软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(4):987−1001 [doi: 10.13328/j.cnki.jos.005411]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

针对特定测试样本的隐写分析方法*

张逸为[1,2], 张卫明[1,2], 俞能海[1,2]

[1](中国科学技术大学 信息科学技术学院,安徽 合肥 230027)

[2](中国科学院 电磁空间信息重点实验室(中国科学技术大学),安徽 合肥 230027)

通讯作者: 张卫明, E-mail: zhangwm@ustc.edu.cn

摘 要: 现今主流的图像隐写分析方法主要聚焦于设计检测特征,用以提高通用盲检测(universal blind detection,简称 UBD)模型的检测准确率,这类检测方法与待测图像无关,难以做到精准检测.在拥有大数据训练资源的前提下,研究了隐写对图像特征的影响,找出了隐写分析与图像特征之间的重要关系,基于此提出了一种为测试样本选择专用训练集的隐写分析方法.以经典的 JPEG 隐写算法 nsF5 和主流的 JPEG 隐写分析特征(CC-PEV、CC-Chen、CF*、DCTR 和 GFR)为例组织实验,结果表明,该方法的检测准确率高于其他同类方法.

关键词: 信息隐藏;隐写分析;特定测试样本;高精度;机器学习

中图法分类号: TP391

中文引用格式: 张逸为,张卫明,俞能海.针对特定测试样本的隐写分析方法.软件学报,2018,29(4):987−1001. http://www.jos.org.cn/1000-9825/5411.htm

英文引用格式: Zhang YW, Zhang WM, Yu NH. Specific testing sample steganalysis. Ruan Jian Xue Bao/Journal of Software, 2018,29(4):987−1001 (in Chinese). http://www.jos.org.cn/1000-9825/5411.htm

Specific Testing Sample Steganalysis

ZHANG Yi-Wei[1,2], ZHANG Wei-Ming[1,2], YU Neng-Hai[1,2]

[1](School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

[2](Key Laboratory of Electromagnetic Spatial Information of the Chinese Academy of Sciences (University of Science and Technology of China), Hefei 230027, China)

Abstract: Nowadays, the steganalysis of digital image mainly focuses on the design of steganalysis features to improve the universal blind detection (UBD) model's detection accuracy. However it has nothing to do with the testing images and is difficult to achieve high-precision detection. Based on large data training resources, this article studies the influence of steganography on image features to uncover the important relationship between steganalysis and image feature. Furthermore, the article proposes a steganalysis method for testing samples to select specialized training sets. The classical JPEG steganography algorithm nsF5 and the mainstream JPEG steganalysis features, such as CC-PEV, CC-Chen, CF*, DCTR and GFR, are used as an example to organize the experiments. The results show that the accuracy of this method is higher than that of other similar methods.

Key words: information hiding; steganalysis; specific testing sample; high precision; machine learning

隐写术(steganography)[1−3]是一种将消息嵌入到数字载体(如图像、视频、音频、文本等)进行隐蔽通信的技术.近十几年来,隐写术作为一种保护通信安全的手段迅速发展,其中以数字图像为载体的隐写技术最为成

* 基金项目: 国家自然科学基金(U1636201, 61572452)

Foundation item: National Natural Science Foundation of China (U1636201, 61572452)

本文由"多媒体大数据处理与分析"专题特约编辑赵耀教授、李波教授、华先胜研究员、文继荣教授、蒋刚毅教授、常冬霞副教授推荐.

收稿时间: 2017-04-30; 修改时间: 2017-06-26; 采用时间: 2017-10-13; jos 在线出版时间: 2017-12-01

CNKI 网络优先出版: 2017-12-04 11:50:47, http://kns.cnki.net/kcms/detail/11.2560.TP.20171204.1150.022.html

熟.数字图像隐写按照载体压缩方式可分为空域图像隐写和频域图像隐写.空域图像隐写由传统的隐写方法,如LSB 替换[4]、LSB matching[5,6]等,发展到自适应隐写算法,如 HUGO[7]、WOW[8]、S-UNIWARD[9]、MG[10]、MVG[11]和 MiPOD[12]等,这类方法定义修改像素的失真,给失真小的像素赋予较高的修改概率以增加隐写的隐蔽性.由于 JPEG 图像格式的广泛使用,频域图像隐写主要以 JPEG 图像为载体,经典的方法有 F5[13]、nsF5[14]、OutGuess[15]、MB1[16]、MB2[17]、PQ[18]、MME[19]、YASS[20]等;自适应隐写方法有 J-UNIWARAD[9]、UED[21]及其改进 UERD[22];还有边信息 JPEG 隐写 SI-UNIWARD[9]以及 SI-UERD[22],此类方法利用 JPEG 图像压缩过程的量化取整信息提升隐写安全性.然而,这些技术也成为了不法分子暗地传递秘密信息的有效渠道,在 2001 年的 911 恐怖袭击、2007 年哥伦比亚毒枭以及 2011 年全能神邪教等案件中都出现了隐写术的影子.

能够与之抗衡的技术被称为隐写分析(steganalysis)[23].隐写分析是针对隐写术的一种分析技术,对于待测载体,隐写分析工作分为几个不同层次,主要分为:隐写载体检测、隐写算法分析、秘密信息提取、隐写明文获取等内容.其中,隐写载体检测旨在检测载体是否被嵌入秘密信息;隐写算法分析是在前一步的基础上,分析被隐写载体的秘密信息嵌入方法和嵌入率;秘密信息提取的任务是在前两步工作的基础上,确定秘密消息嵌入的位置并提取出隐写密文;最后将密文解密为隐写明文即完成了隐写分析工作.

然而,现今主流隐写分析工作集中在分析过程的第 1 步,也就是隐写载体检测,主要研究如何高精度地确定载体是否含有秘密信息,并且通常假设隐写方法与嵌入率已知.目前,数字图像隐写分析技术的主流思路是:设计数字图像特征提取方法,利用机器学习训练分类器区分载体和载密对象,近年来流行使用 Ensemble[24]分类器.常用的隐写分析特征有马尔可夫[25]、共生矩阵[26]、直方图高阶距[27]等,以这些特征为基础发展出了很多隐写分析算法[28−40]:空域中有 SPAM[31]、CSR[36]和基于富模型(RichModel)的高维特征[32,33,35],频域的代表特征有 PEV[28]、CHEN[29]、CC-CHEN[30]、CC-PEV[30]、J+SRM[32]以及近几年提出的高效特征 PHARM[37]、DCTR[38]、GFR[39]等.针对自适应隐写设计的自适应隐写分析方法是最新趋势之一,通过对自适应隐写嵌入路径的估计可以预测出最可能的隐写位置,从而更有针对性地检测自适应隐写,基于这一思想,Tang 等人[41]和 Denemark 等人[42,43]改进了以往的特征,Zhang 等人[44]使用高斯偏导数滤波器也得到了很好的检测效果.此外,近年来,深度学习的成果也逐渐开始应用于隐写分析工作当中[45].

然而,隐写分析在从实验室环境向现实场景过渡的过程中出现了很多困难,其中,最突出的是载体来源失配(cover source mismatch,简称 CSM)问题.CSM 是测试集与训练集不匹配时隐写分析效果显著下降的一种现象.Lubenko 等人[46]为解决 CSM 问题,使用众多不同数据来源的图像训练分类器,然而,该方法中与测试样本不相关的训练数据会干扰检测结果;Kodovský等人[47]针对 CSM 问题提出了 3 种解决方案,第 1 种方案使用混杂的载体图像训练一个综合分类器,第 2 种方案为每一类载体图像训练一个分类器,测试集会被与之最匹配的分类器进行分类,第 3 种方案与第 2 种类似,对每一个测试样本都投入到最匹配的分类器中,但预先训练好的分类器并不能很好地适用于所有测试样本;Lerch-Hostalot 等人提出了 ATS[48]方法,该方法使用测试集与双重隐写(对图像重复隐写两次)后的测试集作为训练数据训练分类器,使用该分类器检测隐写一次的测试集,将该检测结果作为对应测试集图像的隐写分析结果,该方法通过绕过训练数据来避免 CSM 问题,但对测试集有一定的载密样本比例要求,这在真实场景中难以得到满足.

本文考虑需要精准隐写分析少量测试样本的应用场景,这种场景需要很高的检测精度,但是由于待测样本量很小,可以采用高复杂度的检测方法.对少量特定测试样本的精准隐写分析问题来源于真实应用的实际需求,典型的应用场景有如下两种(如图 1 所示).

(1) 考虑一个面向海量图像数据的隐写分析监控场景.为了能够处理大量数据,采用层级化的隐写分析系统.首先使用简单的特征训练出分类器放在层级的第 1 层,输入的图像仅需要提取简单的特征即可进行隐写分析检测,利用这一层过滤掉大部分图像,将疑似载密的图像移交给下一层;第 2 层分析需要付出更大的计算代价,但相应地也具有更准确的分析功能,可以再次过滤掉部分非载密图像,留下更少量的可疑图像给下一层.如此设计若干层分析器,过滤得到高可疑图像,这时便需要特定测试样本隐写分析模块给出最终的精准分析结果.

