

基于位置直方图的单次 JPEG 压缩质量因子检测

张军华, 吕东辉

(上海大学 通信与信息工程学院, 上海 200072)

摘要: 提出了一种针对经过 JPEG (joint photographic experts group, 联合图像专家小组) 压缩但保存为 bmp 或 tiff 格式图像的质量因子检测新方法. 通过统计待测图像灰度直方图, 获取最大值所对应的指定灰度级. 然后将待测图像进行 8×8 分块, 统计每个 8×8 小块中指定灰度级所对应的位置直方图. 最后将所有块对应的位置直方图累加, 作为待测图像中指定灰度级的位置直方图, 根据压缩前后的位置直方图差异估计质量因子. 实验结果表明, 本方法可以有效地检测图像质量因子且正确率较高. 最后, 分析了单次高质量因子压缩图像的相似度曲线在高质量因子处的导数特点, 并提出了改进方案. 实验结果表明, 改进方案较大地提高了高质量因子的检测正确率.

关键词: JPEG 压缩; 质量因子; 位置直方图; 灰度级

中图分类号: TN 911.73

文献标志码: A

文章编号: 1007-2861(2013)02-0120-06

Quality Factor Estimation for Single JPEG Compression Based on Position Histogram

ZHANG Jun-hua, LÜ Dong-hui

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China)

Abstract: This paper proposes a new method for estimation of JPEG (joint photographic experts group) quality factor for images once compressed but saved in the bmp or tiff format. By calculating a gray histogram of the test image, the gray value was set at the maximum of the gray histogram occurs. The test image is split into 8×8 blocks, and a position histogram of the specified gray value g in each 8×8 block is obtained. The position histograms of all blocks are combined to produce a position histogram of the entire image, from which the quality factor of the test image is estimated. Experiments show that the method has a high correct rate. In addition, derivative characteristics of similarity curves of images having undergone single JPEG compression with high quality are analyzed at some high quality factors, and an improved version of a previous method is proposed, and verified in experiments.

Key words: JPEG (joint photographic experts group) compression; quality factor; position histogram; grayscale

随着计算机技术和图像处理技术的发展, 数字图像在网络或者现实生活中随处可见. 不仅如此, 还可以方便地利用软件来处理各种各样的数字图像, 这给日常生活带来了巨大的便利和改变. 但是, 肉眼对利用各种软件对数字图像进行的修改、合成的篡改图像以及利用计算机图形技术生成的图像往往难以辨别. 近年来, 不法分子利用数字图像处理软件制造出对社会具有不良影响, 甚至严重危害社会安定和国家利益的篡改图像, 因此研究图像取证课题已成为当务之急.

图像取证课题包括主动取证和被动取证^[1-2], 主动取证要求取证者对图像预先嵌入脆弱水印或者预先计算hash, 具有一定的局限性; 被动取证是在没有待测图像的附加信息的情况下, 仅通过分析图像的统计特性等来辨别真伪, 因此具有更高的应用价值, 但取证难度要大于主动取证.

JPEG (joint photographic experts group, 联合图像专家小组)是目前应用最广泛的图像压缩格式, 而其相关研究一直是学术界的热点. 在通常情况下,

收稿日期: 2012-07-10

基金项目: 国家自然科学基金资助项目(61073190)

