

Research on Steganalysis of Spatial Rich Model Based on Image Texture

Zixuan Liu, Jianbin Wu*, Qian Wu

Central China Normal University, Wuhan Hubei
Email: *wujianbin@mail.ccnu.edu.cn

Received: Apr. 15th, 2020; accepted: Apr. 30th, 2020; published: May 7th, 2020

Abstract

The complexity of image texture is often concerned when steganalysis is used for attacking adaptive steganography, while spatial rich model is often used to attack the traditional image steganalysis. But the spatial rich model does not consider the texture features of image texture. Considering the accuracy of the spatial rich model and the effect of the image texture complexity on the adaptive algorithm, considering both efficiency and accuracy, this paper combines two ideas, introduces the image texture complexity into the spatial rich model, and proposes a steganalysis method based on the image texture. The test environment is built by MATLAB. The experimental results show that this method can improve the detection accuracy of the adaptive steganography algorithm.

Keywords

Steganography, Steganalysis, Spatial Rich Model, Texture Complexity

基于图像纹理的空域富模型隐写分析研究

刘子璇, 吴建斌*, 伍 迁

华中师范大学, 湖北 武汉
Email: *wujianbin@mail.ccnu.edu.cn

收稿日期: 2020年4月15日; 录用日期: 2020年4月30日; 发布日期: 2020年5月7日

摘 要

图像纹理复杂度是隐写分析在攻击自适应隐写时常常关注的问题, 而空域富模型(Spatial Rich Model)常用于针对传统非自适应图像空域隐写的攻击, 但空域富模型却没有考虑图像纹理的纹理特征。考虑到空域富

*通讯作者。

模型的准确率及图像纹理复杂度对自适应算法的作用, 兼顾效率与准确率, 本文结合两种思路, 将图像纹理复杂度引入到空域富模型中, 提出了一种基于图像纹理的空域富模型隐写分析方法。用MATLAB搭建了测试环境进行测试, 通过实验结果可以看出, 该方法能提高对于自适应隐写算法的检测准确率。

关键词

隐写术, 隐写分析, 空域富模型, 空纹理复杂度

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

图像隐写是将秘密信息按照一定的规律嵌入到载体图像中, 生成含密图像, 使得秘密信息传输难以被人察觉。传统的空域隐写算法 LSB [1]或者 JPEG 域图像隐写方法 Outguess [2]、nsF5 [3]等, 他们通过修改空域或 DCT 域系数, 将二进制比特流嵌入到图像中。然而这样的隐写算法难以抵挡 SRM [4]、DCTR [5]等方法的攻击。

为抵抗隐写分析者的检测, 隐写算法的设计者在设计隐写算法时引入 STC 编码方法[6]及嵌入失真函数[7], 将秘密信息嵌入到图像中难以被统计建模的位置, 而这些位置大多为纹理复杂度较高的区域, 如 HUGO [8]、WOW [9]、UNIWARD [10]等方法。为了针对此类自适应隐写算法, 基于 DCTR 的 SCA-DCTR [11]隐写分析方法也相应产生, 此方法是由 DCTR 加入对于像素嵌入概率的估计, 从而使得针对自适应隐写方式有着较高的检测准确率。但是基于 SRM 隐写分析方法却没有类似的方式提高其对于自适应隐写算法的分析准确率。

其中 SRM 是一种准确度较高的空域高维度图像特征, 通过各种滤波器对图像进行滤波处理后得到相应的特征量。然而 SRM 只考虑的像素点及其相邻位置像素的相关性, 并没有考虑图像的纹理特征, 没有对隐写时的未知进行预测, 故在对于自适应隐写方法进行分析的时候对于准确率会有一定的降低。

针对此问题, 本文在 SRM 特征上进行一定的改进, 将其与图像的纹理特征进行快速结合, 使其对于自适应隐写算法进行隐写分析时准确率有所提高。

2. 空域富模型

SRM 是于 2012 年由 Fridrich 等人提出的一种高维度的空域隐写分析图像特征, 是一种传统的通用隐写分析方法。通用隐写分析是不针对一种隐写算法的隐写分析方法, 其步骤是通过提取已知图像的统计特征训练分类器, 再利用训练好的分类器判别未知图像是否含密, 隐写分析的步骤如图 1 所示。其中特征提取及分类器训练为传统通用隐写分析中较为重要的步骤, SRM 即为一种建模的特征。

