpnc 和 0.20 bpnc 时，文献[5]和文献[9]算法几乎将所有的 F5 隐密图像误判为 PQ 隐密图像。从总体上来看，本算法在不同信息嵌入率下的平均误判率均较低，而文献[5]和文献[9]算法的平均误判率随嵌入率的增加将急剧升高，本算法的总体性能优于已有算法。上述结果表明，在从多类隐密图像中识别 PQ 隐密图像时，尽管其他类隐密图像均不参与分类器的训练，本算法仍能较为准确地识别 PQ 隐密图像和非 PQ 隐密图像。

为了进一步验证 3 种算法对 PQ 隐密图像识别的整体性能，图 2 给出了相应的 ROC 曲线对比结果，ROC 曲线面积、平均识别正确率（所有类别隐密图像正确分类的数量与隐密图像总数的比率）及平均时间耗费对比结果如表 7 所示。

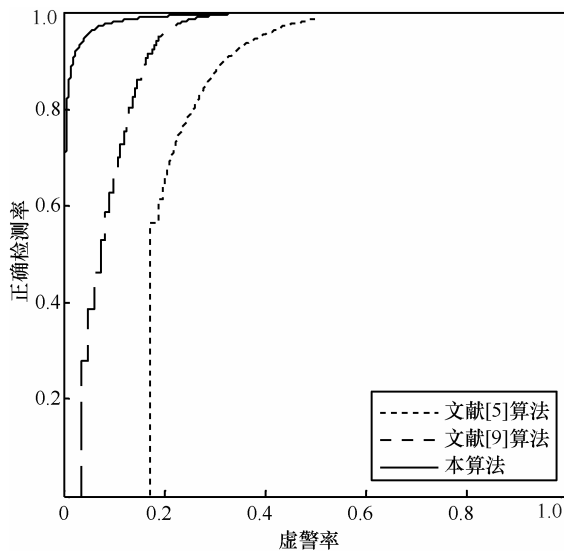


图2 所有隐写算法均未知时的PQ隐密图像识别结果ROC曲线对比

表7 所有隐写算法均未知时的隐密图像识别平均正确率、平均时间耗费(min:sec)及ROC曲线下面积对比

检测算法	平均正确率	平均时间耗费	ROC曲线下面积
文献[5]算法	60.72	14:12	0.781 1
文献[9]算法	73.46	13:12	0.910 9
本算法	91.72	04:34	0.988 7

图 2 中的 ROC 曲线图及表 7 中的平均正确率和 ROC 曲线下面积对比结果进一步反映了本算法整体性能优于文献[5]和文献[9]算法，验证了即使多类隐密图像中的其他类隐密图像均不参与分类器的训练，本算法仍能较为可靠地从多类隐密图像中识别 PQ 隐密图像和非 PQ 隐密图像。同时，本算法的时间耗费明显小于文献[5]和文献[9]算法，具有更快的分类识别速度。

由 3.1~3.3 节的实验结果可知，无论隐密图像集所使用的隐写算法集合是部分已知，还是所有隐写算法均未知，即除待识别的 PQ 隐密图像外，其他类隐密图像是否参与分类器的训练，本算法均能可靠地从多类隐密图像中识别 PQ 隐密图像和非 PQ 隐密图像。尽管已有的典型算法在特定隐写检测中可能具有较高的分类正确率，但在部分类别隐密图像不参与分类器训练的情况下，已有算法很容易将未参与训练的其他类隐密图像误判为 PQ 隐密图像，本算法相比已有算法具有更高的识别正确率和更快的识别速度。上述结果验证了基于 PQ 隐写原理分析所提取的统计特征更适合作为 PQ 隐密图像区别与其他类隐密图像的辨识性统计特征。

4 结束语

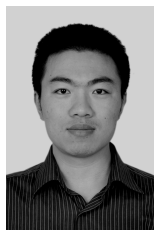
针对多类隐密图像中的特定类隐密图像识别问题，基于相邻系数的统计相关性差异提取了 PQ 隐密图像的辨识性统计特征，给出了一种从多类隐密图像中识别出 PQ 隐密图像的识别算法。基于 JPEG 图像中常见的 PQ、F5、nsF5、MB1 和 MOD 等 5 类隐写的实验结果表明：本算法能够较为可靠地将 PQ 隐密图像从多类隐密图像集中识别出来；即使在假设待检测隐密图像集中存在未知隐写算法生成隐密图像的情况下，仅基于提取的 PQ 隐写辨识性特征，也能够将 PQ 隐密图像从多类隐密图像集中有效地识别出来；与已有典型隐写检测算法相比，本算法具有更高的正确识别率和更快的检测速度。然而，对 PQ 隐密图像识别的辨识性统计特

征进行了研究, 但仍有待进一步改进, 同时对 PQ 隐写的改进算法 PQ_t 和 PQ_e, 即 PQ 系列隐写的隐密图像识别, 以及现有的其他类多种多样的隐写算法, 在基于辨识性统计特征分析和提取的隐密图像识别方面, 尚有待开展进一步的深入研究。