(2) 在刑侦工作中,分析人员锁定了隐写术的疑似使用者,通过某种手段获取其传输或存储的图像,这些图

像往往是高可疑的,值得花更大的计算代价进行有针对性的精准分析.

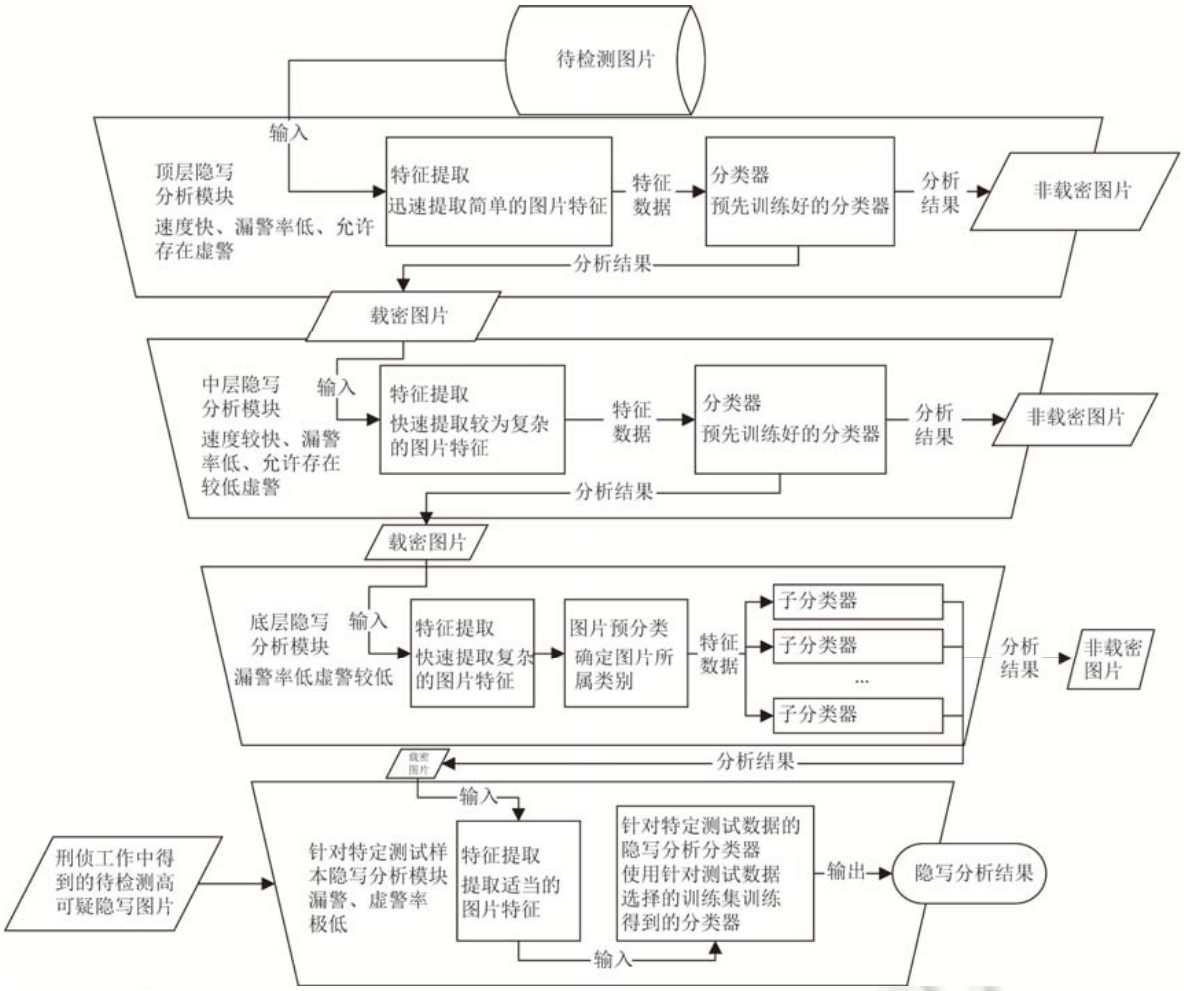


Fig.1 Typical application scenarios for precise steganalysis

图 1 精准隐写分析的典型应用场景

针对上述场景,本文聚焦隐写分析工作中的载体检测任务,假设隐写算法与嵌入率已知,以准确判断样本是否藏有秘密信息为目的,首先研究隐写对图像特征的影响,随后找出隐写分析与图像特征之间的重要关系,基于此提出一种为测试样本选择专用训练集、训练专用分类器的隐写分析方法,我们称其为“特定测试样本隐写分析(specific testing sample steganalysis,简称 STSS)”,该方法度量每一个测试样本与当前训练数据库中的训练样本之间的相似程度,选择与测试样本相似度最高的训练数据训练分类器,排除不相关训练数据的干扰,从而大幅度地提升隐写分析的准确率.形象地讲,如果将训练分类器检测样本比作顾客(测试样本)买衣服(分类器)的过程,衣服越合身则(隐写分析)效果越好,文献[47]的方法好比是“成衣铺”,事先制作好各种尺码的衣服,顾客来了可以按照尺码范围选择大致合身的衣服.而本文方法可以比作为每位顾客“量体裁衣”.显然,后者虽然有较高的成本,但衣服最为合身.

本文第 1 节分析隐写操作对载体图像的影响.第 2 节提出影响隐写分析的两个主要因素.第 3 节设计特定测试样本隐写分析框架.第 4 节讨论核心实验参数的确定.第 5 节、第 6 节分别为实验结果和结论.

1 隐写操作对载体图像的影响

隐写分析结果往往由分类器给出,而分类器的输入只是图像的特征,因此本节讨论隐写对图像特征的影响.本节讨论的隐写方法以 nsF5 为例;对于隐写分析特征,现今有很多 JPEG 图像特征可用,本文使用比较有代表性的 5 种进行说明,分别为 CC-PEV、CC-CHEN、CF*[32]、DCTR 和 GFR,讨论从“不同嵌入率隐写”和“重复隐写”两个角度进行.

1.1 不同嵌入率对图像的影响

Ker 等人[49]阐述了隐写操作前后图像特征变化的一种规律:不同嵌入率的隐写会使得图像特征沿相同方向移动不同距离,该工作中使用的图像特征为 CF*.为验证该结论在不同隐写分析特征下是否可以扩展,我们随机选取 5 000 幅 ImageNet 数据库中的 JPEG 图像作为载体图像集,用符号 Ψ_0 表示.使用 0.05bpac、0.1bpac、0.2bpac、0.3bpac、0.4bpac、0.5bpac 共 6 种嵌入率下的 nsF5 算法生成 6 个不同嵌入率的载密图像集 $\widetilde{\Psi}$, 包括载体图像共 7 组图像集,用符号 Ψ表示.用选用的特征提取方法提取Ψ中的图像特征,用符号 F 表示.

使用 PCA 方法将 F 降维至 2 维,计算所有载密图像与载体图像特征向量差的均值并归一化.得到在 6 种嵌入率的 nsF5 隐写算法作用下,5 种隐写分析特征的移动向量,示意图如图 2 所示.可以看出,在各种隐写特征下,隐写操作使得图像特征在特征域上沿着某个方向运动,且方向与嵌入率无关,而移动距离与嵌入率呈简单的线性关系,即以 i 为运动方向的单位向量,α为嵌入率,特征平均的运动向量 $\Delta_F \approx \lambda\alpha i$, 其中,$\lambda$为固定常数.

这种现象表明,图像经过隐写后,其特征的位置发生了变化,而且不同图像有着相似的运动方向.同时可以发现,特征移动距离随着嵌入率等比例地增大,这可以解释隐写分析中大嵌入率隐写比较容易检测的现象.

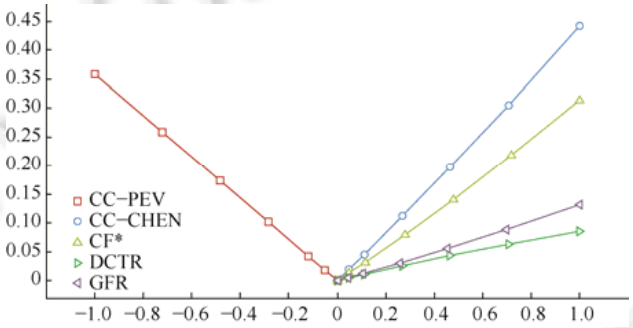


Fig.2 The average displacement of the features obtained by different extraction methods at different payloads