通信作者: 吕东辉(1966—), 男, 副教授, 博士, 研究方向为图像认证、医学图像处理和3D图像重建等. E-mail: dhlu@shu.edu.cn

对图像进行篡改或者修改后往往都要进行保存操作, 而此操作将导致图像的 JPEG 压缩, 这样就会出现 JPEG 压缩痕迹, 从而给研究人员提供有利的取证线索. 在根据这些线索进行 JPEG 图像取证的过程中, 通过 JPEG 文件头^[3]获取 JPEG 图像相关信息, 或者人为地估计经过 JPEG 压缩但保存为位图格式图像的相关信息, 均有利于进一步的取证工作. 例如, 迭代估计出待测图像的原始量化表, 以大致定位出篡改区域^[4]. 到目前为止, 已有研究者提出了多种检测单次 JPEG 压缩质量因子和检测重压缩^[5-6]的方法, 并将其应用于合成图像的检测^[1,4,7]. Luo 等^[8]对 JPEG 压缩所产生的误差进行了深入分析, 提出忽略压缩中的饱和块, 通过比较压缩前后图像的对应非饱和块的相似度, 来估计待测图像的质量因子. 更重要的是, 文献^[8]的方法对 32×32 块仍然具有比较好的估计效果, 但对于质量因子超过 97 的压缩, 该方法会出现误判. Fu 等^[9]提出了基于广义 Benford 定律的单次 JPEG 压缩质量因子检测方法, 指出未压缩图像离散余弦变换 (discrete cosine transform, DCT) 系数首位数字的分布服从 Benford 定律, 但压缩图像并不严格服从 Benford 定律. 通过修正 Benford 曲线的相关参数, 可以使 JPEG 压缩图像 DCT 系数首位数字的分布曲线与修正后的 Benford 曲线一致, 即广义的 Benford 定律. 但是, 要确定广义 Benford 曲线的相关参数来建立模型是较困难的, 而这些参数会影响质量因子检测结果. 根据 JPEG 图像的 DCT 系数直方图所具有的周期性的特点, Ye 等^[10]计算出了 8×8 块中指定位置的量化步长, 其应用价值更高, 但因高频部分的量化步长较大, 使得 DCT 域中高频部分的绝大多数系数为 0 且不具周期性, 因此文献^[10]的方法对此种情况不具检测能力.

根据已有研究可以发现 JPEG 压缩的一个重要特性, 即图像压缩前后必然存在空间域和频率域的差异. 基于此特点, 本研究提出利用位置直方图来估计待测图像的 JPEG 压缩质量因子. 首先, 计算可疑图像的灰度直方图, 获取灰度直方图最大值所对应的灰度级, 计算该灰度级所对应的位置直方图. 然后, 对待测图像进行一系列不同质量因子的二次 JPEG 压缩, 同样计算出每幅经二次 JPEG 压缩的图像相应的位置直方图. 最后, 根据待测图像位置直方图与二次 JPEG 压缩图像位置直方图间的差异, 估计出单次 JPEG 压缩质量因子. 实验结果表明, 该方法能够较准确地估计出单次 JPEG 压缩质量因子.

1 JPEG 压缩与解压缩

在整个图像 JPEG 压缩的过程中, 编码过程可视为对图像质量没有影响, 而对图像质量和压缩率起关

键作用的是 DCT 系数的量化. 求取 DCT 系数的方法是将原始图像分割为不重叠的 8×8 小块, 在每个小块中作二维 DCT 变换:

$$F(u, v) = \frac{2}{N} C(u, v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cdot \cos \frac{\pi u(2x+1)}{2M} \cos \frac{\pi v(2y+1)}{2N}, \quad (1)$$

式中, $C(0, 0) = 0.5$, $C(u, v) = 1$ ($u, v = 0, 1, \dots, 7$), $f(x, y)$ 表示原始图像块中 (x, y) 处的灰度值, M, N 分别表示原始图像块的高和宽.

通过式 (1) 求取 DCT 系数后, 对 DCT 系数的量化操作会导致 DCT 系数的失真, 即量化误差, 使得图像空间域在压缩前后发生变化.

在解压缩过程中, DCT 反变换过程将产生两种特殊的数值, 一种是小于 0 的数值, 另一种是大于 255 的数值. 对于这两种特殊的数值, 通常的做法是将小于 0 的数值记为 0, 将大于 255 的数值记为 255, 这导致了一定误差, 称为截尾误差. 由于 DCT 反变换之后的结果并不是整数, 因此必须对 DCT 反变换之后的结果取整, 这会导致另一种误差, 称为取整误差.

质量因子检测算法本质上是根据量化误差来进行质量因子检测的. 不同质量因子 JPEG 压缩导致的量化误差越大, 压缩前后图像空间域差异越大, 越有助于准确估计质量因子, 反之, 则质量因子检测难度越大. 而截尾误差和取整误差会加剧压缩前后图像空间域差异, 从而影响基于量化误差的质量因子检测算法, 因此本方法不考虑包含像素值为 0 或者 255 的图像块, 这样可以有效地消除截尾误差的影响. 由于本方法是基于位置直方图的检测方法, 只利用每个 8×8 图像块中的部分像素, 从而在一定程度上减轻了取整误差的影响, 提高了算法的有效性.

