

## 基于分割的空域图像隐写分析

汪然<sup>1</sup> 许漫坤<sup>1</sup> 平西建<sup>1</sup> 张涛<sup>1</sup>

**摘要** 提出一种基于图像内容的空域隐写分析方法, 该方法对图像进行分割, 使分割得到的每一类子图像具有相同的统计特性, 并对每一类子图像提取更加敏感的隐写分析特征, 分别构造分类器进行训练和测试, 由此对分割所得到的每一幅子图像都可以得到一个检测结果. 对整幅图像的判决结果通过加权融合得到. 实验结果表明, 该方法具有良好的性能, 尤其是针对自适应隐写方法, 该算法的检测准确率提高更加明显.

**关键词** 信息隐藏, 隐写分析, 图像分割, 纹理复杂度

**引用格式** 汪然, 许漫坤, 平西建, 张涛. 基于分割的空域图像隐写分析. 自动化学报, 2014, 40(12): 2936–2943

**DOI** 10.3724/SP.J.1004.2014.02936

## Steganalysis of Spatial Images Based on Segmentation

WANG Ran<sup>1</sup> XU Man-Kun<sup>1</sup> PING Xi-Jian<sup>1</sup> ZHANG Tao<sup>1</sup>

**Abstract** A new spatial image content based steganalysis algorithm is proposed. The given images are segmented to get the sub-images with similar statistical characteristics. More effective steganalysis features are extracted from each category of segmented sub-images respectively to build a classifier. Thus, a steganalysis result for each segmented sub-image can be obtained. The classifying results of all categories of sub-image are weighted fusing to obtain the overall steganalysis result. The performances shown by experiments demonstrate that the proposed algorithm can significantly improve the detection accuracy, especially for adaptive embedding methods.

**Key words** Information hiding, image steganalysis, image segmentation, texture complexity

**Citation** Wang Ran, Xu Man-Kun, Ping Xi-Jian, Zhang Tao. Steganalysis of spatial images based on segmentation. *Acta Automatica Sinica*, 2014, 40(12): 2936–2943

近年来, 信息隐藏作为信息安全领域的重要课题, 引起了国际学术界的重视. 隐写分析技术根据图像、视频、音频等载体的统计特性, 判断其中是否含有额外的隐蔽信息, 其与数字隐写的对抗成为了网络时代信息战的重要内容之一.

针对空域图像, 研究者提出了一系列隐写算法, 主要包括最不important比特位匹配 (Least significant bitmap matching, LSBM) 隐写方法<sup>[1]</sup> 和自适应隐写方法<sup>[2–4]</sup>. 随着空域图像隐写系统安全性的不断提高, 盲检测成为空域图像隐写分析的研究主流. 目前比较有代表性的盲检测特征主要包括图像质量度量<sup>[5]</sup>、像素或小波系数的概率密度函数 (Probability density function, PDF) 矩<sup>[6–7]</sup> 和特征函数 (Characteristic function, CF) 矩<sup>[8]</sup> 以及像

素联合分布<sup>[9–10]</sup> 等. 近年来, 研究者通过使用线性估计器对图像像素差分值求取高阶残留子带得到高维检测特征集<sup>[11–13]</sup>, 这些高阶统计特征能够有效地反映和刻画图像的纹理和边缘特性, 具有较好的性能. 同时通过构造整体分类器<sup>[14]</sup> 大大提高了训练、测试的效率.

研究者针对空域图像隐写分析问题提出了非常有效的隐写检测特征和丰富多彩的特征提取方法, 能够获得良好的检测效果, 但是仍然存在一些难以解决的问题. 目前, 大部分隐写分析方法都是针对隐写对图像所造成的统计特性变化提取特征, 对整幅图像采用相同的特征提取过程, 忽略了图像本身所具有的统计特性, 这导致隐写检测性能会受到图像内容和质量的严重影响, 并且极度依赖实验所选用的图像库. 同时, 在实验条件下通常会选择相同的训练和测试图像库, 而在实际网络情况中往往难以满足此条件. 当训练和测试库不匹配时, 几乎所有隐写检测算法的性能均会发生明显下降, 因而难以应用于实际. 另一方面, 自适应隐写方法能根据图像的内容自适应地选择嵌入秘密信息的位置和强度, 具有较高的安全性, 如何提高针对自适应隐写的检测精

收稿日期 2014-01-04 录用日期 2014-06-25  
Manuscript received January 4, 2014; accepted June 25, 2014  
国家自然科学基金 (61272490) 资助  
Supported by National Natural Science Foundation of China (61272490)  
本文责任编辑 章毓晋  
Recommended by Associate Editor ZHANG Yu-Jin  
1. 中国人民解放军信息工程大学信息工程学院 郑州 450001  
1. School of Information Systems Engineering, PLA Information Engineering University, Zhengzhou 450001

度也是一个亟待解决的问题.

图像是具有实际内容的实体, 由于相邻像素间具有较强的相关性, 而具有不同内容的图像区域像素间相关性不同, 因此图像具有区域平稳性, 通常可以将其看作马尔科夫信源. 隐写所造成的特征改变不仅与秘密信息嵌入机制有关, 还与载体数据的统计特性密切相关. 一些研究者已经注意到了载体、载密图像统计特性与隐写分析结果之间的关系, 并提出了一系列基于图像内容的隐写分析方法. 针对 JPEG 图像, 文献[15]将图像按照内容复杂度分类至不同的子图像库, 对每个子图像库分别进行训练和测试, 文献[16]将图像分成固定大小的子块, 并根据隐写检测特征对图像块进行分类, 对每类图像块分别进行检测, 通过加权融合得到最终的检测结果. 在相同的实验条件下, 上述方法的检测性能均较传统方法有明显提升. 文献[17]则通过对空域图像进行局部线性变换 (Local linear transform, LLT) 提取图像的纹理特征, 实现隐写分析.