图 2 多种提取方法得到的特征在不同嵌入率下的平均位移示意图

1.2 重复隐写对图像的影响

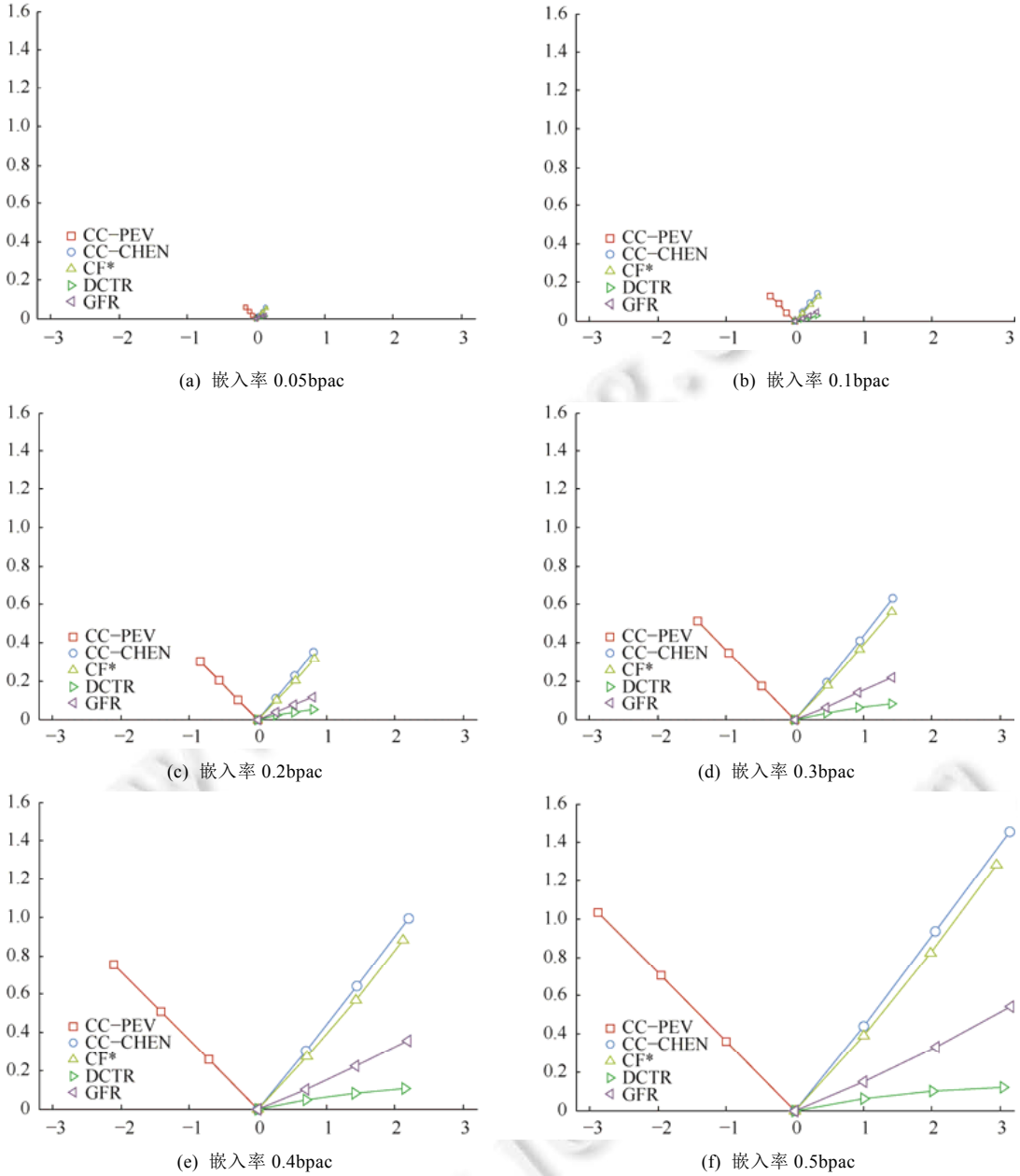
隐写是以隐藏秘密消息为目的进行的文件改写,然而,以载密文件作为载体重新进行隐写得到的图像对隐写分析可以起到一定的辅助作用.这里,我们使用了第 1.1 节中的 Ψ_0 作为载体图像数据,$\widetilde{\Psi}$ 为单次隐写的载密图像,随后以使用 $\widetilde{\Psi}$ 为载体图像,分别使用对应嵌入率的 nsF5 算法再次进行隐写,得到“双嵌”载密图像集 $\widetilde{\widetilde{\Psi}}$, 以此为载体再次隐写得到“三嵌”载密图像集 $\widetilde{\widetilde{\widetilde{\Psi}}}$. 分别提取特征得到 $F,\widetilde{F},\widetilde{\widetilde{F}},\widetilde{\widetilde{\widetilde{F}}}$, 将特征使用 PCA 方法降至 2 维,计算每一组特征相对原始载体图像特征差异的平均值,在二维空间中的表示如图 3 所示.

从图 3 可以看出,在重复隐写操作下,图像特征沿着相同的方向运动相同的距离,距离大小和嵌入率呈正相关.对于一幅图像 C 来说,假设已知隐写算法 E 以及嵌入率α,可以对该样本进行同样的操作得到其对应的载密图像 $S = E_\alpha(C)$, 使用图像特征提取算法 A 提取二者的特征, $f = A(C), \tilde{f} = A(S) = A(E_\alpha(C))$, 计算二者之间的特征变化 $\Delta = \tilde{f} - f$ 作为 C 的特征随着该隐写操作变化的估计,即有:

$$\Delta = A(E_\alpha(C)) - A(C) \tag{1}$$

对于测试样本,使用公式(1)可以计算其在特定隐写操作下的特征运动估计Δ,可以利用Δ作为测试样本的先

验知识,提高对该样本分析的准确率.



(a) 嵌入率 0.05bpac (b) 嵌入率 0.1bpac

(c) 嵌入率 0.2bpac (d) 嵌入率 0.3bpac

(e) 嵌入率 0.4bpac (f) 嵌入率 0.5bpac

Fig.3 Diagram of image feature's change under repeatedly embedded with different payload

图 3 重复隐写操作在不同隐写嵌入率下的图像特征变化示意图

2 针对测试样本的隐写分析

区别于以往的隐写分析框架,本文提出的 STSS 框架旨在挖掘测试样本可利用的信息,为每一个测试样本生成专用的检测分类器,最终给出当前条件下最好的分析结果.我们从机器学习分类器出发,探讨设计特定测试样本专用分类器的策略.

2.1 隐写分析中的机器学习

隐写分析使用带有标签的载体载密数据训练分类器,就当前使用最为广泛的 Ensemble 分类器来说,为了用简单的 FLD 分类器刻画复杂的回归问题,对于给定的大小为 N^{trn} 的 d 维训练数据 $\mathbf{X}^{trn}=\{\boldsymbol{x}_m,\tilde{\boldsymbol{x}}_m\}_{m=1}^{N^{trn}}$,其中,$\boldsymbol{x}_m,\tilde{\boldsymbol{x}}_m\in\mathbb{R}^d$,该分类器将训练数据与特征维度随机分成 L 份,编号 $1\leqslant l\leqslant L$,标签按照载体数据、载密数据两种类型被映射到 {0,1} 上,训练数据子集为 $\mathbf{N}_l\in\{1,...,N^{trn}\}$,维度子集 $D_l\in\{1,...,d\}$,每一个子分类器在训练集 $\{x_i^{(D_l)},\tilde{x}_i^{(D_l)}\mid i\in N_l\}$ 上训练得到分类面,其计算方法如下[24]:

$$\mathbf{v}_l=(\mathbf{S}_W+\lambda\mathbf{I})^{-1}(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}}) \tag{2}$$

其中,$\boldsymbol{\mu}$ 和 $\tilde{\boldsymbol{\mu}}$ 为样本中载体数据和载密数据特征的平均值:

$$\boldsymbol{\mu}=\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\boldsymbol{x}_m^{(D_l)},\tilde{\boldsymbol{\mu}}=\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\tilde{\boldsymbol{x}}_m^{(D_l)} \tag{3}$$

$$\mathbf{S}_W=\sum_{m\in\mathbf{N}_l}\left(\boldsymbol{x}_m^{(D_l)}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}_m^{(D_l)}-\boldsymbol{\mu}\right)^{\mathrm{T}}+\sum_{m\in\mathbf{N}_l}\left(\tilde{\boldsymbol{x}}_m^{(D_l)}-\tilde{\boldsymbol{\mu}}\right)\left(\tilde{\boldsymbol{x}}_m^{(D_l)}-\tilde{\boldsymbol{\mu}}\right)^{\mathrm{T}} \tag{4}$$

对于测试样本 $\mathbf{y}\in\mathbb{R}^d$,子分类器计算 $\mathbf{v}_l^T\mathbf{y}^{(D_l)}$ 并与阈值比较得到该子分类器的判决结果.所有子分类器用投票的形式为判决结果投票,以票数最多的类别作为输出判决.

从上述计算过程中可以看到,子分类器判决是否准确会对判决结果产生重要影响.而每个子分类器表达的信息都包含在向量 $\mathbf{v}_l$ 中.我们拟将测试样本的信息加入到其生成公式中,使得 $\mathbf{v}_l$ 对测试样本做出正确的判决.具体到 $\mathbf{v}_l$ 的生成表达式,两个多项式 $(\mathbf{S}_W+\lambda\mathbf{I})^{-1}$ 和 $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 的乘积构成了 $\mathbf{v}_l$ 生成的两个分量,下面分别从这两个分量着手分析其优化方法.

