软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2012,23(7):1869−1879 [doi: 10.3724/SP.J.1001.2012.04107]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于主成分分析进行特征融合的 JPEG 隐写分析*

黄 炜[1+], 赵险峰[1], 冯登国[1], 盛任农[2]

[1](中国科学院 软件研究所 信息安全国家重点实验室,北京 100190)

[2](北京电子技术应用研究所,北京 100191)

JPEG Steganalysis Based on Feature Fusion by Principal Component Analysis

HUANG Wei[1+], ZHAO Xian-Feng[1], FENG Deng-Guo[1], SHENG Ren-Nong[2]

[1](State Key Laboratory of Information Security, Institute of Software, The Chinese Academy of Sciences, Beijing 100190, China)

[2](Beijing Institute of Electronic Technology and Application, Beijing 100091, China)

+ Corresponding author: E-mail: weihuang@is.iscas.ac.cn

Huang W, Zhao XF, Feng DG, Sheng RN. JPEG steganalysis based on feature fusion by principal component analysis. *Journal of Software*, 2012,23(7):1869−1879 (in Chinese). http://www.jos.org.cn/1000-9825/4107.htm

Abstract: To solve problems in the existing JPEG steganalysis schemes, such as high redundancy in features and failure to make good use of the complementarity among them, this study proposes a JPEG steganalysis approach based on feature fusion by the principal component analysis (PCA) and analysis of the complementarity among features. The study fuses complementary features to reflect the statistical differences between cover and stego signals in the round, isolates redundant components by PCA, and finally achieves the goal of improving accuracy. Experimental results show that in various datasets and embedding rates, this scheme provides more accuracy than the main JPEG steganalysis schemes against steganographic methods of high concealment (e.g. F5, MME and PQ) and greatly reduces the time cost of the existing fusion methods on feature level.

Key words: feature fusion; PCA (principal component analysis); feature reduction; steganalysis

摘 要: 为了解决现有 JPEG 隐写分析方法特征冗余度高和未能充分利用特征间互补关系的问题,提出了一种基于主成分分析(principal component analysis,简称 PCA)进行特征融合的 JPEG 隐写分析方法,并分析所选特征之间的互补性.通过融合将互补特征结合在一起,更全面地反映载体和隐写信号间的统计差异,并用 PCA 分离出冗余成分,最终达到进一步提升准确率的目的.实验结果表明,在不同数据集和嵌入率情况下,该方法分析高隐蔽性隐写(如 F5,MME 和 PQ)的准确率高于主要 JPEG 分析方法,在耗时上较现有特征层融合降维方法大为缩短.

关键词: 特征融合;主成分分析;特征降维;隐写分析

中图法分类号: TP309 **文献标识码**: A

随着计算机和多媒体技术的发展和普及,信息隐藏技术得到国内外的广泛研究.作为信息隐藏重要分支之一,隐写(steganography)及其分析近年来发展迅速.隐写是一种保密通信的方法,它在载体数据中嵌入秘密信息

* 基金项目: 国家自然科学基金(61170281); 北京市自然科学基金(4112063); 中国科学院战略性先导专项课题(XDA06030601); 中国科学院信息工程研究所创新课题(Y1Z0041101,Y1Z0051101)

收稿时间: 2011-02-27; 定稿时间: 2011-08-31

© 中国科学院软件研究所 http://www.jos.org.cn

以隐藏通信事实.相应地,隐写分析(steganalysis)[1]是指对截获的信号进行分析以判断其是否含有秘密信息的技术.JPEG 图像作为最常见的多媒体之一,其隐写及相关分析受到了较多的重视.

目前,JPEG 图像的隐写方法多为在 JPEG 量化系数上进行最低有效位(least significant bits,简称 LSB)替换(LSB replacement)或 LSB 匹配(LSB matching).普通的 LSB 算法隐蔽性较差,但其衍生算法通过密钥控制和自适应等技术选择嵌入位置来降低隐写扰动量.这增加了隐写分析的难度,因而具有较高的隐蔽性.F5[2]通过矩阵编码降低嵌入秘密信息所需的扰动位置数量,从而提高嵌入效率.修改的矩阵编码(modified matrix encoding,简称 MME)[3]在 F5 的基础上寻找失真最小的位置组合进行嵌入,进一步降低扰动量.扰动量化(perturbed quantization,简称 PQ)[4]选取 JPEG 量化过程中失真较小的系数进行嵌入,达到减小扰动量的目的.湿纸编码(wet paper codes)[5]结合差错控制编码方法将嵌入信息分散到满足给定条件的系数中,如量化失真最小原则.

隐写分析方法可分为专用分析方法和通用分析方法.专用隐写分析方法是针对专门的隐写算法或工具所设计的分析方法;而通用分析算法是分析一类或多类隐写算法的方法,因而日益受到研究者的重视.通用分析算法多从信号处理的角度提取待测样本信号特征作为判决依据.在常见的 JPEG 通用分析方法中,准确率较高的有基于 Markov 过程(Markov process based,简称 MPB)的特征[6]、偏序 Markov 模型(partially ordered Markov model,简称 POMM)特征[7]和基于校准的特征集[8]等.MPB 特征计算了 JPEG 量化系数在各个方向的一阶转移概率矩阵(transition probability matrix,简称 TPM).然而,该方法没有采用校准技术,准确率相对较低.此外,MPB 特征冗余度高,绝对值小于 T 的 JPEG 量化系数所形成的特征向量已达 $4(2T+1)^2$ 维.POMM 方法统计 JPEG 量化系数差值在各种构成情况下的概率,但该方法未考虑量化系数差值之间的关系,其准确率还可以继续提升.基于校准技术特征集的方法结合了 7 种隐写分析特征并采用图像校准技术,准确率相对较高.但该组合并不完整,如果将该方法融合其他互补特征,则准确率仍有提升的空间.一方面,上述 3 组特征侧重点不同,互为补充.例如:POMM 和 MPB 特征在 Markov 模型下的统计对象不同,可以相互补充;PEV274 的全局特征部分和作为局部特征的 POMM 之间也有互补关系.另一方面,上述特征之间来自相近的模型,因而具有一定的冗余度.加之特征内部各维之间也存在着相关性,这些冗余对于分类效果贡献微弱,却需要额外的运算量.

