

文章编号:1007-2780(2018)06-0490-07

基于特征排名的图像隐写分析算法

张兴春¹, 孙寿健^{2,3*}

- (1. 武警黑龙江总队司令部, 黑龙江 哈尔滨 150000;
2. 武警黑龙江总队佳木斯支队司令部, 黑龙江 佳木斯 154000;
3. 网络与信息安全武警部队重点实验室, 陕西 西安 710086)

摘要:为提高用于隐写分析的集成分类器的检测精度, 提出一种基于特征排名的隐写分析算法。首先计算每维检测特征的互信息得分并根据得分高低将特征进行排名, 然后设置分界点将特征分为重要特征区域与普通特征区域, 依据设定的抽样比例从两个区域随机抽取特征组成不同的特征子空间并训练集成分类器。最后使用集成分类器进行分类。实验结果表明, 针对使用 nsF5 及 S-UNIWARD 算法进行隐写的频域及空域图像, 本算法较传统分类器在检测错误率方面分别平均下降约 0.006 5 和 0.006 2, 具有较好的检测效果。针对频域与空域中两种不同的隐写算法, 与传统的集成分类器相比, 该算法具有更高的检测精度。

关键词: 隐写分析; 集成分类器; 特征排名; 互信息得分

中图分类号: TP391 **文献标识码:** A **doi:** 10.3788/YJYXS20183306.0490

Image steganalysis algorithm based on feature ranking

ZHANG Xing-chun¹, SUN Shou-jian^{2,3*}

- (1. *Command of Heilongjiang Crops of PAP, Harbin 150000, China;*
2. *Command of Jiamusi Detachment, Heilongjiang Crops of PAP, Jiamusi 154000, China;*
3. *Network and Information Security Key Laboratory of PAP, Engineering University of the PAP, Xi'an 710086, China*)

Abstract: In order to enhance the detection rate of ensemble classifier, an algorithm based on feature ranking is proposed. First, the original feature is sorted according mutual information score. Then the sorted feature is divided into important feature part and common feature part according to the divide point. The feature subset is formed by selecting features randomly in each part according to the given sampling rate. Experimental results show that, our method detects two format images after embedding nsF5 and S-UNIWARD, the false detection is lower than classical ensemble classifier 0.0065 to .jpeg images, and the false detection is lower than 0.0062 to .bmp images. Compared with typical ensemble classifier, the proposed method is more effective than the different stego algorithms in

收稿日期: 2018-01-12; 修订日期: 2018-03-23.

基金项目: 国家自然科学基金(No.61403417)

Supported by National Natural Science Foundation of China(No.61403417)

* 通信联系人, E-mail: zynwj08@163.com

frequency domain and spatial domain.

Key words: steganalysis; ensemble classifier; feature ranking; mutual information scores

1 引言

目前,图像的隐写分析主要是通过监督机器学习来实现的,具体方法是使用训练样本训练一个分类器,然后使用分类器对未知图像做出判断^[1],特征提取与分类器的选择是其中两个重要方面^[2]。随着 HUGO (highly undetectable steganography)^[3]、S-UNIWARD (spatial-universal wavelet relative distortion)^[4]等基于伴随式编码 (Syndrome Trellis Coding, STC)^[5]与最小失真代价函数^[6]的新型隐写算法以及可抗专有隐写分析算法^[7]的出现,传统的低维特征已经对这些算法实现有效的检测。为应对新型隐写算法带来的挑战,隐写分析者提取的特征维数也越来越高,例如频域中利用笛卡尔校准构造的 CC-PEV (Cartesian calibration based feature proposed by Pevny)^[8]特征与空域中基于共生矩阵的空域富模型 (Spatial Rich Models, SRM)^[9]特征。高维特征虽能够更加精度地对图像失真进行建模,但由于维数过高,也造成了“维数灾难”问题^[10]。为解决维数灾难问题, Kodovsky 在文献[11]中提出了基于随机森林算法的集成分类器。该分类器使用 Bagging 算法与随机特征子空间的方法生成若干样本子集训练若干费歇尔线性分类器 (Fisher Linear Discriminate, FLD) 作为基分类器,并使用简单投票方法将分类器结果进行集成。由于采用集成学习的思想,并且将高维特征分成多个低维子集,所以与传统分类器相比,集成分类器具有更好的检测效果与更少的训练时间。