2.2 特征空间距离

分类器根据训练数据的标签标记特征空间区域所属的类别,而计算 $(\mathbf{S}_W+\lambda\mathbf{I})^{-1}$ 的值就是训练数据标记特征的过程.对于带有标签 K 的训练数据 $\boldsymbol{x}$,分类器倾向于标记 $\boldsymbol{x}$ 周围的测试样本为 K.隐写分析致力于区分图像隐写前后的特征,而隐写使图像在特征空间定向小幅移动,这种移动相对于距离近的特征显得较为明显,使用距离近的特征作为训练集会更加有效;相反地,在特征空间上距离很远的训练数据与测试样本的相关性不高,若使用这种不相关的训练数据得到的分类器则并不可靠,图 4 形象地描述了上述情况的一种典型特征分布.

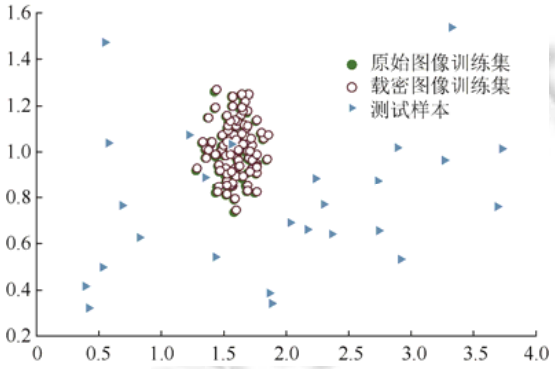


Fig.4 Diagram of steganalysis feature distribution

图 4 隐写分析特征分布示意图

图 4 中的特征数据来源于 ImageNet 数据库,训练集为其中的 100 幅图像,测试集为其中 25 幅图像.对载体训练图像进行隐写得到载密训练图像,提取 CC-CHEN 特征,使用 PCA 方法降维到 2 维,圆圈表示训练样本特征投影,三角表示测试样本特征投影.现实分类问题中,上述情况也会时常出现,训练数据与测试数据在特征空间的分布相差很大,使得分类器计算出的分类面无法正确分类测试样本;相反地,集中的训练数据对于与其分布接

近的测试样本有很强的分析能力,因此,得到足够量的有效训练数据成为隐写分析的关键.在实际隐写分析过程中,可以利用待检测图像特征这一先验知识,充分考虑特征空间中的特征位置分布,选择与测试样本接近的特征数据组成训练集,去除不相关训练数据的干扰,提高分类准确率.

从分类器优化角度来说,由公式(2)我们知道,子分类器 $\mathbf{v}_l$ 含有两个因子 $(\mathbf{S}_W+\lambda\mathbf{I})^{-1}$ 和 $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$, 对于 $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 有如下推导:

$$\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}}=\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\boldsymbol{x}_m^{(D_l)}-\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\tilde{\boldsymbol{x}}_m^{(D_l)}=\frac{1}{N^{trn}}\left(\sum_{m\in\mathbf{N}_l}\boldsymbol{x}_m^{(D_l)}-\sum_{m\in\mathbf{N}_l}\tilde{\boldsymbol{x}}_m^{(D_l)}\right)=\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\left(\boldsymbol{x}_m^{(D_l)}-\tilde{\boldsymbol{x}}_m^{(D_l)}\right) \tag{5}$$

根据第 1.1 节中的结论,在固定嵌入率、隐写算法和特征提取方法的情况下,随机选择的图像特征有朝着一个固定方向移动的趋势,移动幅度与嵌入率α有关,对于固定的α来说,

$$\frac{1}{|\mathbf{N}_l|}\sum_{m\in\mathbf{N}_l}\left(\boldsymbol{x}_m^{(D_l)}-\tilde{\boldsymbol{x}}_m^{(D_l)}\right)=\Delta_F\approx\lambda\alpha i \tag{6}$$

因此有:

$$\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}}=\frac{1}{N^{trn}}\sum_{m\in\mathbf{N}_l}\left(\boldsymbol{x}_m^{(D_l)}-\tilde{\boldsymbol{x}}_m^{(D_l)}\right)=\frac{|\mathbf{N}_l|}{N^{trn}}\frac{1}{|\mathbf{N}_l|}\sum_{m\in\mathbf{N}_l}\left(\boldsymbol{x}_m^{(D_l)}-\tilde{\boldsymbol{x}}_m^{(D_l)}\right)=\frac{|\mathbf{N}_l|}{N^{trn}}\Delta_F\approx\frac{|\mathbf{N}_l|}{N^{trn}}\lambda\alpha i \tag{7}$$

即对于未特定选取位移方向的训练数据来说, $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 项的值仅与隐写行为相关.而在载体与载密特征中心距离固定的情况下,训练特征越集中,则越能强化附近区域的标注,测试数据如果处于被强化标注的区域,则会被更准确地分类.

为研究特征距离度量不同的训练集对隐写分析会产生怎样的影响,我们进行了相关实验,具体实验设置如下:测试集为 ImageNet 数据库中 1 000 幅图像,训练集为不同于训练集的 95 000 幅图像;生成载密图像的隐写算法为 nsF5 算法,设置 5 种嵌入率分别为 0.1bpac、0.2bpac、0.3bpac、0.4bpac、0.5bpac;根据第 1.1 节、第 1.2 节中的结论,各种特征在隐写操作下有相似的变化规律,为了简单明了,这里以 CC-CHEN 特征为例进行实验;使用 Ensemble 分类器分类测试样本;错误率 P_E 使用如下公式进行计算,其中,P_{FA} 表示虚警概率,P_{MD} 表示漏警概率:

$$P_E=\min_{P_{FA}}\frac{1}{2}(P_{FA}+P_{MD}(P_{FA})) \tag{8}$$

实验中为每个测试样本选择固定距离范围内的数据作为训练集,每次从指定范围内的训练数据中随机选取 N=600 组样本作为训练分类器,规定特征距离从 0~7.8,以 0.3 为间隔划分区间,距离度量使用欧氏距离.即对于测试数据 $y\in\mathbb{R}^d$, 训练数据 $x\in\mathbb{R}^d$, 距离度量 $Dis=\|y-x\|$. 由于带标签的训练数据成对出现,定义测试数据到训练数据的距离 d 为其到一对训练特征距离之和,即:

$$d=\|\mathbf{y}-\mathbf{x}\|_2+\|\mathbf{y}-\tilde{\mathbf{x}}\|_2 \tag{9}$$

计算 $\mathbf{y}$ 与 $\mathbf{X}^{trn}$ 中所有训练数据的距离,得到 $\mathbf{D}_{\mathrm{y}}=\{d_i\}_{i=1}^{N^{trn}}$, 其中, $d_i=\|\mathbf{y}-\mathbf{x}_i\|_2+\|\mathbf{y}-\tilde{\mathbf{x}}_i\|_2$, 将 $\mathbf{D}_{\mathrm{y}}$ 按照元素由小到大排序得到 $\dot{\mathbf{D}}_{\mathrm{y}}=\{\dot{d}_{a_i}\}_{i=1}^{N^{trn}}$, $\dot{\mathbf{D}}_{\mathrm{y}}$ 中对任意 $1\leqslant a_i\leqslant a_j\leqslant N^{trn}$, 有 $\dot{d}_{a_i}\leqslant\dot{d}_{a_j}$. 对于指定的特征距离间隔,找到符合距离要求的集合并随机选择 N 个组成 $\{\dot{d}_{a_i}\}_{i=1}^{\mathbf{N}}$, 其对应的训练数据即为 $\mathbf{y}$ 的专用训练集 $\mathbf{N}_{\mathrm{y}}=\{\mathbf{x}_{a_i},\tilde{\mathbf{x}}_{a_i}\}_{i=1}^{\mathbf{N}}$, 实验结果如图 5 所示.

由图 5 可以得知,在不同嵌入率下,随着训练集与测试数据特征距离变大,隐写分析错误率逐渐提高.这说明,特征距离的缩短使得训练数据更好地为测试数据提供分类决策依据,用接近测试数据的特征训练的分类器更加有效.

同时我们做了相同条件下随机选择训练集对同样的测试集进行分类的实验,得到对应不同嵌入率的结果,在图 5 中对应的线上用红色的点突出标注.可以看到,随机数据的分类错误率与特征距离的中间水平相当,这个现象可以解释为随机数据距离测试样本的平均距离处于中间水平,因此,其分类准确率与中间水平的某个固定距离段接近.同时我们注意到,随着嵌入率的增加,与随机数据准确率相当的特征距离在增加,这说明,随着嵌入

率的增加,载密图像特征运动幅度增大,随机训练集的平均距离也随之增加.我们也可以看出,距离测试样本近的训练集分类准确率明显优于随机训练集.上述结果说明,训练集与测试数据之间的特征距离是影响隐写分析结果的重要因素之一,而且相同条件下,距离越近,则隐写分析效果越好.