在文献[4]中指出, SRM 隐写分析特征的抽取步骤分为: 提取残差图像、残差图像的量化及截断、提取共生矩阵。

2.1. 提取残差图像

由于图像隐写是将二进制比特流嵌入到载体图像的空域或 JPEG 域中, 等同于是将一个微弱的噪声与载体图像叠加, 生成一个含密图像。这样将噪声叠加进来就会改变原始图像中相邻像素的相关性, 残差图像亦然, 故使用残差图像作为特征抽取的基底, 这样可以减少图像内容对于隐写分析特征抽取的影响。

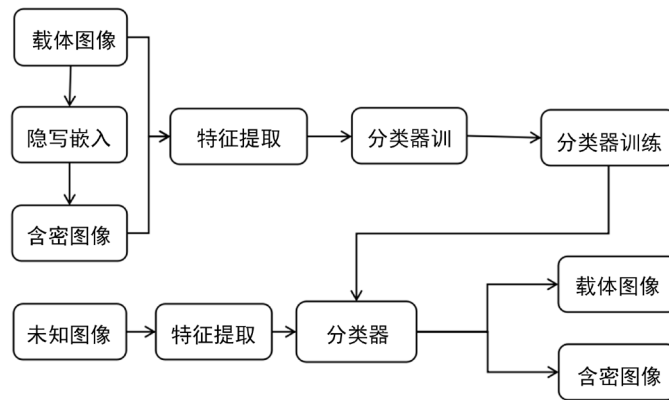


Figure 1. Steps of general steganalysis
图 1. 通用隐写分析步骤

若灰度图像为 $X = X_{i,j} \in R^{n_1 \times n_2}, i = 1, \dots, n_1, j = 1, \dots, n_2$ ，经过 SRM 高通滤波器滤波可以得到残差图像 $R = (R_{i,j}) \in R^{n_1 \times n_2}$ ：

$$R_{i,j} = \hat{X}_{i,j} - X_{i,j}, i = 1, \dots, n_1, j = 1, \dots, n_2 \quad (2.1)$$

公式 2.1 中即为对图像进行高通滤波后得到的残差图像，其中 $\hat{X}_{i,j}$ 是根据图像中像素点 $X_{i,j}$ 的领域对 $X_{i,j}$ 的估计。

