• 安全技术 • 文章编号：1000—3428(2013)12—0148—04 文献标识码：A 中图分类号：TP391

# 基于 Boosting 算法融合的图像隐写分析方法

万宝吉，张 涛，侯晓丹，朱振浩

(解放军信息工程大学信息系统工程学院，郑州 450002)

**摘 要：**现有盲检测技术在实际检测中，由于嵌入算法未知导致检测困难。为此，提出一种基于 Boosting 算法融合的图像隐写分析方法。通过训练分类器建立不同隐写算法下的分类器模型，利用 Boosting 算法计算各分类器的分类性能，对各分类器的概率输出进行融合，得到最终检测结果。基于典型空间域隐写算法和 JPEG 隐写算法的实验结果表明，该方法实现了对多种隐写算法的有效检测，应用 Boosting 算法融合后整体检测性能提升了约 2%。

**关键词：**信息隐藏；数字隐写；隐写分析；Boosting 算法；分类器融合；支持向量机

# Image Steganalysis Method Based on Boosting Algorithm Fusion

**WAN Bao-ji, ZHANG Tao, HOU Xiao-dan, ZHU Zhen-hao**

(Institute of Information System Engineering, PLA Information Engineering University, Zhengzhou 450002, China)

**【Abstract】**The existing detection algorithms are difficult to obtain high detection accuracy when applied to the condition, in which the embedding algorithm of the stego-images is unknown. Therefore, this paper proposes a steganography-unknown image steganalysis method based on Boosting fusion. It obtains various classifying results by establishing steganography algorithm classifier models in the training phase, and acquires the performance of these classifies according to the Boosting algorithm. The final detection result is obtained by combinational rule based on probability output. The detection work is presented to attack the current different spatial domain and JPEG steganographic algorithms. Extensive experimental results show that this proposed method is effective for multi-steganographic algorithms, and Boosting takes advantage of the individual strengths from each detection system and whole detection performance is probably increased by 2%.

**【Key words】**information hiding; steganography; steganalysis; Boosting algorithm; classifier fusion; Support Vector Machine(SVM)

**DOI:** 10.3969/j.issn.1000-3428.2013.12.032

## 1 概述

自 20 世纪 90 年代初以来，信息隐藏技术逐渐成为信息安全领域的研究热点[1]。数字隐写与隐写分析是信息隐藏的主要分支。数字隐写的目的是将秘密消息隐藏在载体中进行传递而不引起第三方怀疑，以实现隐蔽通信，而隐写分析是对秘密消息进行检测、提取、恢复和破坏。

数字隐写的载体包括图像、视频、文本、音频等。在基于图像的数字隐写技术中，根据秘密信息嵌入数据域的不同，可分为空间域隐写和变换域隐写。基于空间域隐写的经典算法有最不重要比特位(Least Significant Bit, LSB)替换[1]和 LSB 匹配(LSB Matching, LSBM)方法[2]、像素值差分(Pixel-Value Differencing, PVD)方法[3]、位平面复杂度分割(Bit-Plane Complexity Segmentation, BPCS)隐写方法[4]和图像边缘的自适应(Adaptive data hiding in Edge areas-Least Significant Bit, AE-LSB)隐写方法[5]等。基于变换域的隐写将秘密消息隐藏在变换域系数中，如离散余弦变换(Discrete Cosine Transform, DCT)系数和离散小波变换(Discrete Wavelet Transform, DWT)系数等。由于 JPEG 图像是最常见的图像类型之一，因此许多典型的变换域隐写算法都将秘密消息隐藏在 JPEG 图像的 DCT 系数中，如 JSteg[6]、F5[7]、MB1(Model-Based)[8]、MB2[9]以及基于湿纸码去除收缩效应的 nsF5 算法[10]等。

目前，隐写分析主要集中于对秘密消息存在性检测的研究，已有大量的检测算法被提出。随着特征提取的有效性和分类器分类能力的提高，通用盲检测方法的检测正确率逐渐提高，且能够检测多种隐写算法，因而其重要性越来越突出。通用盲检测方法主要采用基于机器学习的方法，

