

基于 JPEG 量化失真的合成图像盲检测*

周响金, 李 峰, 熊 兵

(长沙理工大学 计算机通信工程学院, 长沙 410114)

摘 要: 主要针对 JPEG 图像合成伪造, 提出了一种基于量化失真的合成图像盲检测算法。首先针对合成图像以 JPEG 和非 JPEG 不同的存储方式, 分别估计原始量化矩阵; 然后用估计的原始量化矩阵对合成图像再压缩, 计算压缩前后的量化失真; 最后通过判断合成图像不同区域量化失真的大小, 实现篡改区域的自动检测和定位。实验结果表明: 算法能有效地检测 JPEG 和非 JPEG 两种不同存储方式的合成图像。

关键词: JPEG; 合成图像检测; 量化失真

中图分类号: TP309.2 **文献标志码:** A **文章编号:** 1001-3695(2012)06-2346-04

doi: 10.3969/j.issn.1001-3695.2012.06.092

Blind detection of synthetic image based on JPEG quantization distortion

ZHOU Xiang-jin, LI Feng, XIONG Bing

(College of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China)

Abstract: This paper proposed an algorithm to blindly detect synthetic image based on JPEG quantization distortion. Firstly, it estimated the original quantization matrix of the synthetic image in terms of its storage format, JPEG or non-JPEG. Secondly, it recompressed the synthetic image using the estimated original quantization matrix, calculated the quantization distortion of the synthetic image before and after compression. Finally, it judged the size of quantization distortion for different region of synthetic image, to realize automatic detection and location of its tampered regions. Experimental results show that the proposed algorithm is effective in detecting synthetic image stored in both JPEG and non-JPEG.

Key words: JPEG(joint photographic experts group); synthetic image detection; quantization distortion

0 引言

随着图像处理软件的广泛应用和计算机处理性能的不断提升, 人们可以轻易地对数字图像进行各种篡改, 而在视觉效果上表现得非常逼真。伪造的图像若被滥用于法庭证物、科学发现、媒体等各方面, 则对社会可能导致难以想象的后果。因此, 图像的真实性和完整性鉴别研究具有重要意义。目前, 数字图像取证分为主动取证和被动取证^[1]。主动取证需要事先在图像中嵌入数字签名或数字水印, 以达到检测图像篡改的目的。被动取证也叫盲取证, 它不需要事先在图像中嵌入信息, 因而具有更广泛的应用性, 成为目前数字图像取证的热点研究方向。

目前, 数字图像盲取证技术^[2]得到了国内外研究机构的广泛关注, 并提出了许多盲取证方法。JPEG 作为一种图像压缩标准, 因其压缩比高、处理速度快以及图像质量好等优点, 是目前最流行的一种图像压缩格式。因此, 对 JPEG 图像的真伪性鉴别具有重要的实用价值。目前, 针对 JPEG 图像篡改检测的研究还相对较少, 已有的检测方法主要包括二次压缩检测技术和块效应检测技术。Lukás 等人^[3]提出从双重压缩的 JPEG 图像中估计原始量化矩阵的三种方法, 并分析了两次压缩造成 DCT 系数直方图出现缺失值和双峰等特征的存在。Popescu^[4]

利用两次 JPEG 压缩反映在 DCT 系数直方图中的周期性现象来检测图像是否经过篡改。Fu 等人^[5]提出了一种 Benford 定律模型, 进而以单次 JPEG 压缩后的 DCT 系数符合该定律而两次压缩后的 DCT 系数破坏了这种定律作为依据, 来判断图像是否经过两次 JPEG 压缩。He 等人^[6]分析了双重压缩效应, 并利用双重压缩实现图像篡改的检测。魏为民等人^[7]提出了基于对图像二阶差分的频谱分析, 构造了一个新的图像块效应测量指标, 并将其应用到合成图像的检测方面。Ye 等人^[8]提出了利用 DCT 系数直方图的功率谱来估计图像压缩的量化表, 并进一步利用量化表定义图像块效应强度值, 根据待测图像不同区域块效应值的不同, 判断出图像的篡改区域。