集成分类器中,训练基分类器所使用的特征子空间是从原始特征中随机选取的特征,随机选取的子特征空间包含许多分类信息较少或不具备分类信息的特征,因此影响了基分类器的分类精度^[12]。为提高集成分类器在高维特征中的检测精度,本文提出一种基于特征排名的图像隐写分析算法。该方法利用互信息得分作为特征重要性的评价标准,并根据得分的高低将原始特征空间分为重要特征区域与普通特征区域。最后,根据设定的抽样比例从两个空间中抽取特征生成特征

子空间。该方法在不降低基分类器多样性的情况下使生成的特征子空间包含更多的有效特征,所以可以提高基分类器性能,进而提高整个集成分类器的泛化能力。实验结果表明,本文算法在检测错误率上要低于传统的集成分类器。

2 集成分类器

Kodovsky 在文献[11]中采用了集成学习的方法来克服高维特征带来的“维数灾难”问题。该算法使用 FLD 分类器来作为基分类器,并采用随机森林的方式来生成训练 FLD 分类器所用的训练样本与特征空间。假设原始特征空间的特征维数为 d , C^{tm} 与 S^{tm} 代表训练基分类器所用的载体样本特征集与隐写样本特征集,两个样本集中的载体特征与隐写特征是一一对应的, B_i 代表每一个独立的基分类器。集成分类器使用 Bagging 算法中的重采样方法抽取训练样本,同时用随机抽样的方法生成不同的特征子空间。集成分类器伪代码如下:

算法 1: 集成分类器

输入: 载体样本特征集 C^{tm} , 隐写样本特征集 S^{tm}

For $l=1 : L$

{

(1) 使用随机抽样的方法从原始特征空间中抽取部分特征组成特征子空间,其中子空间特征维度为 d_{sub} ($d_{\text{sub}} \ll d$)

(2) 使用 Bagging 方法产生训练样本 $C = \text{Bootstrap}(C^{\text{tm}})$, $S = \text{Bootstrap}(S^{\text{tm}})$

(3) 取 C 、 S 和 d_{sub} 对应的样本集合与特征集合训练基分类器 B_l

}

输出: L 个基分类器 $\{B_1, B_2, B_3, \dots, B_L\}$

在获取 L 个基分类器之后,使用简单投票方法对分类结果进行集成,即首先由每个基分类器对未知样本做出判断,然后取大多数基分类器做出的判决结果为最终分类结果。集成分类器的流程图如图 1 所示。

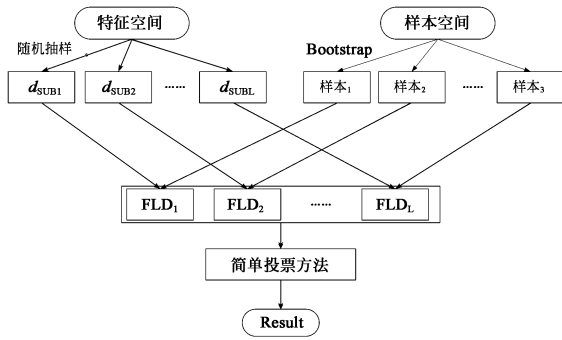


图 1 集成分类器流程图

Fig.1 Flow chart of ensemble classifiers

3 基于特征排名的图像隐写分析算法

集成学习理论认为提高基分类器分类精度与基分类器差异度可以提高集成分类器的整体泛化能力^[13]。因此,本算法的设计主要集中在保证多样性的情况下增大基分类器的分类精度。目前所用的隐写分析特征虽然设计比较复杂,特征维数较高,但每一维特征的重要程度并不相同。高维特征中有包含分类信息的重要特征,有普通特征,也有冗余特征。在基分类器的训练过程中,如果提高有效特征所占的比例,就可以提高基分类器的精度,进而提高集成分类器整体的泛化能力。