不同的高通滤波器可以对像素点的不同领域进行估计，所以 Fridrich 设计了多种多样的高通滤波去使对像素的评估模型更加全面和丰富，滤波器模型如图 2 所示。

在图 2 中 • 表示滤波器的中心，也就是待估计的像素点，其余的符号则为各式滤波器，相同符号表示相同的滤波器。

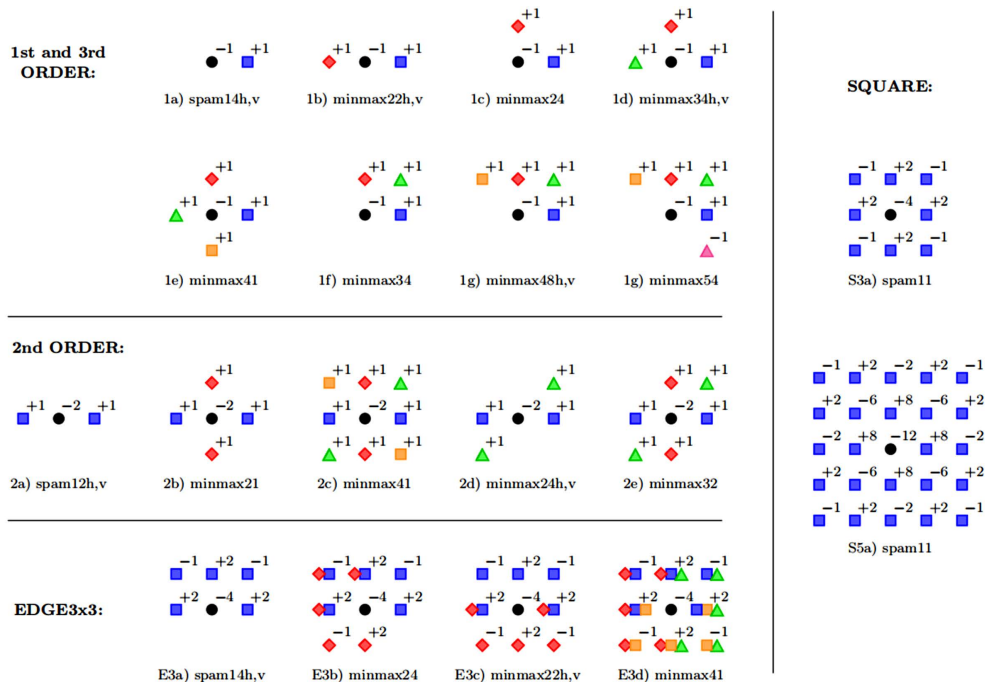


Figure 2. High pass filter (Fig. 2 in reference [4])
图 2. 高通滤波器(文献[4]中的 Fig. 2)

2.2. 提取共生矩阵生成特征

残差图像中一个像素点的幅值越大,则说明该像素点对应于领域像素的相关性较弱,由于残差图像中每个像素点的幅度各不相同,首先对图像进行量化及截断,希望能够将残差图像的范围限定在某一个区间,这样能够降低之后步骤的计算量。其量化及截断的步骤如下:

$$R_{i,j} \leftarrow \text{trunc}_T \left(\text{round} \left(\frac{R_{i,j}}{Q} \right) \right), i=1, \dots, n_1, j=1, \dots, n_2, \text{round}(\cdot) \quad (2.2)$$

式 2.2 中 Q 为量化因子, T 为截断阈值, $\text{trunc}_T(\cdot)$ 表示截断, $\text{round}(\cdot)$ 表示取整。根据不同的量化因子及截断阈值,可以改变 SRM 的检测性能,本文中取 $Q=1$, $T=2$ 。

共生矩阵是统计某一距离上的一类像素的数量。由量化截断后的残差图像提取共生矩阵,得到的共生矩阵即为 SRM 最终想要抽取的隐写分析特征。所提取共生矩阵阶数的大小会直接影响特征抽取的结果,阶数过高会使得特征的统计无意义;阶数如果过低,又会使得特征不够丰富,降低分析准确率。本文中提取残差图像的四阶共生矩阵作为隐写分析特征,且 $Q=1$, $T=2$, 在文献[4]中被称作 SRMQ1, 即 $Q=1$, $T=2$, $d=4$ 。

3. 论加入图像纹理的空域富模型隐写分析方法

图像纹理[12]是一种能够反映物体表面的变化的属性,图像纹理复杂的区域像素变化大、周期性小,纹理平滑区域则像素变化平缓,周期性大,如图 3。纹理在性质上来说是一种难以描述的视觉属性,但现在多采用由 1973 年 Haralick [13]等人提出的灰度共生矩阵来描述图像的纹理特征。灰度共生矩阵与共生矩阵的生成方法相同,是衡量图像空域中相隔某距离的两个像素之间的灰度关系,其中方向包括水平、竖直、对角线、逆对角线方向。最后再通过计算生成的灰度共生矩阵的特征参数,由此来描述图像的纹理信息。