参考文献:

- [1] 沈昌祥, 张焕国, 冯登国等. 信息安全综述[J]. 中国科学 E 辑: 信息科学, 2007, 37(2): 129-150.
SHEN C X, ZHANG H G, FENG D G, *et al.* Survey of information security[J]. Science in China Series: E, 2007, 37(2): 129-150.
- [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] KER A D, BAS P, BÖHME R, *et al.* Moving steganography and steganalysis from the laboratory into the real world[A]. Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security[C]. Montpellier, France, 2013. 45-58.
- [4] PEVNÝ T, FRIDRICH J. Towards multi-class blind steganalyzer for JPEG images[A]. Proceedings of the 4th International Workshop on Digital Watermarking[C]. Siena, Italy, 2005.39-53.
- [5] PEVNÝ T, FRIDRICH J. Multiclass detector of current steganographic methods for JPEG format[J]. IEEE Transactions on Information Forensics and Security, 2008, 3(4):635-650.
- [6] WANG P, LIU F L, WANG G D, *et al.* Multi-class steganalysis for JPEG stego algorithms[A]. Proceedings of the 15th IEEE International Conference on Image Processing[C]. San Diego, CA, 2008. 2076-2079.
- [7] DONG J, WANG W, TAN T N. Multi-class blind steganalysis based on image run-length analysis[A]. Proceedings of the 8th International Workshop on Digital Watermarking[C]. Guildford, UK, 2009. 199-210.
- [8] LUO X Y, WANG D S, WANG P, *et al.* A review on blind detection for image steganography[J]. Signal Processing, 2008, 88(9): 2138-2157.
- [9] KODOVSKÝ J, FRIDRICH J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3):868-882.
- [10] 罗向阳, 刘粉林, 杨春芳等. 基于最优小波包分解的图像隐写通用检测[J]. 中国科学: 信息科学, 2010, 40(2): 327-339.
LUO X Y, LIU F L, YANG C F, *et al.* Universal steganalysis based on best wavelet packet decomposition [J]. Science in China Series F: Information Sciences, 2010, 40(2): 327-339.
- [11] FRIDRICH J, GOLJAN M, SOUKAL D. Perturbed quantization steganography with wet paper codes[A]. Proceedings of the 2004 ACM Workshop on Multimedia and Security[C]. Magdeburg, Germany, 2004. 4-15.
- [12] GUL G, KURUGOLLU F. SVD-based universal spatial domain image steganalysis[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 349-353.
- [13] 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.
- [14] FRIDRICH J, PEVNÝ T, KODOVSKÝ J. Statistically undetectable JPEG steganography: dead ends challenges, and opportunities[A]. Proceedings of the 9th ACM Workshop on Multimedia and Security[C]. 2007.3-14.
- [15] ZHANG Y, LIU F L, LUO X Y, *et al.* Steganalysis of PQ based on the histogram difference between coefficient adjacent positions[A]. Proceedings of the 4th International Conference on Multimedia Information Networking and Security[C]. Nanjing, China, 2012. 955-958.
- [16] WESTFELD A. F5—A steganographic algorithm—high capacity despite better steganalysis[A]. Proceedings of the 4th International Workshop on Information Hiding[C]. Pittsburgh, PA, USA, 2001. 289-302.
- [17] SALLEE P. Model-based steganography[A]. Proceedings of the 2nd International Workshop on Digital Watermarking[C]. Seoul, Korea, 2003.154-167.
- [18] FILLER T, FRIDRICH J. Design of adaptive steganographic schemes for digital images[A]. Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia XIII[C]. 2011.1-14.
- [19] LUO X Y, LIU F L, LIAN S G, *et al.* On the typical statistic features for image blind steganalysis[J]. IEEE Journal on Selected Areas in Communications, 2011, 29(7): 1404-1422.
- [20] 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 F F, *et al.* Steganalysis of LSB matching based on regional correlation [J]. Journal on Communications, 2010, 31(3): 1-11.
- [21] MALIK H, SUBBALAKSHMI K P, CHANDRAMOULI R. Non-parametric steganalysis of QIM steganography using approximate entropy[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2): 418-431.
- [22] SALLEE P. Model-based methods for steganography and steganalysis[J]. International Journal of Image Graphics, 2005, 5(1): 167-190.
- [23] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3):1-27.

作者简介:



卢记仑 (1985-), 男, 河北大名, 解放军信息工程大学讲师, 主要研究方向为信息隐藏与检测。



刘粉林 (1964-), 男, 江苏溧阳人, 解放军信息工程大学教授、博士生导师, 主要研究方向为网络信息安全、信息隐藏与检测。

罗向阳 (1978-), 男, 湖北荆门人, 解放军信息工程大学副教授、博士生导师, 主要研究方向为网络信息安全、信息隐藏与检测。

张轶 (1987-), 男, 河南漯河人, 解放军信息工程大学博士生, 主要研究方向为信息隐藏与检测。