基于共生特征和集成多超球面 OC-SVM 的 JPEG 隐密分析方法

郭艳卿[①] 孔祥维[①] 尤新刚[①②]
[①](大连理工大学信息安全研究中心 大连 116024)
[②](北京电子技术应用研究所 北京 100091)

摘 要：隐密是指将秘密信息以不可察觉的方式隐藏于其他载体之中的技术。隐密分析的目的是检测秘密信息的存在并最终提取秘密信息。目前基于二类或多类分类器的盲隐密分析方法可有效检测已知隐密算法，但无法对未公开隐密算法的生成图像进行检测。该文提出了一种新的 JPEG 盲隐密分析方法，对已知或未公开隐密算法都可检测。基于共生特征和多超球面 OC-SVM 分类器，本方法利用能有效对载体 JPEG 图像的统计分布边界建模。为进一步提高检测性能，还应用 Bagging 集成学习算法提高分类器的泛化能力。实验结果表明，该文方法能较为准确地检测出典型 JPEG 隐密算法生成的含密图像，性能优于已有的同类隐密分析方法。

关键词：隐密分析；共生特征；多超球面；一类支持向量机；Bagging
中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2009)05-1180-05

JPEG Steganalysis Based on Co-occurrence Features and Ensemble Multiple Hyperspheres OC-SVM

Guo Yan-qing[①] Kong Xiang-wei[①] You Xin-gang[①②]
[①](*Information Security Research Center*, *Dalian University of Technology*, *Dalian* 116024, *China*)
[②](*Beijing Institute of Electronic Technology and Application*, *Beijing* 100091, *China*)

Abstract: Steganography is the technology of hiding a secret message in plain sight. The goal of steganalysis is to detect the presence of embedded data and to eventually extract the secret message. Current blind steganalytic methods, which relied on two-class or multi-class classifier, have offered strong detection capabilities against known embedding algorithms, but they suffer from an inability to detect previously unknown forms of steganography. In this paper, a new JPEG blind steganalytic technique for detecting both known and unknown steganography is proposed. On the basis of co-occurrence features and multiple hyperspheres One-Class SVM(OC-SVM) classifier, the proposed method can effectively model the statistics distribution boundary of innocent JPEG images. Bagging ensemble learning algorithm is also used to achieve higher detecting performance. Experimental results show the superiority of the method over other analogous steganalytic techniques.

Key words: Steganalysis; Co-occurrence feature; Multiple hyperspheres; OC-SVM; Bagging

1 引言

隐密技术(steganography)是指将秘密信息隐藏于其他载体之中并应用于秘密通信等领域的技术。隐密的目的是将经伪装后的含密数据混迹于图像、音频、视频等海量日常数据中，利用公开信道以不被察觉的方式进行秘密通信。迄今为止，互联网上已经出现了百余种隐密软件。这些软件在为用户个人隐私提供保障的同时也带来了严重的安全隐患。有证据表明，恐怖分子也曾应用此项技术进行秘密联络。因此，必须采取有效的隐密分析(steganalysis)技术，检测、提取、阻止甚至破坏互联网上可能存在的隐密信息，以打击对隐密技术的非法滥用。

隐密分析技术首先要解决的是隐密判决问题，即检测待测数据是否含有秘密信息。针对图像隐密判决问题，Avcibas 等人[1]最早提出了基于训练的隐密分析方法。此方法分别提取载体图像和含密图像的统计特征送入 Fisher 分类器进行训练，并将训练好的二类分类器用于判决待测图像是否含密。沿用此“提取统计特征+分类器”的设计思想，众多学者在如何提取对隐密敏感的统计特征和如何提高分类器的检测性能方面做了大量工作。Lyu 等人[2]利用正交镜像滤波器对图像进行分解，提取分解后不同分解级和分解方向上子带系数的均值、方差、偏度和峰度为统计特征，并结合 SVM 二类分类器提高了隐密分析方法的检测能力。宣国荣等人[3]提取图像及小波子带的直方图特征函数的重心作为统计特征，提出了一种基于贝叶斯二类分类器的隐密分析方法。Goljan 等人[4]利用估计出的隐密噪声的统计特性作为统计特