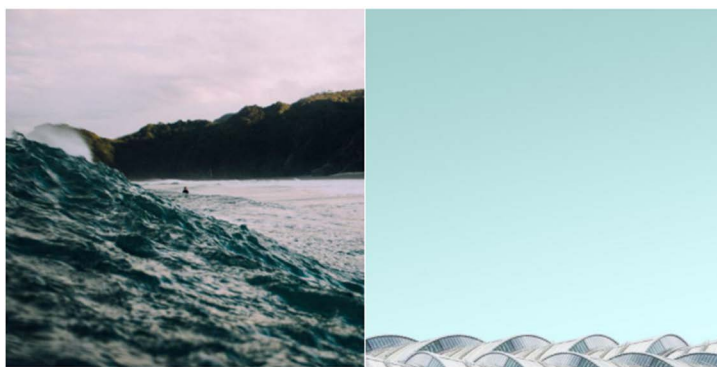


Figure 3. Image of complex texture (left) and image of smooth texture (right)
图 3. 复杂纹理图像(左)与平滑纹理图像(右)

Haralick 等人定义了 14 个由灰度共生矩阵生成的特征参数来描述图像的纹理特征,而 14 个特征参数中有 4 个是非相关的,多被用于纹理特征的提取,本文中还加入另一个特征参数作为纹理特征,分别为能量、熵、对比度、相关性、均质性。通过取各个方向上灰度共生矩阵各个特征参数的均值及方差,就可以得到图像一个有较高分类精确度的 10 维特征。

如今自适应隐写多采用 STC 编码或损失函数来估计隐写对图像造成的改动,而在纹理复杂区域的改动往往对图像整体的影响较小,不易与被统计特征发现,这样损失函数最终确定的位置多位于纹理复杂的区域,所以自适应隐写的最终嵌入区域多处于纹理复杂区域。例如对图 3 中两幅图像进行 S-UNIWARD 方法以 0.1 嵌入率嵌入秘密信息,嵌入点如图 4 所示,其中黑色点为进行-1 修改,白色点为进行+1 修改。

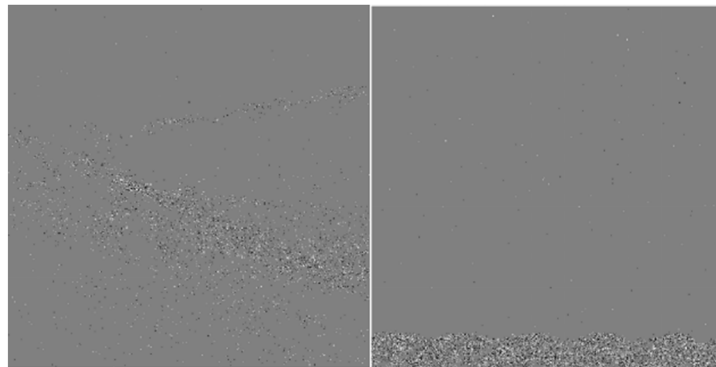


Figure 4. Embedded location of S-UNIWARD
图 4. S-UNIWARD 嵌入位置

由于 SRM 无法对图像的纹理信息进行辨识, 遗漏了图像的纹理信息, 而自适应隐写改动多位于纹理复杂区域。对此本文采用将图像进行 2×2 分割, 对每一块图像进行 SRM 判别。判别结果 -1 代表载体图像, +1 代表含密图像, 再将判别结果乘以相依块的 10 维纹理特征, 最后按顺序拼接为, 则可得到图像的 40 维纹理特征, 最后再按照 40 维纹理特征进行分类。同理可将图像进行 4×4 或更高数量的切割, 以提高对局部纹理特征的辨识度。其具体隐写分析步骤如图 5。

本文所使用的特征分类器使用由 Fridrich 等人提出的集成分类器(Ensemble Classifier) [14] [15], 此分类器对于大样本训练集训练速度快, 分类效果好, 能够明显缩短训练分类器的时间。集成分类器是以 FLD