本文提出一种基于图像内容的空域图像隐写分析方法. 该方法对图像进行分块, 采用图像块的相邻像素绝对差值的均值衡量图像的纹理复杂度, 并根据纹理复杂度度量特征将图像块分为若干类别, 实现图像分割. 对分割所得到的每一类子图像块分别提取基于差分像素联合分布的隐写检测特征, 通过加权融合得到最终的检测结果. 实验结果表明, 本文算法展现出良好的性能, 尤其是针对自适应隐写方法, 本文算法检测准确率的提高更加明显.

## 1 差分像素联合分布特征

文献[10]提出了差分像素邻接矩阵 (Subtractive pixel adjacency matrix, SPAM) 特征实现对 LSBM 匹配算法的检测, 本文在其基础上进行改进, 提出了差分像素联合分布 (Subtractive pixel joint distribution, SPJD) 特征. 本章分别对 SPAM 和 SPJD 特征进行介绍.

给定一幅尺寸为  $W \times H$  的图像  $I$ , 定义水平正向 (从左至右) 方向的差分矩阵:

$$D_{i,j}^{\rightarrow} = I_{i,j} - I_{i+1,j} \quad (1)$$

其中  $i = 1, \dots, W, j = 1, \dots, H$ . 同理可定义水平反向 (从右至左)、垂直正向 (从下至上)、垂直反向 (从上至下)、对角正向 (从左上角至右下角)、对角反向 (从右下角至左上角)、反对角正向 (从右上角至左下角)、反对角反向 (从左下角至右上角) 等八个方向的差分矩阵, 分别记为  $D_{i,j}^{\leftarrow}, D_{i,j}^{\downarrow}, D_{i,j}^{\uparrow}, D_{i,j}^{\searrow}, D_{i,j}^{\swarrow}, D_{i,j}^{\nearrow}, D_{i,j}^{\nwarrow}$  和  $D_{i,j}^{\swarrow}$ . 对水平

正向 (从左至右) 方向的差分矩阵  $D_{i,j}^{\rightarrow}$  提取 Markov 特征, 定义 1 阶概率转移矩阵:

$$M_{u,v}^{\rightarrow} = P(D_{i,j+1}^{\rightarrow} = u | D_{i,j}^{\rightarrow} = v) \quad (2)$$

其中  $u, v = -T_1, \dots, T_1, T_1$  为设置的差分像素值的截断阈值. 定义 2 阶概率转移矩阵:

$$M_{u,v,w}^{\rightarrow} = P(D_{i,j+2}^{\rightarrow} = u | D_{i,j+1}^{\rightarrow} = v, D_{i,j}^{\rightarrow} = w) \quad (3)$$

其中  $u, v = -T_2, \dots, T_2$ . 同理可定义其他七个方向的 1 阶、2 阶概率转移矩阵. 则 SPAM 特征定义为:

$$\begin{aligned} F_{1,\dots,k} &= \frac{1}{4}(M^{\rightarrow} + M^{\leftarrow} + M^{\downarrow} + M^{\uparrow}) \\ F_{k+1,\dots,2k} &= \frac{1}{4}(M^{\searrow} + M^{\swarrow} + M^{\nearrow} + M^{\nwarrow}) \end{aligned} \quad (4)$$

分别对 1 阶和 2 阶概率转移矩阵求取相应阶数的 SPAM 特征. 对于 1 阶 SPAM 特征,  $k = (2T_1 + 1)^2$ , 对于 2 阶 SPAM 特征,  $k = (2T_2 + 1)^3$ .

通过定义可以看出, SPAM 特征计算差分矩阵和概率转移矩阵的方向是一致的. 对于 1 阶 SPAM 特征而言, 每次计算转移概率时, 相当于考虑了同一方向上的三个相邻像素的相关性. 为了能反映更多像素的相关性, 本文所定义的 SPJD 特征在提取 Markov 特征时选择和差分不同的方向, 具体来说即对  $D_{i,j}^{\rightarrow}, D_{i,j}^{\downarrow}, D_{i,j}^{\searrow}, D_{i,j}^{\swarrow}$  分别提取垂直、水平、反对角、对角方向的 Markov 特征. SPAM 和 SPJD 特征在水平方向上参与运算的像素位置关系如图 1 所示. 可以看出, 1 阶 SPJD 特征能反映相邻四个像素的相关性, 并且像素间的最大距离为  $\sqrt{2}$ , 小于 SPAM 特征像素间的最大距离 2. 也就是说, 与 SPAM 相比, SPJD 特征能够捕捉到更多更邻近的像素间的相关性, 因此在提取同样维数特征的前提下, 本文特征能更好地反映像素的相关性.

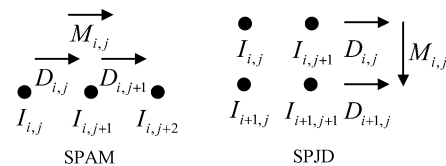


图 1 SPAM 特征和 SPJD 特征的像素位置

Fig. 1 The pixels location of SPAM and SPJD feature set

另外, 在提取 SPJD 特征时, 用联合概率矩阵代替概率转移矩阵, 即将式 (2) 和 (3) 改为:

$$M_{u,v}^{\downarrow} = P(D_{i,j}^{\downarrow} = u, D_{i+1,j}^{\downarrow} = v) \quad (5)$$

$$M_{u,v,w}^{\downarrow} = P(D_{i,j}^{\downarrow} = u, D_{i+1,j}^{\downarrow} = v, D_{i+2,j}^{\downarrow} = w) \quad (6)$$