上述大部分文献都缺乏一种对 JPEG 图像的原始量化步长的有效估计算法, 或算法的检测效果不够准确。本文通过分析 JPEG 图像量化后的 DCT 系数直方图表现出的周期特征, 并针对篡改后的图像以不同压缩格式存储的情况, 提出了一种基于 JPEG 量化失真的合成图像盲检测算法, 以实现篡改区域的自动检测和定位。

1 合成图像检测原理

1.1 JPEG 图像合成伪造模型

一般来说, JPEG 图像的合成伪造过程主要包括以下几个

收稿日期: 2011-10-10; **修回日期:** 2011-11-24 **基金项目:** 国家自然科学基金资助项目(60973113); 湖南省自然科学基金资助项目(09JJ3120)

作者简介: 周响金(1987-), 女, 硕士研究生, 主要研究方向为图像处理与模式识别、信息安全(irisuser226@126.com); 李峰(1964-), 男, 教授, 博士, 主要研究方向为图像处理与模式识别、信息安全; 熊兵(1981-), 男, 讲师, 博士, 主要研究方向为图像处理与模式识别、网络与信息安全。

步骤,如图1所示:

- 输入一幅 JPEG 图像 I_1 , 作为背景图像;
- 从另一幅 JPEG 图像或非 JPEG 图像 I_2 中复制一块区域, 粘贴到图像 I_1 的某个区域, 得到图像 I_3 , 称为合成图像;
- 将合成图像 I_3 再次以 JPEG 或非 JPEG 格式存储, 并输出。

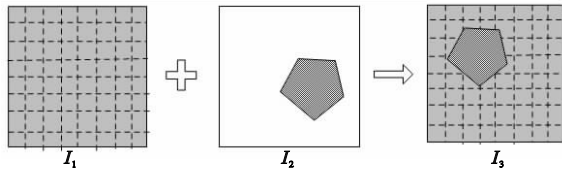


图1 JPEG图像合成伪造模型

1.2 JPEG 图像量化特性分析

JPEG 压缩过程主要包括预处理、DCT 变换、量化、熵编码等几个部分^[1]。其中量化是关键的一个步骤, 因为在对图像每一块的 DCT 系数进行量化时, 是将 DCT 系数除以量化步长后取整, 丢弃对图像影响不大的高频分量, 达到降低码率的目的。而量化这一步是不可逆的, 量化步长越大, 图像在进行反量化时, 丢失的高频信息就会越多, 出现量化失真就会越明显。

图像进行 JPEG 压缩后, 会表现出明显的量化效应。量化效应是指用量化矩阵量化 DCT 系数时, 对小数部分采用四舍五入的方法取整, 从而会产生信息丢失的现象。用量化的 DCT 系数的直方图表示, 就会表现出间断性、周期性的波峰, 如图2所示。图2(a)是未压缩图像的 DCT 系数直方图, 图2(b)是 JPEG 压缩后的 DCT 系数直方图。其中横坐标表示 DCT 系数值, 纵坐标表示 DCT 系数值对应的数量。

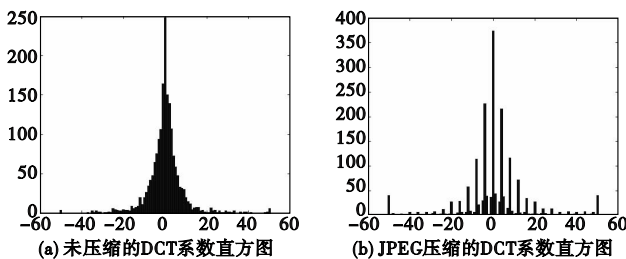


图2 量化的DCT系数直方图

量化矩阵通常与一定的质量因子一一对应。质量因子是一个 1~100 的整数。当图像进行 JPEG 压缩时, 通常需要指定一个质量因子 q 。一旦 q 确定, 量化矩阵 Q_y 就可通过式(1)计算得到:

$$Q_{i,j} = \text{floor}((Q_{i,j}^C \times \alpha(q) + 50)/100) \quad i, j \in \{1, 2, \dots, 8\} \quad (1)$$