为提高有效特征在训练基分类器时所占的比例,本文算法首先计算了各维特征的互信息得分,得分越高,代表特征的区分度越高,特征的分类性能也就越好^[14]。然后将各维特征按照互信息得分的高低由高到低进行排列,并根据设定的分类点将原始特征空间分为重要特征区域与普通特征区域。为每个区域设定不同的抽样比例然后随机

抽样生成特征子空间进而训练基分类器。另外,通过提高重要特征区域的抽样比例可以提高特征子空间中有效特征的数量进而提高基分类器的分类精度。最后,使用简单投票方法将各基分类器的分类结果进行集成。

与传统分类器相比,本文算法具有以下几个优点。(1)由于采用随机抽样的方式生成特征子空间,所以并不会降低基分类器之间的差异性。(2)相比于传统的集成分类器,本文算法的特征子空间中包含有更多的有效特征,所以可以提高基分类器的分类精度。(3)本文算法可以通过设定不同的抽样比例生成不同的特征子空间,使基分类器的训练更加灵活。

3.1 互信息得分

互信息包括单维特征互信息与特征间互信息,可以有效衡量两个变量间相互依赖强弱程度。互信息得分越高,说明特征包含的分类信息越多,特征分类效果越好。

单维特征互信息用来表示图像隐写前后的特征相关性,单维互信息越大,特征相关性越好,特征的区分度越低。假设 X 与 Y 代表隐写前后对应的特征, $P(x_i)$ 与 $P(y_i)$ 代表一维概率密度函数, $P(x_i y_i)$ 代表二维概率密度函数。单维特征互信息定义为:

$$I_{\text{inter}}(X, Y) = \sum_{x_i} \sum_{y_i} P(x_i y_i) \log \frac{P(x_i y_i)}{P(x_i) P(y_i)}, \quad (1)$$

除单维特征互信息外,还存在着特征间互信息,其主要反映特征与类别之间的相关性。特征间互信息越大,特征与类别之间联系越紧密,特征效果越好。特征 X 与特征空间 D 之间的互信息可以表示为:

$$I_{\text{intra}}(X, D) = \frac{1}{2} \sum_{D \in \{C, S\}} \frac{1}{d-1} \sum_{\substack{Y \in D \\ Y \notin X}} P(x_i y_i) \log \frac{P(x_i y_i)}{P(x_i) P(y_i)}, \quad (2)$$

其中: D 为特征空间, C 为载体特征集, S 为隐写特征集, d 为特征集中的特征维数。单维特征互信息中包含了各个单一维度特征的可区分性,其计算结果反映载体与隐密图像特征的类间差异,特征间互信息则反映了类内的差异。

最后,将单维特征互信息与特征间互信息进

行结合,特征 X 的互信息得分为

$$\text{score}(X) = \frac{(1 - I_{\text{inter}}(X, Y)) + I_{\text{intra}}(X, Y)}{2}. \quad (3)$$

3.2 算法流程

本文算法使用了新的方法来生成训练基分类器的特征子空间。具体过程如下:

输入:载体样本特征集 C^{im} , 隐写样本特征集 S^{im} , 原始特征空间 $F = \{F_1, \dots, F_d\}$, 特征子空间维数 d_{sub} , 基分类器数量 L , 两区域分界点 dp , 两区域的抽样比例为 p

For $i = 1 : d$

{

(1) 计算第 i 维特征的互信息得分

}

END

(2) 将特征按照得分高低由高到低进行排列

(3) 根据分界点的 dp 将原始特征空间分为两个区域, 重要特征区域 $[F'_1, F'_{dp}]$ 与有效特征区域 $[F'_{dp+1}, F'_d]$

For $l = 1 : L$

{

(4) 使用 Bagging 方法产生训练样本 $C = \text{Bootstrap}(C^{\text{im}})$, $S = \text{Bootstrap}(S^{\text{im}})$