同理可定义其他方向的联合概率矩阵. 为了降低算法复杂度, 本文方法只计算  $D_{i,j}^{\rightarrow}$ 、 $D_{i,j}^{\downarrow}$ 、 $D_{i,j}^{\searrow}$  和  $D_{i,j}^{\swarrow}$  四个正向的 Markov 特征. 则 SPJD 特征定义为:

$$\begin{aligned} F_{1,\dots,k} &= \frac{1}{2}(M^{\rightarrow} + M^{\downarrow}) \\ F_{k+1,\dots,2k} &= \frac{1}{2}(M^{\searrow} + M^{\swarrow}) \end{aligned} \quad (7)$$

同时由于像素差分值为 0 具有对称性, 可以认为  $P(x, y) = P(-x, -y)$ , 据此对特征进行降维. 对于 1 阶 SPJD 特征,  $k_1 = \frac{1}{2}((2T_1 + 1)^2 + 1)$ , 对于 2 阶 SPJD 特征,  $k_2 = \frac{1}{2}((2T_2 + 1)^3 + 1)$ . 本文提取 1 阶和 2 阶 SPJD 特征作为隐写检测特征, 特征维数共计  $2(k_1 + k_2) = (2T_1 + 1)^2 + (2T_2 + 1)^3 + 2$  维.

## 2 基于分割的空域图像隐写分析算法

通常, 隐写对图像统计特性所造成的改变小于图像内容的差异对隐写检测特征分布的影响, 因此隐写分析是一个“类内分散, 类间聚合”的分类过程. 图像是具有区域平稳性的 Markov 信源, 具有相同内容复杂度的图像区域有相似的像素分布特征. 对自然图像进行隐写之后, 具有强随机性的秘密信息对统计特性不同的图像子区域的会产生不同的影响, 而图像内容复杂度的差异可通过纹理复杂度反映出来. 因此, 可将图像按图像纹理复杂度分割为若干子图像, 使每类子图像的统计特性更加相近, 隐写检测特征分布更加聚合, 使载体与载密图像特征具有更佳的可分性, 降低图像内容差异对隐写检测性能的影响. 对非自适应隐写方法而言, 嵌密对图像平坦区域所造成的影响更易于检测, 而对自适应隐写方法, 根据图像的纹理复杂度提取相应的隐写检测特征, 能够更加敏感地检测秘密信息嵌入所造成的影响, 从而提高检测性能.

本文提出一种新的空域隐写分析方法. 将图像分为尺寸为  $S \times S$  的图像块, 根据子块的纹理复杂度特征对给定图像进行分割. 在训练过程中, 对分割所得到的每一类子图像分别提取隐写分析特征, 构建相应的分类器进行训练, 并根据训练准确率计算各分类器的权值. 在测试过程中, 将对每一幅子图像所提取的隐写检测特征送入相应的分类器, 通过加权融合得到对整幅图像的判决结果. 算法流程如图 2 所示.

### 2.1 图像分割

本算法中的图像分割作为隐写分析的预处理步骤, 不能有过的计算复杂度, 也不需要获得特别高的分割精度. 由于图像相邻像素间具有较强的相关

性, 距离近的像素点的统计特性会趋于一致, 因此本文采用对图像进行分块和分类的方法实现图像分割.

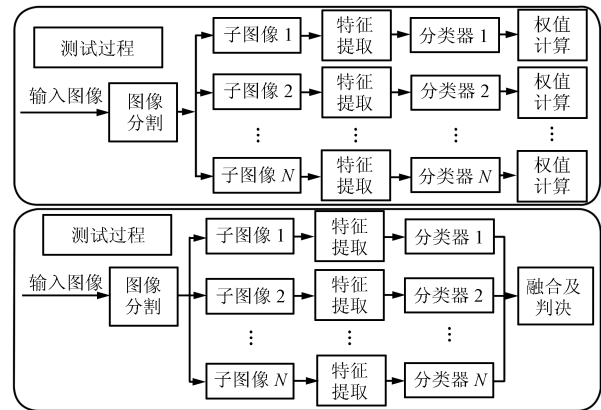


图 2 算法流程图

Fig. 2 The block-diagram of the proposed steganalyzer

设给定一幅尺寸为  $W \times H$  的图像  $I$ , 对每个像素点计算其与八邻域像素绝对差值的均值:

$$C(i, j) = \frac{1}{8} \sum_{m,n=-1}^1 |f(i, j) - f(i + m, j + n)| \quad (8)$$

其中  $i = 1, \dots, W$ ,  $j = 1, \dots, H$ ,  $|m| + |n| \neq 0$ . 将图像分成互不重叠的尺寸为  $S \times S$  的图像块, 对图像块  $B$  定义纹理复杂度度量特征:

$$TC = \frac{1}{S \times S} \sum_{i,j=1}^S C(i, j) \quad (9)$$

则可通过该特征值衡量图像块的纹理复杂度, 图像块的纹理越复杂, 该特征值越大.

