

JPEG 二次压缩的分析与检测

戴蒙¹⁾ 林家骏¹⁾ 毛家发^{1),2)}

¹⁾(华东理工大学信息科学与工程学院,上海 200237) ²⁾(上饶师范学院,上饶 334001)

摘要 为了对 JPEG 二次压缩后的图像进行有效检测,通过研究 JPEG 图像经二次压缩后所引起的 DCT 统计特性的变化及二次压缩前后的质量因子 Q_1 、 Q_2 对二次压缩后图像的 DCT 统计所引起的变化的分析,提出了 JPEG 二次压缩的新的检测方法——抖动模式,并给出了满足 DCT 直方图抖动模式的充分必要条件。实验证明,这种新的检测方法是有效的。

关键词 JPEG 双压缩 质量因子 抖动模式

中图分类号: TN918.91 **文献标识码**: A **文章编号**: 1006-8961(2006)11-1619-04

The Analysis and Detection of Double JPEG Compression

DAI Meng¹⁾, LIN Jia-jun¹⁾, MAO Jia-fa^{1),2)}

¹⁾(College of Information Science & Engineering, ECUST, Shanghai 200237) ²⁾(Shang Rao Normal University, Shangrao 334001)

Abstract After JPEG double compressed, the DCT statistic of JPEG image will be changed, which is the focus in this paper. The distributions of DCT coefficients are changed with the double compression factors: Q_1 and Q_2 . The difference of the distributions is analyzed in this paper. A new method—dithering pattern is proposed to detect double JPEG compression. Meanwhile, the necessary and sufficient conditions about dithering pattern are obtained. Experimental results show that this method is effective.

Keywords JPEG double compressing, quality factor, dithering pattern

1 引言

许多隐写算法的流程中都有对原来 JPEG 图像进行再压缩的处理过程,如 F5、OutGuess^[1,2]。本文即从这里入手,研究检测图片是否被双压缩(简称双压)作为图片是否被隐写的手段之一。

文献[3]提到了 JPEG 格式文件经过双压后,图像的统计特性发生了改变,并以此作为图片是否被篡改的衡量指标。通过实验发现,文献[3]中所提到的周期模式是在一定条件下才出现的。所谓周期模式是指图像的 AC 系数直方图具有周期性的特征,其与自然图像(经过一次压缩)的 AC(alternating current)系数直方图满足 Laplacian 分布有着明显的区别。当第 2 次压缩采用的质量因子小于第 1 次压缩的质量

因子时,双压后的 DCT(discrete cosine transform)系数直方图仍然近似满足 Laplacian 分布,没有出现上述周期模式。因此,本文提出另一种 JPEG 双压检测方法——抖动模式,其具体定义见 4.1 节。

2 JPEG 双压缩流程

JPEG 算法分为有损编码和无损编码两个阶段。在有损编码阶段,首先将输入的图像分成若干个 8×8 的子块,然后对每一个子块进行 DCT 变换,并将变换后的系数用量化表量化,量化在 JPEG 压缩中起关键作用,大量的冗余信息在量化后被丢掉。无损编码阶段使用熵编码,不会造成信息的丢失。

JPEG 双压缩流程是指使用不同质量因子 Q_1 , Q_2 对图片进行压缩,图 1 为 JPEG 双压缩流程图。

收稿日期:2006-06-30;改回日期:2006-09-15

第一作者简介:戴蒙(1979~),女,2004 年获上海师范大学硕士学位,现为华东理工大学信息科学与工程学院博士研究生。目前研究方向为图形图像处理、图像的隐写分析。E-mail:damon_1111@163.com