其中:

$$\alpha(q) = \begin{cases} \frac{5000}{q} & \text{if } 1 \leq q < 50 \\ 200 - 2q & \text{if } 50 \leq q \leq 100 \end{cases}$$

$$Q_{i,j}^C = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

$Q_{i,j}^C$ 是 JPEG 标准推荐的量化矩阵亮度分量, $\text{floor}(\cdot)$ 函数表示向下取整。从式(1)中可以看出: 压缩质量越低, 质量因

子就会越小, 量化矩阵系数就会越大; 压缩质量越高, 则与之相反, 当量化因子为 100 时, 量化矩阵中所有系数都为 1。

JPEG 图像按 8×8 分块进行量化编码。如果伪造者在放置对象到背景图像时, 篡改区域和背景区域的分块位置往往出现分块位置不一致的情况, 并且合成图像中的背景区域和篡改区域原始压缩质量因子也不一定相同, 这就为合成图像的检测提供了重要线索。

若一幅未压缩图像用相同的量化矩阵两次压缩图像, 则产生的量化失真最小。用不同的量化矩阵两次压缩图像时, 两个量化矩阵差值越大, 则产生的量化失真也越大。针对合成图像的篡改区域与背景区域存在分块位置和量化矩阵不一致现象, 可估计背景区域的原始量化矩阵, 用估计的量化矩阵对合成图像再压缩, 并计算图像的量化失真。通过判断合成图像不同区域的量化失真大小, 可确定合成图像的篡改区域。

2 检测算法描述

2.1 原始量化矩阵估计

2.1.1 合成图像以非 JPEG 格式存储

首先讨论合成图像以非 JPEG 格式存储的情况, 比如 BMP 格式。这种情况下, 背景图像只经过了一次 JPEG 压缩。这时不能直接从合成图像的头文件中获取当前压缩的量化表。但合成图像的 DCT 系数的直方图仍然表现出周期性的波峰, 并且与篡改之前 JPEG 图像的 DCT 系数直方图保持一致。因此, 可以从 DCT 系数的直方图中直接估计出原始量化矩阵。

由于 DCT 变换是正交变换, 对某一个 DCT 系数量化并不会影响到其他 DCT 系数, 所以可以对每个 DCT 系数分别估计其量化步长。而对于高频系数, 由于量化后 DCT 系数大部分都量化为 0, 要准确估计这些高频分量的量化步长比较困难。因此, 本文只对低频交流系数进行分析, 按 ZigZag 扫描顺序, 估计前面 20 个量化矩阵系数。然后用估计的前 20 个量化系数与标准的量化表进行匹配, 找到最接近的质量因子, 再通过式(1)计算出其他所有量化系数。

在量化后的 DCT 系数直方图中, 周期性波峰的出现位置均对应于量化步长的倍数, 如图3所示。图3(a)是 Lena 图像, 原始压缩质量因子为 85; (b)是所有 8×8 块 DCT 系数的第 (2,1) 个 DCT 系数值统计的直方图, 其量化步长 $Q(2,1) = 4$; (c)是第 (1,4) 个 DCT 系数对应的直方图, 其量化步长 $Q(1,4) = 5$ 。从图3中可以看出: (b)和(c)中的 DCT 系数的直方图波峰分别对应应在 4 和 5 的倍数上。如果只考虑直方图的右半部分, 波峰的频率都是从 0 往右依次递减。除了 0 对应的频率外, 频率最高的 DCT 系数值就是它的量化步长。通过统计各个 DCT 系数直方图频率关系, 就可以估计出原始量化矩阵。

