

一种新的自适应盲数字水印算法^{*}

彭求明 杨小帆 黄松 李思静 柏森

(重庆大学计算机学院 重庆 400044) (重庆通信学院研究生管理大队 重庆 400035)

摘要 本文提出了一种新颖的基于 DCT 域的自适应盲数字水印算法,该算法由三个阶段组成。首先,根据人眼视觉系统(HVS)和图像的局部特性来选择水印嵌入区域;然后充分利用 JPEG 压缩量化阶段的舍入误差;第三个阶段是采用一组公式来嵌入多个版本的水印,水印是经位扩展和二值随机序列调制得到的。水印的检测不需要原始图像。实验结果表明,该算法在保证水印不可见性的同时,对常见的几种图像处理如缩放、椒盐噪声、滤波特别是 JPEG 压缩和剪切有很高的鲁棒性。

关键词 数字水印,离散余弦变换(DCT),人眼视觉系统(HVS),JPEG 压缩,盲检测

A Novel Adaptive Blind Digital Watermarking Algorithm

PENG Qiu-Ming YANG Xiao-Fan HUANG Song LI Si-Jing BAI Sen

(College of Computer, Chongqing University, Chongqing 400044)

(Graduate Education Division, Chongqing Communication Institute, Chongqing 400035)

Abstract In this paper, an adaptive blind digital watermarking scheme is proposed. The algorithm consists of three phases. First, the embedding region of watermark is selected via the human visual system model and local characteristics of image; then the round-off errors is analysed during the quantization step of JPEG compression; at the third phase, several simple formulas are assumed to embed several watermarks that are generated by bit-spreading and binary random sequence's modulation. The watermark is detected without any reference to the original image. Experimental results show that the embedded watermark is invisible; while at the same time the algorithm is robust to some typical kinds of image processing such as scaling, salt & pepper noise, filtering, especially, JPEG compression and cropping attacks.

Keywords Digital watermarking, DCT, HVS, JPEG compression, Blind detection

1 引言

近年来,随着互联网和多媒体信息处理技术的迅猛发展,网上多媒体信息(图像、视频、音频、文本等)的免费获取、复制、分发、传送、修改已变得极为快捷、方便。然而,网络给人们带来便利的同时也暴露了越来越多的安全问题,如何防止非法复制、有效地保护版权已成为一个极为迫切的问题。数字水印技术是一种有效的保护版权的手段,已经引起了人们的高度重视和广泛关注,数字水印技术的实现主要考虑两个因素:稳健性和不可见性^[1]。这两者是相互矛盾的,一个可靠的数字水印系统应该是稳健性和不可见性的最佳折衷。

数字水印技术一般可以分为两类:空域水印^[2]和变换域水印。空域水印直接修改图像的像素,该类算法的最大特点是算法简单、计算复杂度低,但鲁棒性差。变换域算法是对图像进行各种变换(常见的变换有 DCT^[3-6]、DWT^[7]、DFT)后嵌入水印。相对于空域水印算法,变换域水印算法有如下优点:(1)在变换域内嵌入的水印信号能量可以分布到空间域的

所有像素上,有利于提高鲁棒性;(2)可以较方便地结合 HVS,有利于保证水印的不可见性;(3)变换域算法与现今大多数国际图像和视频压缩标准兼容,可直接实现压缩域水印嵌入。正因如此,变换域水印是当今水印技术研究的主流。

许多文献^[8-13]提出了自适应水印算法,它们都是利用 HVS 和图像自身的特点自适应地嵌入水印,自适应主要体现在:嵌入位置的选取、嵌入水印强度的调节、嵌入水印位置个数的选择。文[8]根据空域 8×8 图像子块边缘点密度把所有的图像子块分为两类:弱纹理和强纹理,水印嵌在每一 DCT 系数块的 DC(direct component)分量上,利用一组公式来量化修改它,通过拉伸因子来调节水印嵌入的强度,强纹理块用大的拉伸因子,弱纹理块用小的拉伸因子,是盲检测。文[9]是根据亮度掩蔽特性和纹理掩蔽特性把所有的 8×8 DCT 子块分为两类:适合嵌入水印和不适合嵌入水印。由 DC 分量的大小来评价亮度特性;把每一个 DCT 系数块用 JPEG 压缩量化矩阵来量化,根据非零个数的多少来判断纹理的复杂程度。水印嵌在中频区较大的系数上,不同的块嵌入不同能量

