

基于主成分分析的泛盲掩密分析方案

陈丹 张键红 王育民

(西安电子科技大学 ISN 国家重点实验室, 西安 710071)

摘要 为了能够对掩密后图像进行有效的检测,首先从分类特征维数、单个特征的有效性和特征相关性等3个方面对 Farid 泛盲掩密分析算法的缺陷进行了分析,然后提出了采用主成分分析技术对分类特征进行降维处理的方法,并基于 RBF(radial basis function)网络构造了新的泛盲掩密分析方案。该方案不但大大降低了用于分类的图像特征的维数,而且提高了掩密分析的检测性能。利用该方案和 Farid 的方案分别对用 JSteg 等软件掩密后的图像进行的检测比较实验表明,经主成分分析预处理后,该方案的样本集特征矢量维数比 Farid 方案分别减少了 174(Jsteg)维、163(EzStego)维和 180(S-Tools)维,而特征数目的减少又大大简化了分类器的设计,而且,该方案能够有效检测嵌入消息占可嵌入容量的比例达 60%(Jsteg)、80%(EzStego)、50%(S-Tools)以上的掩密图像,并获得了比 Farid 算法更高的检测率。

关键词 掩密分析 泛盲掩密分析 主成分分析

中图法分类号: TP309 文献标识码: A 文章编号: 1006-8961(2006)03-0394-07

A Universal Blind Steganalysis Based on Principle Component Analysis

CHEN Dan, ZHANG Jian-hong, WANG Yu-min

(State Key Laboratory on ISN, Xidian University, Xi'an 710071)

Abstract A new universal blind steganalysis scheme is presented based on analysing the limitation of Farid's universal blind steganalysis on the dimension of the feature vector, the validity of the single feature, and the correlation of the features. Preprocessing is proposed using principle component analysis on the image statistics features and steganalysis classifier is constructed using RBF network. The scheme not only reduces the dimension of the feature vector enormously, but also improves the performance of the detection. Apply our scheme and Farid's scheme to detecting the stego images produced by JSteg, EZStego, and S-Tools respectively, and the comparison of these simulation results shows that after the preprocessing using principle component analysis the dimension of the feature vector in our scheme decreases 174(Jsteg), 163(EzStego), 180(S-Tools), and therefore simplifies the design of the steganalysis classifier. Furthermore, our scheme is quite more efficient because the stego image that the proportion of the embedding message to the maximal embedding capability is more than 60%(Jsteg), 80%(EzStego), 50%(S-Tools) can be detected efficiently by our scheme.

Keywords steganalysis, universal blind steganalysis, principle component analysis

1 引言

近几年,信息隐藏学科(information hiding)得到了广泛的关注,其重要分支——掩密术(又称隐写术, steganography)也在不断地发展。掩密术^[1]是一

种隐蔽通信技术,其目的在于隐藏秘密信息的存在,而不像密码术那样仅隐藏秘密信息的内容。毋庸置疑,掩密术不仅开辟了信息安全的新思路,并且弥补了传统加解密系统的一些缺陷。它的发展,使得人们可以利用它来进行敏感信息的秘密传递而不引起别人的注意,但是由于别有用心的人也可能利用同

基金项目:“十五”通信预研项目(41001040303)

收稿日期:2005-01-21; 改回日期:2005-05-31

第一作者简介:陈丹(1976~),女,2003年获西安电子科技大学硕士学位,现为西安电子科技大学博士研究生。主要研究方向为信息隐藏、网络安全。E-mail:chenndann@hotmail.com