(5) 根据抽样比例从两个区域中随机抽取维数为 d_{sub} 的特征组成特征子空间

(6) 训练基分类器 B_l

}

END

输出: L 个基分类器 $\{B_1, B_2, B_3, \dots, B_L\}$

本文算法的流程图如图 2 所示。

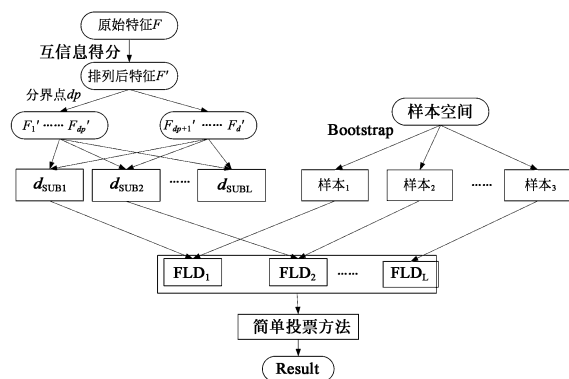


图 2 本文算法流程图

Fig.2 Flow chart of proposed method

4 实验结果及分析

4.1 实验环境

选择从 BossBase-1.01 作为实验图像库, 选取其中的 10 000 幅图像用于进行隐写与隐写分析实验。其中, 50 00 幅图像做为训练样本, 5 000 幅图像做为测试样本。为全面验证算法的性能, 本文从频域与空域两个方面考虑。在频域中, 首先将图像转化为质量因子 75 的 JPEG 图像, 然后采用嵌入率为 0.4bpac (bit per no-zero AC DCT coefficient) 的 nsF5 算法进行隐写, 提取 548 维的 CC-PEV 特征做为隐写分析特征。在空域上, 采用嵌入率为 0.4 bpp (bit per pixel) 的 S-UNIWARD 算法进行隐写, 提取 686 维的 SPAM 特征作为隐写分析特征。实验环境为 Win7 操作系统, 仿真软件为 Matlab2012a。

实验将本文算法与传统集成分类器进行比较。由于特征子空间维度 d_{sub} 对分类器性能有重要影响, 实验中设置的 d_{sub} 为 100、150、200、250、300。实验中训练的基分类器数量为 100 个。在检测 nsF5 算法时, 算法设置的分界点 $dp = 250$, 在检测 S-UNIWARD 算法时, 设置的分界点 $dp = 343$ 。针对重要特征与普通特征, 设置的抽样率分别为 $P = \{0.5, 0.5\}$ 、 $P = \{0.6, 0.4\}$ 、 $P = \{0.7, 0.3\}$ 、 $P = \{0.8, 0.2\}$ 。为便于区分, 本文算法使用符号 IF-RS_ (P) 代表。

4.2 频域图像隐写分析

实验将本文算法与传统的集成分类器进行比较。使用检测错误率作为衡量分类器检测效果的定量指标。检测错误率计算公式为:

$$P_E = (P_{FA} + P_{MD}) / 2, \quad (5)$$

其中: P_{FA} 为虚警率, 表示载体图像被误判的概率, P_{MD} 为漏检率, 表示隐写图像被误判的概率。

表 1 检测错误率对比表

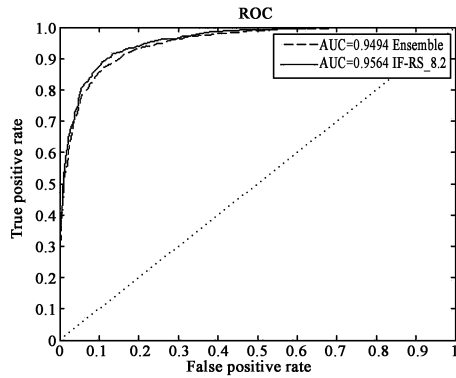
Tab.1 Comparison of P_E of different methods in frequency domain

d_{sub}	IF-RS_5.5	IF-RS_6.4	IF-RS_7.3	IF-RS_8.2	Ensemble
100	0.117 2	0.116 9	0.112 1	0.111 3	0.116 7
150	0.105 1	0.102 6	0.101 1	0.096 3	0.104 1
200	0.099 5	0.097 6	0.094 4	0.096 6	0.101 8
250	0.100 6	0.097 7	0.094 2	0.095 6	0.099 5
300	0.100 9	0.098 0	0.094 5	0.098 4	0.101 1