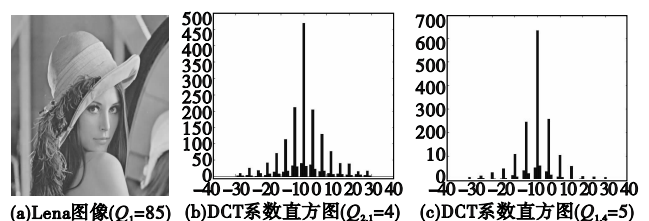


图3 合成图像和DCT系数直方图

量化矩阵估计的具体步骤如下:

- a) 计算所有 8×8 图像块的 DCT 系数, 取绝对值;
- b) 统计前面 20 个低频交流系数的直方图, 记为 $H_i(k)$ ($i=1, 2, \dots, 20$), k 表示取整的 DCT 系数;
- c) 对每个直方图 $H_i(k)$ 按降序排列, 频率排在第二位的 DCT 系数就是该点对应的量化步长 $q(k)$;
- d) 通过式(1), 计算不同的质量因子对应的量化矩阵, 与估计的 20 个量化步长对应相减, 绝对值之和最小的质量因子对应的量化矩阵就是要估计的原始量化矩阵, 如式(2)所示。

$$Q_1 = \arg \min_{Q \in [30, 100]} \sum_{k=1}^{20} |q(k) - Q(k)| \quad (2)$$

2.1.2 合成图像以 JPEG 格式存储

接下来讨论合成图像以 JPEG 格式存储的情况。在这种情况下, 背景图像一般经过两次压缩, 而篡改区域可能经过一次压缩, 也可能经过多次压缩。一般来说, 篡改区域要比背景图像区域小。因此, 估计原始量化矩阵主要针对背景图像。可以采用搜索遍历的方法确定背景图像第一次压缩的质量因子。如果某个质量因子正好是背景图像第一次压缩的质量因子, 则用该质量因子对应的量化矩阵对合成图像进行再压缩后, 计算压缩前后的 DCT 系数差值, 得到的量化误差最小。

量化矩阵估计的具体步骤如下:

- a) 首先从 JPEG 头文件中读取第二次压缩的量化矩阵, 并求得对应的量化因子 Q_2 。
- b) 将合成图像平均分成四个子图像, 对每个子图像再按 8×8 大小进行 DCT 变换, 得到 DCT 系数 $D_i(k)$ 。其中: $k=1, 2, \dots, 64$; i 表示块数。
- c) 分别用质量因子 ($30 < Q < Q_2$) 对应的量化矩阵 $Q(k)$ 对 $D_i(k)$ 进行量化取整, 然后再用 $Q(k)$ 进行反量化, 得到 $D_i'(k)$ 。
- d) 计算原 DCT 系数 $D_i(k)$ 与量化后的 DCT 系数 $D_i'(k)$ 差值, 得到相应的量化误差, 如式(3)所示 (其中 N 表示图像的总块数, $p=1, 2, 3, 4$) :

$$E(p, Q) = \frac{4}{N} \sum_{i=1}^{N/4} \sum_{k=1}^{64} |D_i(k) - D_i'(k)| \quad (3)$$

e) 去掉 $E(1, Q)$ 、 $E(2, Q)$ 、 $E(3, Q)$ 、 $E(4, Q)$ 中最大的一个误差值, 取其他三个误差值的平均值, 得到 $E(Q)$ 。由于篡改区域在这四个子图像中的某一块所占区域最大, 因而导致的量化误差也会最大。去掉最大误差值可以减少篡改区域对原始量化矩阵估计的影响。

f) 最后判断每个质量因子对应的 $E(Q)$, 找到误差值对小的 $E_{\min}(Q_1)$, 则 Q_1 就是原始量化矩阵对应的质量因子, 如图 4 所示。该图是图 3(a) 经过质量因子为 95 再压缩后的图像的量化误差曲线。从图 4 中可以看到: 误差最小值对应的质量因子就是第一次压缩的质量因子。

