第 30 卷 第 6 期 自 动 化 学 报 Vol.30, No.6
2004 年 11 月 ACTA AUTOMATICA SINICA Nov., 2004

基于感兴趣区域的图像认证与自恢复算法[1)]

任 娟 王蕴红 谭铁牛
(中国科学院自动化研究所模式识别国家重点实验室 北京 100080)
(E-mail: {jren, wangyh, tnt}@nlpr.ia.ac.cn)

摘 要 提出了一种基于感兴趣区域的图像认证与自恢复算法，该算法将图像分为两部分：感兴趣区域和背景区域. 算法首先利用了基于 4×4 分块的编码算法对感兴趣区域进行编码，然后通过映射函数将此码流嵌入到背景图像的离散余弦变换系数中. 实验表明该算法可以对整幅图像进行完整性认证，同时对被恶意篡改的感兴趣区域也能很好地恢复原来信息.

关键词 图像自恢复，感兴趣区域，半脆弱性水印，完全脆弱性水印
中图分类号 TP39

A Self-recovery Algorithm Based on Region of Interest

REN Juan WANG Yun-Hong TAN Tie-Niu
(*National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing* 100080)
(E-mail: {jren, wangyh, tnt}@nlpr.ia.ac.cn)

Abstract A self-recovery algorithm based on region of interest is proposed in this paper. This algorithm divides a host image into two parts: region of interest and region of background. The region of interest is coded based on 4×4 blocks, and the codes are then embedded into the DCT coefficients of the region of background by a novel mapping function. Experimental results indicate that this scheme can authenticate the whole image, at the same time, if the region of interest is tampered maliciously, the scheme can recover the original content with good quality.

Key words Self-recovery, region of interest, semi-fragile watermarking, complete fragile watermarking

1 引言

网络时代给人们的生活带来很多便利，人们可以方便地从网上得到图像，但是该图像是否被篡改过却是用户很难确定的. 目前有很多图像工具，可以很容易地篡改图像，甚至达到以假乱真的地步. 因此，认证图像内容完整性显得非常重要. 脆弱性数字水印技术是在不影响原数据的视听效果的情况下，将一些数字信息隐藏到图像、声音和视频等多媒体

1) 国家自然科学基金 (60172054, 69825105) 和国家 "863" 高技术研究发展计划基金 (2003AA144080) 资助
Supported by National Natural Science Foundation of P R. China (60172054, 69825105) and National "863" Hi-tech Projects (2003AA144080)
收稿日期 2003-06-18 收修改稿日期 2004-06-05
Received July 18 2003; in revised form July 5, 2004

数据中，并通过提取隐藏信息来确认数据是否经过恶意篡改，从而达到对数据完整性保护的目的．根据认证的要求，水印算法可以分为两类：完全脆弱性水印和半脆弱性水印．完全脆弱性水印算法[1,2] 能够检测出对图像的任何篡改包括有损压缩；半脆弱性水印算法具有一定的鲁棒性，既可以抵抗一定程度的压缩、滤波和加噪声等攻击，同时又能检测出对图像内容的恶意篡改，并对篡改位置进行定位．我们在知道图像篡改的位置后，希望能够还原图像的真面目，特别是感兴趣区域 (region of interest, ROI). 考虑到现在一般的图像都会经过压缩保存，所以重点讨论半脆弱性水印下的图像认证和感兴趣区域的自恢复问题．

半脆弱性水印的概念是 Schneider[3] 在 1996 年提出的，到目前为止已经出现了很多针对有损压缩的认证算法．Xie[4] 将图像的边缘信息嵌入到其小波系数的低频子带中，该算法在一定程度上能抵抗压缩和低通滤波，缺点是不能精确定位篡改．Bhattacharjee[5] 从图像中提取出特征点，并将其存在另外的一个单独文件中，但是对图像认证时需要这个文件作参考，在实际应用中受限．Marve[6] 提出将原图进行 DCT 变换并对 DCT 系数量化，利用每个图像块量化后的 DCT 系数和的奇偶性作为水印嵌入，但存在的问题是系数和的奇偶性经过压缩后是改变的，因此他的算法并不能很好地对图像进行认证．Wu[7] 提出 DCT 域的查表法，可以抵抗 JPEG 压缩，并定位到 16×16 块大小，但该算法不能抵抗矢量量化 (vector quantization, VQ) 攻击[8]．Kundur[9] 提出基于小波的查表法，可以抵抗 JPEG2000 压缩，但对篡改的定位能力太差．

以上算法均不能很好地解决图像认证这一难题．在对被篡改区域自恢复方面，Lin[10] 从理论上证明了图像的 DCT 系数经过量化后有两个不变的规律．他利用这些规律设计认证算法，同时可以对整幅图像进行自恢复．所采用的编码方法是利用可变长度 (variable length coding, VLC) 的 Huffman 编码对图像进行基于 16×16 分块编码，每块得到不超过 88 比特的码流．VLC 编码虽然可以得到较好的编码率，但是存在一个严重的问题，即一位比特出错，整个码流将不可解；且这种方法恢复后的图像质量不高，最好的情况下相当于质量因子为 50 时 JPEG 压缩的图像质量．对于自恢复算法，Fridrich[11] 将图像首先进行基于 8×8 块编码，然后将水印信息嵌入到图像的最低一位或两位中．该算法是完全脆弱性水印算法，不能抵抗压缩，因此在实际应用中受限．同时和 Lin[10] 的算法一样，该算法恢复后的图像质量也比较差．

事实上，在实际应用中，用户往往只对感兴趣区域的篡改要求定位和恢复．感兴趣区域是指图像中最能引起用户兴趣、最能表现图像内容的区域[12]．当信息传输中带宽受限时，牺牲背景图像保持感兴趣区域图像的质量是非常有必要的．同样在图像认证中，如果篡改的位置无关紧要，用户一般不会太在意，但如果是感兴趣区域则希望恢复原来的内容．JPEG2000 专门有针对感兴趣区域的编码算法[13]．但与 VLC 编码一样，在传输过程中如果一位编码出错，将导致整个码流不可解．对于感兴趣区域而言，算法对图像恢复的质量和对篡改定位的能力要求都比较高．为此，本文使用了一种基于 4×4 分块的图像编码算法，恢复后的图像质量明显要好于文献 [10] 和 [11]．同时利用映射函数，可以检测出对图像的恶意篡改，并可以精确定位篡改位置到 16×16 小块．

2 算法基础

本文所提的基于感兴趣区域 (ROI) 图像认证算法基于 Lin[10] 提出的 JPEG 压缩的不变定理．这里简要介绍一下这两个定理．具体证明请参考文献 [10].