**基金项目：**国家自然科学基金资助项目(60903221, 61272490)

**作者简介：**万宝吉(1986－)，男，硕士研究生，主研方向：信息隐藏技术，信息融合技术；张 涛，副教授、博士；侯晓丹、朱振浩，硕士研究生

**收稿日期：**2012-11-19 **修回日期：**2012-12-25 **E-mail：**dirker2012@163.com

其关键是寻找能有效区分载体和载密图像的特征，因此各方法的差别也在于所提取的分类特征不同。典型的特征包括: 文献[11]提出的基于局部线性变换系数的概率密度函数矩 (Local Linear Transform-Probability Density Function, LLT-PDF)；文献[12]提出的基于像素差分 Markov 状态转移概率矩阵(Subtractive Pixel Adjacency Matrix, SPAM)的特征等。但传统的盲检测方法在训练分类器时通常需要指定嵌入算法，而实际中嵌入算法是未知的，因而该类分类器在实际应用中的测试结果与实验结果偏差较大。

近年来，不少学者应用信息融合技术来解决隐写分析中的一些问题。文献[13]对不同隐写分析的分类结果进行决策级最大值融合和均值融合；文献[14]利用贝叶斯的独立二值分类模型，提出一种通过调整参数改善分类效果的隐写分析算法。上述研究结果表明，在隐写分析中应用信息融合技术能够提高检测性能。为实现对未知隐写算法的有效检测，本文借鉴近年来将信息融合技术应用于隐写分析的思路，利用不同的分类器对分类模式有互补信息的优势，研究基于 Boosting 算法的多分类器融合技术来实现对未知隐写算法下隐藏信息的有效检测。本文研究的未知隐写算法是指在同一数据域嵌入的不同嵌入方式的隐写算法。

## 2　Boosting 多分类器融合的图像隐写分析

### 2.1　基本框图

本文研究基于 Boosting 方法的多分类器融合技术以实现对未知隐写算法载密图像的有效检测。整个过程分为 2 个阶段：训练阶段(图 1)和测试阶段(图 2)。

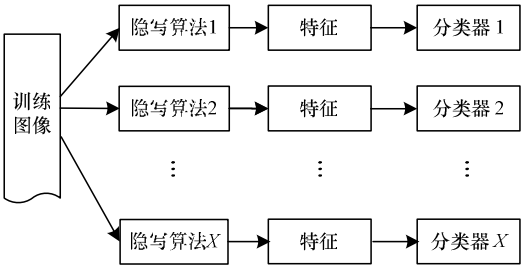


**图 1　未知隐写算法的隐写分析方法训练框图**

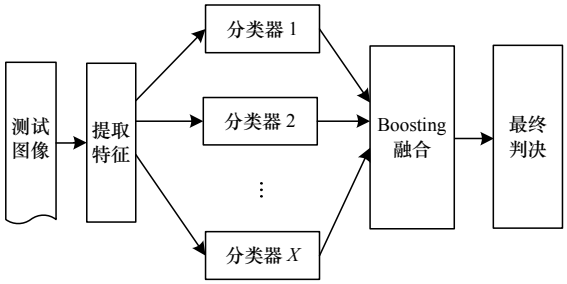


**图 2　未知隐写算法的隐写分析方法测试框图**

训练阶段：首先用 $X$ 种隐写算法(秘密信息嵌入的数据域必须是一样的)对训练载体图像以某种嵌入率 $r$ 嵌入秘密消息，得到对应隐写算法的训练载密图像。然后对不同隐写算法的训练集提取盲检测特征(这里以提取一种盲检测特征为例)，并将得到的特征送至相应的分类器中进行训练，得到对应隐写算法下的分类器训练模型。这里提取的特征是这几种隐写算法的通用盲检测特征。最后应用 Boosting 方法根据各个分类器的分类性能得到一个权值。

