

Semi-Fragile Image Watermarking Method Based on SVD and Block DCT*

Shaoyu Chen[#], Xinwei Wang

Department of Computer Science & Technology, East China Normal University, Shanghai
Email: [#]chenshaoyu86@126.com, wangxw@cs.ecnu.edu.cn

Received: Nov. 24th, 2012; revised: Dec. 16th, 2012; accepted: Dec. 23rd, 2012

Abstract: With the certification of the integrity of the multimedia content, the solutions are mainly concentrated on fragile digital watermarking technology. This paper present a semi-fragile watermarking algorithm based on SVD and sub-block DCT. The watermarking signals are embedded into the second biggest singular values of the image. Watson model is used to control the robustness of embedding. Experiments demonstrate that the algorithm meets the basic goals of semi-fragile watermarking: imperceptible, blind detection, distinguishing legitimate distortions from illegitimate ones, and the algorithm has some practical use.

Keywords: Semi-Fragile Digital Watermark; DCT; SVD; Watson Model

基于 SVD 和分块 DCT 的半脆弱图像水印算法*

陈韶钰[#], 王新伟

华东师范大学计算机科学与技术系, 上海
Email: [#]chenshaoyu86@126.com, wangxw@cs.ecnu.edu.cn

收稿日期: 2012 年 11 月 24 日; 修回日期: 2012 年 12 月 16 日; 录用日期: 2012 年 12 月 23 日

摘要: 随着多媒体内容的完整性认证, 尤其是图像的完整性验证, 解决方案主要集中在脆弱数字水印技术。本文提出基于 SVD 和分块 DCT 的半脆弱水印算法。算法将水印信号嵌入图像第二大奇异值中, 并利用 Watson 模型控制嵌入强度。实验表明算法达到了不可见性、盲检测性、区分合理与不合理失真等半脆弱水印的基本要求, 并能较准确识别篡改区域, 具有一定实用性。

关键词: 半脆弱数字水印; 离散余弦变换; 奇异值分解; Watson 模型

1. 引言

随着计算机和网络技术的快速发展, 各种数字媒体得到广泛的传播和使用。但是, 数字化在给人类带来便利的同时, 也出现了诸如侵权、拷贝等一系列问题。虽然目前的数字水印算法多种多样, 但是根据实现过程不同数字水印可分为空域算法和变换域算法^[1]。因为变换域算法具有良好的不可感知性和鲁棒性, 并且对数据压缩具有兼容性, 所以针对变换域算法已

经取得许多成果。此外, Liu 等较早提出了利用奇异值分解(SVD)来嵌入水印的策略, 随后一些针对该策略的优化或改进算法被提出。

由于数字水印的半脆弱方法较之脆弱方法具有灵敏度低的特点, 已经引起人们的关注。在实际应用中我们常常面临论证图像内容的真实性, 以及容忍对图像的非恶意修改等问题。半脆弱方法是用来辨别图像是否受到故意攻击的有效技术^[2,3]。本文在已有工作的基础上, 探讨基于 DCT^[4]和 SVD^[5]二者结合的数字图像水印算法, 并进行新的设计和分析。

*基金项目: 国家自然科学基金(No. 61273298)。

[#]通讯作者。

2. 数字图像的奇异值分解及其性质

设 $A \in R^{M \times N}$ 表示一个图像矩阵, 其中 R 表示实数域。对 A 进行奇异值分解: $A = U\Sigma V^T$, 其中 $U \in R^{M \times M}$, $V \in R^{N \times N}$ 都是正交阵,

$$\Sigma \in R^{M \times N}, \Sigma = (\sigma_{i,j})\sigma_{M \times N}, \sigma_{i,j} = 0, i \neq j,$$

记 $\sigma_i = \sigma_{i,i}$, 称作 A 的奇异值^[5]。

$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_M = 0$, r 是矩阵 A 的秩。

给定一个矩阵, 它的奇异值是确定的。但不同矩阵可以有相同的奇异值, 这是矩阵与奇异值之间的多对一性质。从图像处理的角度来看, 奇异值分解^[5,6]具有以下特性: 1) 不同图像最前面的若干奇异值集中了大部分矩阵的能量, 后面的奇异值很小。2) 较大的奇异值及其与 $U(V)$ 矩阵中对应的行(列)对图像的影响较大。通常多是选择最大的奇异值来嵌入水印信息。本文尝试选择第二大的奇异值来嵌入水印, 来提高水印的脆弱性能, 并且可以增加宿主图像中嵌入信息的容量。3) 通过对数字图像的奇异值分解得到矩阵 Σ 、 U 、 V 。矩阵 $U(V)$ 的行(列)具有良好的抗干扰能力, 也就是说, 当图像受到轻微的扰动时, 它的结构不会发生明显的改变。