2 基于位置直方图的单次 JPEG 压缩质量因子

2.1 JPEG 压缩分析

为了对 JPEG 压缩进行更深入的研究, 引入以下实验: 不妨假设给定的待测图像是质量因子为 37 的单次 JPEG 压缩图像, 对其进行质量因子分别为 36, 37, 38 的二次压缩, 然后对比这 3 幅二次压缩图像与对应待测图像间的差异大小. 例如, 给定一幅质量因子为 37 的单次压缩 Pepper 图像, 取该图像中第 67 块灰度值矩阵作为待测图像块, 设 G 为对应于待测图像块的灰度值矩阵, D_j 为二次压缩图像块与对应待测图像块间的绝对差异, 即对应图像灰度矩阵间的绝对差异, 其中 $j = 36, 37, 38$, 对比结果如下.

$$G = \begin{bmatrix} 203 & 209 & 206 & 194 & 188 & 191 & 187 & 176 \\ 207 & 213 & 209 & 196 & 191 & 194 & 191 & 182 \\ 205 & 210 & 205 & 191 & 186 & 190 & 189 & 180 \\ 198 & 203 & 197 & 183 & 177 & 181 & 178 & 169 \\ 198 & 204 & 200 & 186 & 178 & 178 & 172 & 160 \\ 206 & 214 & 213 & 200 & 190 & 184 & 171 & 154 \\ 208 & 219 & 221 & 209 & 197 & 185 & 164 & 143 \\ 203 & 216 & 221 & 210 & 195 & 179 & 154 & 130 \end{bmatrix},$$

$$D_{36} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$D_{37} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$D_{38} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

通过比较可以发现, 当前后压缩质量因子相同时, 对应图像块间的不同像素个数最少, 即对应 D_j 中值为 0 的个数最多. 通过大量实验可以进一步证明该观点: 对未压缩彩色图像库 (uncompressed colour image database, UCID) 中 1 338 幅图像作质量因子为 50 的 JPEG 压缩, 获取 1 338 幅单次压缩图像. 然后, 对所有单次压缩图像作以 5 为间隔的 5~100 的一系列二次压缩, 并求出单次压缩图像与所对应的二次压缩图像间不同像素个数与图像像素总数的比值, 即变化率 d . 对二次压缩质量因子相同的所有图像计算 d 的平均值, 记为 \bar{d} , 结果如图 1 所示.

2.2 使用位置直方图进行压缩因子估计

首先, 阐述位置直方图的概念. 类似于 DCT 系数直方图^[11], 所谓位置直方图就是针对一幅图像中的一

个指定灰度级, 统计该灰度级在所有 8×8 块中 64 个位置的出现次数所形成的映射图.

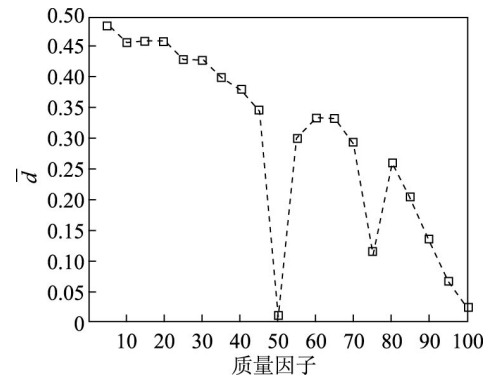


图 1 压缩前后像素平均变化率

Fig. 1 Average ratio of the differences between the compressed images and the recompressed images

取图像尺寸为 256×256 的 Pepper 图像中的第一个 32×32 块作为目标图像块, 统计该图像块中灰度级 184 所对应的位置直方图. 具体做法如下: 首先对目标图像块进行 8×8 分块, 每个小块记为 $b_k, k = 1, 2, \dots, 16$. 设 $f(x, y)$ 表示 8×8 块中 (x, y) 处的像素值, $h_k(n)$ 是第 k 个 8×8 块的位置直方图, $n = 1, 2, \dots, 64$, 即

$$h_k(n) = \begin{cases} 1, & f(x, y) = 184, \\ 0, & f(x, y) \neq 184, \end{cases} \quad (2)$$