设需将每幅图像的图像块分为  $N$  类, 则从给定图像库中随机选择  $K$  幅图像, 记做  $I_1, I_2, \dots, I_K$ , 对每幅图像的所有图像块计算纹理复杂度度量特征, 并将其平均分为  $N$  类, 使每类图像块的数目相等, 将此分类点记为  $P_i$ ,  $i = 1, \dots, N - 1$ . 对  $K$  幅图像计算聚类向量  $\mathbf{V}$ .

$$[v_1 \ v_2 \ \dots \ v_{N-1}] = [\text{MED}_{j=1}^K(P_1^j) \text{MED}_{j=1}^K(P_2^j) \ \dots \ \text{MED}_{j=1}^K(P_{N-1}^j)] \quad (10)$$

其中 MED 表示求中值. 根据聚类向量  $\mathbf{V}$  把每幅图像的图像块分为  $N$  类, 将每一类图像块合并为一个区域, 由此实现对图像的分割. 本文所采用的分割方法能够保证分割所得到的每幅子图像都是由形状规则的区域所构成, 这有利于后续对隐写分析特征的提取. 图 3 为对若干内容差异较大的图像进行分割得到的结果, 其中图像块尺寸为  $8 \times 8$ , 分割数目

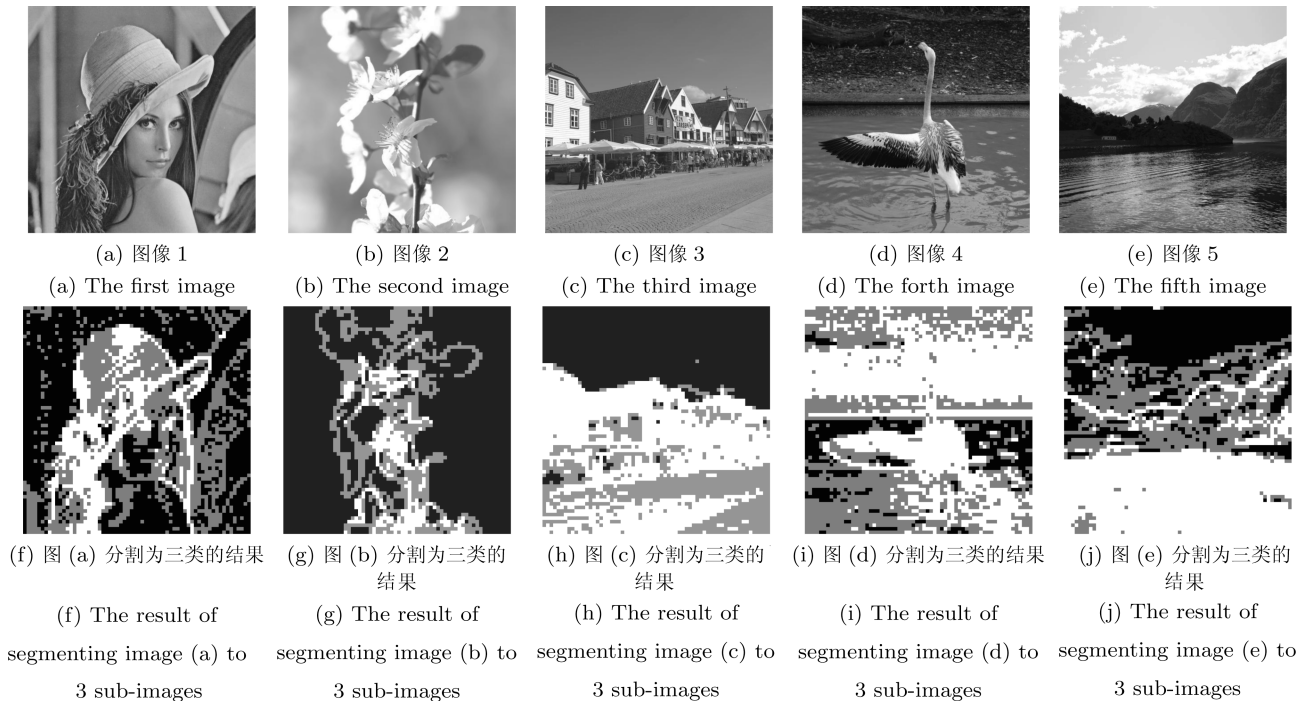


图 3 图像分割结果

Fig. 3 Results of image segmentation

为 3, 灰度值相同的像素构成了一幅子图像, 随着像素值由小至大, 子图像的纹理复杂度依次增高.

## 2.2 基于图像分割的 SPJD 特征

对于平坦图像而言, 进行差分后的像素有更多的 0 值和较小的差分像素值, 其分布会更加集中, 因此对纹理复杂度不同的子图像应设置不同的特征截断阈值, 提取不同维数的特征. 设将给定图像分割为  $N$  幅子图像, 并将子图像的序号按照纹理复杂度由低至高进行排列, 则将前  $\lfloor \frac{N}{3} \rfloor$  幅子图像判定为平坦区域, 后  $\lfloor \frac{N}{3} \rfloor$  幅子图像判定为复杂区域, 其余子图像为中等区域. 例如, 当  $N$  取值为 4 时, 认为第 1 幅子图像为平坦区域, 第 2、3 幅子图像为中等区域, 第 4 幅子图像为复杂区域. 在提取特征时, 根据子图像的纹理复杂度提取相应的特征. 对每类子图像分别提取第 1 节定义的 SPJD 特征, 并设定不同的特征阈值, 即 1 阶特征的  $T_1$  和 2 阶特征的  $T_2$ . 对于平坦子图像,  $T_1 = 2, T_2 = 1$ ; 对于中等子图像,  $T_1 = 3, T_2 = 2$ ; 对于复杂子图像,  $T_1 = 4, T_2 = 3$ . 由此可知, 本算法对平坦、中等、复杂三类子图像所提取的特征维数分别为 54 维、176 维和 426 维.