二维离散余弦变换不但能够将图像的主要信息集中到最少的低频系数上, 而且引起的块效应最小, 能够实现信息集中能力和计算复杂性的良好折中。在数字图像处理中使用的是, 对一幅图像 $f(u, v)$, 它的二维离散余弦变

换^[1,7](DCT)为:

$$F(u, v) = \alpha_u \alpha_v \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) \cos\left[\frac{(2i+1)u\pi}{2N}\right] \times \cos\left[\frac{(2j+1)v\pi}{2N}\right]$$

相应的逆变换为:

$$f(i, j) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha_u \alpha_v F(u, v) \cos\left[\frac{(2i+1)u\pi}{2N}\right] \times \cos\left[\frac{(2j+1)v\pi}{2N}\right]$$

$$\text{其中, } \alpha_w = \begin{cases} \sqrt{\frac{1}{N}}, & w = 0 \\ \sqrt{\frac{2}{N}}, & w = 1, 2, \dots, N-1 \end{cases}.$$

1993 年, Watson 提出基于 DCT 的视觉模型, 并提出 JND 的概念。这一视觉模型包括敏感度、掩蔽效应和综合。在 Watson^[8]感知模型中, 对原始图像以 8×8 方式进行分块, 计算 JND 门限, 如

$$s[i, j, k] = \max\left\{t_L[i, j, k], |C_0[i, j, k]|^{\omega[i, j]} t_L[i, j, k]^{1-\omega[i, j]}\right\}$$

$$t_L[i, j, k] = t(i, j) \left(C_0[0, 0, k] / C_{0,0}\right)^{\alpha T}$$

其中, $C_0[0, 0, k]$ 是原图中第 k 块的 DC(直流分量)系数, αT 为常数且常取为 0.649, $C_{0,0}$ 整幅图的直流分量的平均值, $t(i, j)$ 为 Watson 视觉模型定义的敏感度表, $w[i, j] \in (0, 1)$, 通常取值 0.7。

3. 水印嵌入算法

在奇异分解的框架下, 大多数的水印嵌入算法是将水印信息嵌入最大的奇异值所在的位置。这样的嵌入策略限制了嵌入容量。本文算法设计的嵌入策略从理论上讲, 允许水印信息嵌入所有的非零奇异值所在的位置。将水印信息嵌入不同的奇异值, 需要注意的是嵌入信息以后的奇异值必须保持原来奇异值的大小顺序。本文提出的算法是可以保证不改变奇异值的原来排序。采用本文的算法实现水印信息的嵌入是不会改变宿主图像奇异值的顺序, 这样就可以保持图像的品质。由于, 水印信息嵌入的位置不仅仅局限在最大奇异值得位置, 因此可以大大提高水印的嵌入容量。

本文采用的水印信号是二值有意义图像, 这样在评价图像完整性上, 变得更直观、更简单。为了提高水印安全性, 我们首先对原始水印信号进行 Arnold 置乱^[9], 这样不仅能提高嵌入信号的随机性, 而且也降低水印信号对载体图像不可感知性影响。文献[9]给出了 Arnold 置乱变换的基本算法和周期。假设 (x, y) 是 $N \times N$ 图像矩阵的某个象素点, 通过一次 Arnold 置乱后该象素点应该被移动到 (x', y') 处, 两者满足下式:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \bmod(N), \quad x, y \in \{0, 1, \dots, N-1\}$$

该变换具有周期性, 并且周期由 N 决定。假设 N 对应的周期为 C , WC 为嵌入水印前预处理置乱次数, KC 为提取水印后恢复的置乱次数, 那么必须满足 $WC + KC = C$; 同时 WC 的选取能达到原始水印图像置乱后随机化程度尽可能地高的目的。