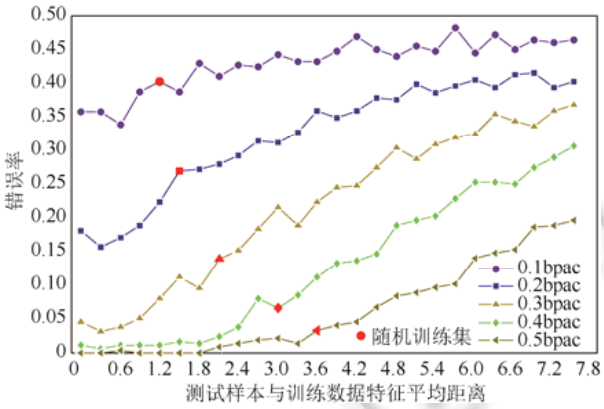


Fig.5 The relationship between the testing and training feathers' distance and the error rate under different embedding payloads

图 5 不同嵌入率下测试与训练特征距离与错误率关系示意图

2.3 特征在隐写操作下的运动模式

隐写分析的目的是检测测试图像是否载密,分类器使用的训练数据是机器学习何为载体图像、何为载密图像的根本来源,二者的区别仅为隐写前后的特征变化 $\mathbf{d}=\tilde{\mathbf{x}}-\mathbf{x}$, 分类器在这微小的变化当中寻找区分载体、载密特征的分界面.在 $\mathbf{v}_l$ 的计算过程中, $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 项表示了隐写前后特征运动信息,为了将训练数据中的载体、载密数据区分开,需要 $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 尽可能地大,这需要训练数据特征沿着同一个方向运动,此时 $(\boldsymbol{\mu}-\tilde{\boldsymbol{\mu}})$ 最大,训练错误率较低.为了准确分析测试样本,我们需要训练数据与测试数据在隐写操作下特征运动模式接近,理想情况为训练数据在隐写操作下的特征运动模式与测试图像完全相同.为了获取测试样本的隐写运动模式,使用第 1.2 节中的结论,特征在多次隐写下会沿相同的方向移动同样的距离,使用公式(1)可以计算出测试样本 $\mathbf{y}$ 的运动模式估计 $\Delta_{\mathbf{y}}$.定义测试数据与训练数据运动模式的欧氏距离 m 为二者运动模式的差异性度量:

$$m=\left\|\Delta_{\mathbf{y}}-\Delta_{\mathbf{x}}\right\|_2=\left\|(\tilde{\mathbf{y}}-\mathbf{y})-(\tilde{\mathbf{x}}-\mathbf{x})\right\|_2 \tag{10}$$

为了验证特征移动模式对隐写分析效果的影响,我们进行了如下实验.使用与第 2.2 节相同的测试集、训练数据库、训练集大小、隐写方法、嵌入率、错误率度量和特征提取方法.对每一个测试数据,计算 $\mathbf{y}$ 与 $\mathbf{X}^{trn}$ 中所有数据的运动模式近似度量,得到 $M_{\mathbf{y}}=\{m_i\}_{i=1}^{N^{trn}}$, 其中, $m_i=\left\|\Delta_{\mathbf{y}}-\Delta_{\mathbf{x}_i}\right\|_2$, 将 $M_{\mathbf{y}}$ 按元素由小到大排序得到, $\dot{M}_{\mathbf{y}}=\{\dot{m}_{b_i}\}_{i=1}^{N^{trn}}$, $\dot{m}_{\mathbf{y}}$ 中对任意 $1\leqslant b_i\leqslant b_j\leqslant N^{trn}$, 有 $\dot{m}_{b_i}\leqslant\dot{m}_{b_j}$.实验中以 5%为间隔在每个嵌入率下选取 19 个度量分位数,得到 20 段间隔,在每段间隔内随机选择 N=600 个度量值组成集合 $\{\dot{d}_{b_i}\}_{i=1}^{N}$, 其对应映射的训练数据即为 $\mathbf{y}$ 的专用训练集 $\mathbf{N}_{\mathbf{y}}=\{\boldsymbol{x}_{b_i},\tilde{\boldsymbol{x}}_{b_i}\}_{i=1}^{N}$, 实验结果如图 6 所示.

从图 6 所示结果可以看出,总体来说,随着运动模式偏差的欧式距离的增大,隐写分析的错误率逐渐升高,这种现象在大嵌入率下尤为明显,基本上验证了上述结论,与测试数据运动模式接近的训练数据更有助于样本的正确分类.但是,该现象在较低嵌入率情况下表现得并不稳定,这是因为,在嵌入率很小的情况下,训练数据之间没有能够很好地分离开,相互之间会产生干扰,导致错误率结果曲线不够稳定.

类似于第 2.2 节的实验,我们从相同条件下 95 000 幅训练图像中随机选择数据量 N=600 幅作为训练集,对同样的测试集进行分类,得到对应不同嵌入率的错误率,在图 6 中对应的线上用红色的点突出标注.可以看到,随机数据的结果仍然相当于偏差距离的中间水平,总体来说,偏差距离越小,隐写分析效果越好.

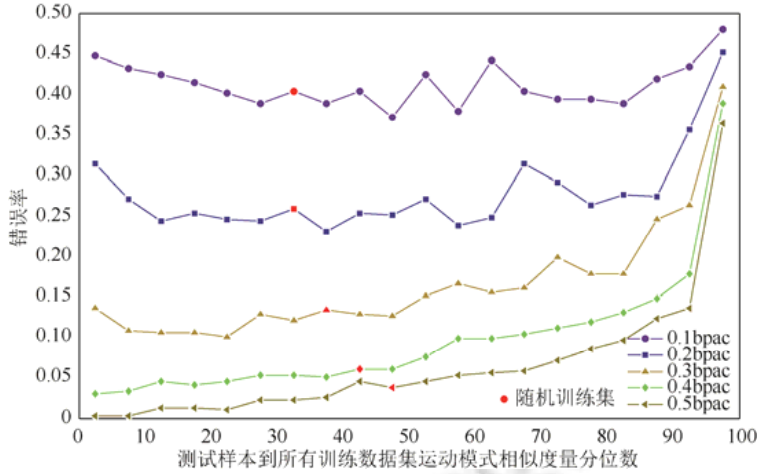


Fig.6 The relationship between the testing and training feathers' moving pattern similarity and the error rate under different embedding payloads

图 6 不同嵌入率下测试与训练特征运动模式差异性度量与错误率关系示意图

3 特定测试样本隐写分析

根据上文所述,我们已经掌握了训练数据选择的重要依据,那么面对特定测试样本隐写分析的高精度需求,分析者可以结合特征空间距离和特征运动模式差异性度量两个指标,针对每个测试样本在当前的图像数据库中寻找与之最匹配的训练集,对每个特定测试样本进行深度隐写分析.

3.1 特定测试样本隐写分析框架

根据上文所述,完整的 STSS 框架结构如图 7 所示.

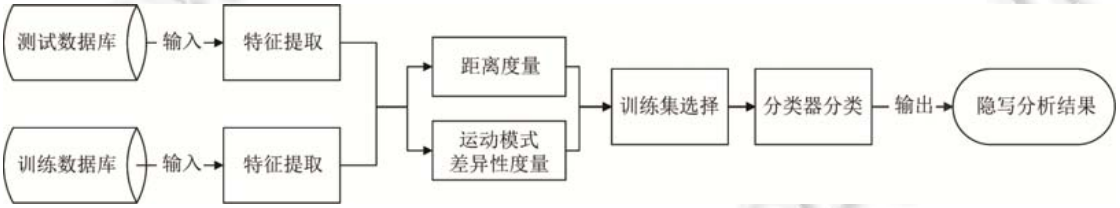


Fig.7 Framework of specific testing sample steganalysis

图 7 特定测试样本隐写分析框架

该方法需要首先准备好训练数据库.分析开始时提取待检测图像特征,与其他隐写分析框架最大的不同之处在于:该框架计算测试样本与每个训练数据的特征距离与运动模式相似度,选择最为匹配的训练集,用来训练专门的隐写分析分类器.

3.2 差异性度量设置

我们已知特征空间距离与隐写特征运动方式是影响隐写分析的重要因素,现在将二者结合成统一的度量特征相似程度的指标.定义两个特征数据 $\mathbf{x}$ 与 $\mathbf{y}$ 之间的差异性度量 S 为

$$S(\mathbf{x},\mathbf{y})=\|\mathbf{x}-\mathbf{y}\|_2+\lambda\|\Delta_{\mathbf{x}}-\Delta_{\mathbf{y}}\|_2 \tag{11}$$

其中,S 为特征向量 $\mathbf{x}$ 与 $\mathbf{y}$ 之间差异性度量的函数,$\lambda\in\mathbb{R}^*$ 为特征距离与运动模式相似度两个指标之间的权重参数,λ越大,差异性度量越趋向运动模式距离,反之,λ越接近 0,则度量越趋向于特征距离.对于隐写分析工作中

的度量,训练数据隐写前后成对出现,我们采用隐写前后的训练数据到测试数据的距离之和为特征空间距离,此时,度量可以表示成

$$S(\mathbf{x},\mathbf{y})=\left(\|\mathbf{x}-\mathbf{y}\|_2+\|\tilde{\mathbf{x}}-\mathbf{y}\|_2\right)+\lambda\left\|(\tilde{\mathbf{x}}-\mathbf{x})-(\tilde{\mathbf{y}}-\mathbf{y})\right\|_2 \tag{12}$$

经过式(12)计算得到的差异性度量值越小,说明两个向量之间越相似,我们认为选择相似的训练数据会提高分类准确率,因此,度量值小的训练数据会被优先选择作为专用分类器的训练数据.然而,λ的取值和训练集数量 N 的取值还没有确定,我们在第 4 节对其进行讨论.