针对以上特征冗余性和互补性的问题,研究人员已经开始在隐写分析中采用特征融合技术.Kodovský 等人[9]提出了隐写分析完备特征集(complete feature set)的概念,即对任何载体图像 $c\in C$,完备特征集 $\boldsymbol{x}$ 内的任意特征 x_i 均将其判定为载体图 $x_i(c)=0$.这说明不同的特征都有助于判断隐写的存在性,在隐写分析中有相互补充的作用.虽然当前对一个完备的隐写分析特征集需要包含哪些特征尚未有文献明确论述,但将不同特征进行融合的方法已逐渐得到了研究人员的重视.现有的融合方法分为 3 个层次[10]:数据层融合、特征层融合和信息层融合.其中,特征层既能保留参与融合的特征之间的有效信息,又能去除特征冗余,对提高隐写分析分类准确度具有重要意义.然而,现有基于特征层融合的隐写分析在特征的优选和分析效率方面仍然存在一些不足.Dong 等人[10]提出了一种基于 Boosting 算法的特征融合方法——提升特征选择(boosting feature selection,简称 BFS)法.该方法利用 Boosting 算法从弱分类器构造强分类器过程中所赋的权值对特征进行取舍,提高了分析准确率.然而,该方法并未综合考虑特征之间的关系,舍弃部分特征维度的同时也丢失了一些有效成分;而且 BFS 对特征进行遍历,每步搜索都要做分类,构造强分类器的过程耗时较长.Miche[11]提出了一种基于后向搜索的隐写分析方法,该方法需要 $N(N+1)/2$ 倍的训练和分类时间.Dash 等人[12]指出,以分类结果为准则的方法效率较低,而且在通用性方面表现较弱,即以一种分类器的分类结果作准则的权值在其他类型分类器上并非最优.在信息层,Kharrazi 等人[13]提出了一种基于分类结果指标均值或最大值的特征融合隐写分析方法.该方法在信息层进行组合,同样未考虑特征之间的冗余和互补关系.近年来,PCA 作为分类器的预处理过程逐渐用在图像和音频隐写分析领域[14,15],取得了较好的成效.

为了更好地综合现有特征的长处,进一步提高隐写分析的准确率和运行效率,本文提出了一种基于主成分分析(principal component analysis,简称 PCA)进行特征融合的通用 JPEG 隐写分析方法.该方法将当前分类准确度较高的特征进行组合,利用 PCA 削弱原始特征之间的相关性,达到了特征融合及优选的效果.实验结果表明,该方法处理后的特征性能稳定,主要表现在准确率相比融合前的分析方法有不同程度的提升.同时,经 PCA 处理

© 中国科学院软件研究所 http://www.jos.org.cn

后的特征降维效果稳定,准确率在特征维度降低至 1/3 时并无明显下降,下降幅度不超过 0.2%.与现有特征层融合的隐写分析相比,本文方法的准确率在大多数情况下提高 1%~4%;与现有信息层融合的隐写分析相比,本文方法的准确率在大多数情况下提高 1%~3%,耗时方面基本持平.

本文第 1 节介绍主要的 JPEG 隐写分析方法以及基于特征融合的隐写分析方法.第 2 节介绍本文方法的隐写分析方案,并分析本文选用特征之间的互补性.第 3 节是实验及其分析,包括将本文方法和现有方法对比,体现本文方法在隐写分析的准确率、融合降维的稳定性和运行效率等方面的优势.第 4 节总结全文.

1 主要 JPEG 隐写分析方法和特征融合的运用

1.1 主要的JPEG隐写分析方法

现有 JPEG 隐写通用分析方法主要有以下几种:

(1) Shi 等人[6]提出的基于 Markov 过程模型的 MPB 分析方法.该方法提取 JPEG 量化系数矩阵,对其按水平、垂直、主对角和副对角 4 个方向求差,得到 4 个差值矩阵.然后,计算各个差值矩阵在相应方向上的一阶 TPM.该方法只计算$[-T,T]$范围内的系数,最终得到 $4\cdot(2T+1)^2$ 维特征向量;

(2) Davidson 等人[7]提出的基于偏序 Markov 模型 POMM 的分析方法.该方法在水平、垂直、主对角和副对角 4 个方向上计算满足同一差值 d 的相邻 JPEG 量化系数$\{(c_1,c_2)|c_1-c_2=d\}$可能组合概率,其中,只统计 $c_1,c_2\in[-T,T]$的情况.接着,对这 4 个方向的概率求和,得到$(2T+1)^2$ 维特征向量.最后,对图像裁剪最外层 4 行 4 列并重新 JPEG 压缩作校准,校准前后的块内和块间特征差值为最终特征;

(3) Kodovský 等人[8]基于校准技术提出的 PEV 分析方法.该方法使用了一组特征集,包括各种直方图、方差、分块特性、共生矩阵和 Markov 特征等,分别计算:亮度部分的 DCT 系数直方图矩阵 $\boldsymbol{H}$;AC 系数直方图矩阵 $\boldsymbol{h}^{ij},(i,j)\in\{(1,2),(2,1),(3,1),(2,2),(1,3)\}$;双直方图矩阵 $\boldsymbol{g}_{ij}^{d}$;方差 V;块内分块特性 B_α;共生矩阵 $\boldsymbol{C}_{st}$;Markov 模型转移概率矩阵 $\boldsymbol{M}_{m,n}$.最后,对图像最外层四周各裁剪 4 行 4 列后重新 JPEG 压缩作为校准,校准前后的特征 F_r,F_c 各为 274 维.本文将 F_r-F_c 称为 PEV274,$\{F_r,F_c\}$称为 PEV548.

1.2 已有的基于融合的隐写分析方法和局限

Dong 等人[10]提出一种 BFS 的隐写分析方法,该方法结合交叉验证(cross validation,简称 CV)结果将一些弱分类器按照准确率的高低赋予相应的权值,组成一个强分类器.该方法通过删除权值较小的特征维度来实现降维.但是,以分类结果为准则的分类算法对分类器的通用性较弱而且效率较低[12],使用一个分类器构造的权值,对另一个分类器未必适用.加之该方法并未充分考虑特征内部的冗余性与互补性,因而优选过程中的稳定性不足.本文通过对照实验发现,该方法的准确率受降维的影响较大,降维后维度越低,准确率下降得越明显.如对 F5 0.05bpac 的分析,其特征降维至 1/3 时,准确率下降约 4%,其他情况约下降 2%.