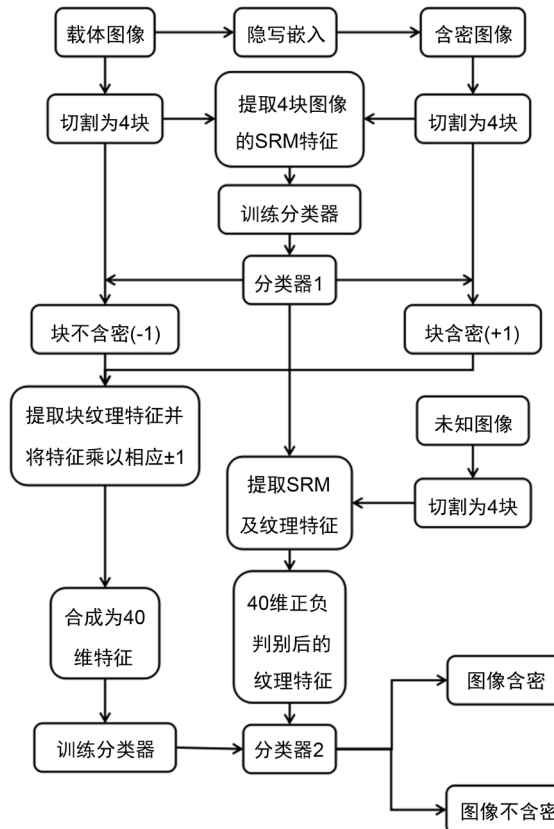


Figure 5. Steganalysis steps combining texture complexity and SRM
图 5. 结合纹理复杂度的 SRM 隐写分析步骤

(Fisher Linear Discriminator)为基本分类器, 通过有微弱差异的训练集对每个 FID 进行训练, 最后将通过所有分类器的判决进行无加权投票, 以投票数量为多数的结果作为最终结果。

4. 实验与结果分析

4.1. 实验配置

本文的采用的图片源自不同手机型号拍摄的原始未经处理的 JPEG 格式的照片(图片源自 <http://www.cihw.org.cn/index.php/9/>), 原始照片按最短边裁剪成正方形, 然后统一缩放成 1024×1024 大小, 筛选出品质因子分布在 90~95 之间的图像。将图像库统一使用 matlab 转换为 BMP 图像格式, 并生成两组图像相同、品质因数分别为 75 和 95 的图像库。隐写方法使用 S-UNIWARD 及 WOW, 类似文献 [4], 检测嵌入率使用 0.1~0.5 及 0.05。由于修改式隐写方法在嵌入率低于 0.05 时已较难被发现, 且在嵌入率大于 0.5 时多种隐写分析方法均能检测出图像已被隐写, 故不使用低于 0.05 和大于 0.5 的嵌入率。

对于不同品质因数下的不同嵌入率, 分别使用相同的 2000 张图像作为训练集, 1000 张图像作为测试集, 分别嵌入后进行得到载体图像和含密图像, 测试模拟已知隐写环境分类器效果。

4.2. 结果分析

通过测试对 1000 张图片进行分类后的错误率来对分类器性能进行评断, 其错误率:

$$P = \frac{N_F + N_M}{N} \quad (4.1)$$

在式 4.1 中, N_F 、 N_M 和 N 分别代表虚警样本数、漏警样本数、总样本数, 这样就可以得到相应的分类器结果, 从而判别分类器性能。

将通过纹理特征结合的 SRM 和原始 SRM 隐写分析性能进行对比, 其中模拟已知隐写环境在 S-UNIWARD 隐写算法下, 在载体图像品质因数 QF 为 75 及 95 的情况下错误率如图 6 及图 7 所示; 其中模拟已知隐写环境在 WOW 隐写算法下, 在载体图像品质因数 QF 为 75 及 95 的情况下错误率如图 8 及图 9 所示。