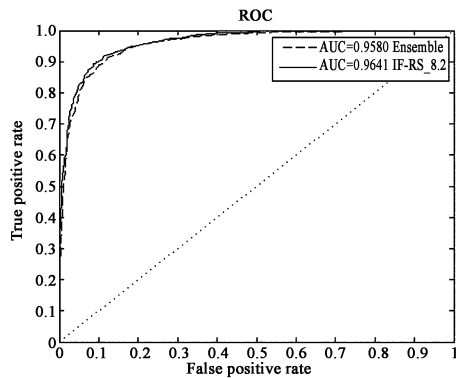
从表 1 中可以看出,本文算法的检测错误率总体要低于传统的集成分类器,说明本文算法可以提高集成分类器的检测精度。另外,从表 1 中可以看出,在不同的子空间维数 d_{sub} 下,取得最好效果时的抽样率也不同。当 d_{sub} 较小时抽样率在 $P=\{0.8,0.2\}$ 较好,当 d_{sub} 增大时,抽样率在 $P=\{0.7,0.3\}$ 较好。说明,当 d_{sub} 较小时,提高子特征空间中有效特征的比例,可以较好地提升集成分类器的泛化能力。当 d_{sub} 较大时,在保证一定比例的有效特征的情况下,提高普通特征比例,可以使基分类器的多样性增加,对提升集成分类器性能更加明显。

为更加直观地评价两种分类器。实验比较了本文算法与传统集成分类器在检测隐写算法的 ROC(Receiver Operating Characteristic) 曲线并计算了 AUC(Area Under The Curve) 值,如图 3 所示。ROC 曲线越靠近左上角, AUC 值越大,代表分类器性能越好。

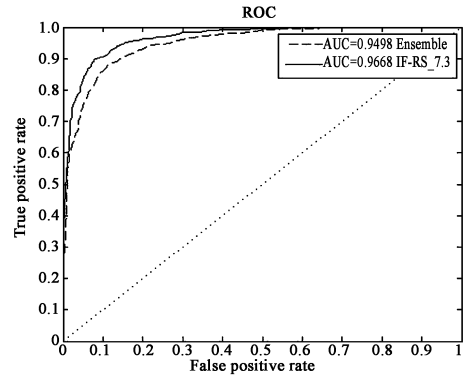
从 ROC 曲线图上看,本文算法的 ROC 曲线



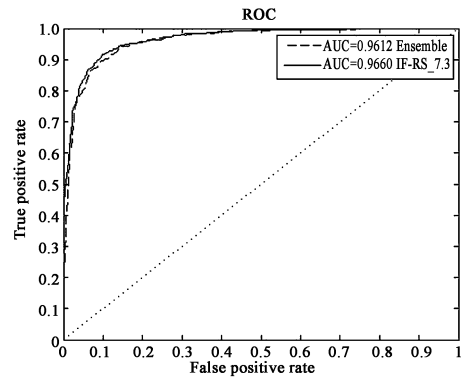
(a) $d_{\text{sub}}=100$



(b) $d_{\text{sub}}=150$



(c) $d_{\text{sub}}=200$



(d) $d_{\text{sub}}=250$

图 3 两种算法 ROC 曲线图比较

Fig. 3 Comparison of ROC curve of different methods in frequency domain

均要优于集成分类器,并且曲线的 AUC 值也要高于集成分类器,说明本文算法在保证相同虚警率的前提下,检测性能上要优于集成分类器。

4.3 空域图像隐写分析

在检测 S-UNIWARD 算法时,提取 686 维的 SPAM 特征作为隐写分析特征。设置的特征分界点 dp 为 340。

从 ROC 曲线图与 AUC 值上来看,本文算法要优于传统的集成分类器,如图 4 所示。

通过检测错误率的对比与 ROC 曲线图的分析可以得到,无论在频域还是空域,本文算法通过特征排名与按比例抽样的方法,可以在保证基分类器差异度的同时,提高有效特征在特征子空间中所占的比例,从而提高基分类器的分类性能,结果如表 2 所示。较传统的集成分类器,本文算法具有更高的检测精度。