^{*}基金项目:本文工作受到重庆应用基础研究项目(编号:8028)资助。彭求明 硕士研究生,研究方向:数字水印、图像处理。杨小帆 教授,博导,研究方向:并行计算、容错及故障诊断、人工神经网络、数理方程。

作,对大量数据进行手工标记在具体实现中是很不现实的。为了解决该问题,本文改进了现有的贝叶斯分类算法,提出了利用未标记数据提高贝叶斯分类器性能的方法。实验表明,这种方法取得了很好的效果。

参考文献

1 Lee W. A data mining framework for constructing features and

- models for intrusion detection systems:[dissertation of Doctor of Philosophy]. Columbia University, 1999
- 2 Lee W, Stolfo S J. Data mining approaches for intrusion detection. the Seventh USENIX Security Symposium (SECURITY '98), San Antonio, TX, Jan. 1998
- 3 Jacobson V, Leres C, McCanne S. tcpdump. Available via anonymous ftp to ftp. ee. lbl. gov, June 1989
- 4 薛静锋,曹元大. 基于贝叶斯分类的分组入侵检测技术研究. 见:第三届全国 CSCW 暨第一届全国 AIN 学术会议论文集, 2002

的水印。文[10]中指出拉伸因子控制水印图像的保真度和检测率,拉伸因子越大,检测率越高,但图像的视觉质量越差。文[11]是利用 HVS 在不影响视觉质量的同时嵌入最大的水印信号,在低频和高频部分分别嵌入两个相同的水印。文[12]利用 HVS 在空域中嵌入水印。

本文提出了一种新的 DCT 域水印算法,该算法根据人眼视觉系统(HVS)和图像的局部特性,按照标准 JPEG 量化的结果自适应地选择水印嵌入区域。嵌入水印时利用 JPEG 量化阶段的舍入误差和一组公式,对系数的修改很小,因此水印的检测是可以不需要原始图像的,是盲检测。并且本算法嵌入了多个版本的水印,水印是经位扩展和二值随机调制产生的。实验结果表明,该算法在保证水印不可见性的同时,对常见的图像处理操作例如 JPEG 压缩、剪切、中值滤波、高斯低通滤波、缩放、椒盐白噪声等都有很好鲁棒性。

2 基于人类视觉系统的自适应频带选择

从视觉掩蔽的角度来看,水印的嵌入可以看作在强背景(原始图像)上叠加一个弱信(水印)。只要信号低于 HVS 的对比度门限(contrast sensitivity threshold),HVS 就无法感觉到水印信号的存在。根据 HVS 的视觉特点,背景的亮度和纹理影响嵌入水印信号的可见性^[14]:背景越亮,所嵌入水印的可见性越低,即所谓的亮度掩蔽特性;背景的纹理越复杂,嵌入水印的可见性越低,即所谓的纹理掩蔽特性。所以,水印应该嵌入在图像的背景照度亮,纹理复杂的位置。DCT 交流系数经 JPEG 量化后非零个数越多系数变化就越大,该块的纹理就越复杂,HVS 对这些系数的改变就越不敏感,就可以嵌入水印。

通常情况下图像的低频带是感知最重要的分量^[15],但不同图像的总特性不可能完全相同,同一幅图像的局部特性也是随空间位置而变化的,对同一图像的不同子块,能量的分布并不一样,能量将集中于不同的系数上,这些幅值较大的系数是感知重要的系数^[5,14],感知重要的系数在低频带较多,但也有可能集中于中频带,这将由图像的具体特性而定。因此固定的水印嵌入位置对不同总体特性的图像以及同一图像的不同局部特性并不是自适应的,嵌入的水印也不可能有的不可见性和稳健性。