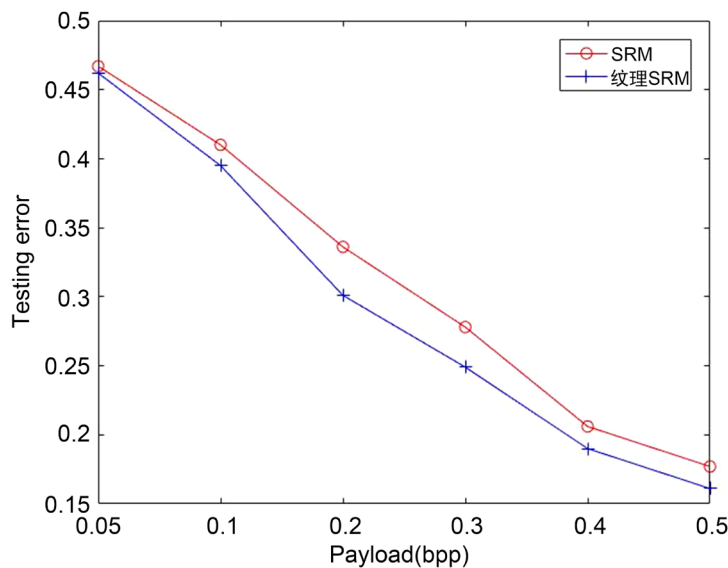


Figure 6. Error rate of steganalysis for S-UNIWARD when QF = 75
图 6. QF = 75 时对 S-UNIWARD 隐写分析错误率

图 6 中横轴为隐写嵌入率，纵轴为分类器进行分类后的错误率。可见本文结合图像纹理复杂度后的 SRM 隐写分析方法，在嵌入率为 0.2、0.3 时，对于采用 S-UNIWARD 为嵌入算法、载体图像品质因数 $QF = 75$ 的图像对，检测错误率降低较为明显；而在 0.05、0.1 及 0.5 时降低不明显。

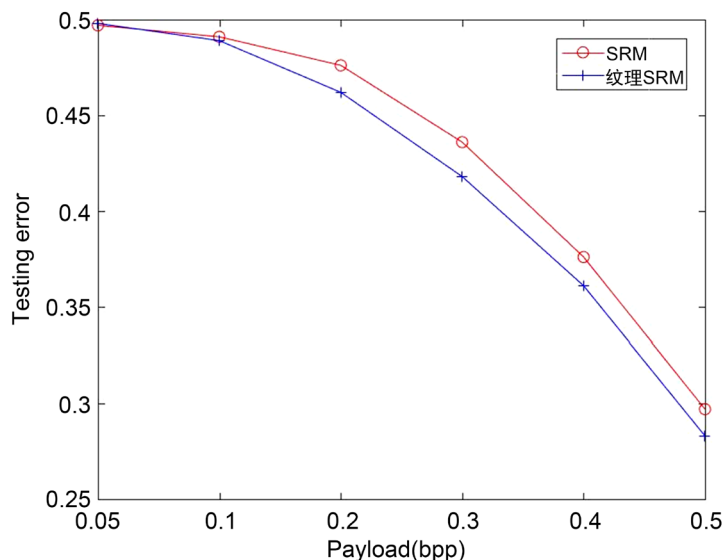


Figure 7. Error rate of steganalysis for S-UNIWARD when $QF = 95$
图 7. $QF = 95$ 时对 S-UNIWARD 隐写分析错误率

类似于图 7，可见本文算法在嵌入率为 0.05 和 0.1 时，相对于 SRM 错误率降低较少，其他嵌入率下错误率降低更为明显。且相比于选择载体图像品质因数 $QF = 75$ 时，本文方法及 SRM 方法错误率均有所上升，且本文算法相对于 SRM 错误率的降低没有品质因数 $QF = 75$ 时那么明显。所以品质因数高的图像更加能够抵抗隐写分析检测，本文的算法对于高品质因数下图像的隐写方法检测性能提升一般。