2008-04-07 收到，2008-07-21 改回
国家自然科学基金(60572111)资助课题

征，提出了较为通用的隐密分析方法。Li 等人[5]利用图像局部线性变换提取变换系数归一化直方图统计特征，构建了基于 Fisher 线性分类器的隐密分析方法。Wang 等人[6]对统计特征的优选问题进行了研究。对于 JPEG 图像，Pevny 等人[7]基于 DCT 系数直方图、块不连续性等统计特征及多类 SVM 分类器提出了多类盲隐密分析方法，可在判决待测图像是否含密的同时,判别出待测图像所使用的隐密算法(或软件)。上述基于二类或多类分类器的隐密分析方法的局限性在于：有限个参与训练的隐密算法生成的含密图像无法涵盖所有含密图像的统计特性，导致训练出的分类器仅能对来源于已训练隐密算法的含密图像进行正确判决，而一旦待测图像来源于未公开隐密算法，则只能判其为载体图像或由已训练隐密算法生成，这样能否将其判决为含密图像是无可靠依据的。因此，如何克服此局限性、设计出可将未公开隐密算法生成图像准确判为含密图像的隐密分析方法，是目前亟待解决的问题之一。

对此，Lyu 等人[8]提出了基于多超球面 OC-SVM 分类器的隐密分析方法。此方法直接利用多超球面 OC-SVM 分类器确定载体图像统计特征的分布边界，将落在此分布边界内的图像样本判为载体图像，落在此分布边界外的图像样本判为含密图像，因此可用于对未公开隐密算法生成图像的判决。经实验验证，此方法对典型 JPEG 隐密算法生成的高隐藏量含密图像具有一定的检测正确率，而对低隐藏量的含密图像几乎无检测能力。

为设计可灵敏检测任意 JPEG 含密图像的通用隐密分析方法，本文在文献[8]的基础上做了以下两方面工作：(1)提出能有效反映 DCT 系数之间相关性的多维共生特征，以提高隐密分析方法对各种 JPEG 隐密操作的灵敏性；(2)利用 Bagging 集成学习算法，进一步提高多超球面 OC-SVM 分类器的检测能力。统计实验结果表明，本文提出的基于共生特征和集成多超球面 OC-SVM 的隐密分析方法能较为准确地检测出典型 JPEG 隐密算法生成的含密图像。当虚警率为 10%时,对最大隐藏量 10%的 JSteg[9],Outguess[10],F5[11],MB1[12]含密图像，检测正确率可分别达到 99.8%，87.3%，66.8%和 81.9%。本文第 2 节将介绍多超球面 OC-SVM 分类器的基本原理；第 3 节将给出基于共生特征和集成多超球面 OC-SVM 分类器的隐密分析方法；在第 4 节中，我们对本文所提隐密方法的检测性能进行了统计实验并对实验结果进行了讨论分析；最后在第 5 节给出本文的结论。

2　多超球面 OC-SVM 分类器

基于二类或多类分类器的隐密分析方法在本质上仅能将待测图像判决为已训练的样本类别，因此从原理上无法将由未知隐密方法生成的图像判决为含密图像。对此，Lyu 等人[8]提出了基于多超球面 OC-SVM 分类器的隐密分析方法。其基本思想为：利用一类支持向量机(One Class SVM，OC-SVM)在高维特征空间中寻找可覆盖载体图像集的最小超球面，令其作为载体图像的分布边界，并将落在此分布边界内的图像判为载体图像，落在此分布边界外的图像判为含密图像。为解决由载体图像内容差异导致的单一超球面很难准确描述载体图像分布边界的问题，文献[8]预先采用 K-均值聚类方法对载体图像进行聚类，再对聚类后的载体图像寻找分布边界，提出了基于多超球面 OC-SVM 分类器的隐密分析方法。本小节将简要介绍多超球面 OC-SVM 分类器的基本原理。

2.1 OC-SVM 分类器

一类支持向量机(OC-SVM)是一种用于估计载体高维统计分布支撑边界的分类器。其基本思想是在特征映射后的高维统计特征空间中寻找能覆盖一定数量载体样本的最小超球面。即对于 l 个 n 维载体训练样本 $x_i \in R^n$， $i=1,\cdots,l$，求解如下优化问题：

$$\min_{c,r,\xi} r^2+\frac{1}{\nu l}\sum_i \xi_i$$
$$\text{s.t.}\quad \left\|\phi(x_i)-\overline{c}\right\|^2 \le \overline{r}^2+\xi_i,\ \xi_i \ge 0,\ i=1,\cdots,l \qquad (1)$$