基于以上的分析,本文提出了一种新的自适应频带选择 DCT 域水印算法,水印嵌入区域自适应的选择为中频带和低频带中频率信息和能量信息最丰富的一段频带。自适应频带选择的具体算法如下:设 8×8 图像块的 DCT 系数为 $\{x_{u,v}, 1 \leq u \leq 8, 1 \leq v \leq 8\}$,按标准 JPEG 量化矩阵(如图 1 所示)对系数量化,量化后系数大部分为零,不为零的系数主要集中在低频和中频带,当系数为零时,对其进行修改,频率信息从无到有,很容易导致图像失真,因此,最好不要修改为零的系数,我们定义量化块中不为零的系数为感知重要的系数。不考虑 DC 分量,按图 2 扫描顺序可得到量化的系数序列 $F = \{f_1, f_2, \dots, f_{63}\}$,如果量化系数序列 $F'_i = \{f_i, f_{i+1}, \dots, f_{i+L-1}\}$ 满足 $NZ(F'_i) = \max_k \{NZ(F'_k)\}$,其中 $F'_k = \{f_k, f_{k+1}, \dots, f_{k+L-1}\}, 1 \leq k \leq 63-L+1$,表示从序列 F 中选取的从第 k 个系数开始的长度为 L 一段频带,函数 $NZ(\cdot)$ 表示序列 F'_k 中非零元素的个数,则选取下标 $\{i, i+1, \dots, i+L-1\}$ 所对应的 DCT 系数作为水印的嵌入区域。当有多个区域同时满足上述条件时,如果算法强调水印的不可见性,则应该选择处于中频段的区域,而如果算法更注重水印的稳健性,则应该优先考虑低频区域。本文中优先考虑低频段作为水印的嵌入区

域。

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

图 1 标准 JPEG 量化表

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

图 2 zig-zag 扫描顺序

3 本文的水印算法

由以上的分析,本文提出了一种新的水印嵌入和水印提取算法,具体如下:

3.1 水印嵌入算法

(1) 水印信号的生成:先读入二值水印图像,得二维矩阵 $W, W = \{W(i, j), 1 \leq i \leq m, 1 \leq j \leq n, W(i, j) \in \{0, 1\}\}, m \times n$ 为水印图像大小,然后按行扫描顺序将其转换为一维矩阵 $W_i, W_i = \{W_i(i), 1 \leq i \leq m \times n, W_i(i) \in \{0, 1\}\}$,如果 $W_i(i) = 0$,则 $W_i(i) = -1$,其它保持不变,把 W_i 转化为 $\{-1, 1\}$ 。为增强水印系统的稳健性和安全性,先对 W_i 进行位扩展,扩展因子为 cr ,得到 $W_1, W_1 = \{W_1(k) = W_i(i), (i-1) \times cr + 1 \leq k \leq i \times cr\}$,再用密钥 key 产生的二值随机序列 $P(k) \in \{-1, 1\}$ 来调制得最终的水印信号 $WM, WM = \{WM(k) = W_1(k) \times P(k), 1 \leq k \leq m \times n \times cr\}$ 。

(2) 水印嵌入区域的选取:先将原图像 $I(M \times N)$ 按从左到右,然后从上到下的顺序依次 8×8 分块,每块分别 DCT 变换得矩阵 A ,并除以 JPEG 量化矩阵 Q ,得到两个矩阵 B 和 C ,矩阵 B 是没有尾数取舍,矩阵 C 是进行了尾数取舍。矩阵 A, B, C 都按上面图 2 所示的 zig-zag 扫描顺序转化为一维序列 $A_1(j), B_1(j), C_1(j), 1 \leq j \leq 64$ 。同时,矩阵根据 2 的算法自适应确定每一块的水印嵌入区域 $\{i, i+1, \dots, i+L-1\}$ 。

(3) 水印嵌入:这里我们冗余的在原图像中依次嵌入 $num = L \times M \times N / (8 \times 8 \times m \times n \times cr)$ 个水印版本(本文中 $M \times N$ 为原图像的大小, $m \times n \times cr$ 为水印序列的长度),使得水印提取时有更多的数据可用,以进一步保证水印算法有较高的抗攻击能力。 num 个水印版本序列 $WM_i(t) = WM(k), t = (i-1) \times m \times n \times cr + k$,其中 $1 \leq k \leq m \times n \times cr, 1 \leq i \leq num$ 。经上面步骤(2)后,每一块选出的 L 个位置 $l \in \{i, i+1, \dots, i+L-1\}$,对每一个位置 l ,利用 JPEG 量化舍入误差,并参照 $D_i(l)$ 较高位比特的二值逻辑(本文中选中第 3 位比特)按下面公式来嵌入水印,取 $C_1(l)$ 绝对值记为 $D_i(l)$,公式如下:

if $WM_i(j) = 1$ and $bitget(D_i(l), 3) = 1$ $E_1(l) = 2 \times \text{round}((B_1(l)-1)/2) + 1$ //量化为奇数

if $WM_i(j) = 1$ and $bitget(D_i(l), 3) = 0$ $E_1(l) = 2 \times \text{round}(B_1(l)/2)$ //量化为偶数

if $WM_1(j) = -1$ and $\text{bitget}(D_1(L), 3) = 1$ $E_1(L) = 2 \times \text{round}(B_1(L)/2)$

if $WM_1(j) = -1$ and $\text{bitget}(D_1(L), 3) = 0$ $E_1(L) = 2 \times \text{round}((B_1(L)-1)/2) + 1$