测试阶段：首先将待测图像依照训练阶段提取特征的方法提取特征，然后将特征分别输入到已训练好的分类器模型中进行局部决策，得到其属于载体和载密的概率。用训练阶段 Boosting 方法得到权值融合各个分类器的局部决策值，得到最终的结果。

### 2.2　基于 Boosting 融合的图像隐写分析

Boosting 是将弱学习算法提升为强学习算法的一类算法，AdaBoost 算法是目前最流行的一种 Boosting 算法。AdaBoost 是“adaptive boosting”(自适应增强)的缩写。这个方法允许设计者不断地加入新的“弱分类器”，直到达到某个预定的足够小的误差率。在 AdaBoost 方法中，每一个训练样本点都被赋予一个权重，表明它被某个分量分类器选入训练集的概率。如果某个样本点已经被准确地分类，那么在构造下一个训练集中，它被选中的概率就被降低；相反地，如果某个样本点没有被正确分类，那么它的权重就得到提高。通过这样的方式，AdaBoost 方法能够“聚集于”那些信息量更大的样本上。在具体实现上，最初令每个样本的权重都相等。对于第 $m$ 次迭代操作，根据权重大小来选取样本点，进而训练分类器 $y_m(x)$，其中，$x$ 表示原始样本集中的样本；$y$ 表示分类器。根据这个分类器，来提高被它错分的那些样本点的权重，并降低可以被正确分类的样本权重。然后，权重更新过的样本集被用来训练下一个分类器 $y_{m+1}(x)$。整个训练过程如此进行下去。

本文用 $\boldsymbol{x}$ 表示原始样本集中的所有样本的集合，$N$ 表示样本点的数量，$t_n$=1,2,…,$T$ 表示要分成的类别，$M$ 表示迭代次数和分类器的个数，用 $w_n^{(m)}$ 表示第 $m$ 次迭代时全体样本的权重分布。具体的 Adaboost 算法设计流程[15]如下：

(1)初始化数据权重系数 $\{w_n\}$，设置 $w_n^{(1)}=1/N$，其中，$n$=1, 2,…,$N$。

(2)对 $m$=1,2,…,$M$:

1)通过数据权重的错误函数 $J_m$：

$$J_m=\sum_{n=1}^{N} w_n^{(m)} I\left(y_m\left(\boldsymbol{x}_n\right)\neq t_n\right) \tag{1}$$

其中，$I\left(y_m\left(\boldsymbol{x}_n\right)\neq t_n\right)=\begin{cases}1 & y_m\left(\boldsymbol{x}_n\right)\neq t_n\left(\text{分类错误}\right)\\ 0 & y_m\left(\boldsymbol{x}_n\right)=t_n\left(\text{分类正确}\right)\end{cases}$，使 $J_m$ 最小使训练数据和分类器 $y_m(\boldsymbol{x})$相匹配。

2)计算数据集错误权重系数 $\varepsilon_m$ 和分类器权重系数 $\alpha_m$ 的值：

$$\varepsilon_m=\frac{\sum_{n=1}^{N} w_n^{(m)} I\left(y_m\left(\boldsymbol{x}_n\right)\neq t_n\right)}{\sum_{n=1}^{N} w_n^{(m)}} \tag{2}$$

$$\alpha_m = \ln\left\{\frac{1-\varepsilon_m}{\varepsilon_m}\right\} \tag{3}$$

3)更新数据的权重系数:

$$w_n^{(m+1)} = w_n^{(m)} e^{\alpha_m I\left(y_m(\boldsymbol{x}_n)\neq t_n\right)} \tag{4}$$

(3)最后总体判决:

$$g(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m y_m(\boldsymbol{x}) \tag{5}$$

$$Y_M(\boldsymbol{x}) = \mathrm{sign}\left[g(\boldsymbol{x})\right] \tag{6}$$

其中，第 1 个分类器用相等的权重系数 $w_n^{(1)}$ 训练，跟通常训练一个单独的分类器相似。在步骤 3)中，当数据被错分时权重系数 $w_n^{(m)}$ 增加，当数据被正确分类时 $w_n^{(m)}$ 减小。值 $\varepsilon_m$ 表示每个分量分类器对数据集错误的权重。因此，式(3)说明，权重系数 $\alpha_m$ 使分类效果优秀的分类器得到更大的权重。在 AdaBoost 方法中，每个分类器都根据它们的性能得到一个权重。因为本文中已训练好的分类器错误率 $\varepsilon_m < \frac{1}{2}$，所以得到的分类器权重系数 $\alpha_m > 0$。