4 关键参数选取

λ和N的取值在一定程度上影响着隐写分析效果,为了找到实际使用中效果最好的测试参数设置,我们在不同嵌入率下使用多种参数对相同的样本进行分析,观察结果错误率与参数之间的关系,从而找到效果较好的经验参数设置.由于各种特征在隐写操作下行为相似,因此,以 CC-CHEN 特征为例寻找参数,直接将其推广到其他隐写分析特征中.

我们使用 ImageNet 数据库中 1 000 幅图像组成测试集,60 000 幅图像组成训练数据库,使用嵌入率 0.1、0.2、0.3、0.4、0.5(bpac)下 nsF5 算法生成载密图像,提取 CC-CHEN 特征在 Ensemble 分类器下进行实验.在每组嵌入率下使用变化的λ和 N 值进行实验,记录每组参数的分析错误率,实验结果如后文的图 8 所示.

图 8 中两个参数变化导致隐写分析结果发生较大波动,虽然没有稳定在各个嵌入率都最优的参数,但可以看到,总体来说,参数与错误率关系图中有一定的趋势.为了得到平均意义上效果好的参数,将各个图中的错误率数据除以其均值,得到均值归一化的数据,将各个嵌入率下的归一化数据累加得到平均错误度量与参数之间的关系,如后文的图 9 所示.

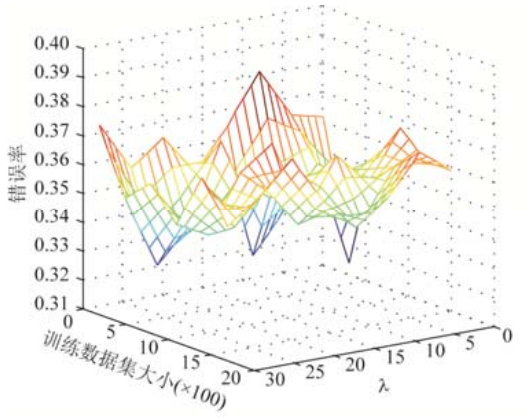
由图 9 中可以看出,当 N 超过 1 000 时,平均错误率开始收敛,并且在λ=5,N=1300 时,平均错误率取得最小值,因此我们使用λ=5 和 N=1300 作为最终实验使用的参数.

5 实验结果

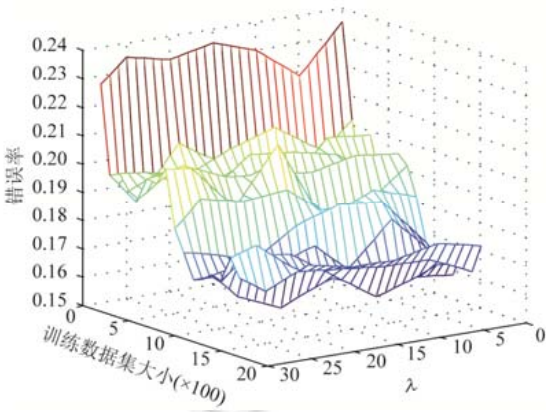
在这一节,我们给出实验结果和分析,以证明本文 STSS 框架的可行性与有效性.实验设置如下:载体图像使用 ImageNet 库中 100 000 幅质量因子 96、尺寸在(300–700)×(300–700)范围内的 JPEG 图像;从 100 000 幅图像的数据库中随机选择 5 000 幅作为测试集,其余 95 000 幅图像为训练数据库;载密图像由 nsF5 算法生成,嵌入率分别使用 0.05bpac、0.1bpac、0.2bpac、0.3bpac、0.4bpac、0.5bpac;实验使用当前主流的隐写分析特征,分别是 CC-PEV、CC-CHEN、CF*、DCTR 和 GFR;错误率使用公式(8)计算,使用 Ensemble 分类器进行分类,参数λ设置为 5,N 设置为 1300.

对比实验采用了经典的 UBD 框架与 ATS 框架.由于本文方法使用全部训练数据库信息进行分析工作,为了使实验条件公平,UBD 使用全部训练库中 95 000 幅图像作为训练集训练分类器.ATS 框架起源于空域图像隐写分析,使用降维后的 SRM[33]特征对测试集进行分析,这里,为了统一条件,将 ATS 框架中的特征替换为本实验指定的特征进行实验.ATS 框架中,对于固定数量的测试集,载体图与载密图像个数越接近,则分类效果越好,因此,我们直接使用 5 000 幅测试图像与其对应的载密图像作为该方法的测试集.图 10(a)~图 10(e)分别表示 5 种隐写分析特征的对比实验结果,横坐标表示嵌入率,纵坐标表示检测错误率,3 种线型分别表示 3 种参与对比的实验框架,具体的数据结果见表 1,粗体为对比实验中最好的结果.

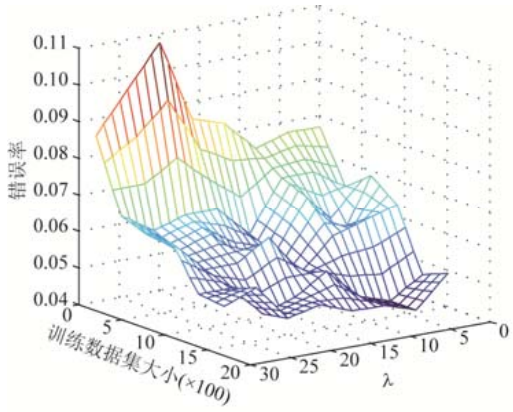
实验结果表明,本文方法基本在所有嵌入率的主流特征下的检测性能均优于其他方法,分析准确率高于其他方法 1%~9%,而且表现稳定.



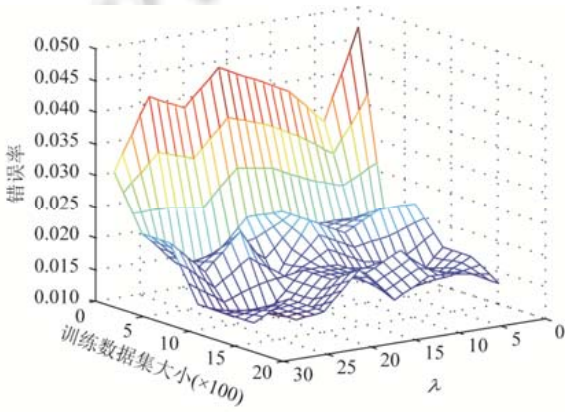
(a) 嵌入率 0.1bpac



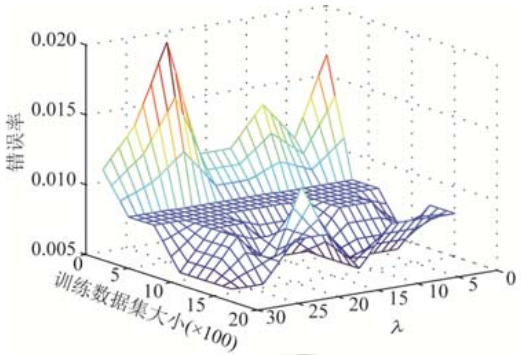
(b) 嵌入率 0.2bpac



(c) 嵌入率 0.3 bpac



(d) 嵌入率 0.4 bpac



(e) 嵌入率 0.5bpac

Fig.8 The relationship between experimental parameters and error rate under different embedding rates