再把 $E_1(L)$ 乘以标准 JPEG 量化矩阵中对应位置的值来反量化得 $F_1(L)$, 然后把 $F_1(L)$ 赋给对应位置的 $A_1(L)$, $A_1(L)$ 中其它位置的值保持不变。再按上面图 2 所示的 zig-zag 扫描顺序把 $A_1(L)$ 还原成 8×8 块。

(4) 经步骤(3)后, 把所得到的 8×8 块作 DCT 逆变换, 然后把该块放入载体图像中原来的位置, 对每一块都执行(2)、(3)、(4)步, 即得到嵌入水印后的图像 I。

3.2 水印提取算法

水印的提取算法与水印的嵌入算法相反, 具体步骤如下:

(1) 水印信号抽取: 把嵌有水印的图像按照从左到右, 然后从上到下的顺序分成多个块, 对每一块先进行 DCT 变换, 再除以标准的 JPEG 量化矩阵并取整, 然后根据 2 中算法来确定每一块的水印嵌入区域 $\{i, i+1, \dots, i+L-1\}$, 由 3.1 节中水印嵌入步骤(3)中的公式来抽取水印, 设 $\{i, i+1, \dots, i+L-1\}$ 中每个量化系数取绝对值后的最低比特位记为 b_1 和第 3 比特位记为 b_3 , 此时提取出的是含有 num 个水印版本的水印信号 $W_a(t)$:

$$W_a(t) = \begin{cases} 1, b_3 \otimes b_1 = 0 \\ -1, b_3 \otimes b_1 = 1 \end{cases} \quad t = 1, 2, \dots, m \times n \times cr \times num$$

(2) 水印图像恢复: 由于图像经过一些处理操作或攻击后, 恢复出的水印信号会有一定程度的失真, 恰好可以利用这 num 个水印版本对其进行校正, 以减少水印信号的失真, 计算公式如下:

$$W_b(i) = \begin{cases} 1, \text{if } \sum_{r=1}^{num} W_a((r-1) \times m \times n \times cr + i) \geq 0 \\ -1, \text{else} \end{cases} \quad i = 1, 2, \dots, m \times n \times cr$$

然后用生成水印时密钥 key 产生相同的长度为 $m \times n \times cr$ 二值随机序列 $P(k) \in \{-1, 1\}$ 对 $W_b(i)$ 进行解调制得解调制后



(a) 原始图像

重庆
大学

(b) 原始水印图像



(c) 嵌水印后的图像
PSNR=40.861dB

重庆
大学

(d) 不加任何攻击
提取出的水印 NC=1

图 3

下面对嵌水印后的图像(c)进行各种攻击, 然后从中提取出水印, 水印的提取不需要原始图像, 是盲检测。

A. JPEG 压缩。实验对加水印后的图像(c)进行多种不同质量因子的 JPEG 有损压缩, 然后盲检测出水印, 其结果如表 1 所示。

表 1

质量因子	90	80	70	60	50	40	35
NC	1	1	0.98773	0.99755	1	0.99509	0.97546

质量因子为 35 时 NC 高达 0.97546, 可见本文的算法能

的序列 $W_c(i), W_c(i) = Wb(i) \times P(i), 1 \leq i \leq m \times n \times cr$, 再对 $W_c(i)$ 进行扩展因子为 cr 的解扩展, 并转换为 $\{0, 1\}$ 的二值水印序列 $W_d(k)$:

$$w_d(k) = \begin{cases} 1, \text{if } \sum_{j=(k-1) \times cr + 1}^{k \times cr} W_c(j) \geq 0 \\ 0, \text{else} \end{cases} \quad k = 1, 2, \dots, m \times n$$