其中 $\phi: R^n \to F$ 为特征映射函数，$v \in (0,1)$ 为超球面外载体训练样本比例的控制参数，ξ_i 为松弛因子，$\overline{c}$,$\overline{r}$ 分别为超球面的圆心和半径。

式(1)等价于：

$$\min_{\alpha}\sum_{i,j}\alpha_i\alpha_j\phi(x_i)^{\mathrm{T}}\phi(x_j)-\sum_i\alpha_i\phi(x_i)^{\mathrm{T}}\phi(x_i)$$
$$\text{s.t.}\quad 0\le\alpha_i\le 1/vl,\quad \sum_i\alpha_i=1 \qquad (2)$$

其中 T 表示转置。

解此二次优化问题，可得到 α_i，进而可得 $\overline{c}$,$\overline{r}$ 表达式分别为

$$\overline{c}=\sum_{i=1}^{l}\alpha_i\phi(x_i) \qquad (3)$$
$$\overline{r}^2=\left\|\phi(y)-\overline{c}\right\|^2 \qquad (4)$$

其中 y 为超球面上的任意点。

这样，判决待测图像统计特征 x 是否含密的公式可表示为

$$f(x)=\left\|\phi(x)-\overline{c}\right\|^2-\left\|\phi(y)-\overline{c}\right\|^2 \qquad (5)$$

当 $f(x)\le 0$ 时,判决待测图像为载体图像;当 $f(x)>0$ 是判决待测图像为含密图像。

2.2 多超球面 OC-SVM 分类器

由于自然图像内容及其存储方式的复杂性，载体图像的统计特征往往聚类性不强，导致载体图像在高维统计特征空间中的分布形状并不能用超球体近似，因此使用单一的 OC-SVM 分类器很容易将含密图像错分入载体图像集内(如图 1 左所示)。为减小此类错误，文献[8]提出了一种多超球面 OC-SVM 分类器，其思路是采用 K-均值方法预先将载体数据聚为 M 个子类，再用各子类训练出的 M 个超球面的组合替代原始的单一超球面进行判决(如图 1 右所示)。实验结

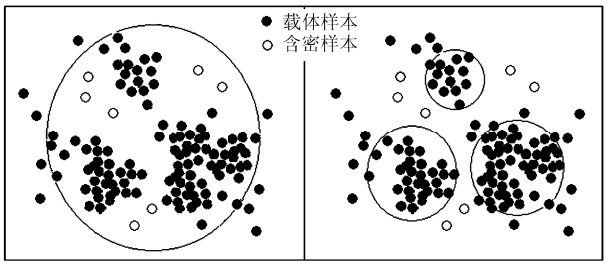

图 1 OC-SVM 与多超球面 OC-SVM 分类器示意图

果表明，此方法较为有效地解决了对含密图像的误判问题。

3 新的 JPEG 隐密分析方法

3.1 JPEG 图像共生特征

隐密分析方法所使用统计特征的敏感性是影响隐密分析方法检测性能的主要因素之一。对于 JPEG 格式图像，由于频域内相邻块之间对应 DCT 系数的相关性在秘密信息嵌入过程往往会遭到一定程度的破坏，因此，为有效度量 DCT 系数相关性在秘密信息嵌入前后的变化，本文将在 DCT 域共生矩阵的基础上提取归一化的 JPEG 图像共生特征，应用于后续的隐密分析方法。

设一幅 JPEG 图像 8×8 块的量化后 DCT 系数矩阵为 $\boldsymbol{J}_{r,c}(i,j)$，其中 $1\le r\le R$，$1\le c\le C$ 分别表示此图像 8×8 块的行列序号，$1\le i,j\le 8$ 分别表示在某一 8×8 块中量化后 DCT 系数的行列序号。定义图像 DCT 域共生矩阵 $\boldsymbol{F}_J$ 为

$$\boldsymbol{F}_J(p_1,p_2)=\frac{1}{2}\sum_{r=1}^{R-1}\sum_{c=1}^{C-1}\sum_{i,j=1}^{8}\left[\prod_{k=1}^{2}\delta(J_{r+k-1,c}(i,j)-p_k)+\prod_{k=1}^{2}\delta(J_{r,c+k-1}(i,j)-p_k)\right] \tag{6}$$