图 8 不同嵌入率下实验参数与错误率关系示意图

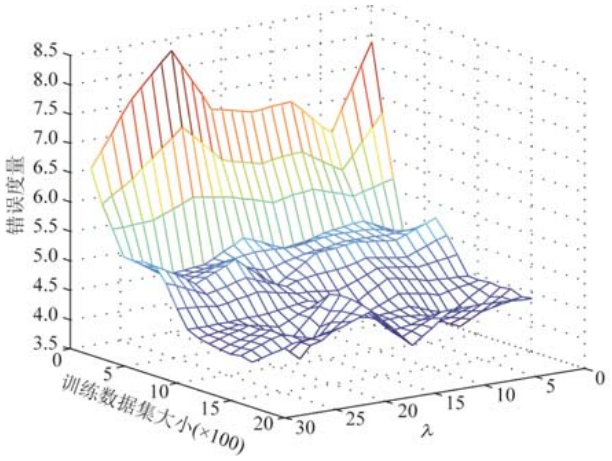
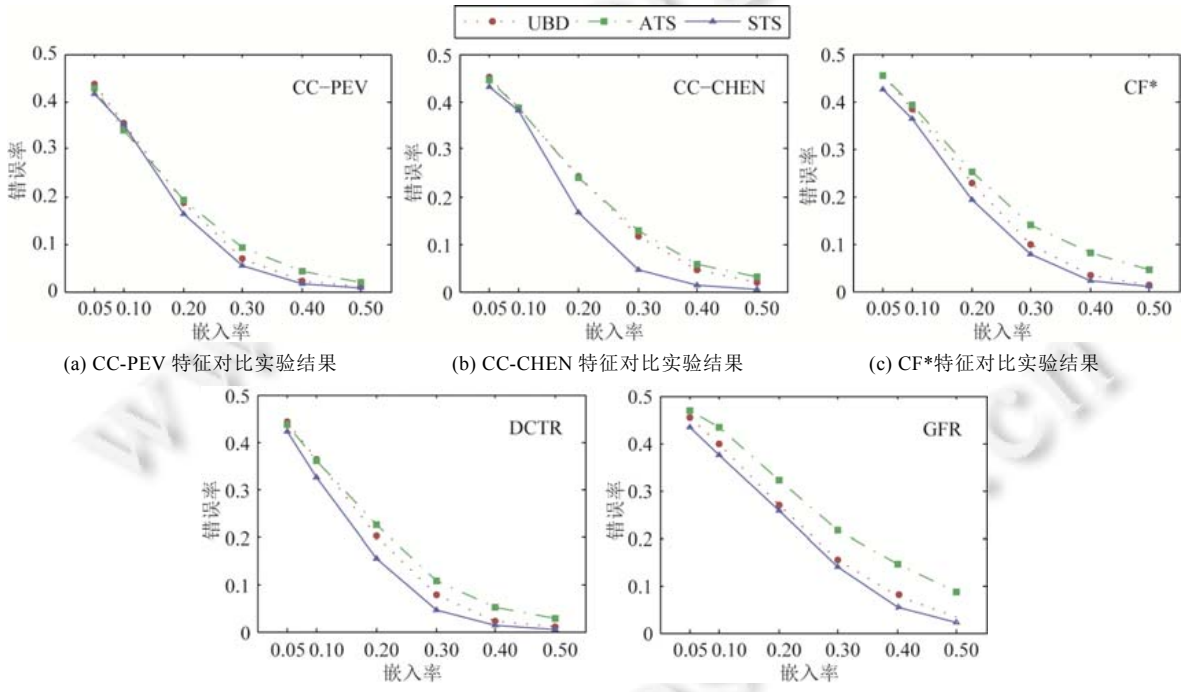


Fig.9 The relationship between the experimental parameters and the average error rate

图 9 实验参数与平均错误率关系示意图



(a) CC-PEV 特征对比实验结果 (b) CC-CHEN 特征对比实验结果 (c) CF*特征对比实验结果

(d) DCTR 特征对比实验结果 (e) GFR 特征对比实验结果

Fig.10 The comparison of experimental results between our method and other methods

图 10 本文方法与其他方法对比结果图

Table 1 The comparison of experimental results between our method and other methods

表 1 本文方法与其他方法对比结果表

隐写分析特征	分析框架	不同嵌入率(bpac)下检测错误率(%)					
		0.05	0.1	0.2	0.3	0.4	0.5
CC-PEV	UBD	43.71	35.14	18.57	6.99	2.33	1.02
	ATS	42.73	**34.03**	19.40	9.25	4.30	1.90
	STSS	**41.47**	35.12	**16.39**	**5.57**	**1.51**	**0.80**
CC-CHEN	UBD	45.05	38.62	24.13	11.71	4.57	1.84
	ATS	44.45	38.60	23.85	12.88	5.73	3.00
	STSS	**43.14**	**38.13**	**16.56**	**4.68**	**1.34**	**0.50**

Table 1 The comparison of experimental results between our method and other methods (Continued)

表 1 本文方法与其他方法对比结果表(续)

隐写分析特征	分析框架	不同嵌入率(bpac)下检测错误率(%)					
		0.05	0.1	0.2	0.3	0.4	0.5
CF*	UBD	45.42	38.31	22.74	9.77	3.41	1.44
	ATS	45.40	39.13	25.15	14.00	8.00	4.55
	STSS	**42.63**	**36.29**	**19.25**	**7.67**	**2.24**	**1.03**
DCTR	UBD	44.38	36.47	20.02	7.80	2.34	0.93
	ATS	43.60	35.95	22.60	10.85	5.08	2.80
	STSS	**42.25**	**32.50**	**15.31**	**4.53**	**1.20**	**0.43**
GFR	UBD	45.52	39.86	26.98	15.49	7.77	3.47
	ATS	46.85	43.45	32.30	21.53	14.48	8.68
	STSS	**43.39**	**37.59**	**25.71**	**13.93**	**5.46**	**2.08**

6 结 论

隐写操作在对数字图像进行修改的同时也改变了图像在特征空间中的位置,不同图像的特征在隐写作用下运动模式也会不同.以往的隐写分析工作没有充分挖掘训练数据与测试样本之间的关系,也没有充分利用测试样本本身的信息来进行分析.本文提出了“特定测试样本隐写分析(STSS)框架”,研究了影响隐写分析的两个重要因素——“特征距离”与“特征运动模式”相似度,基于这两个因素,在训练数据库中针对每个测试样本选择专用的训练集训练分类器,很大程度上解决了隐写分析中的 CSM 问题.实验结果表明,本文方法进一步挖掘了训练数据库的分析潜力,有效利用了测试样本的信息,在特定测试样本的隐写分析场景中表现出的性能优于其他方法.

当拥有大数据训练资源时,我们就有条件对特定测试样本筛选更匹配的训练集,所以,STSS 框架适用于设计大数据环境下的精准隐写分析系统.但本文对 STSS 框架仅进行了初步探索,未来还有许多问题需要进一步研究.

(1) 本文假设隐写算法与嵌入率已知,下一步可以尝试对常见隐写修改模式,如“LSB 替换”和“加减 1”进行未知嵌入率的隐写分析;

(2) 本文目前只对经典的非自适应隐写算法 nsF5 作了分析,设计了两种训练集筛选特征,取得了初步实验效果.对于使用 STC 框架的自适应隐写算法,未来可以考虑融合其他特征(如纹理特征)进行深入研究,设计更好的训练集筛选方法,以进一步拓宽 STSS 框架的应用范围,完善其性能.

References:

[1] Chen, K. Information hiding, digital watermarking and steganography. In: Encyclopedia of Multimedia Technology & Networking. 2005. 382–389. [doi: 10.4018/978-1-59140-561-0.ch055]

[2] Li B, He J, Huang J, Shi YQ. A survey on image steganography and steganalysis. Journal of Information Hiding and Multimedia Signal Processing, 2011,2(2):142–172.

[3] Wang SZ, Zhang XP, Zhang KW. Digital Steganography and Steganalysis: Information Warfare Technology in the Internet Age. Beijing: Tsinghua University Press Co., Ltd, 2005 (in Chinese).

[4] Bender W, Gruhl D, Morimoto N, Lu A. Techniques for data hiding. IBM Systems Journal, 1996,35(3.4):313–336.

[5] Sharp T. An implementation of key-based digital signal steganography. In: Information Hiding. Berlin, Heidelberg: Springer-Verlag, 2001. 13–26. [doi: 10.1007/3-540-45496-9_2]

[6] Mielikainen J. LSB matching revisited. IEEE Signal Processing Letters, 2006,13(5):285–287. [doi: 10.1109/LSP.2006.870357]

[7] Pevný T, Filler T, Bas P. Using high-dimensional image models to perform highly undetectable steganography. In: Proc. of the Int'l Workshop on Information Hiding. Berlin, Heidelberg: Springer-Verlag, 2010. 161–177. [doi: 10.1007/978-3-642-16435-4_13]

[8] Holub V, Fridrich J. Designing steganographic distortion using directional filters. In: Proc. of the 2012 IEEE Int'l Workshop on Information Forensics and Security (WIFS). IEEE, 2012. 234–239. [doi: 10.1109/WIFS.2012.6412655]

[9] Holub V, Fridrich J. Digital image steganography using universal distortion. In: Proc. of the 1st ACM Workshop on Information Hiding and Multimedia Security. ACM, 2013. 59–68. [doi: 10.1145/2482513.2482514]

[10] Fridrich JJ, Kodovský J. Multivariate gaussian model for designing additive distortion for steganography. In: Proc. of the ICASSP. 2013. 2949–2953. [doi: 10.1109/ICASSP.2013.6638198]

[11] Sedighi V, Fridrich JJ, Cogranne R. Content-Adaptive pentary steganography using the multivariate generalized Gaussian cover model. In: Media Watermarking, Security, and Forensics. 2015. 94090H. [doi: 10.1117/12.2080272]

[12] Sedighi V, Cogranne R, Fridrich J. Content-Adaptive steganography by minimizing statistical detectability. IEEE Trans. on Information Forensics and Security, 2016,11(2):221–234. [doi: 10.1109/TIFS.2015.2486744]

[13] Westfeld A. F5-A steganographic algorithm: High capacity despite better steganalysis. In: Proc. of the 4th Information Hiding Workshop. 2001,2137:289–302. https://link.springer.com/chapter/10.1007/3-540-45496-9_21

[14] Fridrich J, Pevný T, Kodovský J. Statistically undetectable JPEG steganography: Dead ends challenges, and opportunities. In: Proc. of the 9th Workshop on Multimedia & Security. ACM, 2007. 3–14. [doi: 10.1145/1288869.1288872]

[15] Provos N. Defending against statistical steganalysis. In: Proc. of the Usenix Security Symp, Vol. 10. 2001. 323–336.