把 $W_d(k)$ 映射成二维矩阵 $W, W = \{W'(i, j), 1 \leq i \leq m, 1 \leq j \leq n, W'(i, j) \in \{0, 1\}\}$, 这样就得到了提取出来的最终水印图像。

(3) 评价水印的准则: 由于提取出的水印可能与原嵌入水印不完全一样, 这就需要有一个评价准则来衡量恢复的水印与原始水印的相似程度。本文除了可以从主观感觉上来评判提取出的水印的质量外, 还可以采用以下归一化相似度计算公式^[5]来客观地评判, 公式为:

$$NC = \frac{\sum_i \sum_j W'(i, j) \times W(i, j)}{\sum_i \sum_j W(i, j)^2}$$

其中 $W'(i, j)$ 表示提取出的水印, $W(i, j)$ 表示原始嵌入的水印。

4 实验结果及性能分析

本文的实验都是在 Matlab6.5、Windows2000 平台上实现的, 实验中使用经典的 $512 \times 512 \times 8$ 的 Lena. bmp 图像作为原始图像, 以一幅 32×32 的“重庆大学”二值图像为水印, 分别如图 3(a)、(b)所示。位扩展因子 cr 取值 4, L 取值 5, 则水印嵌入版本数 num 为 5。由本文算法得到的自适应嵌入水印后的图像如图 3(c)所示, 从视觉效果看, 很难感觉到水印的存在, 其峰值信噪比(PSNR)为 40.861dB, 这说明本文算法有很好的不可见性。在没有任何攻击下提取出的水印图像如图 3(d)所示, 归一化相似度 $NC=1$, 说明可以完全地提取出原水印。

很好地抗 JPEG 压缩, 这正是因为本文利用了 JPEG 量化余误差的结果。

B. 剪切攻击。对嵌水印后的图像从左上角开始分别剪去的整个图像 $1/16, 9/64, 1/4, 25/64, 1/2, 3/4$, 剪去的部分用白像素代替, 然后盲检测, 检测的结果如表 2 所示。

表 2

比例	1/16	9/64	1/4	25/64	1/2	3/4
NC	1	0.99877	0.9865	0.93129	0.84417	0.71043

当剪去整个水印图象的 $1/2$ 时 NC 仍高达 0.84417, 可见

本算法能很好地抗剪切攻击,这正是因为本文嵌入了多个水印版本的结果。

C. 缩放攻击。实验对嵌入水印后的图像分别进行先缩小到原来的 1/4 和 1/2 后再利用双三次插值放大到原图大小;先放大到原图 2 倍、3 倍和 4 倍后,再缩至原图大小。实验结果如表 3 所示,可见,算法对缩放具有很强的鲁棒性。

表 3

缩放比例	1/4	1/2	2	3	4
NC	0.68957	0.98896	0.99877	0.99877	1

D. 叠加椒盐噪声。对嵌入水印后的图像叠加概率为 0.01 的椒盐噪声,盲检测出的水印,NC=0.97178,恢复水印的能力强,这证实了本文的水印算法对椒盐噪声有很好的抵抗能力。

E. 滤波。对嵌入水印后的图像进行 3×3 的中值滤波处理,盲检测出水印,NC=0.98896。对嵌入水印后的图像进行高斯低通滤波后,NC=1。可见,本文的算法能很好地抵抗滤波操作。

结论 本文提出了一种新的数字水印算法,总结其特点如下:

(1) 选取 JPEG 量化取整后连续非零元素个数最多的区域来嵌入水印的,是自适应的,并且检测时不需要原始图像。

(2) 利用 JPEG 量化阶段的舍入误差来嵌入水印,因此能很好地抗 JPEG 压缩,并且对系数的修改很小,所以有很好的不可见性。

(3) 水印是经位扩展和调制产生的,并且嵌入了多个水印,所以能很好地抗剪切攻击。

参 考 文 献

- Hartung F, Kutter M. Multimedia watermarking techniques. In: Proc. IEEE, 1999, 87(7):1079~1107
- Schynndh R, Zrikh A, Osborne C. A digital watermark. In: Proc. IEEE Int. Conf. on Image Processing, 1994, 2:86~90
- Barni M, Barolini F, Cappellini V. A DCT-domain system for ro-