其中 $\delta(x)=\begin{cases}1, & x=0\\ 0, & x\ne 0\end{cases}$，$p_1,p_2\in\{-T,-T+1,\cdots,T\}$ 为量化后 DCT系数值，T 为正整数参数。

由于图像内容的差异，不同载体 JPEG 图像的 DCT 域共生矩阵会存在一定差别，致使隐藏相同长度秘密信息后的含密图像 DCT 域共生矩阵同样存在偏差。为尽可能消除此偏差、突出隐密操作造成的数据变化，本文采用文献[7]中的剪切—重压缩操作估计原始 JPEG 图像 $\boldsymbol{J}_C$，并定义如下的 JPEG 图像共生特征 $\boldsymbol{V}_J$：

$$\boldsymbol{V}_J(p_1,p_2)=\frac{\boldsymbol{F}_J(p_1,p_2)-\boldsymbol{F}_{J_c}(p_1,p_2)}{\boldsymbol{F}_{J_c}(p_1,p_2)} \tag{7}$$

3.2 用 Bagging 算法集成分类器

一个好的隐密分析方法应该具有较强的泛化能力。隐密分析方法的泛化能力是指根据已知数据建立的模型对新数据的检测能力。由于以下原因，上述多超球面 OC-SVM 分类器在泛化能力方面仍有进一步提高的可能：(1)在一般情况下，训练样本都会因数目过少或不够典型等问题而无法完全代表载体图像的真实分布，导致所得模型与真实模型之间存在偏差；(2)在统计特征维数较高但训练样本数较少的情况下，极可能出现的“过拟合”现象会降低分类器的泛化能力。为尽可能提高泛化能力，本文利用 Bagging[13](Bootstrap AGGregatING)集成学习(ensemble learning)算法将源于不同训练子集的多个分类器进行集成，以对上述多超球面 OC-SVM 分类器进行改进，并将其作为本文所提隐密分析方法的分类器。

Bagging 算法的基本思想是对训练样本集进行有放回抽取，进而为每一个多超球面 OC-SVM 分类器都构造出一个跟训练集大小相同但数据不同的训练集进行训练，最后将多个多超球面 OC-SVM 分类器的输出结果投票，得到最终的判决结果。

Bagging 算法的伪代码可描述为

Input:

L:a learning algorithm

N:an integer

For $i=1$ to N

T'= bootstrap sample from training set T

$h_i=L(T')$

End For

Output: $h_f(x)=\arg\max_{y\in Y}\sum_{i}^{N}h_i(x)=y$

对于给定的训练样本集，若集成学习中所有子分类器都给出相同或相近的输出，则集成后分类器的泛化能力接近于各子分类器泛化能力的加权平均；反之，若各子分类器的差异性较大，则集成后分类器的泛化能力将高于各子分类器泛化能力的加权平均。Bagging 算法正是通过重新选取训练样本集的方法构造各子分类器的差异性，从而提高了集成后分类器的泛化能力。

4 实验结果与讨论

结合上述 JPEG 图像共生特征和集成多超球面 OC-SVM 分类器，新的 JPEG 隐密分析方法是一种可灵敏检测任意 JPEG 含密图像的通用隐密分析方法。通过对 JSteg，F5，Outguess，MB1 所生成含密图像的检测实验，本节将验证本文算法在不同参数下的有效性，讨论各参数对算法检测性能的影响，并在相同实验环境下与文献[8]的检测结果进行比较分析。

实验中所使用的载体 JPEG 图像是由 1300 幅高精度数码相机图像(来源于 50 种不同型号的数码相机)经 ACDSee7.0 缩放到 1600×1200 大小、并重新压缩得到(质量因数为 85)。其中 650 幅用于生成训练样本，650 幅用于生成测试样本。含密图像样本由上述 1300 幅载体经 JSteg，F5，Outguess，MB1 等 4 种典型 JPEG 隐密算法生成。

4.1 各参数对检测性能的影响

表 1，表 2 分别给出了不同聚类参数 M 和不同集成参数 N 下，本文方法对 MB1 隐密算法 10%含密图像的检测性能。其中，M 表示超球面的个数，N 表示用 Bagging 算法构造