定理 1. 假设 $\boldsymbol{s}_p$ 是图像 X 的任一 8×8 块的 DCT 系数向量，$\boldsymbol{q}_m$ 是质量因子为 m 的 JPEG 压缩的量化表. 如果 $\boldsymbol{s}_p$ 被修改成 $\boldsymbol{s}'_p$，使 $\frac{\boldsymbol{s}'_p(v)}{\boldsymbol{q}'_m(v)} \in Z$，对 $\boldsymbol{q}'_m(v) > \boldsymbol{q}_m(v)$. 定义 $\tilde{\boldsymbol{s}}_p(v) \equiv \left[\frac{\boldsymbol{s}'_p(v)}{\boldsymbol{q}(v)}\right] \times \boldsymbol{q}(v)$，对任意的 $\boldsymbol{q}(v) \leqslant \boldsymbol{q}_m(v)$，则等式

$$\left[\frac{\tilde{\boldsymbol{s}}_p(v)}{\boldsymbol{q}'_m(v)}\right] \times \boldsymbol{q}'_m(v) = \boldsymbol{s}'_p(v) \tag{1}$$

成立，对任意的 $v \in \{1, \ldots 64\}$，$p \in \{1, \ldots N\}$，N 为图像的总块数，其中 $[\cdot]$ 表示四舍五入取整法.

JPEG 压缩中，质量因子越高，量化系数越小，压缩后的图像质量越好，损失的信息越少. 定理 1 说明如果图像 X 经过质量因子为 m 的 JPEG 压缩得到图像 Y，图像 Y 经过大于 m 的质量因子压缩得到图像 Y1，则 Y1 经过质量因子为 m 的压缩后得到图像 Y. 因为在保持图像质量的前提下，质量因子为 50 的量化系数最大，所以 m 一般取 50. 实际应用中，图像一般都是经过质量因子大于 50 的压缩，利用定理 1 嵌入水印足可满足要求.

定理 2. 假设 $\boldsymbol{s}_p$，$\boldsymbol{s}_q$ 是图像 X 的任意两个 8×8 块的 DCT 系数向量，$\boldsymbol{q}$ 是 JPEG 压缩的量化表. 定义 $\triangle \boldsymbol{s}_{p\,q} \equiv \boldsymbol{s}_p - \boldsymbol{s}_q$，$\triangle \tilde{\boldsymbol{s}}_{p\,q} \equiv \tilde{\boldsymbol{s}}_p - \tilde{\boldsymbol{s}}_q$，$\tilde{\boldsymbol{s}}_p(v) \equiv \left[\frac{\boldsymbol{s}_p(v)}{\boldsymbol{q}(v)}\right] \times \boldsymbol{q}(v)$. 假设一个固定的阈值 $k \in R$，定义 $\tilde{k}_v \equiv \left[\frac{k}{\boldsymbol{q}(v)}\right]$，则有以下规律：

$$\begin{aligned}
&\text{if} \quad \triangle \boldsymbol{s}_{p\,q}(v) > k, \quad \text{then} \quad \triangle \tilde{\boldsymbol{s}}_{p\,q}(v) \geqslant \begin{cases} \tilde{k}_v \times \boldsymbol{q}(v), & \frac{k}{\boldsymbol{q}(v)} \in Z \\ (\tilde{k}_v - 1) \times \boldsymbol{q}(v), & elsewher \end{cases} \\
&\text{elseif} \quad \triangle \boldsymbol{s}_{p\,q}(v) > k, \quad \text{then} \quad \triangle \tilde{\boldsymbol{s}}_{p\,q}(v) \leqslant \begin{cases} \tilde{k}_v \times \boldsymbol{q}(v), & \frac{k}{\boldsymbol{q}(v)} \in Z \\ (\tilde{k}_v + 1) \times \boldsymbol{q}(v), & elsewher \end{cases} \\
&\text{else} \quad \triangle \boldsymbol{s}_{p.q}(v) = k, \quad \triangle \tilde{\boldsymbol{s}}_{p\,q}(v) \geqslant \begin{cases} \tilde{k}_v \times \boldsymbol{q}(v), & \frac{k}{\boldsymbol{q}(v)} \in Z \\ (\tilde{k}_v, \tilde{k}_v \pm 1) \times \boldsymbol{q}(v), & elsewher \end{cases}
\end{aligned} \tag{2}$$

定理 2 中如果 $k = 0$，说明两个 DCT 向量同一位置的系数存在的大小关系，在压缩前后保持不变. 利用这个定理可以生成图像自身的内容水印.

3　基于感兴趣区域 (ROI) 的图像认证算法

当感兴趣区域图像被篡改时，我们需要从背景图像中将它恢复出来. 在对图像进行自恢复的时候，需要从每一小块中提取水印信息，如果这个小块是经过篡改的，显然提取的水印信息是错误的，因此必须对图像进行完整性认证. 以下分别讨论图像认证水印的生成，映射函数的选择和水印嵌入及认证算法.

3.1　认证水印的生成

将图像的每个象素值先减 128，然后分成不重叠的 16×16 大小的块，再将每块分成 8×8 的 4 个小块 A_1，A_2，A_3，A_4. A_1，A_2 和 A_3，A_4 根据定理 2 生成这块的认证水印. 设 $\boldsymbol{s}_i(v)$ 为 A_1 的 DCT 系数经过 zigzag 变换后得到的一维向量的绝对值. 定义 $\triangle \boldsymbol{s}_{i,j}(v) \equiv \boldsymbol{s}_i(v) - \boldsymbol{s}_j(v)$. 对 A_1 和 A_2，如果 $\triangle \boldsymbol{s}_{1\,2}(v) > 0$，取水印值为 1；否则取 0. 对 A_3 和 A_4 同样. 实验中取前 3 个系数的大小关系，一共得到 6 比特的水印信息，按照一定的映射函数将水印嵌入到另一块 DCT 系数中.

3.2 映射函数的选择

映射函数的选择很重要，如果映射函数为单位矩阵，即将自身块生成的水印嵌入本块中，水印图像容易受到 VQ[8] 攻击，因此映射函数要保证恶意篡改的人很难得出其规律. 本文按照如下方法设计映射函数.

设一个映射函数 T, $A = T(B)$ 表示将 B 块中的水印嵌入 A 块中. 这个映射函数应该满足以下条件:

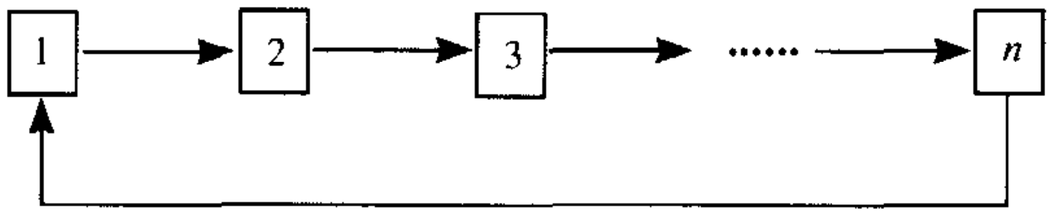


图 1 循环链

Fig. 1 A cycle chain

1) 图像块的对应嵌入块在图像上要相距较远，能够保证这些图像块不被同时修改;

2) 这些块的对应关系应该是单一对应，并且形成一个循环链表，如图 1;

3) 保证安全性，攻击者破解映射函数在实际中是不可能的.