本文对已训练好的 $M$ 个分类器，通过 AdaBoost 方法的步骤(1)和步骤(2)，用训练数据得到各分类器的权重系数 $\alpha_m$，然后用式(5)对各分类器的概率输出进行加权处理，得到如下结果:

$$\begin{cases} Y_{\text{cover}}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m p_{\text{cover}}^m(\boldsymbol{x}) \\ Y_{\text{stego}}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m p_{\text{stego}}^m(\boldsymbol{x}) \end{cases} \tag{7}$$

其中，$p_{\text{cover}}^m(\boldsymbol{x})$ 和 $p_{\text{stego}}^m(\boldsymbol{x})$ 分别为第 $m$ 个分类器输出判为载体和载密的概率。对式(7)的 $Y_{\text{cover}}(\boldsymbol{x})$ 和 $Y_{\text{stego}}(\boldsymbol{x})$ 比较大小可得到最终结果，则最终判决结果为:

$$Test = \begin{cases} \text{cover} & Y_{\text{cover}}(\boldsymbol{x}) \geqslant Y_{\text{stego}}(\boldsymbol{x}) \\ \text{stego} & Y_{\text{stego}}(\boldsymbol{x}) > Y_{\text{cover}}(\boldsymbol{x}) \end{cases} \tag{8}$$

在这 2 个结果中哪个值大，就将待测图像判为哪一类，即得到的最终判决结果应具有最大的可信度。

### 2.3 支持向量机

本文训练和测试的分类器采用支持向量机(Support Vector Machine, SVM)，SVM 是根据统计学习理论的结构风险最小化原则而提出的一种有监督的机器学习方法。其基本思想是通过非线性变换将输入空间变换到一个更高维空间中，在这个新的空间中求取最优分类超平面，使得最优分类超平面与不同类样本集之间的距离最大，从而达到最大泛化能力。

具体实现用 LIBSVM[16]，它是一种 SVM 的程序库。本文选用径向基函数 $K(x,x_i) = \exp(-\|x-x_i\|^2 / 2\sigma^2)$ 作为核函数。惩罚参数 $c$ 和核函数参数 $g$ 采用交叉验证(Cross Validation, CV)的方法获得。通过对 SVM 训练模型参数 $b$(概率估计)的赋值，可得到用于概率估计的支持向量分类器(Support Vector Classifier, SVC)，测试时将测试样本集输入此分类器中并对测试模型参数 $b$ 赋值可得概率输出。

## 3 实验结果及分析

### 3.1 实验设置

实验采用 CAMERA[17]图像库，CAMERA 图像库包含 3 164 幅大小为 512×512 像素的未压缩灰度图像。从中随机选取 1 600 幅进行空间域隐写分析的实验，并对这 1 600 幅灰度图像以质量因子 75 进行一次 JPEG 压缩，得到 JPEG 图像进行变换域隐写分析实验。随机选择 600 幅用于训练，剩余 1 000 幅用于测试。在空间域隐写的嵌入率 0.25 bpp(bit per pixel)和 JPEG 隐写嵌入率为 0.25 bpnc(bit per nozero coefficient)分别生成相应载密图像库。取 5 种常用的经典隐写算法进行实验，其中，空间域采用的隐写算法是 LSB 替换[1]、LSB 匹配[2]、PVD[3]、BPCS[4]和 AELSB[5]，JPEG 隐写算法为 JSteg[6]、F5[7]、MB1[8]、MB2[9]和 nsF5[10]。对空间域隐写算法选取的盲检测特征为 LLTPDF[11]，对 JPEG 隐写算法提取的特征为 SPAM[12]。

