文章编号：1001-0920(2009)08-1239-04

# 基于马尔可夫模型和特征融合的图像隐写分析

孙子文，纪志成
（江南大学 通信与控制工程学院，江苏 无锡 214122）

**摘 要**：提出一种针对JPEG图像隐写的通用隐写分析方法.根据量化后分块DCT系数绝对值构造水平、垂直和zigzag方向的差分数组，利用三向差分数组马尔可夫模型挖掘量化后分块DCT块内邻近系数相关性，提取转移概率矩阵的特征.对三向特征加权融合后进行隐写分析，以提高分类性能.对安全性较高的JPEG隐写OutGuess和F5，在不同嵌入率下进行隐写分析.实验结果显示，引入特征融合后隐写分析的检出率明显提高.

**关键词**：马尔可夫模型；特征融合；隐写分析
**中图分类号**：TP391 **文献标识码**：A

# Image steganalysis based on Markov model and feature fusion

*SUN Zi-wen*, *JI Zhi-cheng*
(College of Communication and Control Engineering, Wuxi 214122, China. Correspondent: SUN Zi-wen, E-mail: sunziwen@jiangnan.edu.cn)

**Abstract**: An universal steganalysis scheme to attack JPEG steganography is presented. Difference arrays along horizontal, vertical and zigzag directions are formed by using the magnitudes of quantized block DCT coefficients. Markov process is applied to model these difference arrays so as to utilize the second order statistics for steganalysis. Feature vectors are derived from the Markov matrix to mining the quantified block DCT neighborhood coefficients' correlation. A weighted feature fusion method is used to fuse these feature vectors to improve the classification accuracy. The experimental results show that the proposed scheme has the advantage in detection rate in attacking OutGuess and F5.

**Key words**: Markov model; Feature fusion; Steganalysis

## 1 引 言

继密码技术之后，隐写术为网络通信提供了一种全新的解决方案，但也为非法传播提供了可能.隐写分析技术是检测载体中是否存在隐藏信息，它可分为专用隐写分析技术和通用隐写分析技术，后者不依赖于具体隐写算法.由于存在数字图像表示的高度冗余性，研究以数字图像作掩体对象的隐写术呈现增长的趋势，相应的图像隐写分析技术已成为研究的热点.

通用隐写分析方法实质上是多维特征空间的模式分类器.Farid等提出了基于小波高阶统计量的通用隐写分析方法[1]，为通用隐写分析指明了方向；Fridrich使用邻近DCT系数共生矩阵为隐写分析特征[2]，实现了对JPEG图像上Jsteg，F5和OutGuess等隐写术的成功检测，但特征提取比较复杂；Sullivan等提出了基于马尔可夫模型的隐写分析方法[3]，通过选取图像共生矩阵主对角线及其附近的系数作为特征，达到了降维的目的，但却丢失了部分特征信息；Shi等定义了差分JPEG 2-D数组[4]，并沿差分数组的水平、垂直和对角方向提取一步转移概率矩阵的特征，进行隐写分析并取得了较好的检测效果，但差分数组的构造不是分块进行的，不能很好地体现DCT块内系数的相关性.

信息融合技术能提高分类性能，在生物特征识别领域得到广泛的应用[5,6]，在信息隐藏领域也得到初步的研究.Cai等提出在隐写分析中引入特征层融合技术[7]，将DCT域特征与空域特征相融合，但融合权值的分配缺乏理论依据；Rodriguez等采用决策层融合，应用多类分类平均融合的方法提高检测率[8]，但权重计算复杂，且检出率有待提高；Kharrazi等对不同隐写方法分类结果进行决策层融

**收稿日期**：2008-09-16；**修回日期**：2008-12-12.
**基金项目**：国家自然科学基金项目(60774030)；江南大学青年科学基金项目(52210754).
**作者简介**：孙子文(1968—)，女，四川大竹人，副教授，从事信息隐藏技术、图像处理的研究；纪志成(1959—)，男，杭州人，教授，博士生导师，从事非线性控制、智能控制等研究.

合[9]. 上述研究结果表明,应用信息融合技术比非融合隐写分析能取得更好的检测性能.