样的方法来进行非法的通信,因此,掩密术的对抗技术——掩密分析技术(steganalysis)也越来越显示出举足轻重的作用。

掩密分析^[2]是利用掩密术固有的弱点,即信息的嵌入会引起载体数据的分布特性或统计特性的改变,这样通过分析各种可能的载体信息,便可检测、提取或破坏隐藏的秘密消息。目前多数的掩密分析方法都是针对某种特定的掩密算法的,例如,对于掩密术中最常见的LSB(least significant bits)算法,许多学者都提出了相应的分析对策,其中Westfeld根据在LSB隐藏信息前后载体图像中的值对(pair of value)的统计特性差异,设计了 χ^2 检验方法^[3];Fridrich利用图像像素值的空间相关性提出了RS(regular singular)统计检测方法^[4];Zhang Tao利用自然图像LSB平面和其他比特平面之间存在的弱相关性,通过直方图差异对比,实现了对隐藏信息的检测^[5]。这些基于某一种掩密算法的检测方法虽然能够取得较准确的分析结果,但是灵活性和可扩展性都很差。由于隐藏学科自身的特点,可能使掩密算法并不公开,而且掩密算法将层出不穷,所以无法得知要攻击算法的具体细节。Farid首先将掩密分析看成是两类模式的识别问题,即通过合理选择特征和设计分类器来区分没有隐藏消息的载体图像和隐藏了消息的掩密图像,并由此提出泛盲掩密分析(universal blind steganalysis)的概念^[6,7]。这一概念的提出,使得人们可以针对某一类而不是某一个掩密算法来检测隐藏消息的存在,这样就大大提高了掩密分析算法的适用范围和检测效率。

但是,Farid提出的算法(简称Farid算法)对于掩密图像的检测效率并不高,究其原因它是没有充分考虑所选择的分类特征对模式检测的有效性,因为在模式识别中,分类特征选取的好坏将直接影响到最终检测性能的优劣。本文将分别从特征维数、单个特征的有效性和特征之间的相关性3个方面对该算法所选分类特征进行有效性分析。研究表明,这些特征不但维数过大,容易造成分类器高度复杂,而且特征之间具有很强的相关性,由于该相关性使得所观测的数据在很大程度上有信息的重叠,从而不利于分类器的设计和性能。为此,本文提出采用主成分分析技术^[8]来对分类特征进行预处理,这样不仅可以对原始特征数据进行降维处理,而且还可简化原始数据的统计数字特征,同时还提供了很多对分类极其重要的信息。掩密分析分类器由

RBF(radial basis function)网络^[9]构造,其在解决非线性分类问题中具有收敛速度快、不易陷入局部极小点、鲁棒性好和易于实现等优点。通过利用本文算法和Farid的算法对JSteg、EZStego和S-Tools^[10-12]掩密后的图像进行检测的实验结果显示,对于不同长度的嵌入消息,本文算法较Farid算法都具有更好的检测性能。

2 泛盲掩密分析系统模型

与一般的模式识别系统相同^[9],泛盲掩密分析过程也分为学习和判决两个部分(如图1所示)。其中学习过程包括从大量的训练样本图像(包括载体图像和掩密图像两类样本)中提取并选择有利于分类的特征矢量,然后以此构造掩密分析分类器并对分类器进行训练,直到满足一定的精度要求;判决过程是利用学习过程中建立的分类器来对被测图像进行分类。这样,即使在不知道掩密算法的情况下,也可以通过比较和分析学习样本来察出由于消息嵌入所带来的数据特性变化,从而提取出可用于分类的特征并设计分类器,以便最终获得测试图像掩密与否的信息。

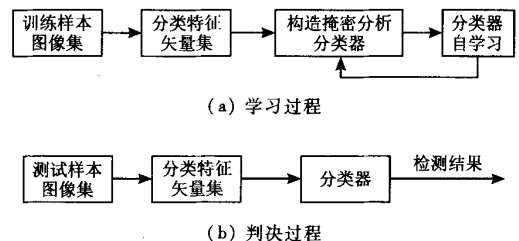


图1 泛盲掩密分析系统模型

Fig.1 Model of universal blind steganalysis system

该模型中,分类特征的选择是非常重要的,它强烈地影响到分类器的设计与性能,因为如果对不同类别所选的特征差别很大,则比较容易设计出高性能的分类器,反之,则有可能使得分类器难于构造或分类效率不高,所以,掩密分析系统成功的关键就是选出利于分类的最有效特征。

3 Farid 泛盲分析算法的缺陷

上节已经指出,分类特征的有效性是影响泛盲掩密分析检测性能的关键因素。这里分别从特征维数、单个特征的有效性和特征之间的相关性3个方

面对 Farid 算法^[6,7]所选特征的有效性进行分析。

(1) 特征维数

一般而言,在样本数不是很多的情况下,高维特征空间无论是从计算的复杂程度,还是从分类器的性能来看,都是不适宜的。通常,在设计分类器之前都要研究如何对其降维,这已经成为一个重要的课题^[9]。Farid 算法首先是针对灰度图像提出的,其选取 $12(n-1)$ 个小波系数统计量和 $12(n-1)$ 个小波系数误差统计量,共 $24(n-1)$ 个统计量作为区分载体图像和掩密图像的特征矢量,其中 n 为小波分解级数。对于彩色图像, Farid 提出分别计算红(R)、绿(G)、蓝(B) 3 个通道的统计量,即由 $24(n-1) \times 3$ 个统计量来构成分类特征矢量。为了获取有效的分类信息, n 不能太小,如 n 取 4, 这样,对于灰度图像就有 72 个特征,而对彩色图像则有 216 个特征,毋庸置疑,这么多的特征将会导致分类器的高度复杂。