从而该 Pepper 图像中的目标块对应的位置直方图为 $H(n)$ (见图 2). $H(n)$ 的计算公式为

$$H(n) = \sum_{k=1}^{16} h_k(n). \quad (3)$$

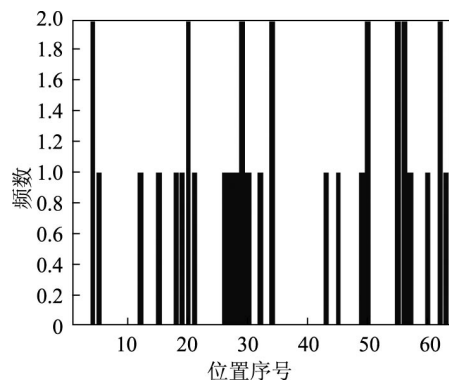


图 2 目标图像块中灰度级 184 所对应的位置直方图

Fig. 2 Position histogram of grayscale level 184 in objected image blocks

具体检测算法步骤如下.

(1) 将待测图像转换为 YCbCr 颜色模式, 取 Y 分量作为灰度图像 I , 并对 I 作适当裁剪, 使得图像的宽和高能整除 8.

(2) 计算 I 的灰度直方图, 取灰度直方图中最大值所对应的灰度级, 记为 g . 为了消除截尾误差的干扰, 若 g 为 0 或 255, 则取 g 为直方图中次大值所对应的灰度级.

(3) 对灰度图像 I 分别进行质量因子 30~100 的二次 JPEG 压缩, 生成 71 幅二次 JPEG 压缩图像. 考虑到实际应用中待测压缩图像的质量因子往往大于 30, 因此将二次压缩质量因子的下限设为 30.

(4) 对上述灰度图像 I 和 71 幅二次 JPEG 压缩图像按式 (2) 和 (3) 计算出每幅图像中灰度级 g 对应的位置直方图, 分别记为 $H_0, H_i, i = 1, 2, \dots, 71$.

(5) 计算 H_0 与 H_i 间的距离 $C_i = \sum_{n=1}^{71} |H_0(n) - H_i(n)|$.

(6) 求出 C_i 中的最小值, 该最小值对应的二次压缩质量因子即为所求值.

为了便于表述, 下面分别用 J_1, J_2 表示单次压缩图像和二次压缩图像.

2.3 实验及结果

在现实生活中, 不法分子对篡改图像进行 JPEG 压缩后一般保存为 bmp 或 tiff 格式, 以避人耳目. 正因如此, 在单次 JPEG 压缩质量因子检测前, 首先需要确定待检测图像已经过 JPEG 压缩, 而目前已经有多种方法^[8-9]可以有效地判定一幅图像是否已经过 JPEG 压缩. 本研究的重点在于单次 JPEG 压缩图像质量因子的估计, 选择 UCID 中的 1 338 幅图像作为

实验素材, 首先将它们转化为灰度图像, 然后进行单次 JPEG 压缩, 质量因子取 50~100 共 51 种情形, 由此获取 14 718 幅单次 JPEG 压缩图像, 并保存为 bmp 格式, 将这些图像作为待测图像. 根据上述步骤, 通过实验得知平均正确率达到 92.3%; 另一方面, 通过实验发现 7.7% 的误判中有 63.2% 的判决结果是 100, 这是由于质量因子为 100 的二次压缩对图像造成的影响很小, 甚至小于相同质量因子的二次压缩所造成的影响, 从而导致误判. 为了提高质量因子检测正确率, 进一步设计以下规则.

规则 1 如果估计结果不是 100, 则将当下估计值作为最终估计结果; 如果估计结果是 100, 则根据规则 2 继续进行估计.

规则 2 将待测图像灰度直方图中次大值对应的灰度级设定为 g , 重新进行估计, 然后再根据规则 1 作判定. 如此循环估计, 且限定估计次数不超过 η , 实验中取经验值 $\eta = 5$.

文献 [8] 提出了图像压缩前后的空间域相似度的数学模型, 即

$$R(J_1, J_2) = \frac{|E|}{MN}, \quad (4)$$

$$E = \{(x, y) | J_1(x, y) = J_2(x, y), 1 \leq x \leq M, 1 \leq y \leq N\}, \quad (5)$$

式中, $|E|$ 表示 E 中像素个数, M, N 表示 J_1 图像的大小. 根据图像压缩前后的空间域相似度 $R(J_1, J_2)$ 估计压缩质量因子, 相似度越大, 则表示对应 J_2 的质量因子是估计结果, 反之, 则不是. 本研究采用同样数据库对文献 [8] 方法进行了仿真, 实验结果如表 1 所示.