对训练的 600 幅图像，分别采用 5 种不同的隐写算法构造训练载密图像集，对每种隐写算法的载体载密图像集提取分类特征，用 SVM 训练得到分类模型。剩余 1 000 幅测试图像以同样的嵌入方法，得到 5 种不同隐写算法的测试载密图像集。将 1 000 幅的测试图像均分成 5 份，每份分别以不同的隐写算法嵌入得到混合隐写算法的测试图像。

### 3.2 评估指标

本文采用下列指标来进行评估:

(1)真阳率(True Positive Rate, TP)：也称正确检测率，是将载密数据正确识别为载密数据的比率。

(2)真阴率(True Negative Rate, TN)：也称正确否定率，是将载体数据正确识别为载体数据的比率。

设待检测的载体数据和载密数据的样本数分别为 $C$ 和 $S$，其中，被正确识别的载体数据和载密数据的样本数分别为 $N$ 和 $P$，则检测正确率(Accuracy, ACC)定义为:

$$ACC = \frac{P+N}{S+C} \tag{9}$$

其中，这 3 个指标越接近 1，说明算法的检测性能越好。

### 3.3 结果分析

#### 3.3.1 固定隐写算法训练的分类器性能

本节采用传统的固定隐写算法训练的分类器，来测试不同隐写算法下的待测图像，并得到相应的检测正确率，通过观察和分析，验证了本文工作的意义和必要性。应用实验设置中训练好的各固定隐写算法分类模型和 5 种不同隐写算法的测试图像集进行实验。

表 1 和表 2 分别为空间域和 DCT 域固定隐写算法训练的分类器测试不同隐写算法的测试图像得到检测正确率(ACC)，其中，粗体表示不同隐写算法训练的分类器测试某一隐写算法时得到的最优检测正确率。

**表 1 空间域固定隐写算法训练的分类器检测正确率** (%)

| 训练隐写算法 | 测试隐写算法 | | | | |
|---|---|---|---|---|---|
| | LSB | LSBM | PVD | BPCS | AELSB |
| LSB | **94.15** | 94.00 | 92.75 | 95.95 | 50.40 |
| LSBM | 94.00 | **94.25** | 92.10 | 96.00 | 50.65 |
| PVD | 56.45 | 57.10 | **97.30** | 98.71 | 50.35 |
| BPCS | 49.90 | 49.90 | 51.80 | **99.00** | 50.05 |
| AELSB | 46.40 | 46.20 | 58.75 | 93.20 | **88.05** |

**表 2 DCT 域固定隐写算法训练的分类器检测正确率** (%)

| 训练隐写算法 | 测试隐写算法 | | | | |
|---|---|---|---|---|---|
| | JSteg | F5 | Mb1 | Mb2 | nsF5 |
| JSteg | **81.90** | 68.70 | 82.15 | 81.15 | 54.70 |
| F5 | 75.65 | **85.40** | 79.15 | 75.65 | 58.50 |
| Mb1 | 80.90 | 69.00 | **89.20** | 88.20 | 53.80 |
| Mb2 | 80.25 | 62.75 | 88.25 | **89.85** | 52.10 |
| nsF5 | 74.70 | 81.80 | 78.15 | 72.15 | **65.65** |

从表 1、表 2 可以看出，当测试图像和训练图像的隐写算法一致时，可以得到最好的检测正确率，如表中粗体所示。然而，当测试图像的隐写算法与训练图像不同时，尤其是当两者的嵌入原理差异较大时，检测效果很不理想。例如，对于空间域隐写算法，若测试载密图像由 LSBM 隐写算法生成，而训练的载密图像由 AELSB 隐写算法生成，其检测正确率比训练和测试图像均由 LSBM 隐写算法生成时下降 48.05%；对于 DCT 域隐写算法，若测试载密图像是由 F5 隐写算法生成，而训练的载密图像由 Jsteg 隐写算法生成，其检测正确率比训练和测试图像均由 F5 隐写算法生成时下降 16.7%。因此，有必要采用融合策略解决训练图像的隐写算法与测试图像不同引起的检测正确率降低的问题。