(2) 单个特征的有效性

为选取有效特征,首先对 Farid 算法的每个分

类特征做方差分析^[8],以检测其区分载体图像和掩密图像的有效性。分析方法为,首先取 500 幅由数码相机拍摄的自然图像(剪裁成 640×480),分别计算它们的小波系数统计量和小波系数误差统计量,并用其构成特征矢量;然后用 JSteg 掩密软件在这 500 幅图像中分别嵌入消息来形成掩密图像,并按照相同的方法计算特征矢量。表 1 给出了部分特征在方差分析中的 F 值(统计量服从 F 分布)以及通过 F 值做假设检验获得的概率值 P (按 R、G、B3 个颜色通道分别排列)。从实验数据可以看出,并非所有的特征对于分类都是有利的,其中约有 50% 的特征(表中的黑体数据)对分类是无效的。

(3) 特征之间的相关性

如果分类特征之间存在很强的相关性,那么所观测的数据将在很大程度上有信息的重叠,这样不但会导致冗余数据过多而效率降低,更不利于分类器的设计和性能。从上面的实验中可以发现一个显著的特点,即表 1 的同一行(相同分解级和相同分

表 1 部分特征在方差分析中的 F 值和 p 值

Tab. 1 F -value and p -value of partial features in the variance analysis

序号	R		G		B	
	F	P	F	P	F	P
1	105.391 3	0	57.312 9	0.000 0	120.157 2	0
2	462.512 6	0	585.595 4	0	417.095 1	0
3	0.099 9	0.752 0	0.095 0	0.757 9	0.100 3	0.751 5
4	7.307 8	0.007 0	7.571 0	0.006 0	7.345 0	0.006 8
5	0.453 0	0.501 1	0.369 4	0.543 5	0.569 3	0.450 7
6	6.131 5	0.013 4	7.517 1	0.006 2	8.167 3	0.004 4
7	1.062 8	0.302 8	2.240 7	0.134 7	1.446 0	0.229 5
8	3.168 4	0.075 4	2.996 4	0.083 8	2.951 4	0.086 1
9	165.736 9	0	92.330 6	0	188.933 2	0
10	279.558 2	0	484.729 9	0	238.761 7	0
11	0.038 1	0.845 4	0.041 4	0.838 7	0.037 0	0.847 5
12	8.328 5	0.004 0	8.929 8	0.002 9	8.371 3	0.003 9
13	0.106 5	0.744 3	0.027 6	0.868 2	0.193 0	0.660 5
14	0.795 9	0.372 5	2.065 2	0.151 0	0.168 3	0.681 7
15	0.091 5	0.762 4	0.218 8	0.640 1	0.147 2	0.701 3
16	0.663 5	0.415 5	0.498 7	0.480 3	0.012 1	0.912 3
17	865.793 8	0	857.456 6	0	852.198 9	0
18	434.075 9	0	423.132 3	0	433.151 1	0
19	2.642 9	0.104 3	2.740 7	0.098 1	2.752 7	0.097 4
20	17.697 2	0.000 0	17.824 9	0.000 0	17.898 5	0.000 0
21	9.845 0	0.001 8	9.843 3	0.001 8	9.868 5	0.001 7
22	18.464 8	0.000 0	14.494 8	0.000 1	9.269 5	0.002 4
23	1.244 3	0.264 9	1.328 9	0.249 3	1.603 9	0.205 6
24	0.692 2	0.405 6	1.463 5	0.226 7	2.164 8	0.141 5

解方向上的统计量)上 R、G、B 所对应的 F 值或 P 值都非常接近,也就是说,这些特征对分类具有相同的效应。这说明 R、G、B 3 个通道的特征数据之间存在着很强的相关性。为了进一步证实这一点,可通过分别求取这些特征之间的相关系数来说明。表 2 给出了 R 和 G 通道、R 和 B 通道、G 和 B 通道相应特征之