表 1 文献 [8] 方法和本方法的质量因子检测正确率

Table 1 Average accuracies for different quality factors using the method of Ref. [8] and the proposed method

| | 质量因子 | | | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
| 本方法 | 0.961 | 0.978 | 0.979 | 0.970 | 0.978 | 0.969 | 0.987 | 0.991 | 0.963 | 0.928 | 1 |
| 文献 [8] 方法 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

从表 1 可以看出, 本方法所得结果略差于文献 [8] 方法. 另外, 在实验过程中发现, 文献 [8] 方法对单次压缩质量因子为 97, 98, 99 的情形并不适用. 同样选择 UCID 中所有图像进行质量因子分别为 96, 97, 98, 99 的单次 JPEG 压缩, 可获得 5 552 幅单次压缩图像, 保存为 bmp 格式, 并将它们作为待测图像, 实验结果如表 2 所示.

观察表 2 可以发现, 文献 [8] 方法在单次压缩质量因子 $q = 97, 98, 99$ 的情况下, 质量因子检测率几乎为 0, 而本方法检测率良好. 通过分析发现, q 对应的量化表在低频部分大多数是 1, 但仍有少部分量化步长大于

表 2 文献 [8] 方法和本方法的高质量因子检测正确率

Table 2 Average accuracies for high quality factors using the method of Ref. [8] and the proposed method

| | 质量因子 | | | |
|-----------|-------|-------|-------|-------|
| | 96 | 97 | 98 | 99 |
| 本方法 | 0.866 | 0.844 | 0.827 | 0.900 |
| 文献 [8] 方法 | 0.991 | 0.031 | 0 | 0 |

1. 而质量因子 100 对应的量化步长均为 1, 因为当量化步长较大时, 量化误差较大, 从而质量因子为 q 的二次压缩造成的量化误差更大, 而质量因子 100 的二次压缩所造成的量化误差更小, 因此文献 [8] 方法将 q 误

判为 100. 通过以下实验可以证明此观点, 即对 64 幅质量因子为 97 的二次压缩图像进行质量因子分别为 97 和 100 的二次压缩, 求取二次压缩前后图像间的空间差异, 即像素的改变个数, 结果如图 3 所示.

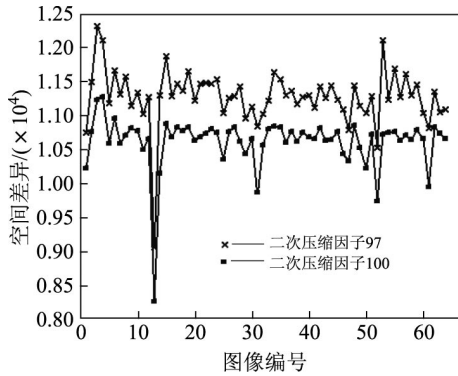


图 3 $q = 97$ 时二次压缩前后空间差异对比

Fig. 3 Comparison of the spatial differences between the compressed images and the recompressed images when q equals to 97

另外, 图 4 给出了压缩质量因子为 97, 98, 99, 100 的图像对应的相似度曲线, 其中横轴表示 J_2 质量因子, R 表示压缩前后图像相似度.

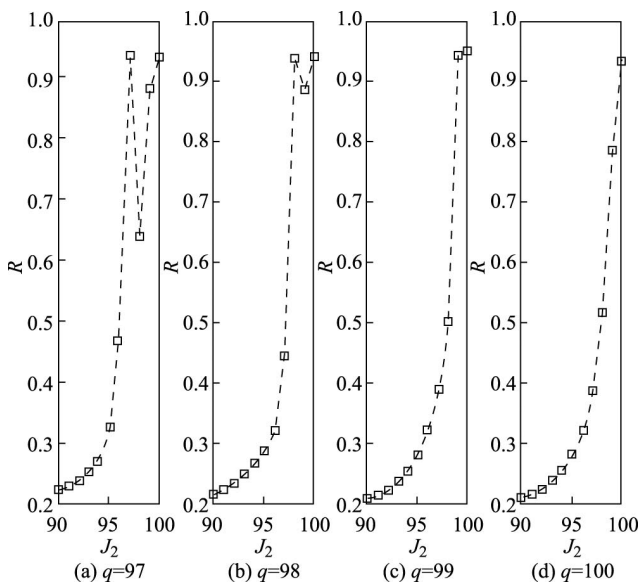


图 4 JPEG 图像压缩前后相似度曲线

Fig. 4 Curves of the similarity between the compressed images and the recompressed images