常用的 JPEG 类隐写方法(如 OutGess 等),都尽量保护一阶统计特征 DCT 系数直方图,以抵抗基于直方图的隐写分析. 本文受文献[4-6]的启发,提出一种基于 DCT 域块内水平、垂直和 zigzag 三向差分数组马尔可夫模型的通用隐写分析方法. 为有效地捕捉图像 DCT 块内邻近系数的相关性,首先对图像分块 DCT 系数绝对值进行三向差分计算,分别提取三向差分数组的一步转移概率矩阵为马尔可夫矩阵;然后运用信息融合技术,按其对分类识别性能的贡献大小,对三向马尔可夫矩阵分配权值进行加权融合. 实验结果表明,本文基于加权特征融合的方法在检测性能上高于只取单向马尔可夫矩阵为特征的识别方法,对于不同隐写容量的 OutGuess 和 F5 隐写方法,检出率均高于 94%.

## 2 特征构造方法

隐写分析问题可看作二类模式识别问题,主要涉及到两方面内容:有效特征的提取和分类器的设计. 本文对测试图像进行 DCT 变换,建立三向差分数组,差分数组可看作马尔可夫随机过程. 根据随机理论,马尔可夫随机过程可由转移概率矩阵来表征. 按其转移方向和转移距离,可分为一步转移概率矩阵和多步转移概率矩阵. 基于 DCT 域的信息隐写直接影响其向邻近系数的转移,因此本文计算沿水平、垂直和 zigzag 扫描方向的三向一步转移概率矩阵,以其特征来揭示隐写的存在.

### 2.1 三向差分数组

为了体现 DCT 块内系数的相关性,本文提出重新设计特征,并结合信息融合技术构造特征,以便有效地提高检出率.

#### 2.1.1 生成三向差分数组

读取 JPEG 图像,进行 8×8 DCT 变换和 JPEG 量化. 对各系数取绝对值,采用水平光栅扫描的方式,扫描一幅图像的各 DCT 块,设共有 $l$ 块. 目前的主流 JPEG 隐写方法仅对非 0 的 AC 系数进行嵌入. 随机抽取 3 幅图像统计,结果表明 21 个低频 AC 系数占全部非零 AC 系数的 92% 以上. 对于每个 DCT 分块 $B_i(u,v)(i=1,2,\cdots,l,\ u,v=1,2,\cdots,8)$ 的 64 维向量,只取低频部分 21 个 AC 系数,分别按水平、垂直和zigzag 3个方向扫描,生成 3 个一维数组 $B_{i_h}(u,v),B_{i_v}(u,v),B_{i_z}(u,v)$.

根据三向一维数组构造三向差分数组. 水平差分数组构造公式为

$$D_{i_h}(k+1)=B_{i_h}(k+1)-B_{i_h}(k),\ k\in[1,20]. \tag{1}$$

同理,构造垂直和 zigzag 差分数组 $D_{i_v}$ 和 $D_{i_z}$.

某一图像经 OutGess0.13 隐写前后,zigzag 差分数组直方图的分布情况如图 1 所示. 隐写前接近 90% 的差分值集中在 0 值,近似拉普拉斯分布;隐写后这种分布特性遭到破坏. 对 400 幅图片作隐写前后差分值的平均分布情况比较分析,结果表明隐写后平均差分值集中在 0 值的下降 10% 左右,说明隐写后 DCT 系数间的相关性减弱. 沿水平和垂直方向的实验也取得了类似的结果. 因此可依据差分数组进行隐写分析.

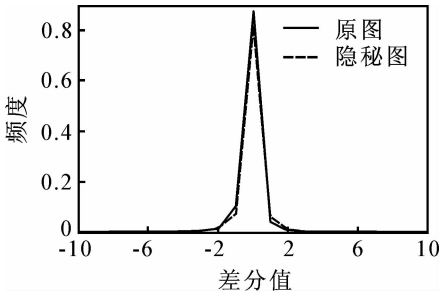


**图 1 图像原图及 F5 隐写后垂直差分矩阵直方图**

#### 2.1.2 差分数组阈值处理

差分数组值的变化范围较广,但大多集中在一个较小的范围内,并且过大的差分值是由图像本身内容引起的,而非隐写嵌入过程导致的. 为去除图像本身内容的影响以及缩小差分值的变化范围,以降低马尔可夫矩阵维数,对差分数组进行阈值处理,将差分值嵌位在$[-T,T]$范围内. 横向差分数组阈值处理规则为

$$D'_{i_h}=\begin{cases}T, & D_{i_h}>T;\\ -T, & D_{i_h}<-T;\\ D_{i_h}, & \text{otherwise.}\end{cases} \tag{2}$$

用相同的方法对垂直和 zigzag 方向的差分数组阈值进行处理,可得到 $D'_{i_v}$ 和 $D'_{i_z}$.