间的相关系数值(因篇幅所限仅列出了部分数据),其中相关系数的绝对值大于等于 0.5 的个数占全部系数个数的 99.31%,而大于等于 0.9 的个数达到全部系数个数的 88.54%。由此可见,从 Farid 算法得到的分类特征数据中含有大量冗余,这些冗余不利于分类性能的提高,应该尽可能予以消除。

表 2 R 和 G、R 和 B、G 和 B 通道相应特征之间的相关系数值

Tab. 2 Correlation between channel R and G, channel R and B, and channel G and B, respectively

序号	R 和 G		R 和 B		G 和 B	
	载体图像	含密图像	载体图像	含密图像	载体图像	含密图像
1	0.9923	0.9759	0.9832	0.9848	0.9875	0.9799
2	0.9173	0.9818	0.8044	0.9595	0.8979	0.9751
3	0.9940	0.9903	0.9932	0.9898	0.9952	0.9932
4	0.9947	0.9952	0.9982	0.9968	0.9957	0.9960
5	0.9401	0.9528	0.8651	0.8824	0.8792	0.8821
6	0.9555	0.9668	0.9020	0.9398	0.9287	0.9383
7	0.8697	0.9335	0.8817	0.8844	0.9157	0.8838
8	0.9691	0.9536	0.9361	0.9347	0.9403	0.9528
9	0.9873	0.9793	0.9868	0.9851	0.9835	0.9747
10	-0.0644	0.7836	0.8257	0.9387	-0.1194	0.7097
11	0.9991	0.9989	0.9963	0.9965	0.9957	0.9961
12	0.9968	0.9965	0.9941	0.9953	0.9879	0.9907
13	0.9663	0.9636	0.9377	0.9424	0.9466	0.9385
14	0.8333	0.8757	0.8869	0.8850	0.7587	0.8160
15	0.9438	0.9581	0.9322	0.9399	0.9491	0.9423
16	0.9437	0.9591	0.9417	0.9304	0.8961	0.9349
17	0.9924	0.9994	0.9909	0.9981	0.9904	0.9982
18	0.9581	0.9987	0.9369	0.9975	0.9478	0.9978
19	0.9999	0.9998	0.9980	0.9971	0.9977	0.9965
20	1.0000	0.9999	0.9993	0.9985	0.9992	0.9982
21	0.9774	0.9884	0.9666	0.9745	0.9746	0.9763
22	0.8993	0.8868	0.8248	0.8354	0.8426	0.8719
23	0.9741	0.9849	0.9573	0.9686	0.9774	0.9692
24	0.9402	0.9700	0.8686	0.9439	0.9053	0.9556

此外,同一颜色通道内,不同的分解级($i = 1, \dots, n-1$)和分解方向(水平、垂直和对角)上的特征之间也具有相关性,例如,本文计算的 R 通道内特征之间的相关系数值,其中绝对值大于等于 0.5 的个数约占全部系数的 10%。

综上所述,有必要对 Farid 算法的分类特征做进一步的筛选或提取处理,以便使得处理后的样本数据不但维数小,而且能够最大限度地反映载体图像和掩密图像两类图像之间的差异。为此,本文提出对原始样本数据进行主成分分析预处理,以便简化分类器设计,并提高总体的检测性能。主成分分析技术不但可

以对原始特征数据进行降维处理,而且能够简化原始数据的统计数字特征,同时还可提供很多对分类极其重要的信息,例如数据群点的重心位置(或称平均水平)、数据变异的最大方向、群点的散布范围等,它们正是看起来杂乱无章的原始数据群的主要特征。下面将详细介绍该掩密分析方案。

4 基于主成分分析的掩密分析方案

4.1 系统框图

掩密分析方案的系统框图如图 2 所示。学习过

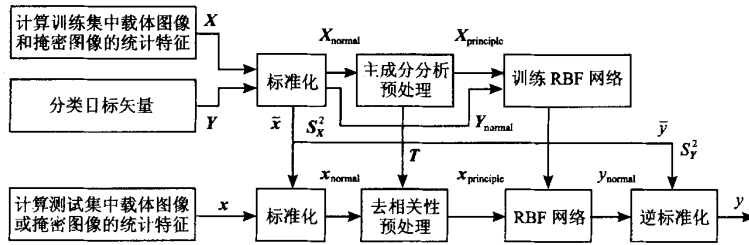


图 2 掩密分析系统结构

Fig. 2 Structure of the steganalysis system proposed

程主要有以下几个步骤:

(1) 根据 Farid 算法分别计算训练集中载体图像和掩密图像的特征矢量, 并将两组特征矢量合并构成原始特征矩阵

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N_1,1} & x_{N_1,2} & \cdots & x_{N_1,p} \\ x_{N_1+1,1} & x_{N_1+1,2} & \cdots & x_{N_1+1,p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N_1+N_2,1} & x_{N_1+N_2,2} & \cdots & x_{N_1+N_2,p} \end{bmatrix}$$

其中, N_1 为载体图像数, N_2 为掩密图像数, $N_1 + N_2 = N$ 为总体样本数, p 是特征数。

(2) 建立分类目标矢量 $Y = [\underbrace{0, \dots, 0}_{N_1 \uparrow}, \underbrace{1, \dots, 1}_{N_2 \uparrow}]^T$, 前面 N_1 个“0”代表载体图像, 后面 N_2 个“1”代表掩密图像。

(3) 对 X 进行标准化处理, 形成标准化特征矩阵 X_{normal} , 即 $X_{normal} = \{x_{normal,k}\}$, $k = 1, \dots, p$, 其中 $x_{normal,k} = \frac{x_k - \bar{x}_k}{\sqrt{S_{X,k}^2}}$, $\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{i,k}$, $S_{X,k}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_{i,k} - \bar{x}_k)^2$ 。因为主成分分析的每个主成分依赖于测量初始变量所用的尺度, 所以当尺度改变时, 会得到不同的特征值。克服这个困难的方法是对初始变量进行标准化处理, 即使其均值为 0, 方差为 1。同理, 对分类目标矢量 Y 也进行标准化处理, 生成标准化目标矢量 Y_{normal} 。

(4) 对 X_{normal} 进行主成分分析, 形成主成分特征矩阵 $X_{principle}$, 并保存转换矩阵 T 。

(5) 将处理后的数据 $X_{principle}$ 和分类信息 Y_{normal} 输入至初始的 RBF 网络, 同时用其来训练网络的权值参数, 直到达到目标误差或训练次数最大值为止。

判决部分主要是计算测试集中某一图像的特征

矢量 x , 首先利用预先计算的样本均值矢量 \bar{x} 和样本方差矢量 S_X^2 进行标准化变换, 以形成标准化特征矢量 x_{normal} , 即 $x_{normal} = \{x_{normal,k}\}$, $k = 1, \dots, p$, 其中

$$x_{normal,k} = \frac{x_k - \bar{x}_k}{\sqrt{S_{X,k}^2}}$$

然后经转换矩阵 T 预处理来获得 $x_{principle} = x_{normal}T$, 再将 $x_{principle}$ 输入到设计好的 RBF 网络分类器, 而网络输出标准化的分类值 y_{normal} , 还需用 \bar{y} 和 S_Y^2 进行逆标准化处理; 最后输出分类值 y , 以检测分类器的性能。

4.2 主成分分析预处理

主成分分析^[8]是一种在保证数据信息损失最小(最小二乘准则)的前提下, 经线性变换和舍弃一小部分信息, 以少数新的综合变量取代原始的多维变量的技术。

预处理的具体过程为:

(1) 计算标准化样本矩阵 X_{normal} 的相关系数矩

$$\text{矩阵 } R = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,p} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p,1} & r_{p,2} & \cdots & r_{p,p} \end{bmatrix} \circ$$

(2) 求特征方程 $|R - \lambda I| = 0$ 的 p 个非负特征值 $\lambda_1 > \lambda_2 > \dots > \lambda_p \geq 0$, I 是单位阵, 而对应于第 i 个特征值 λ_i 的特征向量为 $c_i = \{c_{1,i}, c_{2,i}, \dots, c_{p,i}\}^T$, $i = 1, \dots, p$, 并且有 $c_i^T c_j = \sum_{k=1}^p c_{k,i} c_{k,j} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$, $k = 1, \dots, p$ 。

(3) 计算 $m(m < p)$ 个主成分。首先求取当方差比例 $\alpha = \left(\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} \right)$ 接近于 1 时(一般取 $\alpha \geq 0.85$)的 m , 然后将与前 m 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$ 对应的特征向量 c_1, c_2, \dots, c_m 组成变换矩阵 T , 最后计算主成分特征矩阵 $X_{principle, N \times m} = X_{normal, N \times p} T_{p \times m} \circ X_{principle} = (x_{principle,1}, x_{principle,2}, \dots, x_{principle,m}) (m < p)$ 包