## 2.3 训练和测试

根据图 2 所示算法流程可知, 本算法对  $N$  类子图像的隐写检测特征分别进行训练和测试, 共构造了  $N$  个分类器, 根据训练时的检测正确率设定分类

器的权值, 对每幅子图像均可得到一个检测结果, 因此对一幅图像共得到  $N$  个判决结果. 为了得到一个较为准确的最终判决结果要, 需要将  $N$  个检测结果进行加权融合. 本算法根据每类分类器对训练图像得到的检测正确率 (Detection accuracy, DA) 分配权值. 当某一类子图像的 DA 值较高时, 会有更多的子图像被分至正确的类别, 此时对该类子图像的分类结果设定一个较大的加权重更利于最终的融合结果, 反之亦然. 设第  $i$  个分类器的 DA 值为  $a_i$ , 则其权值为:

$$w_i = \frac{a_i - 0.5}{\sum_{i=1}^N (a_i - 0.5)} \quad (11)$$

其中  $i = 1, \dots, N$ .

在融合阶段, 设每类子图像的判决结果取值域为 0, 1, 即

$$p_i = \begin{cases} 0, & \text{若第 } i \text{ 类图像被判为载体图像} \\ 1, & \text{若第 } i \text{ 类图像被判为载密图像} \end{cases} \quad (12)$$

其中  $i = 1, \dots, N$ . 则根据每类子图像的判决结果和其权值, 设定如下的融合规则:

$$P = \sum_{i=1}^N p_i \cdot w_i \quad (13)$$

当  $P \leq 0.5$  时, 将图像判为载体图像;  $P > 0.5$  时, 则判为载密图像. 通过此方法可以得到对整幅图像的判决结果.

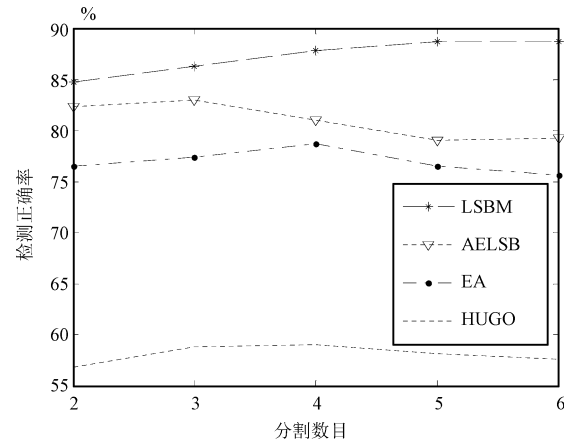
### 3 实验结果

为了更加全面和公平地测试算法性能, 采用在隐写分析中的通用图像库进行实验, 包括: BOSS-base 0.92<sup>[18]</sup>、BOWS2 图像库<sup>[19]</sup>、UCID 图像库<sup>[20]</sup> 以及由三库图像所组成的混合图像库. 其中混合图像库包括从前两个图像库中各随机选取的 3000 幅图像以及从 UCID 图像库中随机选取的 1000 幅图像, 共计 7000 幅图像. 这些图像库包含丰富的来源和内容, 统计特性差异明显. 采用不同的图像库对算法进行检测能够排除某个图像库对性能的影响, 更加全面地验证算法的性能.

对载体图像以不同的隐写方法和嵌入率生成载密图像库, 采用的隐写方法包括: LSBM、AELSB<sup>[2]</sup>、EA<sup>[3]</sup> 和 HUGO<sup>[4]</sup> 算法, 其中后三种为自适应隐写算法. 嵌入率根据每个隐写方法的检测难度进行选择, 以比特/像素值 (Bit per pixel, bpp) 为单位. 采用 SVM 分类器<sup>[21]</sup> 进行分类, 其中 50% 的图像用于训练, 其余 50% 用于测试.

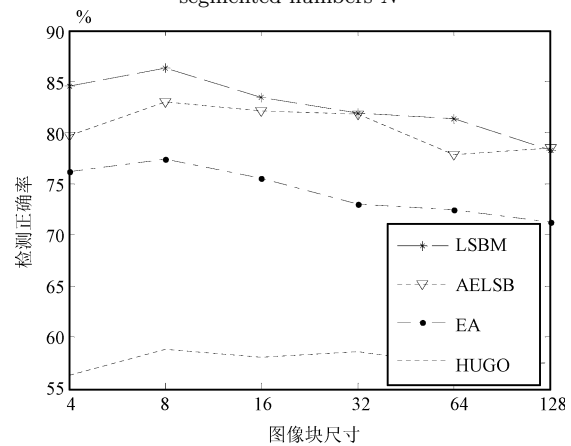
#### 3.1 图像分割参数的影响

本文所提算法的分割参数包括图像分割数目  $N$  和图像块尺寸  $S$ , 它们与算法性能息息相关. 如果  $N$  过小, 可能会使分割所得到的子图像包含纹理复杂度差异较大的区域, 分割不能得到良好的效果; 而当  $N$  变大时, 分割所得到的每一幅子图像的像素数会降低, 由于统计分析的有效性是建立在一定的数据量的基础之上, 如果每一幅子图像的数据量不足, 会影响算法的检测性能. 同时,  $N$  变大时算法复杂度也会随之增大. 当  $S$  过小时, 会产生过多的孤立图像块, 这会影响能够反映像素相关性的 Markov 特征; 当  $S$  过大时, 一幅子图像可能包含纹理复杂度差异较大的图像块. 为了选择最合适的分割参数, 从 BOWS2 库中随机选取 1000 幅图像进行测试, 对 4 种算法的嵌入率分别为 0.1 bpp、0.3 bpp、0.4 bpp 和 0.5 bpp. 在  $N$  和  $S$  取不同值时, 对 4 种隐写方法进行检测的正确率分别如图 4(a)、(b) 所示. 从图 4(a) 可以看出, 当  $N$  增大时, 除了对 LSBM 隐写算法的检测性能递增之外, 对其他算法的检测性能均为上凸曲线. 从图 4(b) 可以看出, 图像块尺寸过大或过小时, 对 4 种隐写方法的检测性能都会降低. 综合考虑算法的性能和复杂度, 图像分割数目  $N$  取 3, 将图像块尺寸  $S$  设为 8.