设一幅图像中有 $N(m \times n)$ 个 16×16 块，按如下公式选择 T: $A_{k+f \bmod M} = T(A_k)$, $1 \leqslant k \leqslant N$, $1 < M \leqslant N$, M 为周期，f 为偏移量. 下面讨论如何选择这些参数.

如果 M 小于 N, 易出现多对应的关系，如 $N = 10$, $M = 6$, $f = 1$, 有 $1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 1$, $7 \to 2$! 块 1 和 7 同时对应 2; f 大于 1 时同样.

如果 M 等于 N, 将会有 3 种情况: 1) $f = 1$, 则满足条件 2), 且 $T^N(A_k) = A_k$, $1 \leqslant k \leqslant N$; 2) f 大于 1, 且 (N mod f)≠0, 条件 2) 也可以满足，如 $N = 7$, $f = 2$, 有 $1 \to 3 \to 5 \to 7 \to 2 \to 4 \to 6 \to 1$; 3) f 大于 1, 且 (N mod f)=0, 条件 2) 不能满足，如 $N = 6$ 且 $f = 2$, 则 $1 \to 3 \to 5 \to 1$, $2 \to 4 \to 6 \to 2$, 出现了两个循环.

从上面的讨论可知, M 应为 N, f 应为 1 或者不是 N 的因子. 但是都难以满足条件 1). 为了不受图像大小的限制，这里选择偏移量 f 为 1, 周期 M 为 N, 使周期最长，攻击的难度加大. 但是这样做了之后，相邻块直接存在对应关系，因此必须首先打乱 N 块的次序，使原来相邻块之间的距离尽量远. 我们采用如下的置乱方法打乱 N 块的次序: 先打乱行的次序 $(i, j) \to ((j + i - 2 + offset) \bmod m, j)$, $i = 1, 2, \ldots m$, $j = 1, 2, \ldots n$; $offset$ 为偏移量; 同理可以打乱列的次序. 打乱的次数 K 由用户决定，也可随机生成. 显然, N 越大，攻击难度越大.

举例说明，假设 $N = 4 \times 4 = 16$, $offset = 1$, 置乱的过程如图 2 所示. 比较 (a) 和 (c), 可以发现相邻两块经过打乱后距离拉开. 当 m 与 n 值较大时，这个距离会更大，这样做可以避免相邻块的水印互相嵌入的情况.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

(a)

1	14	11	8
5	2	15	12
9	6	3	16
13	10	7	4

(b)

1	14	11	8
12	5	2	15
3	16	9	6
10	7	4	13

(c)

(a) 原始块的次序 (the order of original blocks), (b) 打乱列次序后的结果 (the result of confusing row), (c) 在 (b) 的基础上打乱行次序的结果 (the result of confusing column based on (b))

图 2 一个利用映射函数打乱矩阵次序的例子

Fig. 2 An example of confusing orders by mapping function

3.3 水印嵌入及认证算法

按照如上讨论的映射函数，根据定理 1, 将 B 块中生成的 6 比特水印信息嵌入到 A 块的 4 个 DCT 向量中，每 2 个小块中嵌入 3 比特. 设 $\boldsymbol{s}_i(v)$ 为 A_i 的 DCT 系数经过 zigzag 变换后得到的一维向量的绝对值 $\boldsymbol{s}_i(v) = [\boldsymbol{s}_i(v)/\boldsymbol{q}(v)]$, $(i = 1 \sim 4)$, v 表示嵌入水印的 3 个不同于 3.1 节中的生成水印的位置, $\boldsymbol{q}$ 表示质量因子为 50 时 JPEG 压缩的量化矩阵. 我们将水印嵌入到量化后 DCT 系数的最低位. 在 A_1, A_2 块嵌入水印的过程如文献 [10]. 先求 $\boldsymbol{f}_1(v)$ 和 $\boldsymbol{f}_2(v)$ 最低位的异或，如果与水印值相同，则不改变；否则，如果 $\boldsymbol{f}_1(v)$ 和 $f_2(v)$ 中只有一个为 0, 则按照下面的公式

$$(\boldsymbol{f}_1'(v), \boldsymbol{f}_2'(v)) = \begin{cases} (\boldsymbol{f}_1(v), \boldsymbol{f}_2(v) + \mathrm{sgn}(\boldsymbol{s}_2(v)/\boldsymbol{q}(v) - \boldsymbol{f}_2(v))), & \boldsymbol{f}_1(v) = 0, \boldsymbol{f}_2(v) \neq 0 \\ (\boldsymbol{f}_1(v) + \mathrm{sgn}(\boldsymbol{f}_1(v)/\boldsymbol{q}(v) - \boldsymbol{f}_1(v)), \boldsymbol{f}_2(v)), & \boldsymbol{f}_1(v) \neq 0, \boldsymbol{f}_2(v) = 0 \end{cases} \tag{3}$$

调整 DCT 系数的值；如果两者全为 0 或者全不为 0, 则

$$(\boldsymbol{f}_1'(v), \boldsymbol{f}_2'(v)) = \begin{cases} (\boldsymbol{f}_1(v), \boldsymbol{f}_2(v) + \mathrm{sgn}(\boldsymbol{s}_2(v)/\boldsymbol{q}(v) - \boldsymbol{f}_2(v))), & |\boldsymbol{s}_1(v)/\boldsymbol{q}(v) - \boldsymbol{f}_1(v)| \leqslant |\boldsymbol{s}_2(v)/\boldsymbol{q}(v) - \boldsymbol{f}_2(v)| \\ (\boldsymbol{f}_1(v) + \mathrm{sgn}(\boldsymbol{f}_1(v)/\boldsymbol{q}(v) - \boldsymbol{f}_1(v)), \boldsymbol{f}_2(v)), & |\boldsymbol{s}_1(v)/\boldsymbol{q}(v) - \boldsymbol{f}_1(v)| > |\boldsymbol{s}_2(v) - \boldsymbol{f}_2(v)| \end{cases} \tag{4}$$

其中 $\mathrm{sgn}(x) = \begin{cases} 1, & x \geqslant 0 \\ -1, & x < 0 \end{cases}$, 这样调整之后，提取水印的水印值为 $\boldsymbol{f}_1(v)$ 和 $\boldsymbol{f}_2(v)$ 最低位的异或.

同样在 A_3 和 A_4 小块中嵌入水印，从而得到水印图像.

图像的认证过程是嵌入过程的逆过程. 同样，首先按照 3.1 节中的方法生成待测图像各块水印信息. 假设图像中的 A, B, C 3 块存在如下关系：$A = T(B)$, $C = T(A)$. 显然， A 块的内容是否被篡改与 B, C 两块都有关系. 如果 3 块提取的水印信息可以相互对应，则说明 A 块的内容是完整的；如果 3 块提取的水印信息不能相互对应，则说明 A 块的内容是经过恶意篡改的.