含了原特征数据的 m 个主成分,由于其方差和占全部总方差的 85% 以上,基本上保留了原来特征的信息,而特征数目却由 p 个减少至 m 个,从而达到了筛选特征的目的。

4.3 RBF 网络设计

分类器采用 RBF(radial basis function)网络^[9],即径向基函数网络,这是一种较常用的前馈型神经网络。它不仅可以实现对任意非线性函数的逼近,而且用于模式分类时具有收敛速度快、不易陷入局部极小点、鲁棒性好和易于实现等优点。

由于掩密分析属于两类模式识别问题,所以相应的 RBF 网络结构比较简单(如图 3 所示)。

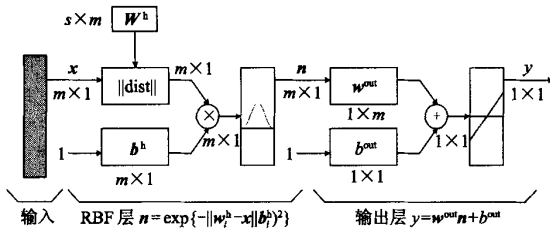


图 3 基于 RBF 网络的掩密分析分类器结构

Fig.3 Framework of the steganalysis classifier based on RBF network

图中 x 为输入矢量, n 表示隐层输出矢量, w^h_i 表示第 i 个隐层神经元的权值矢量,即隐层神经元权值矩阵 W^h 的第 i 行, b^h 是隐层的阈值矢量, w^{out} 是输出权值矢量, b^{out} 是输出阈值。网络的输入维数为 m 、隐层神经元个数为 s 个、输出个数为 1,隐层神经元采用高斯函数作为传递函数,输出层的传递函数为线性函数。

可根据训练样本数量采用迭代方法来确定隐层神经元的个数,即每迭代一次就增加一个神经元,直到平方和误差下降到目标误差以下或者隐层神经元个数达到最大值时,则迭代停止。输出单元的权值可直接由最小二乘法计算。

5 仿真实验

本文用 Jsteg、EzStego 和 S-Tools 软件^[10-12]掩密后的图像分别对本文算法和 Farid 算法进行了仿真测试。仿真测试用的图像是由数码相机拍摄的自然图像。测试时,首先取 500 幅上述图像作为未嵌消息的载体图像,存为 JPEG(或 GIF 或 BMP)格式,并

分别对每幅图像的 R、G、B3 个颜色通道进行 4 级小波分解,再计算 216 维特征矢量;然后用 Jsteg(或 EzStego 或 S-Tools)软件在每幅图像中嵌入消息来形成掩密图像,并计算每幅掩密图像的 216 维特征矢量。这样,1 000 幅图像的特征矢量构成了训练样本集。类似地,另取 1 000 幅图像构成测试样本集。

经主成分分析预处理后,样本集特征矢量的维数从 216 分别降到了 42(Jsteg)、53(EzStego)和 36(S-Tools),这样,特征数目的减少就大大简化了分类器的设计。图 4~图 6 分别给出了本文算法和 Farid 的算法对用 JSteg、EZStego 和 S-Tools 软件嵌入