表 2 检测错误率对比表

Tab.2 Comparison of P_E of different method in spatial domain

d_{sub}	IF-RS_5.5	IF-RS_6.4	IF-RS_7.3	IF-RS_8.2	Ensemble
100	0.327 7	0.328 4	0.326 8	0.321 2	0.329 0
150	0.318 4	0.318 4	0.317 2	0.317 4	0.319 3
200	0.321 4	0.315 8	0.313 7	0.313 0	0.321 0
250	0.328 1	0.325 6	0.323 9	0.327 9	0.330 3
300	0.333 5	0.335 7	0.330 4	0.332 8	0.336 9

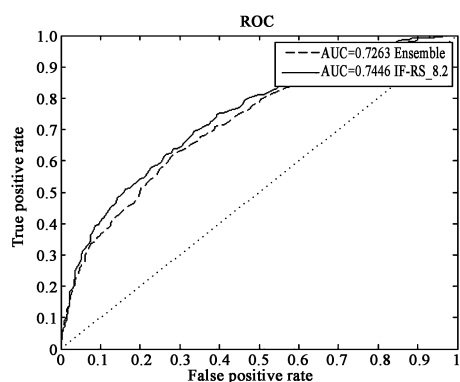
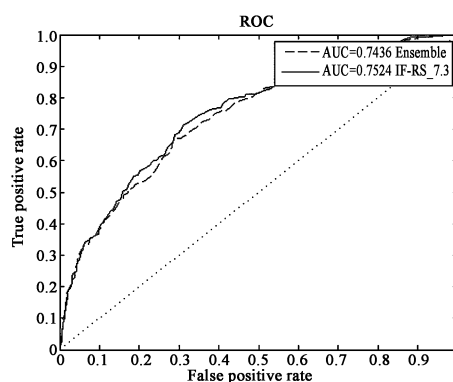
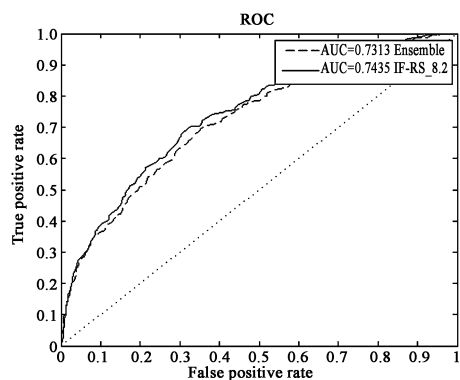
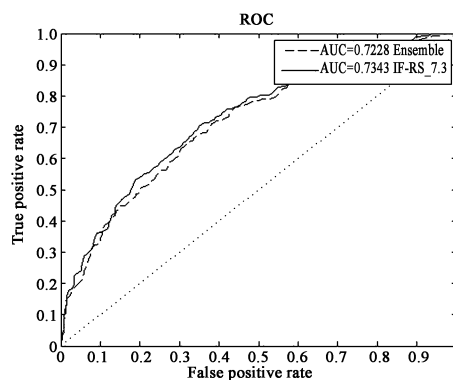
(a) $d_{\text{sub}} = 100$ (b) $d_{\text{sub}} = 150$ (c) $d_{\text{sub}} = 200$ (d) $d_{\text{sub}} = 250$

图 4 两种算法 ROC 曲线图比较

Fig.4 Comparison of ROC curve of different method in spatial domain

5 结 论

目前,高维特征是进行隐写分析时所用的主要特征。通过对特征进行排名,提高有效特征在特征子空间的比例,可以提高集成分类器的泛化能力,取得了较好的实验效果。

本文算法首先计算每维检测特征的互信息得

分并根据得分高低将特征进行排名,然后设置分界点将特征分为重要特征区域与普通特征区域,依据设定的抽样比例从两个区域随机地抽取特征组成不同的特征子空间并训练集成分类器。最后使用集成分类器进行分类。本文算法检测时采用隐写中较好的 nsF5 以及 S-UNIWARD,分别针对频域及空域图像进行隐写,而后进行检测。实验结果表明,针对频域图像,在不同的子空间维数

d_{sub} 下,取得最好效果时的抽样率也不同。当 d_{sub} 较小时抽样率在 $P=\{0.8,0.2\}$ 较好,当 d_{sub} 增大时,抽样率在 $P=\{0.7,0.3\}$ 较好;针对空域图像,最低检测错误率都是在 $P=\{0.7,0.3\}$ 、 $P=\{0.8,0.2\}$ 出现,本算法较传统分类器在检测错误率方