出的训练样本集的个数。表中第一列为本文方法的虚警率(载体图像判决含密图像的个数/载体图像总数)，各行分别代表对应虚警率下隐密分析方法对含密图像的检测率(含密图像判决为含密图像的个数/含密图像总数)，其中黑体数字表示在该虚警率下检测率的最大值，“－”表示在此参数下隐密分析方法无法达到此虚警率。

表 1 N=45 时聚类参数 M 对检测性能的影响(%)

虚警率	M						
(%)	1	2	4	6	8	12	20
30	85.1	88.7	95.5	96.7	97.3	98.3	**98.6**
25	79.1	83.8	94.7	94.3	95.9	**96.7**	—
20	75.4	75.7	92.3	92.6	94.1	**95.0**	—
15	61.1	67.5	88.7	89.0	**89.3**	—	—
10	41.9	51.4	81.7	**81.9**	79.1	—	—
5	27.0	27.5	**61.4**	56.5	—	—	—

表 2 M=6 时集成参数 N 对检测性能的影响(%)

虚警率	N						
(%)	1	5	15	25	45	100	200
30	94.9	95.5	95.8	96.3	**96.7**	96.0	96.4
25	93.1	93.6	94.4	94.4	94.3	94.5	**94.5**
20	91.4	92.1	92.0	92.2	**92.6**	92.1	92.5
15	86.8	87.2	88.8	88.9	89.0	**89.4**	89.1
10	77.9	82.1	82.7	**82.8**	81.9	82.6	82.6
5	51.3	55.0	55.9	56.4	56.5	57.8	**57.8**

从表 1 可以看出，随着聚类参数 M 的增加，本文方法在不同虚警率下的检测率均有所提高，表明随着超球面个数的增加算法能更准确地描述载体图像统计特征分布的边界。但当 M >6 时，本文方法可达到的最低虚警率也随 M 的增加而增大，这说明过大的聚类参数 M 会造成本文方法对载体图像的判决能力下降。因此合适的聚类参数 M 是影响本文算法检测性能的关键因素之一，本文实验中 M =6 较为合适。

从表 2 可以看出，集成参数 N 增加时，本文方法的检测率大体呈增加趋势，且在低虚警率情况下检测率增加相对较大(虚警率 5%时的检测率增加 6.5%)。表明随着集成参数 N 的增加，算法的检测性能也随之增强且逐渐趋于稳定。为在一定检测性能的前提下降低本文方法的计算量，可令集成参数 N =45。

从 DCT 域共生矩阵与 JPEG 图像共生特征的计算公式(6)，式(7)可得，当正整数参数 $T=1,2,\cdots,n$ 时，JPEG 图像共生特征的维数 $D=9,25,\cdots,(2n+1)^2$ 。为测试不同共生特征维数 D 对本文方法检测性能的影响，本文取 $N=45$ ，$M=6$ ，分别给出了 $D=9,25,49,81$ 时本文方法对 MB1 算法 10%含密图像的 ROC 曲线。如图 2 所示，当 $D=9$ 时，由于仅考虑了值为−1,0,1 的 DCT 系数的统计特征，所利用的载体信息过少，因此隐密分析方法的检测性能较差。当 $D=25$ 时，本文方法的检测性能达到峰值，随着 D 的进一步增加逐渐缓慢下降，这是因为当统计特征维数与训练样本个数之比高于一定程度时，各子分类器无法在高维空间中建立较为精确的模型，导致了隐密分析方法检测性能的下降。

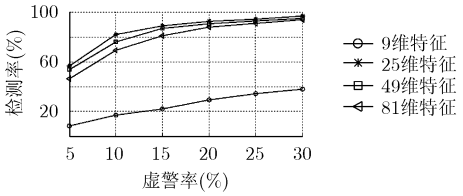

图 2 共生特征维数 D 对检测性能的影响

4.2 与文献[8]算法的检测性能比较

为证明本文方法的先进性，我们将本文方法与文献[8]方法进行了检测性能的对比实验，实验结果如表 3 所示。实验中所用的含密图像分别由 JSteg，F5，Outguess，MB1 隐密算法生成(算法名称后的百分数代表隐藏的秘密信息量与该算法最大隐藏量之比)。从表中可以看出，文献[8]方法对较大隐藏量的 JSteg，Outguess 含密图像具有一定的检测能力，而对于安全性较高的 F5，MB1 隐密算法生成的含密图像几乎无法进行检测。相比之下，由于本文方法在选用敏感统计特征的同时又提高了分类器的泛化能力，因此对上述 4 种隐密算法均具有较高的检测性能。