4 感兴趣区域 (ROI) 的自恢复算法

对待测图像进行完整性认证之后，如果图像经过恶意篡改，在自恢复时本文算法需要判断篡改的图像是否为感兴趣区域的图像，如果是则需要恢复. 因此，嵌入算法应将 ROI 图像内容嵌入到背景图像中. 嵌入的同时应该考虑 ROI 图像在原图像中的位置及大小. 这一点和文献 [10,11] 对整幅图像进行恢复的算法不同，他们的算法不必考虑位置和大小.

4.1 感兴趣区域 (ROI) 水印的生成

由于本文所提算法主要是针对感兴趣的区域的自恢复, 因此首先要得到这部分区域. 比如对 lena 图像 (如图 3) 而言，我们最关心的是她的脸部内容. 对于感兴趣区域可以直接用白色的矩形框标出来，经过预处理之后进行编码. 为确定图 3 中白色矩形框的位置，只需找到第一点和第四点的坐标，不妨设为 (a, b) 和 (c, d). 计算 $a_0 = \lfloor a/8 \rfloor$, $b_0 = \lfloor b/8 \rfloor$, $c_0 = \lfloor c/8 \rfloor$, $d_0 =$

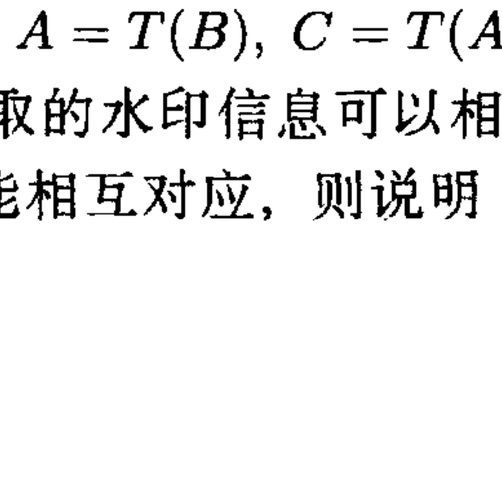


图 3 感兴趣区域

Fig. 3 Lena with a white rectangle

$\lfloor d/8 \rfloor$, 其中 $\lfloor x \rfloor = \max\{a \leqslant x, a \in Z\}$. 令 $a = a_0 \times 8 + 1$, $b = b_0 \times 8 + 1$, $c = c_0 \times 8$, $d = d_0 \times 8$, 将原来标出的区域稍稍调整使之为多个 8×8 的小块, 以方便将感兴趣区域的位置和大小编码. 这里只需将 a_0, b_0, c_0, d_0 编码, 所需编码的比特位为 $\max(\log_2(height/8), \log_2(width/8))$, 其中 *height* 和 *width* 表示原始图像的长和宽. 如对于 256×256 的图像, 它的每个坐标需要 5 比特表示.

确定感兴趣区域的位置 $((a,b),(c,d))$ 之后, 将感兴趣区域内的图像分成 4×4 分块, 按照如下算法进行压缩量化和编码, 每块可得到 50 比特的 ROI 内容水印信息.

目前, 对图像进行恢复一般采用的是 8×8 分块或 16×16 分块的图像压缩编码算法, 但这两种编码方法恢复的图像质量比较差. 本文采用对图像进行 4×4 分块的压缩编码方法, 整个流程与 JPEG 压缩相似. 我们对 200 多幅灰度图像进行实验, 分析了 DCT 系数, 采用一般 JPEG 压缩的量化表中左上角的 16 个系数为基本量化表 Q_0

$$Q_0 = \begin{bmatrix} 16 & 11 & 10 & 16 \\ 12 & 12 & 14 & 19 \\ 14 & 13 & 16 & 24 \\ 14 & 17 & 22 & 29 \end{bmatrix}, \; L = \begin{bmatrix} 7 & 6 & 5 & 4 \\ 6 & 5 & 4 & 0 \\ 5 & 4 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix}, \; Q_r = Q_0 \times r \tag{5}$$

其中 Q_0 为基本量化表, L 表示对压缩量化后的 DCT 系数进行编码所需的编码位数, r 表示量化系数的倍数. 针对此量化表, 本文采用的量化系数是基本量化表中系数的 1.5 倍, 如公式 (5). 矩阵 L 表示对压缩量化后的 DCT 系数进行编码所需的编码位数. 通过这样的编码, 每小块得到 50 比特码流, 这样编码得到的图像质量相当于质量因子为 80 时 JPEG 压缩的图像质量.

4.2 ROI 水印的嵌入算法

由 4.1 节得到的 ROI 水印信息分为两个部分: 感兴趣区域的位置水印和图像内容水印. 以 256×256 的 lena 为例, 位置水印为 20 比特. 为了保证在图像恢复时位置水印完整无误地被提取出来, 本文将它重复嵌入到背景图像的 10 个随机选择的 8×8 块的 DCT 系数中. 具体算法如公式

$$f_i'(v) = f_i(v) + \mathrm{sgn}(\boldsymbol{s}_1,(v)/\boldsymbol{q}(v) - f_i(v)) \tag{6}$$

式中 v 表示不同于 3.1 节的生成水印的位置和 3.3 节中已经嵌入水印的位置.

图像内容水印先按照 3.2 节所述的置乱方法打乱水印的顺序. 根据感兴趣区域的位置和大小, 选择不同的偏移量和打乱的次数, 这样可以增加攻击的难度. 然后按照公式 (6) 将其嵌入到背景图像的 8×8 块的 DCT 系数中, 每块嵌入 25 比特的水印, 则一个 4×4 块的水印信息要嵌入到两个 8×8 的块中. 重复的次数根据感兴趣区域的大小而定.

如果感兴趣区域太大, 要先缩小再进行编码, 然后嵌入到其它区域. 设 h 和 w 为感兴趣区域的长和宽, x 和 y 为长和宽的缩小倍数. 在不考虑位置水印的情况下, 有下列等式:

$$\frac{h \times w}{x \times y \times 4 \times 4} \times 2 = \frac{heigth \times width - h \times w}{8 \times 8} \tag{7}$$

这样可以求得的感兴趣区域不得超过整幅图像的 $xy/(xy+8)$. 因此缩小因子 x 和 y 也应该作为水印, 可将它们作为 4 比特水印与位置水印放在一起嵌入到 10 个随机选择的

8×8 的块中. 一般情况下，为满足大多数情况下的实际需要，取 $x = y = 2$, 使感兴趣区域可达到原图像的 1/3.

综上所述，ROI 水印的嵌入算法可以分为以下几个步骤：

1) 标出感兴趣区域，然后对其进行预处理；2) 生成感兴趣区域的位置与内容水印信息；3) 根据位置和大小的不同，选择不同的偏移量打乱水印的顺序；4) 将感兴趣区域的内容水印、位置水印和缩小因子信息重复嵌入到背景图像中.