假设载体图像为 $M \times N$ 大小的灰度图像, 嵌入的水印信号是一幅 $L \times P$ 大小的二值图像, 基于文献[10]的算法, 我们提出新的水印信号嵌入策略, 具体嵌入步骤如下:

1) 对原始水印信号进行 Arnold 置乱, 可以得到没有意义的二值伪随机序列。根据原始水印图像的大小选择 Arnold 置乱变换时合适的取模参数 N , 并选取置乱效果最好的情况作为预处理时的置乱次数 WC 。

2) 将原始图像以 M/L 和 N/P 的形式分块, 其中水印图像的尺寸参数 L 和 P 作为提取水印的密钥。

3) 然后对其进行 DCT 变换, 变换后的系数矩阵进行 SVD 分解。设某个分块对应矩阵为 B , 则 $B = U_B \times S_B \times V_B^T$, 由此得到对角矩阵 S , 将水印嵌入 S 中第二大的元素 $S(2,2)$ 。

4) 生成一个大小为 $(M/L) \times (N/P)$ 的矩阵 dd , 用来存放分块后每一子块的抖动参数, 用来调整不同子块被嵌入水印信号的强度。抖动参数的计算模型^[11]如下:

$$dd(i, j) = \min(s(2,2) - s(3,3), s(1,1) - s(2,2)) + d \quad (1)$$

$$\text{if } dd(i, j) > s(2,2) \times r, \text{ then } dd(i, j) = dd(i, j) \times rr \quad (2)$$

其中步骤 1 是控制嵌入信号后的奇异值大小的顺序不会改变; d 是偏移值, 适当地选择 d 可以避免水印信号的强度过小而导致水印无法正确提取的现象; 而步骤 2 对 $dd(i,j)$ 的调整可以避免嵌入信号的强度过大而导致宿主图像破坏的问题。通过实验分析, 选择出的合适的抖动系数。

5) 按顺序从已置乱的水印信号序列中取出一个元素, 然后嵌入到每一个分块, 直到每个水印信号元素被嵌入其中一个分块中(块数和水印信号元素个数是一致的)。假设某个水印信号元素为 W , 嵌入块的奇异值是 $s(2,2)$, 嵌入水印以后为 $s'(2,2)$ 。则嵌入算法描述如下:

第一步, 计算 $X = \lfloor s(2,2)/dd(i, j) \rfloor$, $\lfloor \rfloor$ 表示向下取整, $dd(i, j)$ 为该块对应的抖动系数^[12];

第二步分两种情况

情况 1: 如果 X 是奇数

$$s'(2,2) = \begin{cases} X \times dd(i, j) + dd(i, j)/2 & \text{if } W = 1 \\ \left(\left[X - \text{sign}(y) \right] + \frac{1}{2} \right) dd(i, j) & \text{if } W = 0 \end{cases} \quad (3)$$

情况 2: 如果 X 是偶数

$$s'(2,2) = \begin{cases} X \times dd(i, j) + dd(i, j)/2 & \text{if } W = 0 \\ \left(\left[X - \text{sign}(y) \right] + \frac{1}{2} \right) dd(i, j) & \text{if } W = 1 \end{cases} \quad (4)$$

$$\text{其中, } \text{sign}(L) = \begin{cases} 1 & \text{if } L \leq 0 \\ -1 & \text{if } L > 0 \end{cases}$$

$$y = V - \left(X + \frac{1}{2} \right) \times dd(i, j)$$

7) 对每个修改后的分块进行 SVD 逆变换, 再进行 DCT 变换, 最后把所有分块修改后的新矩阵合并, 得到嵌入水印后的图像矩阵。

8) 根据 Watson 视觉模型评价原始图像和嵌入水印的图像^[13], 若其视觉距离 D 在阈值范围内, 则水印嵌入过程结束; 否则适当调整抖动系数 $dd(i, j)$, 重新执行第 4~6 步。视觉距离 D 的阈值选择、抖动系数 $dd(i, j)$ 的选择根据具体应用确定。

4. 水印提取和鉴别算法

本文提出的水印算法还符合半脆弱水印^[12]所要求的盲检测性。提取水印时, 只要知道抖动参数矩阵 dd 、预处理置乱次数 WC 、取模参数 N 和水印图像大小 L 及 P (这些参数可以看作提取水印的密钥)。算法通过判定奇偶性^[10]来提取水印, 提取速度很快, 满足半脆弱水印提取的速度要求。具体算法步骤如下:

1) 假设提取水印的图像大小为 $M \times N$, 而提取的水印图像大小为 $L \times P$ (已知的密钥)的, 将被检测图像进行 M/L 和 N/P 的分块。

2) 对每一分块进行 SVD 变换;

3) 计算 $X = \lfloor s(2,2)/dd(i, j) \rfloor$; 如果 X 是奇数, 提取出一个 1, 再提取出一个 0。

按顺序对每个分块按照(2)和(3)步骤进行提取, 把提取出的序列按水印图像的大小 L 、 P (已知的密钥)依次排成一个二维矩阵 EW 。

4) 根据 N 和 WC , 我们可以知道提取出的 EW 图像还要置乱 KC 次, 即得到提取出的水印图像。

为了实现半脆弱水印识别恶意攻击和评价数字图像失真程度^[10]的目的, 我们对嵌入载体图像的已置乱水印图像和从待测图像中直接提取的还没有再次置乱恢复的水印图像进行评判。设前者为 $A1$, 后者为 $A2$, 定义如下 2 个指标来识别篡改区域和评价失

真程度:

$$\text{篡改区域: } DA = A1 \oplus A2$$

(\oplus 表示 A1, A2 按位异或);

比特误差率: ER = DA 矩阵中为 1 的元素个数/DA 矩阵元素总个数。

5. 实验和结果分析

本文使用 320×320 的标准 Lena 图像作为宿主图像, 嵌入的水印是一幅 40×40 的二值图像。在嵌入算法中, 表 1 反映抖动参数的 r 和 rr 的变化将会影响提取水印的比特误差率(ER), 适当选择可以提高算法的误差性能。由实验数据可以得到: rr 在 0.3~0.6 时, ER 是单调递减的; 而在 0.6~0.8 时, ER 是单调递增的。可以看到选 $rr = 0.6$ 和 $r = 0.6$ 时水印的比特误差率最低。因此, 本文选用的最佳控制抖动参数的值为 $rr = 0.6$ 和 $r = 0.6$ 。

Table 1. Bit error rate (ER) and the dither parameter r, rr relationship
表 1. 比特误差率(ER)与抖动参数 r, rr 的关系

$r \backslash rr$	0.3	0.4	0.5	0.55
0.3	0.0813	0.0344	0.0169	0.0088
0.4	0.0806	0.0344	0.0169	0.0088
0.5	0.08	0.0344	0.0169	0.0088
0.55	0.0794	0.0344	0.0169	0.0088
0.6	0.0781	0.0338	0.0169	0.0088
0.65	0.0781	0.0331	0.0169	0.0088
0.7	0.0794	0.0338	0.0181	0.01
0.75	0.0781	0.0338	0.0181	0.0088
0.8	0.0775	0.0338	0.0187	0.0081
$r \backslash rr$	0.6	0.65	0.7	0.8
0.3	0.0050	0.0094	0.01	0.02
0.4	0.0050	0.0094	0.01	0.02
0.5	0.0050	0.0094	0.01	0.02
0.55	0.0050	0.0094	0.01	0.02
0.6	0.0050	0.0094	0.01	0.02
0.65	0.0050	0.0094	0.01	0.02
0.7	0.0063	0.0106	0.0113	0.0213
0.75	0.0063	0.0106	0.0113	0.0213
0.8	0.0069	0.0113	0.0119	0.0219

对被嵌入水印的图像进行不同品质的 JPEG 压缩后提取水印并计算 ER 可以看到无论压缩品质如何改变, ER 均为 0.005。表明该算法对 JPEG 压缩这一合理失真有很好的鲁棒性, 如表 2 所示。

从试验结果分析可以知道, 在高压缩率的情况下, 本文的所提出的算法具有很低的比特误差率, 优于文献[10]的方法。

图 1 中的(a)是 Lena 原图和(b)嵌入水印后的图像。实验时采用 $n = 40$ 的 Arnold 置乱, 此时 $C = 30$, 取 $WC = 17, KC = 13$ 。(c,d)是原始水印图像和经过置乱的水印图像。