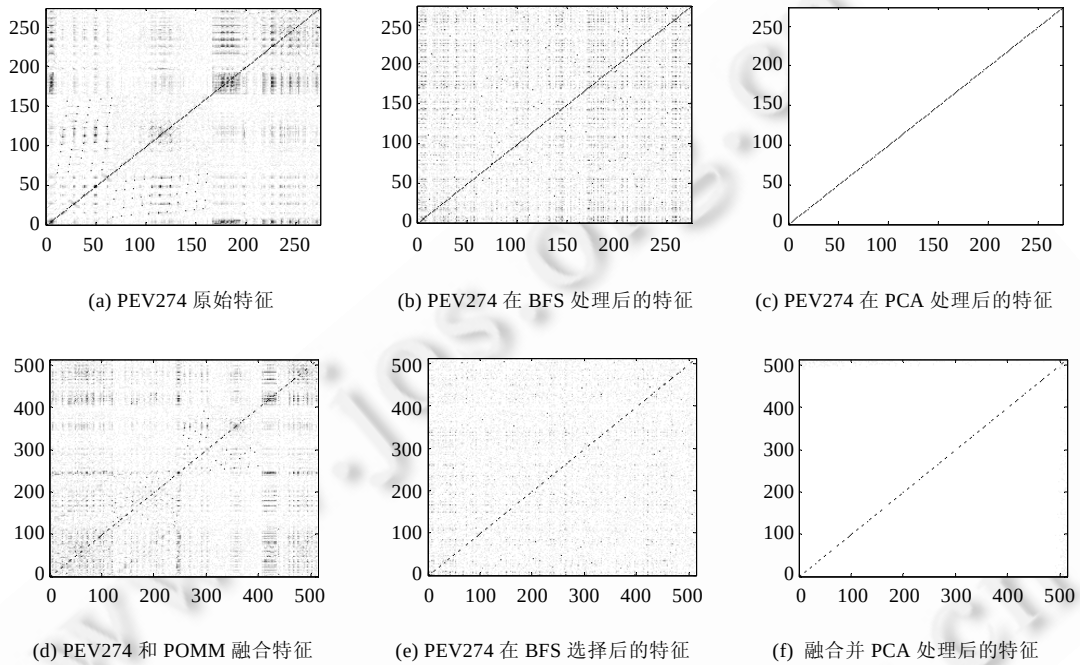
Miche 等人[11]的后向搜索方法从全部特征维度开始,遍历当前特征删去任意 1 维度的情况,将准确率下降最小的特征维度删除.接着,对剩下特征迭代计算,直至删除任意特征维度时准确率下降均超过容忍值.该方法的原理类似于 BFS,因而也存在上述分类器通用性弱和效率低下[12]等问题.

Kharrazi 等人[13]提出一种基于分类结果的决策层信息融合方法.该方法采用均值(mean)和最大值(max)两个准则,对各个分类器的结果作综合决断,即对各分类器的分类概率结果取均值或最大值作为新的分类结果.该方法的各个分类器之间独立性较强,对互补性的利用程度不足,而且容易受到错误率较高的分类器的影响.本文通过对照实验发现,“均值”融合效果远优于“最大值”分类效果,但一般低于特征层融合的分析准确率约 0.5%~3%.另外,本文方法与该方法融合层次不同,可以结合使用,即把训练器参数相近的部分特征用于特征层融合,训练器参数不同的特征用于决策层融合.

我们用自相关系数矩阵来衡量特征间的冗余性.图 1 中横轴和纵轴表示特征维度,每个点(i,j)的灰度值 $h(i,j)$为自相关系数矩阵的绝对值,即第 i 维特征与第 j 维特征之间的相关系数值.图 1(a)表示 700 对图像样本提取 PEV274[8]原始特征的自相关系数矩阵.可以看出,原始特征自相关系数矩阵除对角线外还存在多处系数较大的

© 中国科学院软件研究所 http://www.jos.org.cn

情况.经过 BFS 处理后的特征如图 1(b)所示,BFS 只是对特征按照单个维度的可区分度进行排序,其自相关系数只是重新排列,并不改变数值大小.优选后的自相关系数矩阵,即是图 1(b)左下角的某个区域.无论降维程度如何,优选后的特征还是有一定的相关性.类似地,图 1(d)表示 700 对图像样本提取 PEV274 原始特征合并 POMM[7] 特征的自相关系数矩阵,其中,(0,0)到(274,274)的区域是 PEV274 特征的自相关系数,(275,275)到(514,514)的区域是 POMM 特征的自相关系数,其余区域为 PEV274 和 POMM 两种特征间的相关系数.可见,上述 3 个区域都存在很大的线性相关.经过 BFS 处理后的特征如图 1(e)所示,3 个区域中较大的系数仍较明显地分布在整个图中.由此可见,BFS 方法难以有效降低特征冗余度.此外,决策层融合在特征层面并未作任何处理,其特征间相关性等同于原始特征相关性.



(a) PEV274 原始特征 (b) PEV274 在 BFS 处理后的特征 (c) PEV274 在 PCA 处理后的特征

(d) PEV274 和 POMM 融合特征 (e) PEV274 在 BFS 选择后的特征 (f) 融合并 PCA 处理后的特征

Fig.1 Autocorrelation coefficient of features before and after the process of BFS and PCA

图 1 BFS 和 PCA 处理前后特征自相关系数图

2 基于 PCA 特征融合的 JPEG 隐写分析

如上所述,现有基于融合的隐写分析方法未能较好地处理特征冗余性问题,增加了计算开销;在一定的计算能力和性能要求下,也妨碍了融合其他具有补充作用的特征.为了解决该问题,本文提出了一种基于 PCA 特征融合的 JPEG 隐写分析,融合 PEV(含 PEV274 和 PEV548),POMM 和 MPB 这 3 组特征,分析特征之间的互补关系以及特征内部和特征间的冗余性.

2.1 PCA简介

PCA,又称为 K-L 变换(Karhunen-Loéve transformation)[16],是信号处理中常用的变换之一.其目的是产生最优不相关特征,因而具有很好的信号压缩效果.令 $\boldsymbol{x}=\{x_1,x_2,...,x_{n_x}\}$ 为 n_x 维输入样本向量,PCA 处理的目的即是生成 $\boldsymbol{y}=\boldsymbol{A}^T\boldsymbol{x}$,满足 $E[y_iy_j]=0,i\neq j$.如果 $\boldsymbol{x}$ 是归一化后的矩阵,即 $E[\boldsymbol{x}]=0$,则 $E[\boldsymbol{y}]=0$,从而可得协方差矩阵:

$$\boldsymbol{R}_y=E[\boldsymbol{y}\boldsymbol{y}^T]=E[\boldsymbol{A}^T\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{A}]=\boldsymbol{A}^T\boldsymbol{R}_x\boldsymbol{A} \quad (1)$$

其中,$\boldsymbol{R}_*$表示协方差矩阵.对于多个样本的训练集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_{n_X}\}$,训练集协方差矩阵为

© 中国科学院软件研究所 http://www.jos.org.cn

$$\boldsymbol{R}_x \approx \frac{\sum_k (\boldsymbol{x}_k \boldsymbol{x}_k^T)}{n_X} \tag{2}$$

由于 $\boldsymbol{R}_x$ 是对称的,其本征向量互为正交.将 $\boldsymbol{R}_x$ 的正交本征向量 $\{\boldsymbol{a}_1,\boldsymbol{a}_2,...,\boldsymbol{a}_{n_x}\}$ 组成转换矩阵 $\boldsymbol{A}$,则 $\boldsymbol{R}_y$ 是对角矩阵,即满足最优不相关的要求.一般地,对 $\boldsymbol{y}$ 中的维度按方差降序排列,其中,相对靠前的维度方差较大,成为主成分(principal component,简称 PC).