3.3.2 与其他融合方法的性能比较

为测试本文方法与其他融合方法对未知隐写算法的检测性能，对混合隐写算法的测试图像进行检测。其中：

S1 为用混合隐写算法训练的分类器。即对训练图像集均分成 5 份，每份分别用不同的隐写算法嵌入，用生成的混合隐写算法的图像集提取特征并送至 SVM 训练得到；

S2 为对各分类器的结果用投票融合的方法；

S3 为对各分类器的结果用均值融合的方法；

Proposed 是本文提出的基于 AdaBoost 方法的多分类器融合。

利用以上 4 种方法测试混合隐写算法的测试图像，得到检测正确率(ACC)、正确否定率(TN)和正确检测率(TP)，结果如表 3 所示，其中，粗体表示得到的最优结果。由表 3 可知，无论是空间域隐写还是 JPEG 隐写，本文方法表现出较好的检测正确率，在变换域中各个检测结果更是全面优于其他方法。简单的投票融合 S2 仅根据每个分类器的判决，取多数分类器的意见为最终决策，而没有考虑每个分类器的分类性能，因而检测性能相对较差。均值融合 S3 根据每个分类器判决概率的平均值，取概率平均值最大的类为最终决策，由于考虑了每个分类器的性能，取得了相对较好的检测性能。实验结果表明，本文提出的基于 AdaBoost 算法融合的检测性能全面优于均值融合和投票融合，整体检测性能提升了约 2%。

**表 3 不同方法对混合隐写算法的检测性能** (%)

| 嵌入域 | 检测方法 | ACC | TN | TP |
|---|---|---|---|---|
| 空间域 | S1 | 86.30 | 89.20 | **83.40** |
| | S2 | 71.05 | 99.20 | 42.90 |
| | S3 | 71.10 | **99.60** | 42.60 |
| | Proposed | **88.45** | 95.80 | 81.10 |
| DCT 域 | S1 | 76.65 | 87.60 | 65.70 |
| | S2 | 78.90 | 90.70 | 67.10 |
| | S3 | 78.90 | 91.10 | 66.70 |
| | Proposed | **79.40** | **91.20** | **67.60** |

## 4 结束语

现有的许多盲检测算法虽然针对多种隐写算法有效，但往往需要对不同的嵌入算法分别训练分类器。对于测试图像也假设已知其嵌入算法，这在实际中难以得到满足。针对这一问题，本文提出利用已训练好的多种隐写算法的分类器进行检测，对多分类器的检测结果用 AdaBoost 方法进行融合的思路。实验结果表明，本文方法在嵌入算法未知的条件下，也能取得较好的检测性能，优于投票融合方法和均值融合方法，并且本文算法一旦训练好模型后，可以直接进行测试。下一步工作中将利用组合分类器，并分析各种盲检测特征，设计对未知隐写算法检测性能更好的盲检测算法。

**参考文献**

[1] Petitcolas P, Anderson J, Kuhn G. Information Hiding——A Survey[J]. Proceedings of the IEEE, 1999, 87(7): 1062-1078.

[2] Sharp T. An Implementation of Key-based Digital Signal Steganography[C]//Proc. of the 4th Information Hiding Workshop. Berlin, Germany: Springer-Verlag, 2001: 13-26.

[3] Wu D, Tsai W. A Steganographic Method for Images by Pixel-value Differencing[J]. Pattern Recognition Letters, 2003, 24(9-10): 1613-1626.

[4] Kawaguchi E, Eason R O. Principle and Applications of BPCS Steganography[J]. Multimedia Systems and Applications, 1999, 3528(1): 464-473.

[5] Cheng Hsing-Yang, Chi Yao-Weng. Adaptive Data Hiding in Edge Areas of Images with Spatial LSB Domain Systems[J]. IEEE Transactions on Information Forensics and Security, 2008, 3(3): 488-497.

(下转第 156 页)