图 2 对图像进行对比度调整和椒盐噪声(均值为 0, 方差为 0.0001)等测试。图 2 显示, 对于合理失真的情况, 算法的敏感性很低(ER 分别等于 0 和 0.0063)。文献[10]对图像进行对比度调整和椒盐噪声(均值为 0, 方差为 0.0001)测试, ER 分别为 0 和 0.0019, 也就是

Table 2. The comparison of the algorithm (S1) and the algorithm of the document [10] (S2) for different quality JPEG compression ratio (JC%) produced by the bit error rate (ER%)
表 2. 本文算法(记为 S1)和文献[10]的算法(记为 S2)针对不同品质的 JPEG 压缩率(JC%)所产生的比特误差率(ER%)比较

JC%	90		80		70	
	S1	S2	S1	S2	S1	S2
ER%	0.5	0	0.5	0	0.5	0
JC%	25		15		10	
	S1	S2	S1	S2	S1	S2
ER%	0.5	32.06	0.5	39.12	0.5	61.02

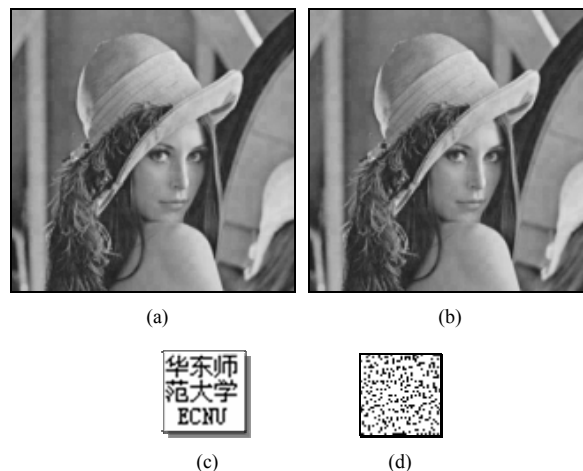


Figure 1. (a) Original image; (b) Watermarked image; (c) The watermark; (d) Scrambling watermark
图 1. (a) 原始图像; (b) 含水印图像; (c) 水印原图; (d) 置乱水印图

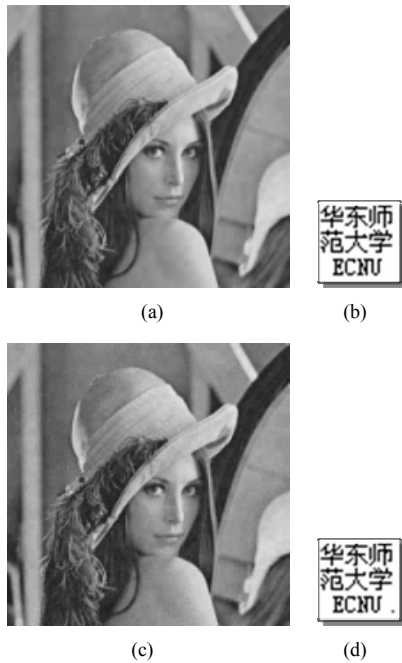


Figure 2. (a) The image after contrast adjustment; (b) The extracted watermark; (c) The image processed by salt and pepper noise; (d) The extracted watermark
图 2. (a) 对比度调整后图像; (b) 提取的水印; (c) 椒盐噪声处理的图像; (d) 提取的水印

说, 对无恶意的图像处理操作, 本文的算法鲁棒性比较强。

图 3 显示, 在通常情况下, 不合理失真体现出很强的脆弱性, 算法还能准确判定篡改区域。算法对剪裁(例如, 剪切右下 10%区域的图像 $ER = 0.623$)和恶意篡改等都会引起 ER 的明显上升, 但是, 在文献[10]中, 剪切右下 10%区域的图像, $ER = 0.04$ 这一分析表明本算法符合半脆弱水印的要求。为了很好的显示在不合理失真情况下, 图像是否遭到破坏, 我们使用 ER 作为评价的标准。通过以上分析, 算法能够比较有效地区分合理和不合理失真, 满足半脆弱水印的性能要求, 对 JPEG 压缩等表现出强鲁棒性, 但是对微小的恶意攻击能准确判定篡改位置。

6. 结论

本文基于 DCT 变换、SVD 变换和 Watson 视觉模型探讨了半脆弱水印算法的设计。本文通过引入抖动参数来控制不同图像子块被嵌入水印的强度, 改进已有算法的策略, 提高半脆弱水印性能。通过仿真实验, 该算法能很好地区分合理失真和不合理失真, 还能较准确识别出恶意篡改位置, 而且提取和定位的速度很