4.3 ROI 自恢复算法

由上面的嵌入算法，我们可以恢复被篡改的感兴趣区域. 由于重点关心感兴趣的区域，首先判断嵌有位置水印的 10 个 8×8 块是否经过篡改. 如果存在大于 1 块没有经过篡改，则按照多数原则提取位置水印. 知道感兴趣区域的位置后，判断感兴趣区域是否完整，如果不完整则需要自恢复. 在对图像进行恢复的时候，如果背景图像中一个 8×8 块经过篡改，则不必提取它的水印信息，否则提取感兴趣区域的内容水印. 最后，根据 4×4 块的 DCT 逆变换，利用 3.2 所述的映射函数，可以得到感兴趣区域的图像. 简单归纳如下：

1) 利用映射函数对水印图像进行完整性认证; 2) 提取 ROI 的位置水印及其缩小因子；3) 判断感兴趣区域是否被篡改；4) 如果 ROI 内容被篡改，从背景图像中提取 ROI 内容水印，恢复 ROI 图像.

整个算法的流程见图 4 和图 5.

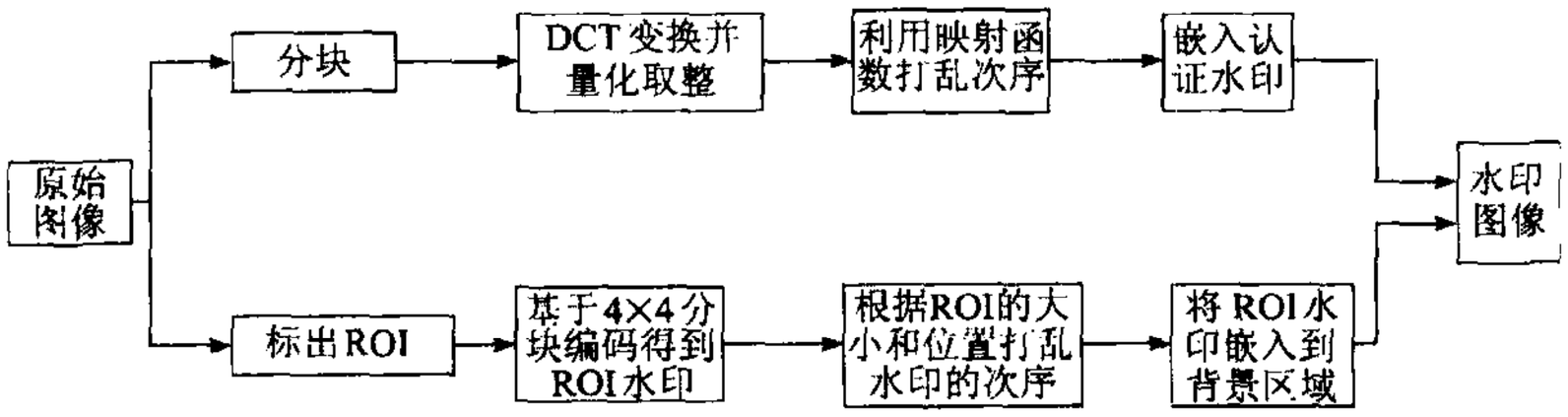


图 4　嵌入水印

Fig. 4 Watermark insertion

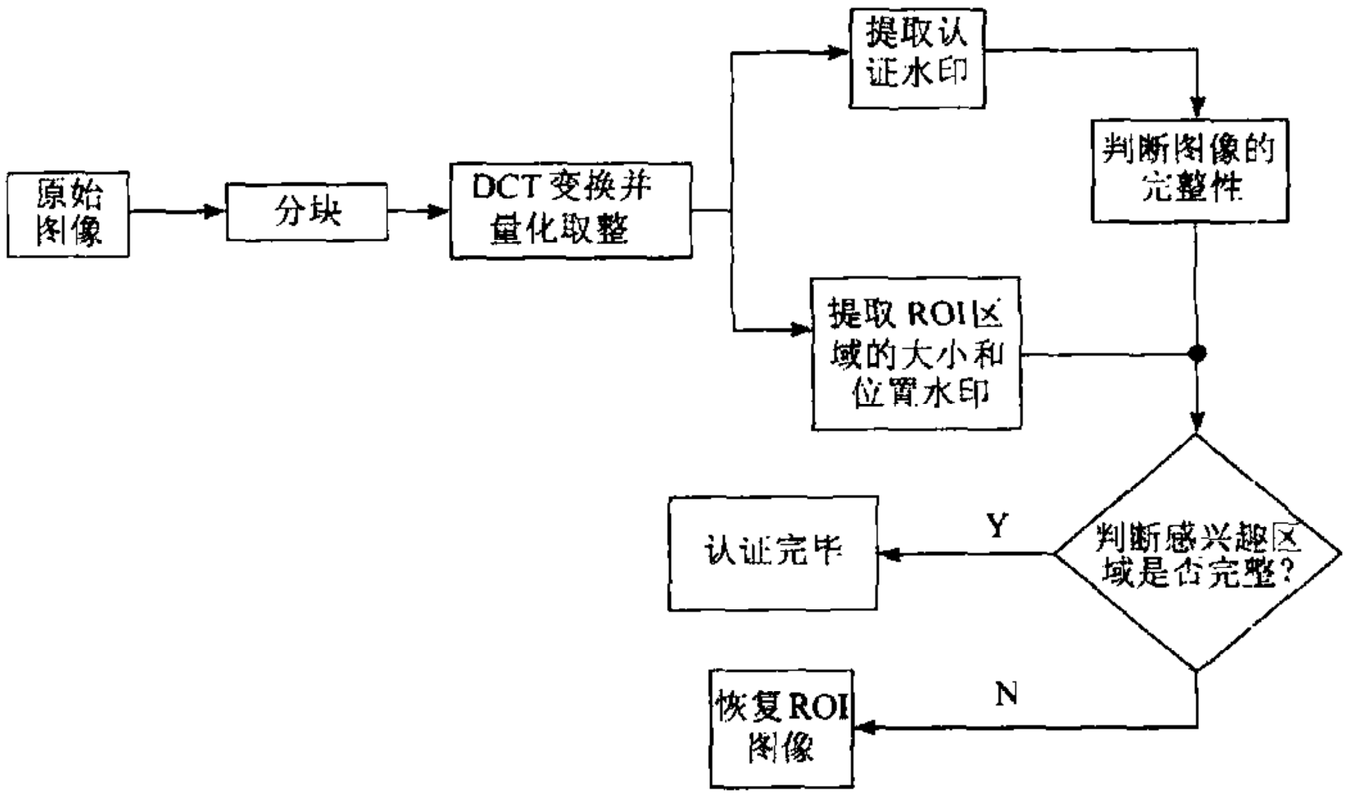


图 5　图像认证及 ROI 自恢复

Fig. 5 Image authentication and ROI recovery

5　实验结果及分析

首先给出基于 4×4 块压缩后的图像质量与经过 JPEG 压缩的图像质量的比较. 图 6 给

出我们实验中所用 200 多幅图像中的 10 幅 (256×256) 比较典型的测试图像示例.