此处只关注 [97, 100] 的区间范围, 因此只绘制 [90, 100] 区间范围的相似度曲线. 观察图 4 可以得出, J_2 质量因子为 100 时相似度最大, 因此估计结果都是 100, 图 4(a)~(c) 均属于误判情况. 通过实验发现, 对于不同的图像, 只要 J_1 质量因子一致, 其对应的相似度曲线的走向趋势就是一致的. 更重要的是, 在 J_1 质量

因子处相似度曲线左导数大于 0, 右导数小于 0, 这也符合压缩前后质量因子一致时相似度最大的特性. 但是, 当 J_1 质量因子为 99 时, 相似度曲线在该处的左右导数均大于 0, 不过右导数特别小, 可以通过设定阈值 τ 来将这种情形与其他情形区分开; 当 J_1 质量因子为 100 时, 相似度曲线在该处不存在右导数. 利用这些特性, 本研究对文献 [8] 方法作出改进, 提出规则 3 和规则 4, 使得该算法对 J_1 质量因子为 97, 98, 99 时的情形同样有效.

规则 3 估计结果不是 100, 即将其作为最终估计结果.

规则 4 估计结果是 100, 则分别计算出相似度曲线在 97, 98, 99 处的左右导数. 若在 97, 98 处左导数大于 0, 且右导数小于 0, 则估计结果分别是 97, 98; 若在 99 处左右导数均大于 0, 且右导数小于阈值 τ , 则估计结果是 99; 否则估计结果是 100.

采用相同的图像库对改进算法进行实验, 实验中取经验值 $\tau = 0.02$, 实验结果如表 3 所示.

表 3 文献 [8] 方法和改进方法的高质量因子检测正确率

Table 3 Average accuracies for high quality factors using the method of Ref. [8] and the improved method

| | 质量因子 | | |
|-----------|-------|----|----|
| | 97 | 98 | 99 |
| 本改进方法 | 1 | 1 | 1 |
| 文献 [8] 方法 | 0.031 | 0 | 0 |

从文献 [8] 中可知, 文献 [9] 方法对于质量因子低于 50 的二次压缩图像的检测正确率均低于 50%, 并且对于质量因子高于 50 的二次压缩图像的检测正确率均低于 90%, 可见本方法明显优于文献 [9] 方法.

下面进一步研究待测图像是二次 JPEG 压缩图像的情形. 对此给出以下一组实验, 首先将 UCID 中所有图像进行质量因子为 80 的二次压缩, 然后对二次压缩图像进行以 5 为间隔, 50~100 的一系列二次压缩, 最后将二次压缩图像作为待测图像, 应用本方法进行检测, 结果如表 4 所示.

实验结果表明, 对于二次 JPEG 压缩图像, 本方法检测所得结果是第二次压缩的质量因子, 检测的正确率与表 1 列出的二次 JPEG 压缩的质量因子的检测结果类似, 但不能对二次 JPEG 压缩图像的第一次压缩的质量因子进行估计.

从算法的复杂度来看, 与文献 [8] 方法比较, 本方法更为复杂. 在实际检测中, 由于较低质量因子单次即可估计正确, 检测过程相对简单; 而高质量因子通过多次选取不同指定灰度级 g 并重新检测以确保检测的

表4 二次 JPEG 压缩图像质量因子检测正确率
Table 4 Average accuracies for images double JPEG compressed

| | 质量因子 | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
| 本方法 | 0.979 | 0.990 | 0.975 | 0.957 | 0.978 | 0.978 | 0.997 | 0.997 | 0.984 | 0.981 | 0.970 |

有效性, 因此检测过程较复杂. 在主频为 3.0 GHz 的 PC 机上, 选取 UCID 中的 300 幅图像进行实验. 通过统计, 文献 [8] 方法检测每幅图像平均耗时 1.348 5 s, 本方法平均耗时分为两种情况: 检测质量因子小于 96 的图像时, 每幅图像平均耗时 1.538 5 s; 检测质量因子大于等于 96 的图像时, 每幅图像平均耗时 3.384 0 s.