对于实验中选取的 400 幅图片,统计差分值在不同$[-T,T]$范围内所占比重,结果差分值在$[-5,5]$范围的比重在 96% 以上. 因此,本文将差分数组值嵌位在$[-5,5]$的范围内,以减小特征向量的维数.

### 2.2 二阶统计量马尔可夫矩阵

上述差分数组具有马尔可夫随机特性. 根据随机过程理论,可用一步转移概率矩阵来表征马尔可夫过程[4]. 差分值嵌位在$[-5,5]$后,马尔可夫矩阵的维数为$(2T+1)\times(2T+1)=11\times11$. 沿水平方向阈值差分数组的局部马尔可夫矩阵计算公式为

$$G_{i_h}(m,n)=$$
$$p\{D'_{i_h}(k+1)=n\mid D'_{i_h}(k)=m\}+$$
$$p\{D'_{i_h}(k)=n\mid D'_{i_h}(k+1)=m\}=$$

$$\frac{\sum_k \delta(D'_{i_h}(k)=m, D'_{i_h}(k+1)=n)}{\sum_k \delta(D'_{i_h}(k)=m)}+$$
$$\frac{\sum_k \delta(D'_{i_h}(k+1)=m, D'_{i_h}(k)=n)}{\sum_k \delta(D'_{i_h}(k+1)=m)},$$
$$k\in[1,20],\ m,n\in\{-5,\cdots,0,\cdots,5\}. \quad (3)$$

其中

$$\delta(A,B)=\begin{cases}1, & A=m, B=n;\\ 0, & \text{otherwise.}\end{cases} \quad (4)$$

用相同的方法计算垂直和 zigzag 方向差分数组的马尔可夫矩阵，可得到阈值三向差分数组的局部马尔可夫矩阵 $G_{i_h}$，$G_{i_v}$ 和 $G_{i_z}$. 以上公式计算的马尔可夫矩阵是对称矩阵，因此只取上三角矩阵共 66 维作为隐写分析的特征.

为了消除 DCT 块的影响，计算阈值横向差分数组的全局马尔可夫矩阵

$$G_h(m,n)=\frac{1}{l}\sum_{i=1}^{l}G_{i_h}(m,n), \quad (5)$$

其中 $l$ 为 DCT 块总数.

用相同的方法可计算出垂直和 zigzag 方向差分数组的全局马尔可夫矩阵.

### 2.3　基于特征融合的隐写分析

#### 2.3.1　信息融合技术

决策是由证据到结论的处理过程. 至少存在两种决策情况：1) 对于同样的证据，采用不同的决策函数可能获得相同或不同的结论；2) 对于不同的证据，采用同样的决策函数可能获得相同或不同的结论. 考虑将多个决策融合成一个最终的决策. 利用信息融合技术可实现数据层、特征层和决策层 3 个层次的融合分类问题.

特征层融合是将由多源数据中分别提取的特征汇总在一起，形成新的特征集，据此新特征集进行决策. 在特征层融合分类中，分类的计算对象是由多源数据分别或融合提取的特征，即由几个传感器获取的与目标相关的参数组合成的特征矢量，或依据单传感器多独立特征组合成的特征矢量，在此特征矢量的基础上进行决策.

融合特征矢量的方法主要有平均方法和加权平均方法. 平均方法取所有 $k$ 个特征的平均值，即

$$C_H=\frac{1}{k}\sum_{i=1}^{k}C_{H_i}; \quad (6)$$

加权平均方法取所有 $k$ 个特征的加权平均值，即

$$C_H=\sum_{i=1}^{k}w_iC_{H_i}. \quad (7)$$

权值满足

$$\sum_{i=1}^{k}w_i=1, \quad (8)$$

其中权值 $w_i$ 依据各特征分类性能的优劣予以分配.