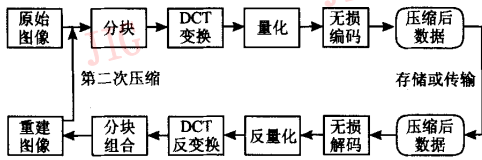


图 1 JPEG 双压缩过程

Fig. 1 Procedure of double JPEG compression

3 JPEG 双压缩质量因子分析

K_1 为经过一次压缩后的 AC 系数, K_2 为 K_1 经过再压缩后的 AC 系数, 它们的量化表和量化因子分别为 q_1, q_2 和 Q_1, Q_2 , 则 K_1, K_2 满足

$$K_2 = \text{integer_round} \left[\frac{K_1 \cdot q_1}{q_2} \right] \quad (1)$$

integer_round 为四舍五入运算, 式(1)等价于

$$K_2 - 0.5 \leq \frac{K_1 \cdot q_1}{q_2} < K_2 + 0.5 \quad (2)$$

设 Q 为质量因子, s 为标准量化表的缩放比例, 则

$$s = \begin{cases} 50/Q & \text{if } Q \leq 50 \\ 2 - Q/50 & \text{if } Q > 50 \end{cases} \quad (3)$$

q_s 表示标准量化表(下角 s 代表 standard), q 表示实际使用的量化表, 则

$$q = q_s \cdot s \quad (4)$$

由式(3)、式(4)可得到 $\frac{q_1}{q_2}$ 的关系满足

$$\frac{q_1}{q_2} = \frac{s_1}{s_2} = \begin{cases} \frac{Q_2}{Q_1} & \text{if } Q_1 \leq 50 \text{ 和 } Q_2 \leq 50 \\ \frac{100 - Q_1}{100 - Q_2} & \text{if } Q_1 > 50 \text{ 和 } Q_2 > 50 \\ \frac{2500}{Q_1(100 - Q_2)} & \text{if } Q_1 \leq 50 \text{ 和 } Q_2 > 50 \\ \frac{(100 - Q_1)Q_2}{2500} & \text{if } Q_1 > 50 \text{ 和 } Q_2 \leq 50 \end{cases} \quad (5)$$

同时由式(5)可得出双压缩后可能出现的各种情况(见表 1)。

若 $K_2 = K_1$ 等价于 $\text{integer_round} \left[\frac{K_1 \cdot q_1}{q_2} \right] = K_1$, 则式(6)必定成立:

$$K_1 - 0.5 \leq \frac{K_1 \cdot q_1}{q_2} < K_1 + 0.5 \quad (6)$$

要使得 $K_2 > K_1$, 则式(7)必定成立:

表 1 约束条件分析表

Tab. 1 Table of analysis conditions

情况 1	情况 2	情况 3	情况 4
$Q_1 > 50$ 和 $Q_2 > 50$	$Q_1 \leq 50$ 和 $Q_2 \leq 50$	$Q_1 \leq 50$ 和 $Q_2 > 50$	$Q_1 > 50$ 和 $Q_2 \leq 50$
(1) $K_2 = K_1$ (2) $K_2 > K_1$ (3) $K_2 < K_1$	(1) $K_2 = K_1$ (2) $K_2 > K_1$ (3) $K_2 < K_1$	(1) $K_2 = K_1$ (2) $K_2 > K_1$ (3) $K_2 < K_1$	(1) $K_2 = K_1$ (2) $K_2 > K_1$ (3) $K_2 < K_1$

$$\frac{K_1 \cdot q_1}{q_2} \geq K_1 + 0.5 \quad (7)$$

要使得 $K_2 < K_1$, 则式(8)必定成立:

$$\frac{K_1 \cdot q_1}{q_2} \geq K_1 - 0.5 \quad (8)$$

下面根据表 1 所分 Q_1, Q_2 的不同情况进行分析:

3.1 $Q_1 > 50$ 和 $Q_2 > 50$

据式(5)~式(8)在此约束条件下可转化为

$$\left(\frac{K_1 + 0.5}{K_1} \right) Q_2 - \frac{50}{K_1} \leq \left(\frac{K_1 - 0.5}{K_1} \right) Q_2 + \frac{50}{K_1} \quad (9)$$

$$Q_1 \leq \left(\frac{K_1 + 0.5}{K_1} \right) Q_2 - \frac{50}{K_1} \quad (10)$$

$$Q_1 > \left(\frac{K_1 - 0.5}{K_1} \right) Q_2 + \frac{50}{K_1} \quad (11)$$

3.2 $Q_1 \leq 50$ 和 $Q_2 \leq 50$

式(6)~式(8)在此约束条件下可转化为

$$\left(\frac{K_1 - 0.5}{K_1} Q_1 \right) \leq Q_2 < \frac{(K_1 + 0.5) Q_1}{K_1} \quad (12)$$

$$Q_2 \geq \frac{(K_1 - 0.5) Q_1}{K_1} \quad (13)$$

$$Q_2 < \frac{(K_1 - 0.5) Q_1}{K_1} \quad (14)$$

3.3 $Q_1 \leq 50$ 和 $Q_2 > 50$

式(6)~式(8)在此约束条件下可转化为

$$\frac{K_1 \cdot 2500}{(K_1 + 0.5)(100 - Q_2)} < Q_1 \leq \frac{K_1 \cdot 2500}{(K_1 - 0.5)(100 - Q_2)} \quad (15)$$

$$Q_1 \leq \frac{K_1 \cdot 2500}{(K_1 + 0.5)(100 - Q_2)} \quad (16)$$

$$Q_1 > \frac{K_1 \cdot 2500}{(K_1 - 0.5)(100 - Q_2)} \quad (17)$$

3.4 $Q_1 > 50$ 和 $Q_2 \leq 50$

式(6)~式(8)在此约束条件下可转化为

$$\frac{(K_1 - 0.5) \cdot 2500}{K_1 \cdot (100 - Q_1)} \leq Q_2 < \frac{(K_1 + 0.5) \cdot 2500}{K_1 \cdot (100 - Q_1)} \quad (18)$$

$$Q_2 \geq \frac{(K_1 + 0.5) \cdot 2500}{K_1 \cdot (100 - Q_1)} \quad (19)$$

$$Q_2 < \frac{(K_1 - 0.5) \cdot 2500}{K_1 \cdot (100 - Q_1)} \quad (20)$$

4 双压的检测方法

4.1 抖动模式及其出现的必要条件

定义 1 如果 x 满足 $p(x)$ 分布,同时存在 x_n 使得 $p(x_n) < p(x_{n-1})$ 及 $p(x_n) < p(x_{n+1})$,且 $p(x_{n+1})$ 与 $p(x_{n-1})$ 有一定的落差,则称 $p(x)$ 具有抖动模式。

由于图像 DCT 系数的 AC 系数 K 是满足 Laplacian 分布或广义高斯分布的^[4],即

$$p(x) = \frac{\lambda}{2} e^{-\lambda|x|}, \lambda = \frac{1}{E(x)} \quad (21)$$

所以当 $|x_1| > |x_2|$,必有 $p(x_1) < p(x_2)$ 成立。

若双压后的 AC 系数不变,即 $K_2 = K_1$,则 AC 系数分布不变;若双压后 $K_2 < K_1$,则有

$$p_2(v_1) = \sum_{u_1=(u_1)_{\min}}^{(u_1)_{\max}} p_1(u_1) \quad (22)$$

$$p_2(v_2) = \sum_{u_2=(u_2)_{\min}}^{(u_2)_{\max}} p_1(u_2)$$

上式中 p_1, p_2 分别为一次压缩和二次压缩后 AC 系数的概率分布; $v < u_{\max}$; u_{\min} 和 u_{\max} 分别代表与 K_2 有关系的 K_1 的最小值和最大值, $(u_1)_{\max} < (u_2)_{\min}$ 。举例说明 v 和 u 之间的关系: $K_1 = 1, 2, 3, 4, 5$, 经过除以 2, 且四舍五入运算后, $K_2 = 1, 2, 3$, 即 $v_1 = 1, v_2 = 2, v_3 = 3$ 。当 $v_1 = 1$ 时, $u_1 = 1, 2$, 其中 $(u_1)_{\min} = 1, (u_1)_{\max} = 2$ 。因为由式(21)、式(22)可得 $p_2(v_1 = 1) > p_2(v_2 = 2)$, 同