2.2 所选特征的互补性和冗余性分析

本文对 PEV(含 PEV274 和 PEV548),POMM 和 MPB 这 3 组特征进行融合和主成分优选.下面我们从互补性的角度阐述选择它们的理由,并从冗余性的角度阐述使用 PCA 的原因.

首先,本文选取的 3 组特征之间存在较显著的互补性,因此,融合这些特征可以提高分析的准确率.这主要表现在以下两个方面.

(1) 局部特征和整体特征的互补

局部特征只关心局部变化情况,其统计样点数量较少;而整体特征统计量较大.因而,局部特征相较于整体特征而言有统计不稳定的缺点,却具备不易被修复的优点.有些隐写算法在局部扰动不明显,如 F5 遇到系数 0 和 1 则重新嵌入,其隐写过程中产生了新的 0 系数,仅从局部特征考察变化并不全面,需要配合整体特征.例如,MPB 特征计算了系数差值的一阶 TPM,属于局部特征,其中纵向的计算公式如下:

$$\boldsymbol{M}_{n,m}^h = \Pr[\boldsymbol{F}_{u+1,v}^h = n \mid \boldsymbol{F}_{u,v}^h = m] = \frac{\sum_{u,v} \delta[\boldsymbol{F}_{u,v}^h = m, \boldsymbol{F}_{u+1,v}^h = n]}{\sum_{u,v} \delta[\boldsymbol{F}_{u,v}^h = m]} \tag{3}$$

其中,$\boldsymbol{F}_{u,v}^h = \boldsymbol{A}_{u,v} - \boldsymbol{A}_{u+1,v}$,$\boldsymbol{A}_{u,v}$ 为位置(u,v)的 JPEG 量化系数.对于 F5 隐写,该特征表现为 m=0 处转移概率有所增加.一些整体特征,如 PEV274 维整体系数直方图 $\boldsymbol{H}$,$H_l \in [-5,5]$可以更明显地反映隐写后 0 系数增长的趋势.另外,在统计上一阶 TPM 属于二阶概率分布,直方图属于一阶概率分布,二者统计独立性较强.有些算法在控制整体特征扰动上采取了一些修补技术,如 PQ 隐写优先选择符合:

$$A_{u,v} - \lfloor A_{u,v} \rfloor \in [0.5-\varepsilon, 0.5+\varepsilon] \tag{4}$$

即小数部分很靠近 0.5 的 DCT 系数,以尽量减小全局误差,但该方法难以保持或修复局部特征变化.本文选择的 PEV 特征中的亮度、直方图、方差和分块特性等特征属于全局特征,共生矩阵、Markov 过程部分以及 MPB 特征属于局部特征.POMM 特征,计算满足同一差值 $A_{u,v}-A_{u,v+1}=d$ 的各种系数组合$(A_{u,v},A_{u,v+1})$概率,是一种更细部的特征.

(2) 特征分布模型不同

源自不同模型的特征,反映隐写扰动的改变不同,特征相关性也较低.高隐蔽性隐写算法常修改扰动量较小的系数位置或者结合编码学选择较少的嵌入位置.从单一分布模型来看,这样的改变相对微弱.简单特征容易被隐写算法设计者修复,复杂特征的修复往往很难实现.加之隐写行为在不同模型下特征值改变程度不同,要兼顾不同分布模型,达到每个模型下的扰动量最小,更是困难重重.因此,将不同模型的特征融合在一起,可以增大隐写扰动在特征维度上的改变量,因而具有互补性.例如,MPB 将 JPEG 量化系数差值视为一阶 Markov 过程,对其计算 TPM,如公式(3)所示.而 POMM 将 JPEG 量化系数差值视为偏序 Markov 过程,在纵向上计算各系数差值的组成情况而非差值的 TPM:

$$\boldsymbol{P}_{n,m}^h = \Pr[(\boldsymbol{A}_{u,v}, \boldsymbol{A}_{u+1,v}) \mid \boldsymbol{F}_{u,v}^h = m] = \frac{\sum_{u,v} \delta[\boldsymbol{A}_{u,v} = m+n, \boldsymbol{A}_{u+1,v} = m]}{\sum_{u,v} \delta[\boldsymbol{F}_{u,v}^h = m]} \tag{5}$$

PEV 特征则基于校准技术组合了 7 种特征.对图像最外层四周各裁剪 4 行 4 列后重新 JPEG 压缩,得到校准图像.该方法认为校准图像能够保持载体图像的统计特性.

其次,本文选择的 3 组特征,无论特征与特征之间还是特征内部维度之间都存在不同程度的冗余性.因此,使用 PCA 压缩特征并降低冗余很有必要.这主要表现在以下两点:

© 中国科学院软件研究所 http://www.jos.org.cn

(1) 特征内部冗余性

现有通用分析特征往往来自于单一手段,特征内部不同维度之间的分布并非相互独立.例如,TPM 中相近的系数之间转移概率大,差异较大的系数之间转移概率小.如我们对 700 对自然图像进行统计发现,公式(3)中|*m*−*n*|较小时, $\boldsymbol{M}_{n,m}^{h}$ 较大,|*m*−*n*|较大时, $\boldsymbol{M}_{n,m}^{h}$ 较小.

(2) 特征间冗余性

现有各种通用隐写分析方法在特征设计上有类似之处,如:PEV[8]中包含了 Markov 过程 TPM 计算,等于对 MPB 所统计的 TPM 求均值并校准;而 MPB 特征与 POMM 特征原理类似,属于二阶概率分布的计算,也存在一定的线性关系.

图 1(a)表示 700 对图像样本提取 PEV274 原始特征的自相关系数矩阵,除对角线外还存在多处系数较大的情况;而经过 PCA 处理后的特征如图 1(c)所示,只有对角线上自相关系数较大,其余各处自相关系数很小.可见,PCA 很好地去除了特征内部的线性相关性.图 1(d)则表示 700 对图像样本提取 PEV274 原始特征合并 POMM 特征后的自相关系数矩阵.可见,上述 PEV274 特征和 POMM 特征的内部以及两种特征间都存在线性相关.而经过 PCA 处理后的特征如图 1(f)所示,线性相关性大为降低.

由于 PCA 融合降维去除的是相关性较小的成分,不是相关性较大的维度,删除其中若干维对其他主成分的可区分度影响很小,降维不会带来准确率的明显下降.