#### 2.3.2　加权特征融合隐写分析

各特征并不具有相同的影响和被重视程度，应依据各自的识别性能确定其权重. 对于检测性能贡献大的应给予更高的重视，即获得大的权值. 根据 JPEG 类隐写方法，水平方向和垂直方向的特征对隐写的敏感性基本相同，zigzag 方向的特征对隐写的敏感性则要低于前两个方向的特征敏感性. 因此在特征融合时，应分配给水平和垂直方向的特征相同的权重，分配给 zigzag 方向较低的权重.

在取得三向差分数组的全局马尔可夫矩阵后，按一定的权重对这些特征进行数据融合，形成新的特征. 融合特征为

$$G=\sum_i w_iG_i, \quad (9)$$

其中 $w_i$ 取 $w_h$，$w_v$ 和 $w_z$，分别为水平、垂直和 zigzag 方向的权重. 加权融合权重之比 $w_h:w_v:w_z=4:4:2$.

## 3　仿真结果及性能分析

选择 800 幅彩色图像，图像内容包括人物肖像、自然风景、人造设施、动物等. 用 OutGuess 方法生成嵌入率分别为 0.05 bpc，0.1 bpc 和 0.2 bpc 的隐写图像各800幅；用F5方法生成嵌入率分别为0.05 bpc，0.1 bpc，0.2 bpc 和 0.4 bpc 的隐写图像各 800 幅. 对三向马尔可夫矩阵进行加权融合构造特征矢量，并用 SVM 作为分类器进行隐写分析. 为得到各向特征对分类识别性能的影响，用各向局部马尔可夫矩阵的特征分别进行分类测试，根据测试性能决定分配权重 $w_h$，$w_v$ 和 $w_z$，并进行三向特征融合.

在用每种特征检测时，针对不同嵌入容量的不同隐写算法，任选 440 幅原图及其对应的 440 幅隐写图片作为训练样本，剩余的 360 幅原图及其对应的 360 幅隐秘图片作为检测样本. 为验证权值分配的合理性，还进行了均值融合和 3∶3∶3 加权融合. 仿真结果如表 1 所示. 其中：$T_P$ 表示正确接受率，$T_N$ 表示正确否定率，$T$ 表示总的检出率.

从表 1 数据可知，4∶4∶2 加权融合方法的检测性能明显高于均值融合和 3∶3∶4 加权融合方法；对于 7 种不同隐写率和不同隐写方法，加权融合方法的检出率均达到 94% 以上. 通过 3 种不同权值融合策略的检测性能对比，表明了权值确定依据的正确性. 采用特征融合的隐写分析方法，检测性能得到明显提高，尤其是在低嵌入率为 0.05 bpc 时，F5 和 OutGess 隐写方法的检出率仍达到 94% 以上.

**表 1 单取某一方向的特征及几种权值三向特征融合的检测率**

| 隐写方法 | 隐写率/bpc | 水平方向 | | | 垂直方向 | | | zigzag 方向 | | | 均值融合 | | | 3∶3∶4 加权融合 | | | 4∶4∶2 加权融合 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $T_P$ | $T_N$ | $T$ | $T_P$ | $T_P$ | $T$ | $T_P$ | $T_N$ | $T$ | $T_P$ | $T_N$ | $T$ | $T_P$ | $T_N$ | $T$ | $T_P$ | $T_N$ | $T$ |
| Out-Gess | 0.05 | 78.9 | 95.5 | 87.2 | 94.4 | 84.4 | 89.4 | 93.3 | 77.8 | 85.6 | 81.1 | 92.2 | 86.7 | 83.3 | 94.4 | 88.9 | 95.3 | 94.1 | 94.7 |
| | 0.10 | 92.2 | 95.5 | 93.9 | 97.8 | 88.9 | 93.4 | 97.8 | 72.2 | 85.0 | 88.9 | 92.2 | 90.6 | 88.9 | 91.1 | 90.0 | 98.9 | 96.7 | 97.8 |
| | 0.20 | 93.3 | 96.7 | 95.0 | 96.7 | 91.1 | 93.9 | 93.3 | 96.7 | 95.0 | 83.3 | 96.7 | 90.0 | 81.1 | 97.8 | 89.5 | 100 | 96.7 | 98.4 |
| F5 | 0.05 | 78.9 | 93.3 | 86.1 | 84.4 | 93.3 | 88.9 | 92.2 | 85.6 | 88.8 | 87.8 | 94.4 | 91.1 | 92.2 | 95.6 | 93.9 | 98.9 | 88.9 | 94.0 |
| | 0.10 | 86.7 | 90.0 | 88.3 | 83.3 | 97.8 | 90.6 | 92.2 | 88.9 | 90.6 | 84.4 | 98.9 | 91.7 | 88.9 | 97.8 | 93.4 | 96.7 | 91.2 | 94.0 |
| | 0.20 | 90.0 | 95.5 | 92.8 | 91.1 | 96.6 | 93.9 | 96.7 | 86.7 | 91.7 | 86.7 | 98.9 | 92.8 | 88.9 | 98.9 | 93.9 | 97.8 | 92.2 | 95.0 |
| | 0.40 | 92.2 | 96.7 | 94.5 | 98.9 | 88.9 | 93.9 | 97.8 | 85.6 | 91.7 | 91.1 | 97.8 | 94.5 | 92.2 | 96.7 | 94.5 | 97.8 | 97.8 | 97.8 |