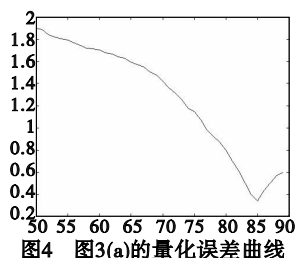


图4 图3(a)的量化误差曲线

2.2 篡改区域的检测与定位

针对不同的存储方式得到相应的原始量化矩阵后, 就可以进一步检测合成图像的篡改区域。对每个 8×8 图像块, 按照如下公式计算量化失真的平方值:

$$QE(i) = \sum_{k=1}^{64} \left| D_i(k) - Q_1(k) \times \left[\frac{D_i(k)}{Q_1(k)} \right] \right|^2 \quad (4)$$

其中: $QE(i)$ 表示第 i 块的量化失真的平方值, $i=1, 2, \dots, N$, N 表示图像总块数; $D_i(k)$ 是篡改后的图像的 DCT 系数; $Q_1(k)$ 是第一次压缩质量因子对应的量化表。如果某个区域经过篡改, 则 QE 的值就会比较大, 而非篡改区域得到的值就会较小。通过判断合成图像不同区域的量化失真大小, 就可以定位出图像的篡改区域。

3 实验

实验从原始量化矩阵估计和篡改区域自动检测与定位两个方面对本文算法进行了评估。从标准图像库中随机选择了 200 幅图像进行测试, 存储格式都为 BMP 格式。首先, 使用质量因子 $QF_1 \in \{50, 60, 70, 80, 90\}$ 分别对这 200 幅图像进行压缩, 生成一次 JPEG 压缩图像。然后通过篡改伪造, 得到 $200 \times 5 = 1000$ 幅合成图像, 并以两种方式保存: a) 以 BMP 格式保存的合成图像; b) 以质量因子 QF_2 为 95 重新压缩保存的合成图像。

3.1 原始量化矩阵估计

为了评估本文算法中原始量化矩阵估计方法的有效性, 将本文算法与文献[8]进行了实验对比, 检测结果如表 1 所示。从表 1 中可以看出: 合成图像以 BMP 格式存储时, 两种算法估计量化矩阵的正确率都相对较高。但是, 本文算法耗时低于文献[8]。对于一幅 512×512 大小的灰度图像, 本文算法与文献[8]的算法估计原始量化矩阵耗时分别为 0.41 s 和 0.85 s。合成图像以 JPEG 格式存储时, 本文算法估计量化矩阵的正确率明显比文献[8]要高。这是因为, 本文算法通过将合成图像进行分块, 去掉量化误差最大值的那块, 一定程度上减少了篡改区域对量化矩阵估计影响。

表1 量化表评估的正确率 /%

格式	算法	QF_1				
		50	60	70	80	90
BMP 格式存储	本文算法	92.4	95.7	94.8	97.3	96.4
	文献[8]	91.5	92.9	90.2	93.2	92.1
JPEG 格式存储	本文算法	89.6	92.7	94.5	95.2	91.8
	文献[8]	65.5	70.3	75.9	78.1	76.5

3.2 合成图像的篡改区域检测与定位

实验首先测试合成图像以 BMP 格式存储的情况, 检测结果如图 5 所示。图 5(a) 是一幅原始的 JPEG 图像, 压缩质量因子为 80, 大小为 300×400 像素; 从另一幅图像中复制一只海鸟粘贴到原始图像后的合成图像如(b)所示; (c) 和(d) 分别是本文算法和文献[8]检测篡改区域的效果图。从(c) 和(d) 可以看出: 本文算法能够准确定位出篡改区域, 而文献[8]算法虽然也能定位出篡改区域, 但受到非篡改区域的影响比较大。