2.3 分析方法

本文提出的基于 PCA 进行特征融合的 JPEG 分析方法框架如图 2 所示,提取特征的具体步骤如下:

(1) 特征提取:对给定的图像分别用 PEV,POMM 和 MPB 等方法提取特征,得到 PEV(含 PEV274 和 PEV548),POMM 和 MPB 等特征集;

(2) 特征合并:将所提取的特征按顺序组合为特征集 $\boldsymbol{f}=\{\boldsymbol{f}_1,\boldsymbol{f}_2,\ldots,\boldsymbol{f}_\alpha\}$,对于多个样本的训练集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$,得到 n 行 α 列的矩阵 $\boldsymbol{F}=\{\boldsymbol{f}_1,\boldsymbol{f}_2,\ldots,\boldsymbol{f}_\alpha\}$,其中,$\boldsymbol{f}_i$ 为 n 个不同的训练样本第 i 维特征值组成的向量,$\boldsymbol{f}_i=\{f_{i,1},f_{i,2},\ldots,f_{i,n}\}^T$;

(3) 预处理(归一化处理):对给定的 n 个训练样本,对每个特征维度求期望值 $\boldsymbol{\mu}=\{\mu_1,\mu_2,\ldots,\mu_\alpha\}$ 和标准差 $\boldsymbol{s}=\{s_1,s_2,\ldots,s_\alpha\}$,其中,对每个 $i\in\{1,2,\ldots,\alpha\}$,有

$$\mu_i=\frac{\sum_{j=1}^{n}f_{i,j}}{n},s_i=\sqrt{\frac{\sum_{j=1}^{n}(f_{i,j}-\mu_i)^2}{n-1}} \tag{6}$$

然后,按下式进行归一化:

$$g_{i,j}=\frac{f_{i,j}-\mu_i}{s_i} \tag{7}$$

得到归一化后的特征矩阵 $\boldsymbol{G}=\{\boldsymbol{g}_1,\boldsymbol{g}_2,\ldots,\boldsymbol{g}_\alpha\}$;

(4) 特征变换(PCA 处理):利用第 2.1 节所述的方法求得转换矩阵 $\boldsymbol{A}$,将 $\boldsymbol{G}$ 转换到变换域上,得到不同维度(即主成分)间相关性较小且按各列方差降序排列的矩阵 $\boldsymbol{H}=\boldsymbol{F}\cdot\boldsymbol{A}$,$\boldsymbol{H}$ 为 n 行 α 列,则特征 $\boldsymbol{f}$ 转换为主成分 $\boldsymbol{h}=\{h_1,h_2,\ldots,h_\alpha\}$.合适的转换矩阵 $\boldsymbol{A}$ 可以通过一次训练过程得到;

(5) 降维:将 $\boldsymbol{H}$ 的前 d 维主成分(即前 d 列数据)共 n 行 d 列的矩阵作为本文方法得到的特征,其中,d 可以通过交叉验证确定,也可以取经验值 $d=\alpha/3$.

在训练过程中,分析者提取出经过 PCA 融合降维的特征,用支持向量机(support vector machine,简称 SVM)训练得到训练结果,并记录 $\boldsymbol{\mu}$、$\boldsymbol{s}$、转换矩阵 $\boldsymbol{A}$ 和降维维数 d,以供分类过程使用.在分类过程中,分析者对待分析图像提取同样的特征,用同样的转换矩阵做对应的变换和降维得到特征向量,并用 SVM 得到最终的分类结果.

© 中国科学院软件研究所 http://www.jos.org.cn

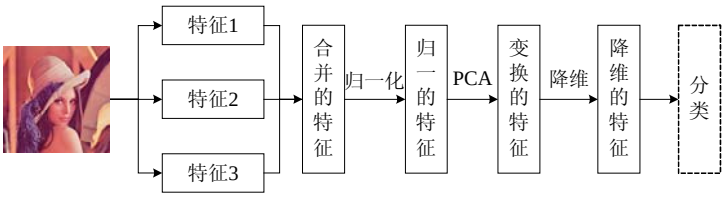


Fig.2 Framework of JPEG steganalysis approach based on feature fusion by PCA

图 2 基于 PCA 进行特征融合的 JPEG 分析方法框架图

3 实验及结果分析

第 3.1 节介绍实验条件和实验参数.第 3.2 节验证特征融合隐写分析性能的设计稳定性,即分析特征融合方法对隐写分析准确率具有提升作用.第 3.3 节验证 PCA 特征降维对准确率的影响,即本文方法在特征降维时不会对正确率造成大的波动,也是 PCA 主成分降维优于现有特征降维之处.第 3.4 节将特征融合结果与融合前作对比,验证特征融合对隐写分析正确率的提高有促进作用,比现有主要分析方法的正确率要高.第 3.5 节将本文方法与其他基于融合的隐写分析方法进行对比,验证本文方法优于其他方法.

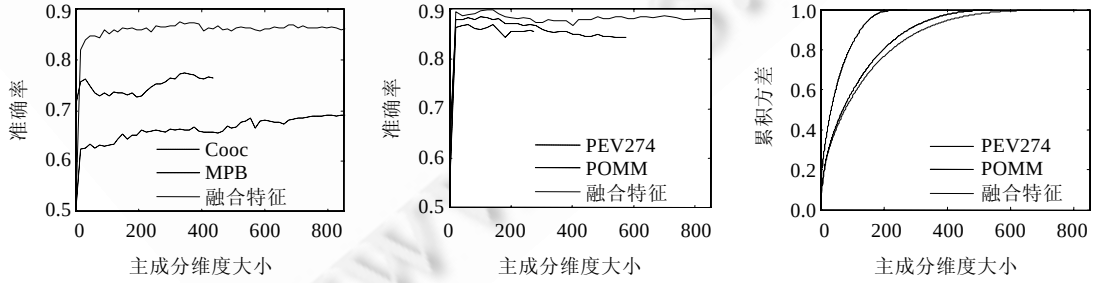
3.1 实验条件和参数

为排除不同编码器差异[17]和双重 JPEG 压缩[18]等因素对准确率的统计影响,本文选取的图像源为 BOSS v0.92 图像库(10 074 幅中随机选取 4 000 幅)[19]和 UCID 图像库(1 338 幅)[20].实验制备的阳性集为 F5 Release 11[2], MME3[3](在 F5 的基础上改进)和 PQ[4]等隐写工具在不同嵌入率下得到的 JPEG 隐写图像;对应阴性集为原始格式使用上述隐写工具所用编码器在相同 JPEG 质量因子下转换得到的 JPEG 图像.其中,BOSS 图库训练样本 3 000 对,测试样本 1 000 对;UCID 图库训练样本 670 对,测试样本 668 对.为了更好地说明各种实验条件情况下 PCA 融合降维的效果,实验选择的嵌入率为 0.05 bpac~0.20 bpac,JPEG 质量因子为 90(PQ 首次压缩质量因子为 80,二次压缩时为 75).