## 4 结 论

本文在图像 DCT 域构造三向差分数组的马尔可夫矩阵作为隐写分析的特征.运用基于信息融合的思想,采用加权融合策略进行三向特征融合构造特征,有效地捕捉到因嵌入信息导致的图像中内在统计特性的改变.实验结果表明,该算法的检测率得到提高,误检率有所降低.

## 参考文献(References)


[1] Farid H, Siwei L. Detecting hidden messages using higher-order statistics and support vector machines[C]. Proc of 5th Int Workshop on Information Hiding. Heidelberg: Springer, 2002: 340-354.

[2] Fridrich J. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes[C]. Proc of 6th Int Workshop on Information Hiding. Heidelberg: Springer, 2004: 67-81.

[3] Sullivan K, Madhow U, Chandrasekaran S, et al. Steganalysis of spread spectrum data hiding exploiting cover memory[C]. Proc of Security, Steganography and Watermarking of Multimedia Contents-Ⅶ. San Jose: SPIE, 2005: 38-46.

[4] Shi Y Q, Chen C, Chen W. A Markov process based approach to effective attacking JPEG steganography[C]. Information Hiding Workshop 2006. Heidelberg: Springer, 2006: 249-264.

[5] 董火明, 高隽, 汪荣贵. 多分类器融合的人脸识别与身份认证[J]. 系统仿真学报, 2004, 16(8): 1849-1853. (Dong H M, Gao J, Wang R G. Fusion of multiple classifiers for face recognition and person authentication [J]. J of System Simulation, 2004, 16(8): 1849-1853.)

[6] Ekenel H K, Sankur B. Multiresolution face recognition [J]. Image and Vision Computing, 2005, 23(5): 469-477.

[7] Cai H, Agaian S S. Spatial-frequency feature vector fusion based steganalysis [C]. IEEE Int Conf on Systems, Man and Cybernetics. Taipei: IEEE Press, 2006: 1866-1870.

[8] Rodriguez B M, Peterson G L, Agaian S S. Multi-class classification averaging fusion for detecting steganography[C]. Proc of IEEE Int Conf on Systems Engineering. San Antonio: IEEE Press, 2007: 1-5.

[9] Kharrazi M, Sencar T H, Memon N. Improving steganalysis by fusion techniques: A case study with image steganography [C]. Proc of Security, Steganography and Watermarking of Multimedia Contents-Ⅷ. San Jose: SPIE, 2006: 123-137.


(上接第 1238 页)


[9] 佟绍成, 王铁超. 一类非线性互联系统的模型参考跟踪模糊 $H_\infty$ 控制[J]. 控制与决策, 2006, 21(6): 612-618. (Tong S C, Wang T C. Fuzzy model reference tracking $H_\infty$ control for nonlinear interconnected systems[J]. Control and Decision, 2006, 21(6): 612-618.)

[10] Wei-Wei Lin, Wen-June Wang, Shu-Han Yang. A novel stabilization criterion for large-scale T-S fuzzy systems [J]. IEEE Trans on Systems, Man and Cybernetics — Part B, 2007, 137(4): 1074-1079.