(a) 图像分割数目  $N$  不同时的检测正确率

(a) Detection accuracy of the proposed algorithm with different segmented numbers  $N$



(b) 图像块尺寸  $S$  不同时的检测正确率

(b) Detection accuracy of the proposed algorithm with different block sizes  $S$

图 4 分割参数对检测正确率的影响

Fig. 4 Relations between the segmentation parameters and detection accuracy

#### 3.2 性能比较

为了验证本文所提取的 SPJD 特征的有效性, 并检测图像分割对隐写检测性能的影响, 首先将本文算法与未采用分割的 SPJD 特征和文献 [10] 的 SPAM 特征进行对比. 在 BOWS 库随机选取 1000 幅图像进行测试, 测试的隐写方法包括 LSBM、AELSB 和 EA 算法, 实验结果如表 1 所示. 可以看出, 相较于 SPAM 特征, SPJD 特征能得到更高的检测率, 而基于图像分割的算法则进一步提高了检测性能. 这充分说明了图像分割和 SPJD 特征的有效性.

为了检验在不同情况下的算法性能, 分别在训练、测试图像库匹配及不匹配的情况下进行了实验. 针对 4 种隐写方法、2 种嵌入率, 在不同图像库上将

本文算法与 SPAM 算法及文献 [17] 算法进行了比较, 其中文献 [17] 用不同的线性模板对图像进行局部线性变换, 对变换残差提取直方图和共生矩阵特征进行隐写分析.

表 1 算法有效性 (%)

Table 1 Validity of the proposed algorithm (%)

隐写方法	嵌入率	SPAM	SPJD	本文算法
LSBM	0.1	77.84	82.83	86.52
AELSB	0.3	69.87	76.77	82.83
EA	0.4	68.67	71.36	77.23

表 2 给出了在训练、测试库完全匹配的情况下几种算法的检测正确率. 为了更公平地对算法性能进行比较, 对所有算法随机选择相同的图像用于训练和测试, 重复进行实验 5 次, 取结果的均值为对比指标. 从表 2 可以看出, 对于 4 种隐写方法, 本文算法在 4 个测试图像库上的性能均优于 SPAM 算法, 并且性能几乎均能提高 4 个百分点以上, 对 AELSB 的检测效果改进尤为明显, 在 BOWS 库嵌入率为 0.2bpp 时, 性能提升能达到 14 个百分点. 与文献 [17] 算法相比, 本算法对于 LSBM、AELSB 和 EA 算法的检测性能均有所提高. 由于 HUGO 算法是针对 SPAM 特征提出的隐写方法, 因此 SPAM 对其检测性能较差. 本文算法特征是基于 SPAM 特征提出的, 与文献 [17] 算法相比性能略差, 但较 SPAM 仍有明显的提高.

表 3 所示为在训练、测试库不匹配的情况下对几种算法的实验结果, 表中第一行所列出的为实验所采用的测试库, 其余两个库的图像混合作为训练库图像. 从表 3 可以看出, 本文算法性能仍明显优于 SPAM 算法和文献 [17] 算法, 并且性能的改进更加明显. 当训练、测试库不一致时, 文献 [17] 算法和本算法受到的影响并不明显, 展现了更优的性能, 这是因为文献 [17] 算法通过局部线性变换捕捉到了秘密信息嵌入对图像纹理所造成的微小改变, 而本算法则充分考虑了图像的内容差异和统计特性不同对隐写检测性能的影响, 这表明本算法更适合应用于实际.

### 3.3 不同区域特征对算法性能的贡献

为了研究分割后得到的不同纹理区域与检测性能之间的关系, 本节给出了分割数目  $N = 3$  时, 三类子图像的检测正确率. 从 BOWS2 库中随机选取 1000 幅图像进行测试, 对 4 种算法的嵌入率分别为 0.1bpp、0.3bpp、0.4bpp 和 0.5bpp, 三类子图像的检测正确率如图 5 所示.

表 2 训练、测试库匹配时的性能比较 (%)

Table 2 Comparison of detection accuracy when training and testing database are the same (%)

隐写方法	嵌入率	对比算法	BOSS	BOWS	UCID	混合库
LSBM	0.1	SPAM	79.25	84.94	64.69	81.90
		文献 [17] 算法	82.23	86.12	67.08	84.93
	0.2	本文算法	85.56	91.96	72.13	86.46
		SPAM	85.19	91.40	75.69	86.98
		文献 [17] 算法	88.92	93.03	79.23	89.08
		本文算法	90.91	95.27	82.06	91.02
AELSB	0.2	SPAM	66.02	65.71	56.94	64.33
		文献 [17] 算法	74.58	72.76	63.62	71.12
	0.3	本文算法	76.58	79.73	64.38	75.72
		SPAM	78.35	73.13	61.63	75.41
		文献 [17] 算法	84.83	79.87	69.46	79.13
		本文算法	85.52	84.73	72.54	82.01
EA	0.3	SPAM	70.40	67.46	53.94	66.36
		文献 [17] 算法	72.18	71.87	57.52	69.72
	0.4	本文算法	75.98	72.54	57.63	71.52
		SPAM	78.96	76.71	60.38	75.38
		文献 [17] 算法	81.53	79.62	62.78	77.72
		本文算法	83.79	80.19	67.88	80.21
HUGO	0.5	SPAM	55.71	58.40	51.31	56.51
		文献 [17] 算法	63.23	67.12	57.83	63.92
	0.6	本文算法	61.19	64.58	54.38	60.68
		SPAM	66.21	64.67	52.44	63.50
		文献 [17] 算法	73.23	75.98	59.12	71.05
		本文算法	71.33	72.31	56.13	68.05