[16] Sallee P. Model-Based steganography. In: Proc. of the IWDW. 2003,2939:154–167. [doi: 10.1007/978-3-540-24624-4_12]

[17] Sallee P. Model-Based methods for steganography and steganalysis. Int'l Journal of Image and Graphics, 2005,5(1):167–189. [doi: 10.1142/S0219467805001719]

[18] Fridrich J, Goljan M, Soukal D. Perturbed quantization steganography. Multimedia Systems, 2005,11(2):98–107. [doi: 10.1007/s00530-005-0194-3]

[19] Kim Y, Duric Z, Richards D. Modified matrix encoding technique for minimal distortion steganography. In: Information Hiding. 2006,4437:314–327. [doi: 10.1007/978-3-540-74124-4_21]

[20] Solanki K, Sarkar A, Manjunath BS. YASS: Yet another steganographic scheme that resists blind steganalysis. In: Proc. of the Int'l Workshop on Information Hiding. Berlin, Heidelberg: Springer-Verlag, 2007. 16–31. [doi: 10.1007/978-3-540-77370-2_2]

[21] Guo L, Ni J, Shi YQ. Uniform embedding for efficient JPEG steganography. IEEE Trans. on Information Forensics and Security, 2014,9(5):814–825. [doi: 10.1109/TIFS.2014.2312817]

[22] Guo L, Ni J, Su W, Tang C, Shi YQ. Using statistical image model for JPEG steganography: Uniform embedding revisited. IEEE Trans. on Information Forensics and Security, 2015,10(12):2669–2680. [doi: 10.1109/TIFS.2015.2473815]

[23] Wang SZ, Zhang XP, Zhang WM. Recent advances in image-based steganalysis research. Chinese Journal of Computers, 2009, 32(7):1247−1263 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2009.01247]

[24] Kodovsky J, Fridrich J, Holub V. Ensemble classifiers for steganalysis of digital media. IEEE Trans. on Information Forensics and Security, 2012,7(2):432–444. [doi: 10.1109/TIFS.2011.2175919]

[25] Shi YQ, Chen C, Chen W. A Markov process based approach to effective attacking JPEG steganography. In: Proc. of the Int'l Workshop on Information Hiding. Berlin, Heidelberg: Springer-Verlag, 2006. 249–264. [doi: 10.1007/978-3-540-74124-4_17]

[26] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans. on Systems Man & Cybernetics, 1973,3(6):610–621. [doi: 10.1109/TSMC.1973.4309314]

[27] Holotyak T, Fridrich J, Voloshynovskiy S. Blind statistical steganalysis of additive steganography using wavelet higher order statistics. In: Proc. of the IFIP TC-6 TC-11 Int'l Conf. on Communications and Multimedia Security. Springer-Verlag, 2005. 273–274. [doi: 10.1007/11552055_31]

[28] Pevny T, Fridrich J. Merging Markov and DCT features for multi-class JPEG steganalysis. In: Proc. of the Int'l Society for Optics and Photonics on Electronic Imaging. 2007. 650503-13. [doi: 10.1117/12.696774]

[29] Chen C, Shi YQ. JPEG image steganalysis utilizing both intrablock and interblock correlations. In: Proc. of the IEEE Int'l Symp. on Circuits and Systems. IEEE, 2008. 3029–3032. [doi: 10.1109/ISCAS.2008.4542096]

[30] Kodovský J, Fridrich J. Calibration revisited. In: Proc. of the 11th ACM Workshop on Multimedia and Security. ACM, 2009. 63–74. [doi: 10.1145/1597817.1597830]

[31] Pevny T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix. IEEE Trans. on Information Forensics and Security, 2010,5(2):215–224. [doi: 10.1109/TIFS.2010.2045842]

[32] Kodovský J, Fridrich JJ. Steganalysis of JPEG images using rich models. Media Watermarking, Security, and Forensics, 2012, 8303:0A–1.

[33] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Trans. on Information Forensics and Security, 2012, 7(3):868–882. [doi: 10.1109/TIFS.2012.2190402]

[34] Huang W, Zhao XF, Feng DG, Sheng RN. JPEG steganalysis based on feature fusion by principal component analysis. Ruan Jian Xue Bao/Journal of Software, 2012,23(7):1869−1879 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4107.htm [doi: 10.3724/SP.J.1001.2012.04107]

[35] Li F, Zhang X, Chen B, Feng G. JPEG steganalysis with high-dimensional features and Bayesian ensemble classifier. IEEE Signal Processing Letters, 2013,20(3):233–236. [doi: 10.1109/LSP.2013.2240385]

[36] Denemark T, Fridrich JJ, Holub V. Further study on the security of S-UNIWARD. In: Media Watermarking, Security, and Forensics. 2014. 902805. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9028/902805/Further-study-on-the-security-of-S-UNIWARD/10.1117/12.2044803.full?SSO=1

[37] Holub V, Fridrich JJ. Phase-Aware projection model for steganalysis of JPEG images. In: Media Watermarking, Security, and Forensics. 2015. 94090T. [doi: 10.1117/12.2075239]

[38] Holub V, Fridrich J. Low-Complexity features for JPEG steganalysis using undecimated DCT. IEEE Trans. on Information Forensics and Security, 2015,10(2):219–228. [doi: 10.1109/TIFS.2014.2364918]

[39] Song X, Liu F, Yang C, Luo X, Zhang Y. Steganalysis of adaptive JPEG steganography using 2D Gabor filters. In: Proc. of the 3rd ACM Workshop on Information Hiding and Multimedia Security. ACM, 2015. 15–23. [doi: 10.1145/2756601.2756608]

[40] Zheng GH, Feng GR, Yu J, Cheng H, Zhang XP. JPEG steganalysis based on LSB detection and enhanced features. Journal of Applied Sciences, 2016,34(6):670–676 (in Chinese with English abstract).

[41] Tang W, Li H, Luo W, Huang J. Adaptive steganalysis against WOW embedding algorithm. In: Proc. of the 2nd ACM Workshop on Information Hiding and Multimedia Security. ACM, 2014. 91–96. [doi: 10.1145/2600918.2600935]

[42] Denemark TD, Boroumand M, Fridrich J. Steganalysis features for content-adaptive JPEG steganography. IEEE Trans. on Information Forensics and Security, 2016,11(8):1736–1746. [doi: 10.1109/TIFS.2016.2555281]

[43] Denemark T, Sedighi V, Holub V, Cogranne R, Fridrich J. Selection-Channel-Aware rich model for steganalysis of digital images. In: Proc. of the 2014 IEEE Int'l Workshop on Information Forensics and Security (WIFS). IEEE, 2014. 48–53. [doi: 10.1109/WIFS.2014.7084302]

[44] Zhang Y, Liu F, Yang C, Luo XY, Song XF, Lu J. Steganalysis of content-adaptive JPEG steganography based on Gauss partial derivative filter bank. Journal of Electronic Imaging, 2017,26(1):013011. [doi: 10.1117/1.JEI.26.1.013011]

[45] Qian Y, Dong J, Wang W, Tan T. Deep learning for steganalysis via convolutional neural networks. Media Watermarking, Security, and Forensics, 2015,9409:94090J. [doi: 10.1117/12.2083479]

[46] Lubenko I, Ker AD. Going from small to large data in steganalysis. Media Watermarking, Security, and Forensics, 2012,8303: 0M01–0M10. [doi: 10.1117/12.910214]

[47] Kodovský J, Sedighi V, Fridrich JJ. Study of cover source mismatch in steganalysis and ways to mitigate its impact. In: Media Watermarking, Security, and Forensics. 2014. 90280J. [doi: 10.1117/12.2039693]

[48] Lerch-Hostalot D, Megías D. Unsupervised steganalysis based on artificial training sets. Engineering Applications of Artificial Intelligence, 2016,50:45–59. [doi: 10.1016/j.engappai.2015.12.013]

[49] Ker AD, Pevný T. A mishmash of methods for mitigating the model mismatch mess. In: Media Watermarking, Security, and Forensics. 2014. 90280I. [doi: 10.1117/12.2038908]

附中文参考文献:

[3] 王朔中,张新鹏,张开文.数字密写和密写分析:互联网时代的信息战技术.北京:清华大学出版社有限公司,2005.

[23] 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.计算机学报,2009,32(7):1247–1263. [doi: 10.3724/SP.J.1016.2009.01247]

[34] 黄炜,赵险峰,冯登国,盛任农.基于主成分分析进行特征融合的 JPEG 隐写分析.软件学报,2012,23(7):1869–1879. http://www.jos.org.cn/1000-9825/4107.htm [doi: 10.3724/SP.J.1001.2012.04107]

[40] 郑国华,冯国瑞,余江,程航,张新鹏.基于 LSB 检测的 JPEG 隐写分析特征增强方法.应用科学学报,2016,34(6):670–676.

张逸为(1991－),男,吉林省吉林市人,学士,主要研究领域为信息隐藏,人工智能.

俞能海(1964－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为多媒体数据处理与分析、检索,互联网信息检索(社区,标注),数字内容安全(云计算与云计算安全).

张卫明(1976－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为信息隐藏,密文域计算.