在训练和分类过程,本文使用径向基函数(radial basis function,简称 RBF)支持向量分类(C-SVC)的 LibSVM[21],其中,缩放因子和代价因子等实验参数通过对训练集采用交叉验证的方法确定.

3.2 特征融合隐写分析性能的设计稳定性

特征融合为分类器提供了更多的信息,其对分类结果的影响表现在融合发挥了特征间的互补作用,有助于区分不同的类,因而准确率提升.本文用准确率来研究融合前后分类效果的变化情况,如图 3 所示.图中特征都经过了 PCA 变换,横轴表示选取 PC 的维度,纵轴表示选取指定长度的 PC 下分类的准确率,每幅图的每个样点由 UCID 图像库统计获得.



(a) 融合分类互补性强的特征 (b) 融合分类互补性弱的特征 (c) 图 3(a)的累积方差比例图

Fig.3 Comparison of the impact of the classification of the new fused features

图 3 融合新特征对分类效果的影响比较图

© 中国科学院软件研究所 http://www.jos.org.cn

从图中可以看出:(1) 融合一组互补性较强的特征可以大大提高准确率.如图 3(a)所示,对 UCID 图库及其 F5 0.10bpac 隐写图提取 PEV548 的共生矩阵特征(Cooc)部分和 MPB[6]特征并融合后,融合特征(fusion)的分类效果有很大提高;(2) 即使融合一组分类互补性差的特征,准确率也会有所提升.如图 3(b)所示.同等的隐写条件下,融合后的特征(fusion)相较于融合前的 PEV274 特征和 POMM 特征,其准确率都有一定的提高.

3.3 PCA特征降维对准确率的影响

经过 PCA 处理后的 PC 之间相关性降低,并按方差降序排列.一般而言,方差较大的特征区分度相对较高,对区分能力的贡献相对较大;方差较小的特征区分度较低,对区分能力的贡献相对较小.PCA 处理(即 KL 变换)相当于将原有坐标系投射到各个成分相互独立的坐标系上.

从图 3 可以看出,去掉带有较少信息的坐标轴,可以达到降维的目的.例如,PEV274 特征经 PCA 变换后的特征准确率随维度变化符合上述趋势,如图 3(b)所示.起初,准确率随着 PC 维度的增加而显著上升;到一定程度后(如图 3(a)和图 3(b)中的拐点所示),准确率的上升速度将变慢.这是因为,在拐点之前的特征主成分对该组特征的分类效果起到决定性的作用,而拐点之后的特征对分类效果帮助不大,甚至起干扰作用.融合后的特征综合了参与融合的各组特征的性能,优势互补而显得稳定,在拐点之后,准确率一般趋于平稳,变化不大.如图 3(c)所示,方差也有类似的变化趋势.一般地,当 PC 上的方差占所有特征方差的 70%以上时,准确率已经基本稳定[22].

从图 3(a)和图 3(b)还可以看到,经过融合以后,降维对特征准确率的影响趋于平和.图 3(a)的 Cooc 在约 40 维时准确率达到一个峰值后,随着 PC 的增加,准确率反而降低;到约 200 维以后开始回升,准确率曲线波动较大.图 3(b)的 POMM 则在约 30 维时出现准确率峰值,比 PEV274 还高;而后,随着 PC 的增加,准确率明显降低,甚至低于 PEV274 的准确率.两图中融合后的曲线随着 PC 的维度变化,经过拐点后其准确率波动更为平缓.

3.4 融合前后效果对比

本文按第 2.3 节方法融合了 PEV274,PEV548,POMM 和 MPB 等特征,对比处理前后的准确率变化,其结果见表 1.其中,“融合后”记录了 PCA 融合后在不降维情况下的准确率,“融合降维”记录了 PCA 融合并降维至原特征维度约 1/2 处时的准确率,单个特征最高准确率用粗体表示.

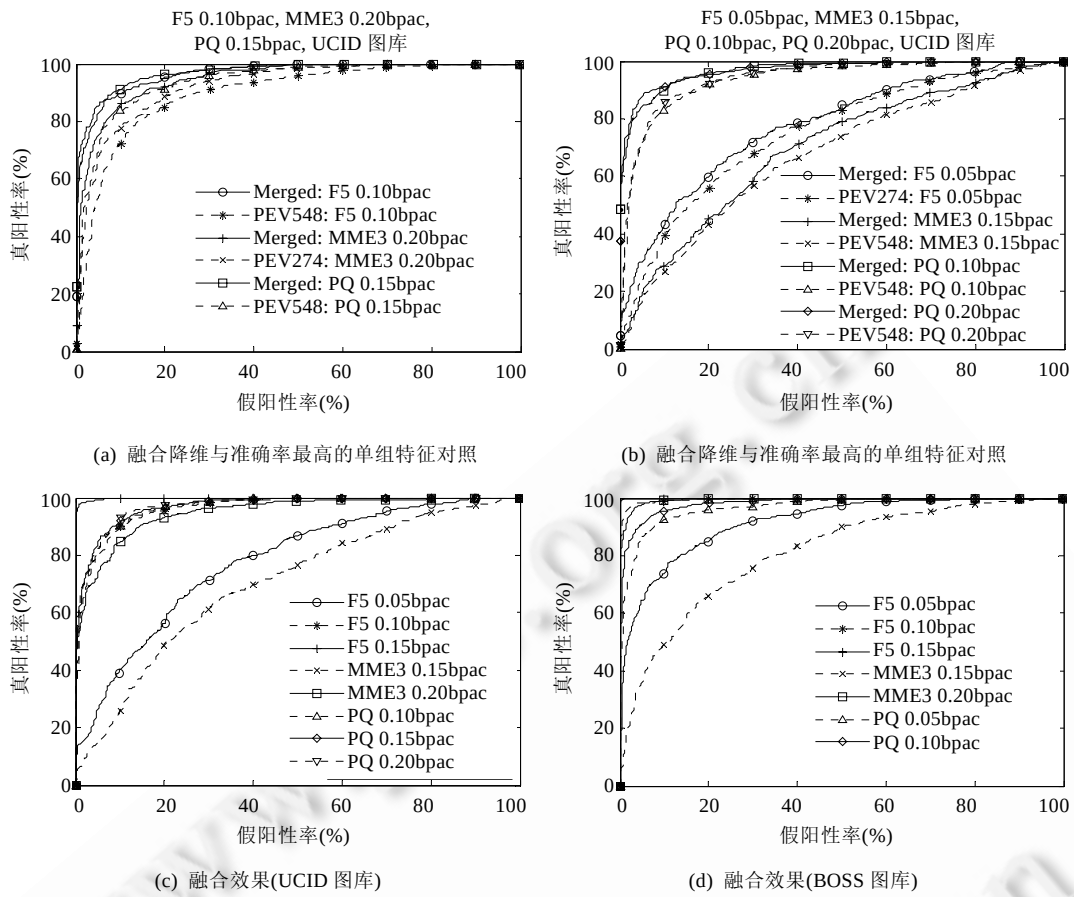
Table 1 Comparison of the effect of our experiment