从图 5 可以看出, 对于非自适应隐写算法 LSBM, 由于秘密信息均匀地分布在图像中, 具有强随机性的秘密信息对平坦区域统计特性的影响高于纹理区域, 因此对平坦子图像的检测结果明显优于其他区域. 对于自适应隐写算法 AELSB, 大量秘密信息自适应地嵌入到了纹理复杂的区域, 因此复杂子图像的检测结果更好. EA 算法将秘密信息嵌入在边缘区域, 因此中等子图像的检测结果更优. 注意到这三种算法对单个类别子图像的检测结果均优于 SPAM 算法, 并且由于单类子图像的检测性能明显优于其他类别子图像, 融合并没有起到提升性能的作用. 基于 SPAM 的特征对 HUGO 算法的检测效果不佳, 三类子图像的检测结果较为平均, 但平坦子图像的性能仍略优于其他纹理区域. 这种情况

下融合有可能提高最终的检测结果。

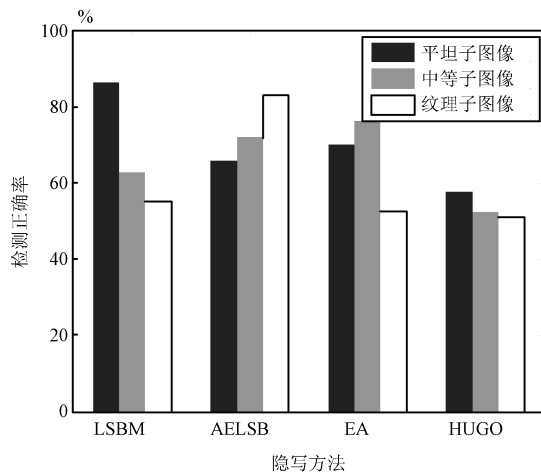


图 5 纹理复杂度不同的子图像的检测结果

Fig. 5 Detection results of each kind of sub-images with the same or close texture complexity

表 3 训练、测试库不匹配时的性能比较 (%)

Table 3 Comparison of detection accuracy when training and testing database are different (%)

隐写方法	嵌入率	对比算法	BOSS	BOWS	UCID
LSBM	0.1	SPAM	75.82	71.93	63.10
		文献 [17] 算法	79.87	75.76	65.79
		本文算法	82.38	77.68	69.55
	0.2	SPAM	81.97	76.88	68.85
		文献 [17] 算法	85.46	81.83	69.72
		本文算法	87.22	82.90	73.45
AELSB	0.2	SPAM	57.23	59.88	56.50
		文献 [17] 算法	69.85	67.62	59.98
		本文算法	75.43	68.87	63.20
	0.3	SPAM	67.75	71.65	60.75
		文献 [17] 算法	76.11	74.42	66.28
		本文算法	77.95	75.82	71.10
EA	0.3	SPAM	63.38	67.40	52.85
		文献 [17] 算法	67.58	68.92	55.43
		本文算法	68.60	72.08	56.75
	0.4	SPAM	73.42	75.08	58.03
		文献 [17] 算法	74.43	77.24	60.19
		本文算法	80.47	81.83	65.92
HUGO	0.5	SPAM	50.03	50.00	51.05
		文献 [17] 算法	57.03	56.19	54.42
		本文算法	57.58	57.10	53.03
	0.6	SPAM	60.78	58.62	51.67
		文献 [17] 算法	66.23	64.18	55.24
		本文算法	65.83	65.23	55.43

### 3.4 算法复杂度

虽然本文算法能够明显提高对各种隐写方法的检测性能,但是由于涉及到图像分割和对多幅子图像进行特征提取,本文算法复杂度要高于 SPAM 算法. SPAM 算法对八个方向的差分矩阵提取二阶转移概率矩阵,而本文算法对分割得到的三幅子图像进行特征提取,对每幅子图像四个方向的差分矩阵提取一阶和二阶联合分布概率. 设图像尺寸为  $M \times N$ , 则本文算法与 SPAM 算法相比约需多进行  $16MN - 33M - 33N + 57$  次加法运算. 文献 [17] 算法用 20 个不同尺寸的模板对图像进行卷积,并且需要提取共生矩阵特征,共需要进行约  $244MN$  次乘法运算和  $8MN - 6M - 6N + 4$  次加法运算. 实验所采用的计算机为联想 M4300t, 采用酷睿 i5-2400 处理器, 4 GB 内存, 采用对比算法作者所提供的源代码进行实验. 对于一幅图像提取 SPAM 特征需要约 0.552 秒, 提取本文所提特征则需要约 0.942 秒, 而提取文献 [17] 算法约需 1.667 秒. 进行训练分类时, 由于原始 SPAM 特征维数为 686 维, 本文算法虽然需要进行 3 次训练, 但训练的特征维数分别为 54 维、176 维和 426 维, 均低于 SPAM 特征, 因此三次训练分类的总时间基本与 SPAM 算法持平, 文献 [17] 算法特征维数为 120 维, 所耗费的训练时间最短. 对以 HUGO 算法嵌入率为 0.5 bpp 的 7000 幅图像进行训练分类, 文献 [17] 算法约需 24 分钟, SPAM 算法时间为 100 分钟, 本文算法为 102 分钟.

## 4 结论

本文提出了一种基于图像分割的空域隐写检测方法. 该算法根据图像块的纹理复杂度对图像  $8 \times 8$  分块进行分类, 实现对图像的分割, 使得分割所得到的每一类子图像具有相同或相近的纹理复杂度, 对每类子图像分别提取改进的 SPAM 特征, 并构造相应的分类器进行训练和测试, 通过加权融合得到对整幅图像的检测结果. 实验结果表明, 本文所提出的改进特征相较于 SPAM 特征有更高的检测正确率. 基于图像分割的方法充分考虑了图像的内容差异, 能够有效提高空域盲检测算法的性能, 尤其是针对自适应隐写方法, 性能提升更加明显.