3 结束语

基于 JPEG 压缩会导致空域和频域变化的原理, 本研究提出了利用位置直方图来估计单次 JPEG 压缩图像的质量因子的新方法. 如果待测图像质量因子不超过 96, 检测估计效果比较理想, 但比文献 [8] 方法略差; 当待测图像质量因子为 97, 98, 99 时, 文献 [8] 方法失效, 而本方法检测效果仍然良好; 最后, 本研究利用相似度曲线在 97, 98, 99 处的导数所具有的特殊性质, 对文献 [8] 方法进行改进. 实验结果表明, 与文献 [8] 方法相比, 改进方法能够有效地检测质量因子为 97, 98, 99 的单次 JPEG 压缩图像, 正确率有了很大的提高.

质量因子检测算法均是针对标准量化, 即图像压缩过程都采用同一个标准量化体系, 而应用视觉特性进行优化的 JPEG 压缩^[12-13], 通常针对不同图像块的特性生成不同的量化表. 在估计单次压缩的质量因子时, 由于图像已被压缩, 因而无法准确推导出其压缩所用的量化表, 故在检测算法中就无法采用原量化表对待测图像进行二次压缩, 也无法进行质量因子估计.

因此, 下一步需要做的工作有以下几方面: ①深入研究位置直方图方法对于高质量因子估计的有效性及其原理; ②在已有的二次压缩研究的基础上, 深入研究压缩前后 DCT 系数的变化特征以及探索 JPEG 二次压缩检测方法; ③探索基于视觉特性优化 JPEG 压缩的原理, 进一步研究优化 JPEG 压缩的质量因子检测.

参考文献:

[1] 李晟, 张新鹏. 利用 JPEG 压缩特性的合成图像检测 [J]. 应用科学学报, 2008, 26(3): 281-287.

[2] 李晟. 基于 JPEG 特性与小波域共生矩阵特征的被动图像取证 [D]. 上海: 上海大学, 2008.

[3] KEE E, JOHNSON M K, FARID H. Digital image authentication from JPEG headers [J]. IEEE Trans Inf Forensics Security, 2011, 6(3): 1066-1075.

[4] 扈文斌, 刘凯. 基于量化表不一致性的 JPEG 图像篡改盲检测 [J]. 中国图象图形学报, 2011, 16(3): 316-323.

[5] CHEN Y L, HSU C T. Detecting recompression of JPEG images via periodicity analysis of compression artifacts for tampering detection [J]. IEEE Trans Inf Forensics Security, 2011, 6(2): 396-406.

[6] HUANG F J, HUANG J W, SHI Y Q. Detecting double JPEG compression with the same quantization matrix [J]. IEEE Trans Inf Forensics Security, 2010, 5(4): 848-856.

[7] FARID H. Exposing digital forgeries from JPEG ghosts [J]. IEEE Trans Inf Forensics Security, 2009, 4(1): 154-160.

[8] LUO W Q, HUANG J W, QIU G P. JPEG error analysis and its application to digital image forensics [J]. IEEE Trans Inf Forensics Security, 2010, 5(3): 480-491.

[9] FU D, SHI Y Q, SU Q. A generalized Benford's law for JPEG coefficients and its applications in image forensics [C]// SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents IX. 2007: 1L1-1L11.

[10] YE S M, SUN Q B, CHANG E C. Detecting digital image forgeries by measuring inconsistencies of blocking artifact [C]// 2007 IEEE International Conference on Multimedia and Expo. 2007: 12-15.

[11] 孙文颖, 张新鹏, 王朔中. 基于直方图差异的模型隐写分析 [J]. 上海大学学报: 自然科学版, 2008, 14(2): 111-115.

[12] 唐跃林, 肖沙里, 姚富光, 等. 一种反映人眼视觉特性的 JPEG 压缩编码 [J]. 微型机与应用, 2004(2): 49-50.

[13] WANG C Y, LEE S M, CHANG L W. Designing JPEG quantization tables based on human visual system [J]. Signal Process: Image Comm, 2001, 16: 501-506.