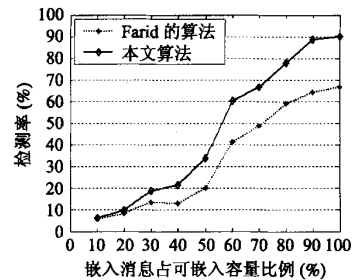


图 4 用 JSteg 软件掩密后图像的检测率

Fig.4 Stego-image detecting rate for JSteg

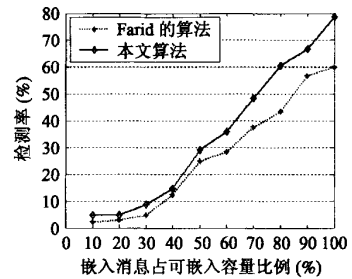


图 5 用 EzStego 软件掩密后图像的检测率

Fig.5 Stego-image detecting rate for EZStegov

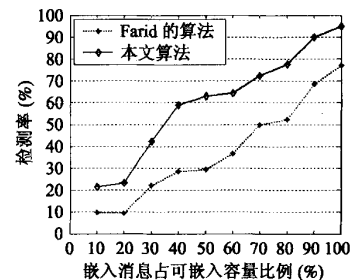


图 6 用 S-Tools 软件掩密后图像的检测率

Fig.6 Stego-image detecting rate for S-Tools

不同长度消息后的掩密图像进行检测的结果。对于用 JSteg 软件嵌入消息,嵌入消息占可嵌容量的 60% 以上时,就能被本文算法有效检测,而 Farid 算法只能检测 80% 以上嵌入消息的图像;对于用 EZStego 软件嵌入消息,只要嵌入消息占可嵌容量的 80% 以上,就可被本文算法检测到,而 Farid 算法则几乎不能检测掩密的图像;对于用 S-Tools 软件嵌入消息,嵌入消息占可嵌容量的 50% 以上时,就可被本文算法检测,而 Farid 算法只能检测 85% 以上嵌入消息的图像。由此可以看出,本文的算法具有更好的检测性能。

6 结 论

本文在 Farid 提出的泛盲掩密分析思想的基础上,提出了基于主成分分析预处理和 RBF 网络的掩密分析方法。之所以采用主成分分析进行预处理是因为主成分分析不仅可以对原始特征数据进行降维处理,还简化了原始数据的统计数字特征,同时还提供了很多对分类极其重要的信息,而分类器采用 RBF 网络则是因为 RBF 网络具有收敛速度快、不易陷入局部极小点、鲁棒性好等优点。为了解本文算法的效果,分别测试了本文算法和 Farid 算法对用 Jsteg、EzStego 和 S-Tools 软件掩密后图像进行检测的能力,实验结果表明,其性能较 Farid 算法有了大幅度的提高。

参考文献 (References)

- 1 Provos N, Honeyman P. Hide and seek: An introduction to steganography[J]. IEEE Security & Privacy, 2003, 1(3): 32~44.
- 2 Fridrich J, Goljan M. Practical steganalysis-state of the art[A]. In: Proceedings of SPIE Photonics West, Electronic Imaging 2002 on Security and Watermarking of Multimedia Contents[C], San Jose, California, USA. 2002: 1~13.
- 3 Westfeld A, Pfitzmann A. Attacks on steganographic systems[A]. In: Pfitzmann A. (Eds.). Proceedings of 3rd Int'l Workshop Information Hiding[C]. Dresden, Germany, Springer-Verlag, 1999: 61~76.
- 4 Fridrich J, Goljan M, Du Rui. Detecting LSB steganography in color and gray-scale images[J]. Magazine of IEEE Multimedia, Special Issue on Security, 2001, 8(4): 22~28.
- 5 ZHANG Tao, PING Xi-jian. Reliable detection of LSB steganography based on the difference image histogram[A]. In: Proceedings of International Conference Asian Symposium Signal Processing[C], Hong Kong, China, 2003: 545~548.
- 6 Farid H. Detecting hidden messages using higher-order statistical models[A]. In: Billene Mercer(Ed.). Proceedings of International Conference Image Processing[C], Rochester, NY, USA: IEEE Press, 2002: 905~908.
- 7 Lyu S, Farid H. Detecting hidden messages using higher-order statistics and support vector machines[A]. In: Petitcolas et al. (Eds.). Proceedings of 5th International Workshop on Information Hiding[C], Noordwijkerhout, the Netherlands: Springer-Verlag, 2002: 340~354.
- 8 REN Ruo-en, WANG Hui-wen. Multiple statistical data analysis-theory, technique, and instance[M]. Beijing: National Defence and Industry Press, 1997. [任若恩,王惠文著.多元统计数据分析——理论、方法、实例[M].北京:国防工业出版社,1997.]
- 9 BIAN Zhao-qi, ZHANG Xue-gong. Pattern recognition (2nd edition)[M]. Beijing: Tsinghua University Press, 2000. [边肇祺,张学工等编著.模式识别(第二版)[M].北京:清华大学出版社,2000.]
- 10 Derek Upham. JPEG-JSteg-V4 [EB/OL]. <http://www.funet.fi/pub/crypt/steganography/jpeg-JSteg-v4.diff.gz>.
- 11 Romana Machado. EZStego [EB/OL]. <http://www.stego.com>.
- 12 Andrew Brown. S-Tools4 [EB/OL]. <http://www.snapfiles.com/get/stools.html>.