表 1 本文实验方法效果对照表

图库	隐写算法	嵌入率(bpac)	准确率(%)					
			PEV274	PEV548	POMM	MPB	融合后	融合降维
UCID 图库	F5	0.05	**69.12**	67.47	67.32	57.19	71.13	70.99
		0.10	**85.46**	82.99	84.41	69.12	89.73	89.74
		0.15	**95.58**	93.85	95.12	78.94	98.35	98.35
	MME3	0.15	62.74	**63.64**	60.57	56.37	65.15	64.99
		0.20	**84.48**	83.96	81.18	73.54	87.78	87.86
	PQ	0.10	85.16	**88.53**	77.74	72.41	90.10	89.96
		0.15	86.51	**88.68**	79.47	74.29	90.78	90.70
		0.20	86.36	**88.75**	78.26	71.36	91.07	91.15
BOSS 图库	F5	0.05	**81.70**	81.25	75.30	66.40	83.45	83.10
		0.10	**95.85**	94.80	91.55	80.35	97.25	97.05
		0.15	**99.10**	98.65	97.65	90.70	99.60	99.50
	MME3	0.15	**71.80**	71.20	66.30	62.80	73.40	72.70
		0.20	**94.75**	94.00	91.10	87.40	96.80	96.35
	PQ	0.05	90.15	**90.60**	74.80	72.80	91.80	91.20
		0.10	92.10	**92.70**	83.25	78.30	93.85	92.75

图 4 表示融合效果对照实验的 ROC 曲线.从图表中可以看出,无论训练样本较少的 UCID 图库还是训练样本较多的 BOSS 图库,融合降维后比融合前的单个最高准确率有 1%~4%的提升,而且比较稳定.

© 中国科学院软件研究所 http://www.jos.org.cn



(a) 融合降维与准确率最高的单组特征对照 (b) 融合降维与准确率最高的单组特征对照

(c) 融合效果(UCID 图库) (d) 融合效果(BOSS 图库)

Fig.4 Contrast of ROC of experiments on the fusion effect

图 4 融合效果对照实验 ROC 图

3.5 与其他融合分析方法效果对比

PCA 融合降维分析方法与 BFS[10],Kharrazi[13]方法的融合对照实验结果见表 2.其中,本文方法和 BFS 方法为在融合降维至原始特征 1/3(即 776 维)条件下的准确率与耗时(不包括提取特征的耗时),Kharrazi 方法的实验为 LibSVM 分类器所输出隐写概率均值和最大值作为分析结果情况下的准确率和耗时.每组测试样本及其隐写算法和嵌入率都一致,都采用公开的隐写工具和公开的图像库(1 335 个样本).此外,实验对照的隐写分析算法和训练器都相同,只有特征优选方法不同;耗时为在 Intel Xeon E7420 2.13GHz 和 8GB 内存硬件环境下统计得到的运行时间.结果显示,本文方法优于其他两种算法,并且优势较稳定.

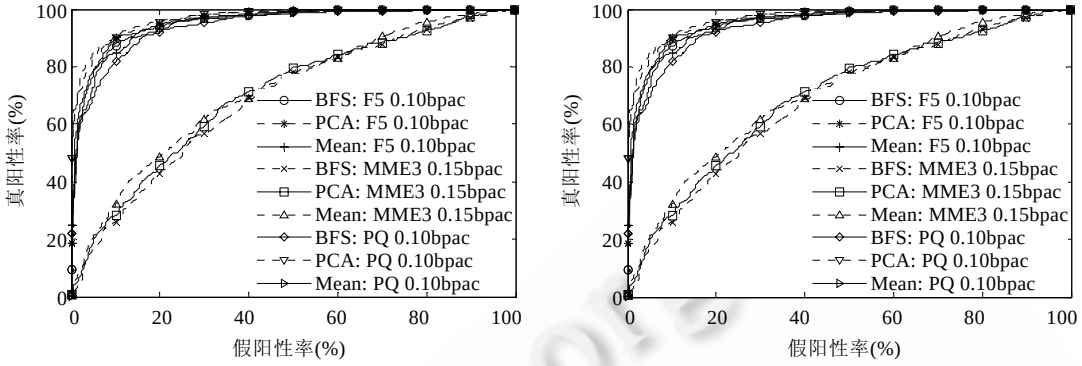
Table 2 Comparison of the effect between the proposed and relevant approaches

表 2 本文方法与相关融合方法效果对照表

隐写算法	嵌入率(bpac)	融合前最高(%)	BFS 方法		Kharrazi 方法			本文方法	
			准确率(%)	耗时(s)	准确率(mean)(%)	准确率(max)(%)	耗时(s)	准确率(%)	耗时(s)
F5	0.05	69.12	67.51	2 272.2	68.56	63.70	54.04	70.99	38.05
	0.10	85.46	88.19	2 284.3	87.05	77.62	35.19	89.74	38.16
MME3	0.15	63.64	63.42	2 277.3	64.90	57.71	27.54	64.99	47.42
	0.20	84.48	86.31	2 261.4	88.92	73.80	33.82	87.86	39.69
PQ	0.10	88.53	87.03	2 570.8	89.15	73.50	38.53	89.96	34.32
	0.15	88.68	91.52	2 513.1	90.27	75.60	35.47	90.70	35.29

© 中国科学院软件研究所 http://www.jos.org.cn

融合效果对照实验的 BFS、Kharrazi 的 mean 方法和本文方法 ROC 曲线图如图 5 所示.从图表中可以看出,本文 PCA 融合降维法比 BFS 方法具有更好的区分能力.这是因为,PCA 对特征整体作变换而非单独考察每一维特征,去除冗余特征的同时也丢失了对分类有用的部分.同时,由于 PCA 计算相对于 BFS 复杂度要低,故耗时大为缩短.另外,Kharrazi 方法在信息层进行融合,未能充分发挥特征间的互补性,故融合对准确率的提升效果并不稳定,例如在 F5 0.05bpac 的情况下,mean 和 max 准则融合后准确率都有不同程度的下降.