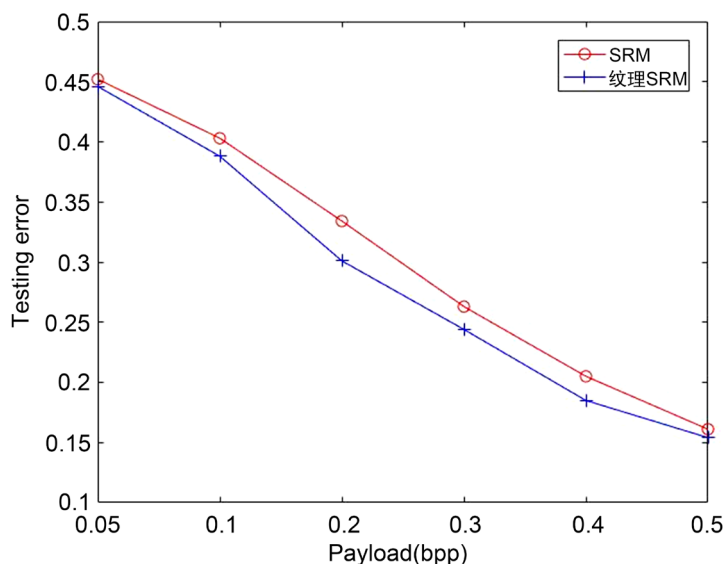


Figure 8. Error rate of steganalysis for WOW when $QF = 75$
图 8. $QF = 75$ 时对 WOW 隐写分析错误率

由图 8 可见, 其隐写分析性能曲线与图 6 曲线类似, 说明 S-UNIWARD 与 WOW 隐写算法的隐写性能较为相近。而本文结合图像纹理复杂度后的 SRM 隐写分析方法, 对于在嵌入率适中时, 性能提升优于在嵌入率较高或较低的情况。因此, 使用本文方法对采用 WOW 进行隐写的图像进行检测时, 类似于 S-UNIWARD, 在嵌入率适中时相对于 SRM 隐写分析方法性能提升更加明显。

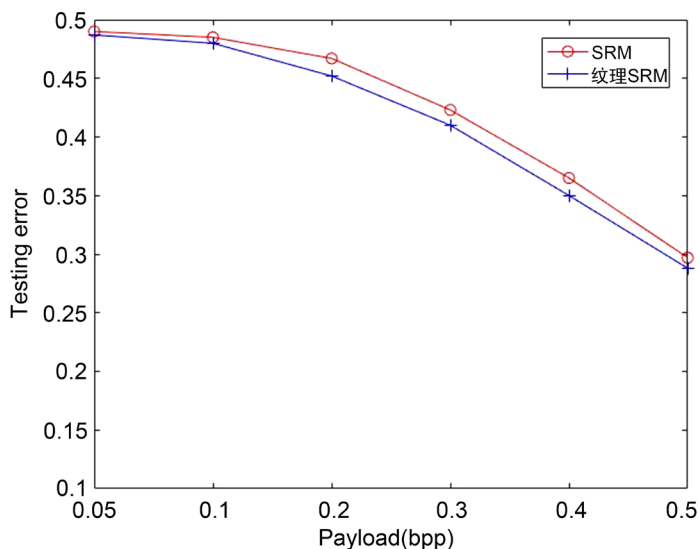


Figure 9. Error rate of steganalysis for WOW when QF = 95
图 9. QF = 95 时对 WOW 隐写分析错误率

由图 9 可见, 在嵌入率大于 0.1 时, 本文方法相对于 SRM 错误率有一定的降低, 且在嵌入率适中时降低更加明显, 与以 S-UNIWARD 作为隐写嵌入算法时的曲线类似。

可见在隐写嵌入率为 0.2 到 0.4 的时候, 本文中的与纹理特征相结合的 SRM 隐写分析方法判别错误率, 相较于 SRM 降低得更加明显。这说明本文提出的方法对于自适应隐写方法下隐写分析的检测性能有一定的提高, 但是对于低嵌入率隐写或者图像高品质因数下的判别性能提升并不明显, 所以本文方法适用于对中等嵌入率、图像品质因数不高的隐写图像进行隐写分析检测。

5. 总结与展望

本文从 SRM 隐写分析方法未考虑自适应隐写算法嵌入信息于图像纹理复杂区域的角度出发, 提出了一种将图像纹理特征与 SRM 快速结合的图像隐写分析方法。先将图像进行分块判别, 并将判别结果与块 12 维纹理特征结合, 最后将纹理特征拼接, 通过与判别结果结合后的纹理特征对图像进行分析。从实验结果中可以看出, 在隐写嵌入率适中、图像品质因数不高时, 此方法对于自适应隐写算法的检测错误率有一定的降低。

下一步, 可以考虑如何将 SRM 特征与 RGB 通道等其他未考虑到的特征进行快速结合, 提高隐写分析系统的性能, 降低判别错误率。

基金项目

本文受到国家自然科学基金(U1736121)资助。

参考文献

- [1] Johnson, N.F. and Jajodia, S. (1998) Exploring Steganography: Seeing the Unseen. *IEEE Computer*, **31**, 26-34.