## References

- 1 Sharp T. An implementation of key-based digital signal steganography. In: Proceedings of 4th International Workshop on Information Hiding. Heidelberg: Springer-Verlag, 2001. 13–26
- 2 Yang C, Weng C, Wang S. Adaptive data hiding in edge areas of images with spatial LSB domain systems. *IEEE*

- Transactions on Information Forensic Security*, 2008, **3**(3): 488–497
- 3 Luo W Q, Huang F J, Huang J W. Edge adaptive image steganography based on LSB matching revisited. *IEEE Transactions on Information Forensic Security*, 2010, **5**(2): 201–214
  - 4 Pevný T, Filler T, Bas P. Using high-dimensional image models to perform highly undetectable steganography. In: *Proceedings of 12th International Workshop on Information Hiding*. Heidelberg: Springer-Verlag, 2010. 161–177
  - 5 Avci I, Memon N, Sankur B. Steganalysis using image quality metrics. *IEEE Transactions on Image Processing*, 2003, **12**(2): 221–229
  - 6 Farid H. Detecting hidden messages using higher-order statistical models. In: *Proceedings of IEEE International Conference on Image Processing*. New York, USA: IEEE, 2002. 905–908
  - 7 Goljan M, Fridrich J, Holotyak T. New blind steganalysis and its implications. In: *Proceedings of SPIE Security, Steganography, and Watermarking of Multimedia*. San Jose, USA: SPIE, 2006. 1–13
  - 8 Xuan G R, Shi Y Q, Gao J H, Zou D K, Yang C Y, Zhang Z P, Chai P Q, Chen C H, Chen W. Steganalysis based on multiple features formed by statistical moments of wavelet characteristic functions. In: *Proceedings of 7th International Workshop on Information Hiding*. Heidelberg: Springer-Verlag, 2005. 262–277
  - 9 Sullivan K, Madhow U, Chandrasekaran S, Manjunath B. Steganalysis for Markov cover data with applications to images. *IEEE Transactions on Information Forensic and Security*, 2006, **1**(2): 275–287
  - 10 Pevný T, Bas P, Fridrich J. Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensic Security*, 2010, **5**(2): 215–224
  - 11 Fridrich J, Kodovský J. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 2012, **7**(3): 868–882
  - 12 Holub V, Fridrich J, Denemark T. Random projections of residuals as an alternative to co-occurrences in steganalysis. In: *Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XV*. San Francisco, CA: SPIE, 2013. 86650L
  - 13 Holub V, Fridrich J. Random projections of residuals for digital image steganalysis. *IEEE Transactions on Information Forensics and Security*, 2013, **8**(12): 1996–2006
  - 14 Kodovský J, Fridrich J, Holub V. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, 2012, **7**(2): 432–444
  - 15 Amirkhani H, Rahmati M. New framework for using image contents in blind steganalysis systems. *Journal of Electronic Imaging*, 2011, **20**(1): 013016-1–013016-14
  - 16 Cho S, Cha B, Wang J, Kuo C C J. Block-based image steganalysis: algorithm and performance evaluation. In: *Proceedings of IEEE International Symposium on Circuits and Systems*, Piscataway. Paris: IEEE, 2010. 1679–1682
  - 17 Xiong G, Ping X, Zhang T, Hou X. Image textural features for steganalysis of spatial domain steganography. *Journal of Electronic Imaging*, 2012, **21**(3): 033015-1–033015-15
  - 18 Filler T, Pevný T, Bas P. BOSS. [Online], available: <http://boss.gipsa-lab.grenobleinp.fr/BOSSRank/>. June 21, 2011
  - 19 Bas P, Furon T. Bows-2. [Online], available: <http://bows2.gipsa-lab.inpg.fr/BOWS2OrigEp3.tgz>. July 21, 2011
  - 20 Schaefer G, Stich M. UCID - An Uncompressed Colour Image Database, Technical Report, School of Computing and Mathematics, Nottingham Trent University, U.K, 2003
  - 21 Chang C, Lin C. LIBSVM: a library for support vector machines. [Online], available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. September 16, 2011



汪然 中国人民解放军信息工程大学信息工程学院博士研究生。主要研究方向为信息隐藏和隐写分析。本文通信作者。

E-mail: wangran721@gmail.com

(WANG Ran Ph.D. candidate at the School of Information Systems Engineering, PLA Information Engineering University. Her research interest covers information hiding and image steganalysis. Corresponding author of this paper.)



许漫坤 中国人民解放军信息工程大学信息工程学院讲师。主要研究方向为图像处理和信息隐藏。

E-mail: maggiexmk@163.com

(XU Man-Kun Lecturer at the School of Information Systems Engineering, PLA Information Engineering University. Her research interest covers image processing and information hiding.)



平西建 中国人民解放军信息工程大学信息工程学院教授。主要研究方向为图像处理, 模式识别和信息安全。

E-mail: pingxijian@163.com

(PING Xi-Jian Professor at the School of Information Systems Engineering, PLA Information Engineering University. His research interest covers image processing, pattern recognition, and information security.)



张涛 中国人民解放军信息工程大学信息工程学院副教授。主要研究方向为图像处理和信息隐藏。

E-mail: dr.zhangtao@gmail.com

(ZHANG Tao Associate professor at the School of Information Systems Engineering, PLA Information Engineering University. His research interest covers image processing and information hiding.)