(a) 第 1 组 (b) 第 2 组

Fig.5 Contrast of ROC of experiments on the fusion effect between proposed approach and related approach

图 5 本文方法与相关方法融合效果对照实验 ROC 图

4 结 论

本文提出了一种基于 PCA 进行特征融合的 JPEG 隐写分析方法,从现有的各种隐写特征中选择 3 种具有互补性质的特征,将它们组合在一起,通过 PCA 分离出冗余成分,最后得到含有有效成分的融合特征.实验结果表明,PCA 融合降维方法对现有的高隐蔽性 JPEG 隐写算法的分类准确率优于现有的 JPEG 隐写分析方法平均约 5%,优于现有融合分析方法平均约 2%;且相较于现有特征层融合分析方法,耗时大为缩短.

References:

[1] Pevný 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]

[2] Westfeld A. F5—A steganographic algorithm: High capacity despite better steganalysis. In: Moskowitz I, ed. Proc. of the 4th Int'l Workshop on Information Hiding (IH 2001). LNCS 2137, Berlin: Springer-Verlag, 2001. 289−302. [doi: 10.1007/3-540-45496-9]

[3] Kim Y, Duric Z, Richards D. Modified matrix encoding technique for minimal distortion steganography. In: Camenisch J, ed. Proc. of the 8th Int'l Workshop on Information Hiding (IH 2006). LNCS 4437, Berlin: Springer-Verlag, 2007. 314−327. [doi: 10.1007/978-3-540-74124-4_21]

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

[5] Fridrich J, Goljan M, Soukal D. Wet paper codes with improved embedding efficiency. IEEE Trans. on Information Forensics and Security, 2006,1(1):102−110. [doi: 10.1109/TIFS.2005.863487]

[6] Shi YQ, Chen CH, Chen W. A Markov process based approach to effective attacking JPEG steganography. In: Camenisch J, *et al*., eds. Proc. of the 8th Int'l Workshop on Information Hiding (IH 2006). LNCS 6387, Berlin: Springer-Verlag, 2007. 249−264. [doi: 10.1007/978-3-540-74124-4_17]

[7] Davidson J, Jalan J. Steganalysis using partially ordered markov models. In: Böhme R, *et al*., eds. Proc. of the 12th Int'l Workshop on Information Hiding (IH 2010). LNCS 6387, Berlin: Springer-Verlag, 2010. 143−157. [doi: 10.1007/978-3-642-16435-4_10]

© 中国科学院软件研究所 http://www.jos.org.cn

[8] Kodovský J, Fridrich J. Calibration revisited. In: Felten E, *et al*., eds. Proc. of the 11th ACM Workshop on Multimedia and Security (MM&Sec 2009). New York: ACM Press, 2009. 63−73. [doi: 10.1145/1597817.1597830]

[9] Kodovský J, Fridrich J. On completeness of feature spaces in blind steganalysis. In: Dittmann J, *et al*., eds. Proc. of the 10th ACM Workshop on Multimedia and Security (MM&Sec 2008). New York: ACM Press, 2008. 123−132. [doi: 10.1145/1411328.1411352]

[10] Dong J, Chen X, Guo L, Tan T. Fusion based blind image steganalysis by boosting feature selection. In: Shi Y, *et al*., eds. Proc. of the 6th Int'l Workshop on Digital Watermarking (IWDW 2007). LNCS 6387, Berlin: Springer-Verlag, 2008. 87−98. [doi: 10.1007/978-3-540-92238-4_8]

[11] Miche Y, Roue B, Lendasse A, Bas P. A feature selection methodology for steganalysis. In: Gunsel B, *et al*., eds. Proc. of the Multimedia Content Representation, Classification and Security. LNCS 4105, Berlin: Springer-Verlag, 2006. 49−56. [doi: 10.1007/11848035_9]

[12] Dash M, Liu H. Feature selection for classification. Intelligent Data Analysis, 1997,1(1-4):131−156.

[13] Kharrazi M, Sencar H, Memon N. Improving steganalysis by fusion techniques: A case study with image steganography. In: Shi Y, *et al*., eds. Proc. of the LNCS Trans. on Data Hiding and Multimedia Security I. LNCS 4300, Berlin: Springer-Verlag, 2006. 123−137.

[14] Qi Y, Wang Y, Yuan J. Audio steganalysis based on co-occurrence matrix and PCA. In: Proc. of the Int'l Conf. on Measuring Technology and Mechatronics Automation 2009 (ICMTMA 2009). IEEE Computer Society, 2009. 433−436. [doi: 10.1109/ICMTMA.2009.342]

[15] Tian Y, Cheng YM, Qian ZX, Wang YL. Image steganalysis based on PCA and SVM. Journal of the Graduate School of the Chinese Academy of Sciences, 2008,25(1):74−79 (in Chinese with English abstract).

[16] Theodoridis S. Pattern Recognition. 4th ed., Beijing: China Machine Press, 2009.

[17] Li B, Huang F, Tan S, Huang J, Shi Y. Effect of different coding patterns on compressed frequency domain based universal JPEG steganalysis. In: Shi Y, ed. Proc. of 6th Int'l Workshop on Digital Watermarking (IWDW 2007). LNCS 5041, Berlin: Springer-Verlag, 2008. 143−157. [doi: 10.1007/978-3-540-92238-4_12]

[18] Pevný T, Fridrich J. Detection of double-compression in JPEG images for applications in steganography. IEEE Trans. on Information Forensics and Security, 2008,3(2):247−258. [doi: 10.1109/TIFS.2008.922456]

[19] Bas P, Filler T, Pevný T. Break our steganographic system v0.92. 2011. ftp://mas22.felk.cvut.cz/PGMs/bossbase v0.92.tar.gz

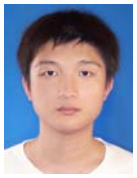
[20] Schaefer G, Stich M. Uncompressed colour image database v2. 2003. http://vision.cs.aston.ac.uk/datasets/UCID/data/ucid.v2.tar.gz

[21] Chang C, Lin C. LibSVM 3.0. 2010. http://www.csie.ntu.edu.tw/cjlin/

[22] Escolano F, Suau P, Bonev B. Information Theory in Computer Vision and Pattern Recognition. London: Springer-Verlag, 2009.

附中文参考文献:

[15] 田源,程义民,钱振兴,汪云路.基于 PCA 及 SVM 的图像信息隐藏检测.中国科学院研究生院学报,2008,25(1):74−79.



黄炜(1985－),男,福建南安人,博士生,主要研究领域为信息隐藏.

冯登国(1965－),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为网络与系统安全.

赵险峰(1969－),男,博士,副研究员,CCF 会员,主要研究领域为信息隐藏.

盛任农(1969－),男,副研究员,主要研究领域为数字信号处理.

© 中国科学院软件研究所 http://www.jos.org.cn