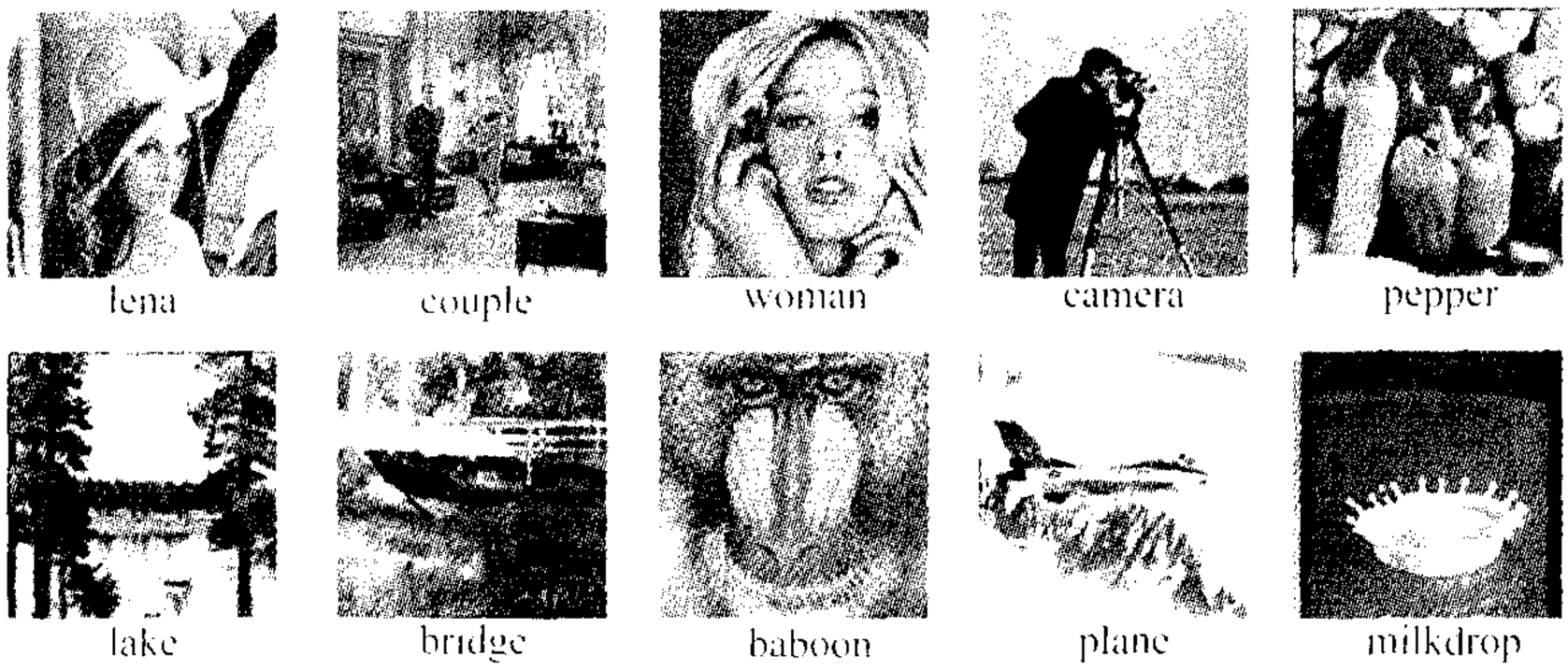


图 6 测试图像

Fig. 6 Test images

图 7 给出了基于 4×4 分块的编码和 JPEG 压缩后图像质量的比较结果. 0.6, 1, 1.5 表示基于 4×4 分块压缩中采用的 r 值; 50, 70 和 80 表示 JPEG 压缩中的质量因子; image 表示图 6 中的图片, 依次为 lena, couple 等. 峰值信噪比 PSNR 用如下公式计算:

$$PMSE = \frac{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(f(m,n)-g(m,n))^2}{255^2} \tag{8a}$$

$$PSNR = -10 \times \log_1 0(PMSE) \tag{8b}$$

从图 7 中可以看出, 本文提出的基于 4×4 分块编码后的图像质量 ($r = 1.5$) 介于质量因子为 70 和 80 之间的 JPEG 压缩的图像质量, 明显好于利用文献 [10] 和 [11] 的方法编码后得到的图像 (相当于质量因子为 50 时 JPEG 压缩图像) 质量. 实验结果 (图 8) 更直观地说明了这一问题.

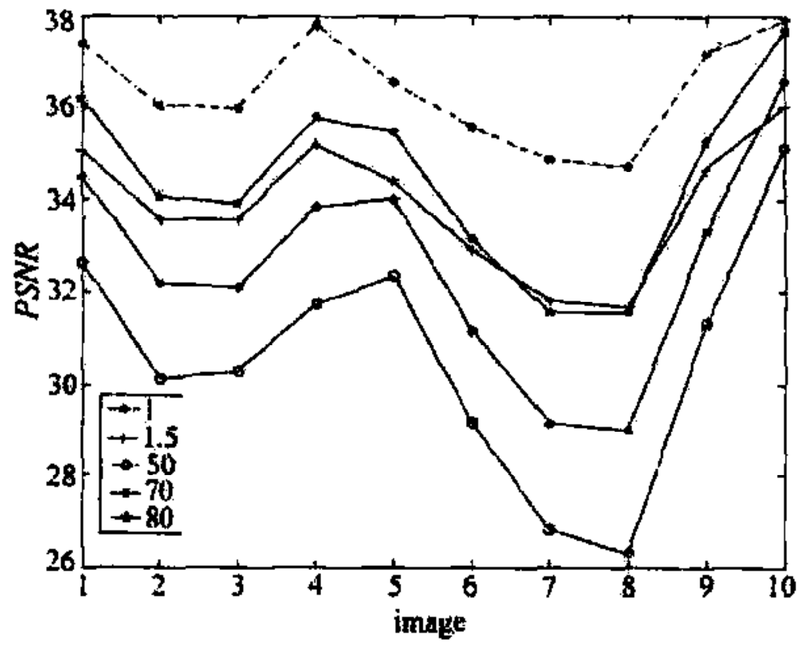
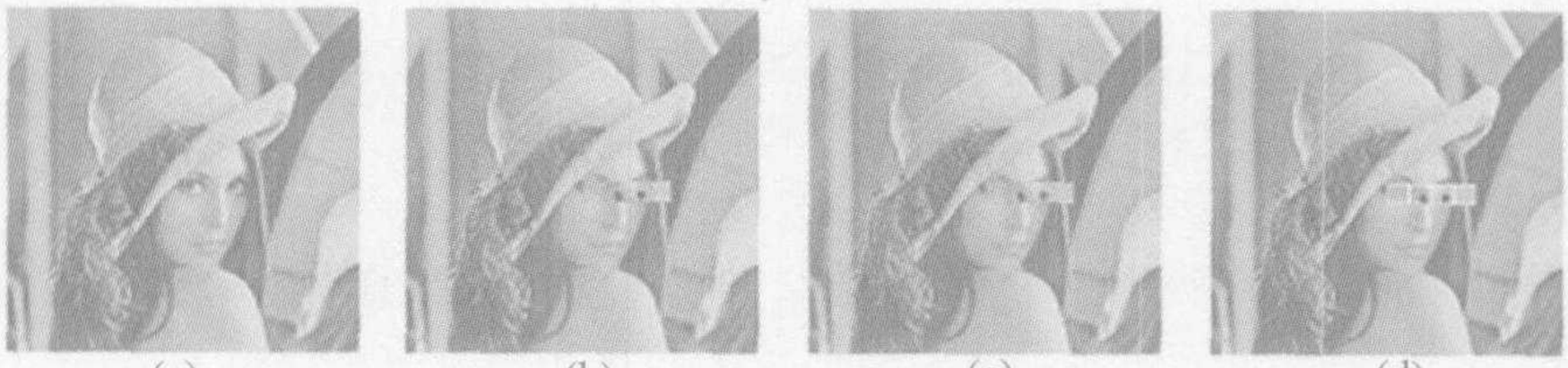