为了测试合成图像为 JPEG 格式的情况, 实验采用两组不

同篡改区域大小的图像分别进行对比,对比结果如图 6 所示。图 6(a)是图 5(a)经过小区域篡改的一幅 JPEG 合成图像,压缩质量因子为 95;图 6(b)和(c)是本文算法和文献[8]算法检测篡改区域结果图。从这组实验中可以看出:对于小区域的篡改,本文算法基本上可以准确定位篡改区域,文献[8]算法无法定位篡改区域。图 6(d)是 Lena 图像经过篡改后的 JPEG 合成图像,背景图像是以质量因子为 70 进行压缩的 JPEG 图像,合成图像以质量因子 95 再次压缩后的 JPEG 图像。采用本文算法和文献[8]的算法检测篡改区域的结果如图 6(e)和图 6(f)所示。从这组实验中可以看出:对于较大区域的篡改,本文算法与文献[8]都能准确定位出篡改区域,但本文算法的检测效果更为明显。

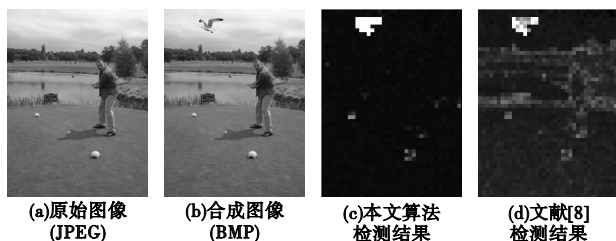


图5 合成图像为非JPEG格式的篡改检测

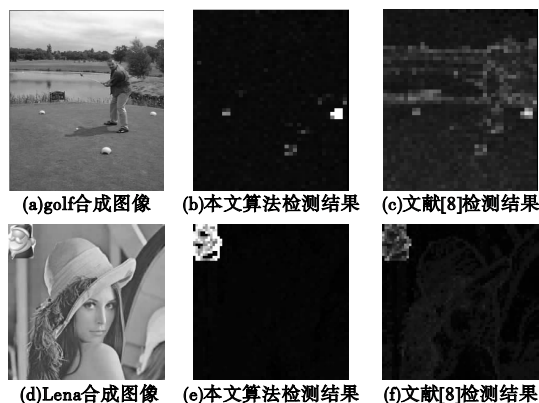


图6 合成图像为JPEG格式的篡改检测

4 结束语

JPEG 合成图像的量化失真可以作为判断图像是否被篡改的重要依据。本文对 JPEG 合成伪造图像进行了分析,提出了一种简单有效的图像盲取证算法。实验表明,该方法简单有效,针对不同篡改区域大小和不同篡改类型,都能够准确找到图像的篡改区域。当然,本文算法也存在一定的局限性。例如,如果第二次压缩的质量因子比第一次压缩的质量因子小,则量化表的评估准确率还不够高。这将是笔者下一步的研究重点。

参考文献:

- [1] 周琳娜. 数字图像盲取证技术研究[D]. 北京:北京邮电大学, 2007.
- [2] NG T T, CHANG S F, LIN C Y, *et al.* Passive-blind image forensics [M]//Multimedia Security Technologies for Digital Rights. [S. l.]: Elsevier, 2006:383-412.
- [3] LUKÁŠ J, FRIDRICH J. Estimation of primary quantization matrix in double compressed JPEG images [C]//Proc of Digital Forensic Research Workshop. 2003.
- [4] POPESCU A C. Statistical tools for digital image forensics [D] Hanover, NH: Department of Computer Science, Dartmouth College, 2004.
- [5] FU Dong-dong, SHI Y Q, SU Wei. A generalized Benford's law for JPEG coefficients and its applications in image forensics [C]//Proc of SPIE, Security, Steganography and Watermarking of Multimedia Contents IX. 2007.
- [6] HE J, LIN Z, WANG L, *et al.* Detecting doctored JPEG images via DCT coefficient analysis [C]//Proc of ECCV. 2006:423-435.
- [7] 魏为民, 唐振军. 利用 JPEG 块效应不一致性的合成图像盲检测 [J]. 中国图象图形学报, 2009, 14(11): 2387-2390.
- [8] YE S M, SUN Qi-bin, CHANG E C. Detecting digital image forgeries by measuring inconsistencies of blocking artifact [C]//Proc of IEEE International Conference on Multimedia and Expo. 2007:12-15.