5 结束语

本文分析了基于二类或多类分类器的隐密分析方法的不足，针对已有 JPEG 隐密分析方法无法对未公开隐密算法

表 3 本文方法与文献[8]算法检测性能对比(%)

虚警率(%)	10		20	
	文献[8]	本文	文献[8]	本文
JSteg100	53.1	**100**	73.3	**100**
JSteg30	14.5	**100**	26.9	**100**
JSteg10	11.3	**99.8**	22.1	**100**
F5 100	12.9	**100**	22.4	**100**
F5 30	10.4	**100**	20.4	**100**
F5 10	11.7	**66.8**	20.1	**100**
Outg 100	16.2	**100**	28.9	**100**
Outg 30	11.2	**100**	22.3	**100**
Outg 10	10.8	**87.3**	20.4	**100**
MB1 100	10.0	**100**	18.9	**100**
MB1 30	10.1	**100**	20.1	**100**
MB1 10	9.9	**81.9**	20.1	**100**

生成图像进行准确判决的问题，提出了一种基于载体统计特征分布的隐密分析方法。此方法提取 JPEG 图像 DCT 域多维共生特征作为统计特征量，在高维特征空间中使用 K-均值聚类及 OC-SVM 方法构建分类器，并利用 Bagging 算法提高分类器的泛化能力。实验结果表明，本文提出方法能较为准确地检测出典型 JPEG 隐密算法生成的含密图像。对于 JSteg，Outguess，F5，MB1 生成的 10%隐藏量含密图像，当虚警率为 10%时，检测正确率可分别达到 99.8%，87.3%，66.8%和 81.9%。

参 考 文 献

[1] Avcibas I, Memon N, and Sankar B. Steganalysis based on image quality metrics. Proc of the IEEE 4th Workshop on Multimedia Signal Processing, Canners, France, 2001: 517–522.

[2] Lyu S and Farid H. Detecting hidden messages using higher-order statistics and support vector machines. 5th International Workshop on Information Hiding. Noordwijkerhout, Netherlands, 2002: 340–354.

[3] Xuan Guorong. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. Information Hiding 2005, Barcelona, Spain, 2005, LNCS 3727: 262–277.

[4] Goljan M, Fridrich J, and Holotyak T. New blind steganalysis and its implications. Proc. SPIE, Security, Steganography, and Watermarking of Multimedia, San Jose, CA, 2006: 1–13.

[5] Li B, Huang J W, and Shi Y Q. Textural features based universal steganalysis. Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, 2008, SPIE Vol. 6819: 681912-681912–12.

[6] Wang Y and Moulin P. Optimized feature extraction for learning-based image steganalysis. *IEEE Trans. on Information Forensics and Security*, 2007, 2(1): 31–45.

[7] Pevny T and Fridrich J. Toward multi-class blind steganalyzer for JPEG images. Proceedings of International Workshop on Digital Watermarking, Lecture Notes in Computer Science, 2005, Vol. 3710: 39–53.

[8] Lyu S and Farid H. Steganalysis using color wavelet statistics and one-class vector support machines. Proceedings of SPIE, 2004, Vol. 5306: 35–45.

[9] Derek U. JPEG-JSteg-V4 [EB/OL]. http://www.funet.fi/pub/crypt/steganography/jpeg-Jsteg-v4.diff.gz, 1995-07–06.

[10] Westfeld A. F5-a steganographic algorithm: high capacity despite better steganalysis. Information Hiding 4th International Workshop, IH'01. Pittsburgh, USA, 2001: 289–302.

[11] Provos N. Defending against statistical steganalysis. Proc 10th Usenix Security Symposium. Washington USA, 2001: 323–335.

[12] Sallee P. Model-based steganography. International Workshop on Digital Watermarking, LNCS 2939. Springer-Verlag, Berlin Heidelberg, 2004: 154–167.

[13] Breiman L. Bagging predictors. *Machine Learning*, 1996, 24(2): 123–140.

郭艳卿： 男，1980 年生，博士生，研究方向为数字信号处理、多媒体信息安全.

孔祥维： 女，1963 年生，教授，博士生导师，主要研究方向为多媒体信息安全、统计图像处理与模式识别.

尤新刚： 男，1963 年生，教授，博士生导师，主要研究方向为多媒体通信、信息安全.