图 7 基于 4 × 4 分块的编码和 JPEG 压缩后图像质量的比较

Fig. 7 Comparisons of image quality

图 8 验证了本文所提出的算法对图像认证的效果. 如果对图像进行 VQ 攻击[8], 即将 lena 的水印图像中眼睛部分用 woman 的水印图像同一位置的图像内容代替. 如果水印图像是利用单位矩阵的映射函数 (即同一块生成的水印直接嵌入到相同块中) 得到的, 认证失败 (见图 8(c)), 所以不能抵抗 VQ 攻击; 如果水印图像是利用所提的映射函数得到的, 图 8(d) 表明该算法能够检测出篡改的位置, 因此可以看出本文算法可以抵抗 VQ 攻击.

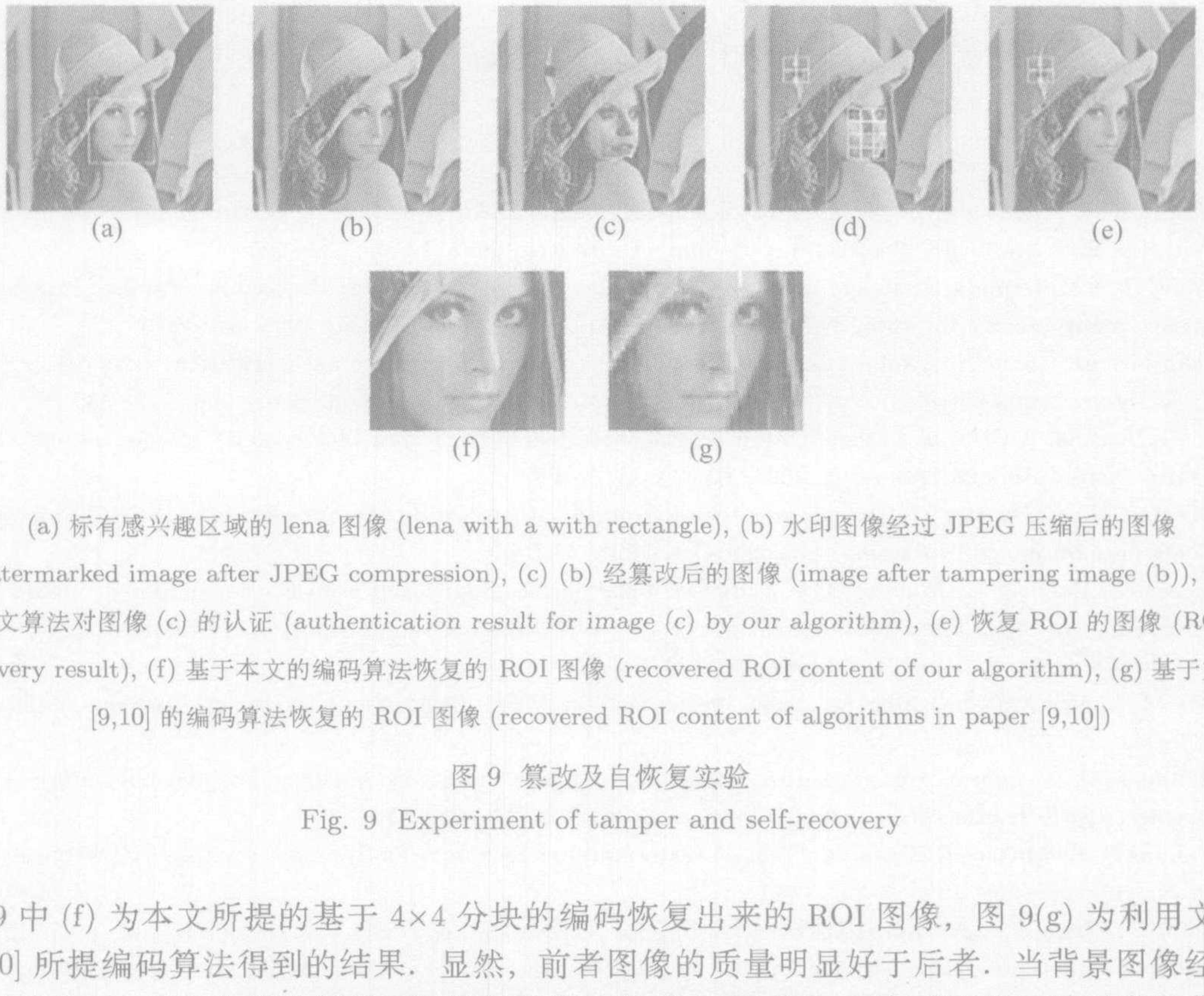


(a) Lena 的水印图像 PNSR=34.4951 (watermarked image), (b) VQ 攻击 (VQ attack), (c) 映射函数为单位矩阵的认证结果 (authentication result if the mapping function is identity matrix), (d) 利用本文所提映射函数的认证结果 (authentication result if the mapping function is out function)

图 8　VQ 攻击实验

Fig. 8　Experiment of VQ attack

图 9 给出本文所提出的算法对感兴趣区域自恢复的实验结果. 实验表明水印图像经过质量因子大于等于 50 的 JPEG 压缩后, 可以很好地认证篡改位置并恢复图像 ROI 部分.