理可得 $p_2(v_2 = 2) > p_2(v_3 = 3) \dots$, 所以不存在整数 n 满足抖动的条件。

若 $K_2 > K_1$, 不考虑图像在 DCT 和 IDCT 过程的损失, 那么在理论上, 则有 $p_2(0) = p_1(0)$, 此时因为式(22)中 $v > u_{\max}$, 所以 $p_2(1) = 0$ 。实际情况是 $p_2(1)$ 不为 0, 但小于其周围的值, 所以双压后的直方图必在 $-1, 1$ 处抖动。

综上所述, 若双压前后的 AC 系数满足 $K_2 > K_1$, 则双压后的图像必出现抖动模式; 若双压后的 AC 系数出现抖动模式, 则双压前后的 AC 系数满足 $K_2 > K_1$, 所以双压后 AC 系数出现抖动模式的充分必要条件是: $K_2 > K_1$ 。根据第 3 节所述可得满足 $K_2 > K_1$ 的 Q_1, Q_2 取值。

4.2 $K_2 > K_1$ 的检测方法

满足 $K_2 > K_1$ 的这类图像的 AC 系数直方图会出现 4.1 节所述的抖动模式。因此只要考察是否存在整数 n 使得下式成立:

$$p(x_n) < p(x_{n-1}) \text{ 并且 } p(x_n) < p(x_{n+1}) \quad (23)$$

其中, x_n 表示 AC 系数等于 n , 则可判别图像经过二次压缩。图 2 及表 2 即为这类双压图像的直方图及部分 AC 系数值。由图 2 可见, AC 系数为 $-4, -1, 1, 4$ 处出现了抖动。

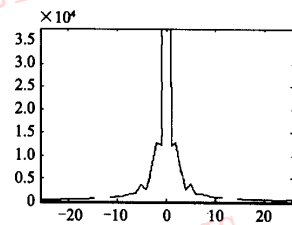


图 2 $Q_1 = 75, Q_2 = 85$ 的 AC 系数直方图

Fig. 2 The histogram of AC coefficients for double JPEG compression with quality 75 followed by 85

表 2 图 2 中部分 AC 系数值与其对应的个数

Tab. 2 Partial AC coefficients and numbers in Fig. 2

AC 值	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
个数	1141	1106	1470	1634	1817	3727	2578	7215	12938	12384	599669	12166	12984	7094	2837	3952	1723	1740	1453	1196	977	857

4.3 $K_2 < K_1$ 的检测方法

对这类图像, 若用文献[3]方法检测, 首先使用扫描仪得到 500 幅位图, 并用质量因子 90 来压缩位图; 然后使用 OutGuess^[5] 隐写软件选择压缩参数为 80 来嵌入信息, 即 $Q_1 = 90, Q_2 = 80$ 。结果 500 幅图像的 AC 系数直方图并没有出现文献[3]所述的周

期模式, 这也印证了第 4 节的理论推导。图 3 及表 3 数据为其中的一幅, 由图 3、表 3 可见, 可以推断周期模式在检测该类图片有一定的局限性。下面给出本文对这类图片的检测方法:

令测试图片 T 量化后的 AC 系数直方图为 H_{r_i} , 其中 $r_i = L_i \dots R_i, L_i = \min(K_i)$ (下角 t 代表 test, 下

同), $R_i = \max(K_i)$; 参照图片 R (只经过一次压缩) 量化后的 AC 系数直方图为 H_r , $r_i = L_r \cdots R_r$ (下角 r 代表 reference), $L_r = \min(K_r)$, $R_r = \max(K_r)$, 其中参照图片 R 的 L_r, R_r 等于或无限接近于 L_i, R_i 。考察下式:

$$H_s = \text{Sub}(H_{r_i} - H_i) \quad s = -c, -c + 1, \dots, c - 1, c \quad (24)$$