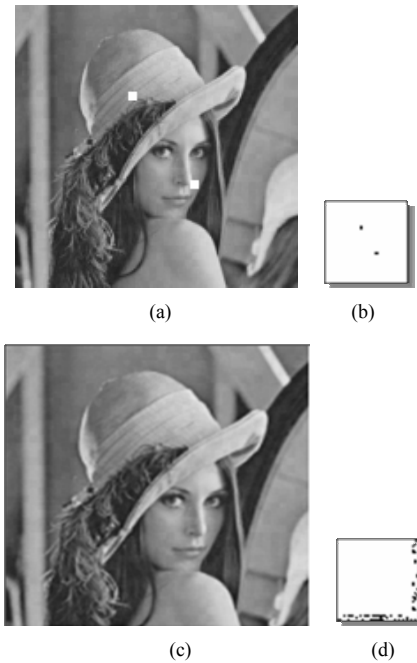


Figure 3. (a) Tampering with the two 8×8 image watermarks; (b) The identified watermark tampering domain; (c) Shearing 10% of the lower right image; (d) The extracted watermark ($ER = 0.623$)
图 3. (a) 篡改 2 个 8×8 块后图像; (b) 鉴别出的水印篡改域; (c) 剪切右下 10%区域的图像; (d) 提取的水印 ($ER = 0.623$)

快。进一步的工作主要包括: 1) 改进 Watson 模型使之能更好的处理算法分块所带来的块噪声干扰; 2) 算法本身的安全性也需要加强, 使它能够有效抵抗诸如几何攻击等恶意攻击。

参考文献 (References)

- [1] I. J. Cox, M. L. Miller 等, 著, 王颖等, 译. 数字水印[M]. 北京: 中国电子工业出版社, 2003.
- [2] C.-H. Lin, T.-S. Su and W.-S. Hsieh. Semi-fragile watermarking scheme for authentication of JPEG images. Tamkang Journal of Science and Engineering, 2007, 10(1): 57-66.
- [3] T. Gopalakrishnan, et al. Semi-fragile watermarking using Gaussian mixture model for malicious image attacks. 2011 World Congress on Information and Communication Technologies, Mumbai, 11-14 December 2011: 120-125.
- [4] I. J. Cox, J. Kilian, F. T. Leighton, et al. Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing, 1997, 6(12): 1673-1687.
- [5] 孙锐, 孙洪, 姚天任. 基于奇异值分解的半易损水印算法[J]. 电路与系统学报, 2002, 7(3): P62-P65.
- [6] V. I. Gorodetski, L. J. Popyack, V. Skormin and V. A. Samoilov. SVD-based approach to transparent embedding data into digital images. Information Assurance in Computer Networks, 2001, 2052: 263-274.
- [7] R. O. Preda. Semi-fragile watermarking for image authentication with sensitive tamper localization in the wavelet domain. Journal of the International Measurement Confederation, 2013, 46(1): 367-373.
- [8] A. B. Watson. DCT quantization matrices visually optimized for individual images. Human Vision, Visual Processing, and Digital

- Display IV, 2002, 19(13):202-216.
- [9] C. L. Li, B. Ma, Y. H. Wang, D. Huang and Z. X. Zhang. A secure semi-fragile self-recoverable watermarking algorithm using group-based wavelet quantization. *Advances in Multimedia Information Processing—PCM*, 2012, 7674: 327-336.
- [10] 李斌, 王新伟. 基于 SVD 和 Watson 模型的半脆弱图像水印算法[J]. *计算机应用与软件*, 2008, 25(6): 254-256.
- [11] A. A. Abdulfetah, X. M. Sun, H. F. Yang and N. Mohammad. Robust adaptive image watermarking using visual models in DWT and DCT domain. *Information Technology Journal*, 2010, 9(3): 460-466.
- [12] 钮心忻, 编著. 信息隐藏于数字水印[M]. 北京: 北京邮电大学出版社, 2004.
- [13] 凌贺飞, 卢正鼎等. 基于 Watson 视觉感知模型的能量调制水印算法[J]. *软件学报*, 2006, 17(5): 1124-1132.