面分别平均下降约 0.006 5(nsF5)和 0.006 2(S-UNIWARD),具有较好的检测效果。

综上,针对频域与空域中两种不同的隐写算法,与传统的集成分类器相比,该算法具有更高的检测精度。

参 考 文 献:

- [1] SEDIGHI V, FRIDRICH J. Effect of saturated pixels on security of steganographic schemes for digital images [C]//*Proceedings of 2016 IEEE International Conference on Image Processing*. Phoenix, AZ, USA: IEEE, 2016: 2747-2751.
- [2] TANG W X, LI H D, LUO W Q, *et al.* Adaptive steganalysis based on embedding probabilities of pixels [J]. *IEEE Transactions on Information Forensics and Security*, 2016, 11(4): 734-745.
- [3] PEVNÝ T, FILLER T, BAS P. *Using High-dimensional Image Models to Perform Highly Undetectable Steganography* [M]//BÖHME R, FONG P W L, SAFARI-NAINI R. *Information Hiding*. Berlin, Heidelberg: Springer, 2010: 161-177.
- [4] DENEMARK T, FRIDRICH J, HOLUB V. Further study on the security of S-UNIWARD [C]//*Proceedings of SPIE Volume 9028, Media Watermarking, Security, and Forensics* 2014. San Francisco, California, USA: SPIE, 2014: 902805.
- [5] 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.
- [6] FRIDRICH J, FILLER T. Practical methods for minimizing embedding impact in steganography [C]//*Proceedings of SPIE Volume 6505, Security, Steganography, and Watermarking of Multimedia Contents IX*. San Francisco, California, United States: SOIE, 2007: 650502.
- [7] 张健,于晓洋,黄海霞,等.可抵抗 SPA 分析的 HSBH 改进算法[J].*光学精密工程*, 2008, 16(7): 1295-1302.
ZHANG J, YU X Y, HUANG H X, *et al.* Improved HSBH algorithm against SPA [J]. *Optics and Precision Engineering*, 2008, 16(7): 1295-1302. (in Chinese)
- [8] KODOVSKÝ J, FRIDRICH J. Calibration revisited [C]// *Proceedings of 11th ACM Workshop on Multimedia and Security*. Princeton, NJ, USA: ACM, 2009: 63-74.
- [9] FRIDRICH J, KODOVSKÝ J. Rich models for steganalysis of digital images [J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(3): 868-882.
- [10] 赵宝琴,袁志民.面向隐写分析的图像富模型特征的改进[J].*科学技术与工程*, 2016, 16(31): 56-60, 65.
ZHAO B Q, YUAN Z M. Steganalysis oriented improved rich image model features [J]. *Science Technology and Engineering*, 2016, 16(31): 56-60, 65. (in Chinese)
- [11] KODOVSKÝ J, FRIDRICH J, HOLUB V. Ensemble classifiers for steganalysis of digital media [J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(2): 432-444.
- [12] 张兴春,孙寿健.基于贝叶斯分类器的图像隐写分析[J].*液晶与显示*, 2017, 32(7): 560-566.
ZHANG X C, SUN S J. Image steganalysis based on Bayesian classifier[J]. *Chinese Journal of Liquid Crystals and Displays*, 2017, 32(7): 560-566.
- [13] 王秀霞.分类器的选择性集成及其差异性研究[D].兰州:兰州理工大学, 2011: 15-18.
WANG X X. Research on classifier selective ensemble method and thire diversity measurement [D]. Lanzhou: Lanzhou University of Technology, 2011: 15-18. (in Chinese)
- [14] LIU Q, SHI S H, ZHU H M, *et al.* A mutual information-based hybrid feature selection method for software cost estimation using feature clustering [C]//*Proceedings of the 38th Annual Computer Software and Applications Conference*. Västerås, Sweden: IEEE, 2014: 27-32.

作者简介:张兴春(1980—),男,黑龙江哈尔滨人,主要从事隐写分析以及图像处理的工作。

孙寿健(1990—),男,山东临邑人,硕士,主要从事信息隐藏以及信号处理等工作。E-mail:zynwj08@163.com