如果测试图片只经过一次压缩, 且 H_r 是满足 Laplacian 分布或广义高斯分布的, 则 H_s (下角 S 代表 sub) 总体上在区间 $-c, -c + 1, \dots, -1$ 和区间 $1, \dots, c - 1, c$ 单调, 不会出现抖动; 如果测试图片经过了二次压缩, H_r 不满足 Laplacian 分布或广义高斯分布, 则 H_s 在区间 $-c, -c + 1, \dots, -1, 1, \dots, c - 1, c$ 有抖动模式出现。图 4 为经过 $Q_1 = 90, Q_2 = 80$ 两次压缩后图像的 H_s 直方图, 根据图 4 的波形及表 4 的数据, 可见 AC 系数为 $-7, -4, -2, 2, 4, 6$ 处出现抖动, 因此可以较容易地检测出该类二次 JPEG 压缩图像。500 幅图像测试图像使用该方法检测的正确率达到 100%。

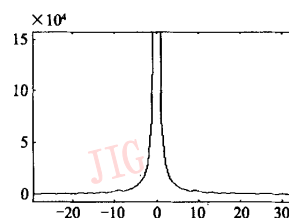


图 3 $Q_1 = 90, Q_2 = 80$ AC 的系数直方图

Fig. 3 The histogram of AC coefficients for double JPEG compression with quality 90 followed by 80

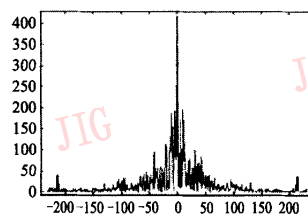


图 4 图 3 的 H_s 直方图

Fig. 4 The H_s histogram of Fig. 3

表 3 图 3 中部分 AC 系数值与其对应个数

Tab. 3 Partial AC coefficients and numbers in Fig. 3

AC 值	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
个数	2835	4035	2681	4362	5631	7655	11117	16224	29285	73404	278193573	038	29578	16240	10864	7737	5636	4462	2865	3952	2852	2332

表 4 图 4 中部分系数值与其对应个数

Tab. 4 Partial AC coefficients and numbers in Fig. 3

AC 值	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
个数	187	123	42	5	19	193	83	170	78	91	416	204	15	80	12	174	8	18	51	154	197	148

5 结论

在图片的实际运用过程中, 为了不引起图片质量的大幅降低, 一般两次压缩因子均大于 50, 本文的检测方法同样适用于 Q_1, Q_2 使得 $K_2 > K_1$ 或 $K_2 < K_1$ 成立的另外 3 种情况。但是, 本文所提出的检测方法并不适用于两次压缩因子 Q_1, Q_2 使得 $K_2 = K_1$ 的情况; 同时, 对于某个 AC 系数出现轻微抖动并不能说明该图片经过二次压缩, 而应看其直方图的总体走向。

实验表明, 本文提出的 JPEG 图片双压缩的检测方法是有效的, 该检测方法还可以用在图片是否被篡改的工作中。下一步将进行如何检测 JPEG 双压缩质量因子相近的图片。

参考文献 (References)

- 1 Fridrich J, Goljan M, Hogeia D, et al. Quantitative steganalysis of digital images: Estimating the secret message length [J]. ACM Multimedia Systems Journal (Special issue on Multimedia Security), 2003, 9(3): 288 ~ 302.
- 2 Fridrich J, Goljan M, Hogeia D. Steganalysis of JPEG images: breaking the F5 algorithm [A]. In: Proceedings of 5th International Workshop Information Hiding [C], Noordwijkerhout, Netherlands: Springer-Verlag, 2002: 310 ~ 323.
- 3 Popescu Alin C, Farid Hany. Statistical tools for digital forensics [A]. In: International Workshop on Information Hiding [C], Toronto, Canada, 2004: 128 ~ 147.
- 4 Eggerton J D, Srinath M D. Statistical distributions of image DCT coefficients [J]. Computers and Electrical Engineering, 1986, 12(3/4): 137 ~ 145.
- 5 Niels Provos. OutGuess2.0 [EB/OL], <http://www.outguess.org/>, 2001-10-02.