- <https://doi.org/10.1109/MC.1998.4655281>
- [2] Westfeld, A. (2001) High Capacity despite Better Steganalysis. *Proceedings of 4th International Workshop on Information Hiding*, No. 2137, 289-302. https://doi.org/10.1007/3-540-45496-9_21
- [3] Fridrich, J. and Pevny, T. (2007) Statistically Undetectable JPEG Steganography: Dead Ends Challenges, and Opportunities. In: *Proceedings of 9th Workshop on Multimedia & Security*, 3-14. <https://doi.org/10.1145/1288869.1288872>
- [4] Fridrich, J. and Kodovsky, J. (2012) Rich Models for Steganalysis of Digital Images. *IEEE Transactions on Information Forensics and Security*, **7**, 868-882. <https://doi.org/10.1109/TIFS.2012.2190402>
- [5] Holub, V. and Fridrich, J. (2015) Low Complexity Features for JPEG Steganalysis Using Undecimated DCT. *IEEE Transactions on Information Forensics and Security*, **10**, 219-228. <https://doi.org/10.1109/TIFS.2014.2364918>
- [6] Filler, T., Fridrich, J. and Judas, J. (2010) Minimizing Embedding Impact in Steganography Using Trellis-Coded Quantization. *Proceedings of SPIE Electronic Imaging, Media Forensics and Security XII5*, San Jose, CA, 18-20 January 2010, 1-14. <https://doi.org/10.1117/12.838002>
- [7] Holub, V. and Fridrich, J. (2013) Digital Image Steganography Using Universal Distortion. *Proceedings of 1st ACM Workshop on Information Hiding and Multimedia Security*, France, Montpellier, 59-68. <https://doi.org/10.1145/2482513.2482514>
- [8] Filler, T. and Fridrich, J. (2011) Gibbs Construction Insteganography. *IEEE Transactions on Information Forensics and Security*, **5**, 705-720. <https://doi.org/10.1109/TIFS.2010.2077629>
- [9] Holub, V. and Fridrich, J. (2012) Designing Steganographic Distortion Using Directional Filters. *Fourth IEEE International Workshop on Information Forensics and Security*, Tenerife, 234-239. <https://doi.org/10.1109/WIFS.2012.6412655>
- [10] Holub, V., Fridrich, J. and Denmark, T. (2014) Universal Distortion Function for Steganography in an Arbitrary Domain. *EURASIP Journal on Information Security*, **2014**, 1-13. <https://doi.org/10.1186/1687-417X-2014-1>
- [11] Chen, M., Sedighi, V., Boroumand, M. and Fridrich, J. (2017) JPEG-Phase-Aware Convolutional Neural Network for Steganalysis of JPEG Images. *Proceedings of IH & MMSec 17*, Philadelphia, PA, 20-22 June 2017. <https://doi.org/10.1145/3082031.3083248>
- [12] 高程程. 基于灰度共生矩阵的纹理特征提取[J]. *计算机系统应用*, 2010, 19(6): 195-198.
- [13] Haralick, R.M. and Shanmugam, K. (1973) Texture Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-3**, 610-621. <https://doi.org/10.1109/TSMC.1973.4309314>
- [14] Kodovsky, J. and Fridrich, J. (2011) Steganalysis in High Dimensions: Fusing Classifiers Built on Random Subspaces. *Proceedings of SPIE, Electronic Imaging, Security and Forensics of Multimedia XIII*, San Francisco, CA, 23-26 January 2011, 1-13. <https://doi.org/10.1117/12.872279>
- [15] Kodovsky, J., Fridrich, J. and Holub, V. (2012) Ensemble Classifier for Steganalysis of Digital Media. *IEEE Transactions on Information Forensics and Security*, **7**, 432-444. <https://doi.org/10.1109/TIFS.2011.2175919>