(a) 标有感兴趣区域的 lena 图像 (lena with a with rectangle), (b) 水印图像经过 JPEG 压缩后的图像 (watermarked image after JPEG compression), (c) (b) 经篡改后的图像 (image after tampering image (b)), (d) 本文算法对图像 (c) 的认证 (*authentication result for image (c) by our algorithm*), (e) 恢复 ROI 的图像 (ROI . recovery result), (f) 基于本文的编码算法恢复的 ROI 图像 (recovered ROI content of our algorithm), (g) 基于文献 [9,10] 的编码算法恢复的 ROI 图像 (recovered ROI content of algorithms in paper [9,10])

图 9　篡改及自恢复实验

Fig. 9　Experiment of tamper and self-recovery

图 9 中 (f) 为本文所提的基于 4×4 分块的编码恢复出来的 ROI 图像, 图 9(g) 为利用文献 [9,10] 所提编码算法得到的结果. 显然, 前者图像的质量明显好于后者. 当背景图像经过少量篡改时, 我们的算法仍然可以恢复被篡改的感兴趣区域. 当与感兴趣区域对应的背景图像也被篡改时, 该区域将不能被很好地恢复. 但是实际中, 因为我们利用了 3.2 节所述的映射函数打乱图像块的次序, 且随着感兴趣的位置和大小的不同, 映射函数的参数也不同, 刻意地去找寻其规律并将背景区域篡改是非常困难的, 并且感兴趣区域的图像水印是

被重复嵌入的. 所以在实际应用中, 感兴趣区域与其对应的背景图像同时都被篡改的几率非常小, 一般情况下都可以将篡改的感兴趣区域很好地恢复出来. 综上所述, 本文所提的基于感兴趣区域认证与自恢复的算法有以下优点:

1) 可以抵抗质量因子大于等于 50 的 JPEG 压缩;

2) 利用 3.2 节所述的映射函数可以抵抗 VQ 攻击, 并定位图像篡改到 16×16 块;

3) 可以对感兴趣图像自恢复, 恢复后图像的质量相当于质量因子为 80 时 JPEG 压缩的图像质量.

6 结论

本文提出了一个基于感兴趣区域的认证与自恢复的算法. 该算法利用映射函数实现了对图像完整性的验证, 并且能够抵抗 VQ 攻击. 对于原来设定的感兴趣区域的篡改, 该算法可以利用背景图像中的隐藏信息自动恢复出感兴趣区域. 由于采用了基于 4×4 块编码方法, 恢复的图像质量明显好于其它文献的图像自恢复算法. 但是当感兴趣区域很大, 且背景图像篡改也很严重的情况下, 算法可以对篡改位置进行定位, 但对 ROI 图像的自恢复可能不完全. 在实际应用中, 背景图像和感兴趣区域同时被大幅度的篡改的情况较少发生. 因此, 算法可以基本满足实际应用需要. 在下一步工作中, 我们将结合 MPEG4 设计水印算法, 以保证在视频压缩过程中保证感兴趣区域的质量和内容的完整性.

References

1 Utku Celik M, Sharma G, Saber E, Murat Tekalp. Hierarchical Watermarking for Secure Image Authentication With Localization. *IEEE Transactions on Image Processing*, 2002, **11**(6): 585∼595

2 Wong P. A Watermark for Image Integrity and Ownership Verification. In: Proceeding of Image Processing, Image Quality, Image Capture, Systems Conference. Portland, Oregon: May 1998. 374∼379

3 Schneider M, Chang S. A robust content based digital signature for image authentication. In: proceeding of IEEE International Conference on Image Processing. Lausanne, Switzerland: Sep. 1996. 227∼230

4 Xie L, Arce G. A Class of Authentication Digital Watermarks for Secure Multimedia Communication. *IEEE Transactions on Image Processing*, 2001, **10**(11): 1754∼1764

5 Bhattacharjee S, Kutter M. Compression tolerant image authentication. In: proceeding of IEEE International Conference on Image Processing, Chicago: Oct. 1998. 4∼7

6 Marvel L, Hartwig G, Boncelet C. Compression-compatible fragile and semi-fragile tamper detection. In: Proceedings of *SPIE-The International Society for Optical Engineering, Security and Watermarking of Multimedia Contents*. San Jose, California: 2000. 140∼151

7 Wu M, Liu B. Watermarking for image authentication *IEEE Transactions on Image Processing*, 1998, **2**: 437∼441

8 Holliman M, Memon N. Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. *IEEE Transactions on Image Processing*, 2000, **9**(3): 432∼441

9 Kundur D, Hatzinakos D. Towards a Telltale Watermarking Technique for Tamper-proofing *IEEE Transactions on Image Processing*, 1998, **2**: 409∼413

10 Lin C Y, Chang S F. Semi-Fragile Watermarking for Authenticating JPEG Visual Content. In: proceeding of SPIE-The International Society for Optical Engineering, International Conf. on Security and Watermarking of Multimedia Contents II. San Jose, USA: Jan 2000, 3971: 140∼151

11 Fridrich J. Security of Fragile Authentication Watermarks with Localization. In: proceeding of ACM Workshop on Multimedia and Security. Orlando, FL: 1999. 19∼23

12 Moghaddam B, Biermann H, Margaritis D. Defining Image Content With Multiple Regions-of-Interest. In: proceeding of IEEE Workshop on Content-Based Access of Image and Video Libraries. June 1999. 89∼93

13 Christopoulos C, Askelof J, Larsson M. Efficient region of interest encoding techniques in the upcoming JPEG2000 still image coding standard. In: proceeding of IEEE Conference on Image Processing. Vancouver, Canada: 2000. 41～44

任　娟　中国科学院自动化研究所硕士，2000 年本科毕业于西安交通大学．研究方向为数字水印及图像处理．

(**REN Juan**　Master in Institute of Automation, Chinese Academy of Sciences, and received her bachelor degree from Xi'an Jiaotong University in 2000. Her research interests include digital watermarking, and image processing.)

王蕴红　中国科学院自动化研究所副研究员，1989 年本科毕业于西北工业大学，1998 年在南京理工大学获得博士学位，1998 年开始在自动化研究所工作．研究方向为生物特征识别，数字水印等．

(**WANG Yun-Hong**　Associate Professor, received her bachelor degree from Northwestern Polytechnical University in 1989, and her Ph.D. degree from Nanjing University of Science and Technology in 1998. During 1998.4-2000.5, she was a postdoctoral fellow in NLPR, Institute of Automation, Chinese Academy of Sciences. Her current research interests include biometrics, statistical pattern recognition and digital image processing.)

谭铁牛　中国科学院自动化研究所研究员，1984 年本科毕业于西安交通大学，1986 年和 1989 年在英国伦敦的帝国理工大学分别获得硕士学位和博士学位．1998 年回国．研究方向为生物特征识别、数字水印、图像视频处理和视频监控等．

(**TAN Tie-Niu**　Professor, receved his bachelor degree from Xi'an Jiaotong University in 1984, and received his master degree (in 1986) and Ph.D. degrees (in 1989) from Imperial College of Science, Technology and Medicine, London, UK. His research interests include visual surveillance and monitoring of dynamic scenes, personal identification based on multiple biometric features, and watermarking of digital multimedia data.)