bust image watermarking. Signal Processing, 1998, 66(3):357~372

- 黄继武, SHI Y, 程卫东. DCT 域图像水印:嵌入对策和算法. 电子学报, 2000, 28(4):57~60
- Hsu C, Wu J. Hiding digital watermarks in image. IEEE Trans. on Image Processing, 1999, (1):58~68
- Lin S, Chen C. A robust DCT-based watermarking for copyright protection. IEEE Trans. on Consumer Electronics, 2000, 46:415~421
- Kunder D, Hatzinkos D. A robust digital image watermarking method using wavelet-based fusion. in Int. Conf. on Image Processing, 1997, 3:544~547
- Deng F, Wang B. A novel technique for robust image watermarking in the DCT domain. IEEE Int. Conf. Neural Networks & Signal Processing, 2003, 2: 1525~1528
- Wu J, Xie J. Adaptive image watermarking scheme based on HVS and fuzzy clustering theory. IEEE Int. Conf. Neural Networks & Signal Processing, 2003, 2:1493~1496
- Eyadat M. Factors that affect the performance of the DCT-block based image watermarking algorithms. In: Proc. IEEE Int. Conf. on information technology: Coding and Computing, 2004, 1:650~654
- Chen L, Lin J. Mean quantization based image watermarking. Image and Vision Computing, 2003, 21(8):717~727
- Karyali I, Berberidis K. Blind image-adaptive watermarking. In: Proc. IEEE Int. Conf. on, Electronics, Circuits and Systems, 2003, 2:894~897
- Yu D, Sattar F, Razul S G. Transparent robust information hiding for ownership verification. In: Proc. IEEE Int. Conf. on, Acoustics, Speech and Signal Processing, 2004, 3:401~404
- Bertran M, Delaigle J F, Macq B. Some improvements to HVS models for fingerprinting in perceptual decompression. In: Proc. 2001 Int. Conf. on Image Processing [C], 2001, 2:1039~1042
- Cox I J, Kilian J, Kilian J, Leighton F T, Shamoon T. Secure spread spectrum watermarking for multimedia. IEEE Trans. on Image Processing, 1997, (12):1673~1686

(上接第 56 页)

我们把基于 URL 垃圾邮件过滤系统,安装在西南教育网信息中心的邮件服务器上,统计了在安装系统前的一个月(6月)和安装系统后的一个月(7月)5个邮箱每天收到的垃圾邮件数,如图 3 所示。6月5个邮箱总共收到 1456 封垃圾邮件,平均每天 48.5 封,7月总共收到 1097 封,平均每天 36.6 封。安装基于 URL 垃圾邮件系统后,平均每天比安装前减少了 24.5% 的垃圾邮件,随着使用时间的增加,URL 列表服务器中收集的垃圾 URL 地址也增多,系统的过滤效果更好,收到的垃圾邮件逐渐减少,在 7 月底的几天内每天收到 20 封左右的垃圾邮件。

结论 随着垃圾邮件发送者的伪装技术增强,使用 HTML 格式含有 URL 的垃圾邮件不断增多。针对其特点,我们构建的基于 URL 的垃圾邮件过滤系统,对此类垃圾邮件具有很好的过滤效果,特别是对于短的、含有 URL 超链接的垃圾邮件;在邮件中大量随机插入无关,像是合法邮件,含有 URL 超链接的垃圾邮件;以上两类垃圾邮件过滤方法难以准确、高效地过滤。本文采用的 URL 地址字符串散列函数,加

快了垃圾邮件的查询速度,减少了邮件的处理时间。试验测试显示出基于 URL 的垃圾邮件过滤系统是种有效的垃圾邮件过滤方法。

参 考 文 献

- 上海艾瑞市场咨询公司. 2004 年中国反垃圾邮件研究报告, 2004, 3
- Sahami M, Dumais S, et al. A Bayesian Approach to Filtering Junk E-Mail. Learning for Text Categorization -Papers from the AAAI Workshop. Madison Wisconsin, 1998
- Levitt M, Burke B E. Choosing the Best Technology to Fight Spam. White Paper. Commtouch Company, April 2004
- Hulten G, Penta A, Opalaks G, Anavm M. Trends in spam products and methods. [Http://www.ceas.cc/papers-2004/165.pdf](http://www.ceas.cc/papers-2004/165.pdf)
- 李晓明, 凤旺森. 两种对 URL 的散列效果很好的函数. 软件学报, 2004, 15(2)