一种基于图像内容的半脆弱数字水印算法*)

高铁杠[1] 顾巧论[2] 陈增强[3]
(南开大学软件学院 天津 300071)[1] (上海交通大学安泰管理学院 上海 200052)[2]
(南开大学自动化系 天津 300071)[3]

摘 要 本文提出了一种新的基于混沌系统的半脆弱数字水印算法,该算法将图像分成两个大小相等的部分,其中一部分用来提取基于图像特征的水印,另外一部分则用来嵌入提取的特征水印,水印嵌入在小波变换的逼近子带中。水印提取算法不需要原始图像,而且通过比较原始水印和水印图像中的特征水印与提取的水印的两个相似度,能够区分图像受到攻击的种类,算法对恶意的篡改攻击能够较好地进行定位。实验结果表明,该算法在保持对常见的JPEG压缩稳健的同时,能有效地区分偶然失真与恶意篡改。

关键词 混沌系统,图像认证,半脆弱数字水印,小波变换

A New Semi-fragile Watermarking Algorithm Based on Image Content

GAO Tie-Gang[1] GU Qiao-Lun[2] CHEN Zeng-Qiang[3]
(College of Software, Nankai University, Tianjin 300071)[1] (Antai School of Management, Shanghai Jiaotong University, Shanghai 200052)[2]
(Department of Automation, Nankai University, Tianjin 300071)[3]

Abstract A new chaos-based semi-fragile watermarking algorithm is proposed in this paper, which first divided the total image into two parts of same size in pixel numbers, one part is used to extract image features which will be used watermarking, the other part is used to embed the watermarking extracted from the first part. The watermarking is embedded in the coefficients of low frequency of wavelet transform, and the watermarking can be extracted without resorting to the original host image. The correlation between the original watermarking and that of extracted and between the original watermarking and that of feature extracted from the watermarked image are used to distinguish the sort of distort imposed on image, and the tampered area of watermarked image can be reflected. Experimental results show that algorithm is robust to noise adding, compress and effective to differentiate occasional distortion from malicious tamper.

Keywords Chaotic system, Image authentication, Semi-fragile watermarking, Wavelet transform

1 引言

网络技术的发展使得数字媒体的分发变得异常简单和方便,同时使得数字媒体的安全成为一个非常严重的问题。因此,作为一种良好的数字媒体版权保护的手段,数字水印技术得到了快速的发展和广泛的关注,并已经成为国际学术界的一个研究热点[1]。

所谓数字水印技术,就是将一种特殊标志信息(伪随机序列或可识别图案文字)嵌入到数字媒体中,用以辨识数据的版权、合法使用者,从而认证或控制数据的使用。而利用数字水印进行多媒体信息的认证则是利用人类知觉系统的冗余,在不影响数字媒体的感官(视或听)质量的前提下,将与媒体内容相关或不相关的标志信息作为水印直接嵌入媒体内容中。当媒体内容需认证时,可将水印提出,鉴定其是否真实完整。认证水印除了具有公开水印的一般特征,如不可感知性、安全性、可靠性以及鲁棒性外,水印本身对篡改必须具有一定的敏感性和脆弱性,这也就是目前水印技术的发展和研究方向之一——半脆弱数字水印技术[2~4]。

半脆弱数字水印就是既能具备一般水印鲁棒的性能,又具有水印的脆弱性,即在容忍一定程度的常见信号处理操作的同时,还可把正常的信号处理与恶意篡改区别对待。篡改发生时,半脆弱水印认证系统不仅可提供篡改的破坏量位置,而且可帮助分析篡改类型。

由Kundur和Hatzinakos于1999年提出的称为非法篡改证明的半脆弱数字水印技术是最早提出的半脆弱水印方案之一[2],该方案借用一个篡改评估函数,能够在空域和频域范围内确定信号的修改程度。Lin等人通过在原始图像的DCT域提取伪随机的白噪声序列,将此序列作为认证序列,并把序列叠加到每个8×8的DCT块上的三角矩阵中,而后对DCT块反变换,并结合水印强度因子合成含水印图像。此算法几乎对所有的图像处理操作都有较强的鲁棒性,不足的是对平滑处理较敏感[3]。张静和张春田根据JPEG压缩过程中的不变参量进行水印生成和嵌入调制,利用小波特性对图像篡改区域进行定位[4]。李春等人利用JPEG压缩对图像小波系数大小关系的影响提出了一种半脆弱水印算法,算法仿真显示出具有很好的抗JPEG压缩性能,对篡改的定位也很精确[5]。陈生潭等则提出了一种DCT域与DWT域相结合的双重认证图像半脆弱数字水印算法,通过定义一个小波系数量化函数以实现水印的嵌入与提取,并根据图像局部特性,该量化函数可动态地进行调整,有效地区分了偶然失真与恶意篡改[6]。王兴元等提出了一种基于图像特征和超混沌迭代的图像认证算法,该算法从原始图像的小波分解中提取低频分量,并从其边缘中找出稳定的特征点,结合水印图像信息进行超混沌迭代,得到索引集。其研究结果表明:该算法对噪声、滤

*)国家自然科学基金(60574036)以及教育部博士点基金(20050055013)资助。高铁杠 教授,博士,主要研究方向为信息安全、软件工程、混沌与复杂系统。

波、压缩、旋转等图像处理方法具有较好的鲁棒性[7]。

上述提出的半脆弱数字水印算法中，有的将水印嵌入到小波变换的高频子带中[5,6]，有的则是在全局图像的小波变换的低频子带嵌入水印图像[7]。本文提出了一种新颖的基于图像特征的半脆弱数字水印方法，利用混沌系统的类随机性和遍历性的特点，将载体图像分成两个部分。一部分用于提取图像特征，进而形成需要嵌入的数字水印；另外一部分则利用小波变换的逼近子带嵌入水印，通过原始水印和提取的水印的相关度以及原始水印和基于特征提取的水印相关度这两个相关度来判别图像的认证和篡改，并给出了版权信息判断与恶意篡改认证的认证规则。实验结果验证了本算法的有效性能。

2 水印嵌入方法

混沌系统由于其本身的类随机性和遍历性的特点，在数字水印的水印生成、算法设计中得到了广泛的应用[7,8]。本文采用基于混沌系统的数字水印嵌入技术。为了使得算法具有一定的鲁棒性，本文不是在每一个分块都提取图像特征水印并进行嵌入[9,10]，而是采取了和传统的整体水印嵌入不同的方法：首先将原始图像进行分块，然后采取将一部分图像块用来生成水印信息，而另外的部分则用来嵌入水印，这样形成在水印提取中可以相互补充的两部分。

2.1 水印生成

对于灰度为 256、分辨率为 256×256 的图像，首先将图像分为 1024 个 8×8 的互相不重叠的子块，并将其给予 1 至 1024 的编号。而后根据 Logistic 混沌映射将其随机地分成两个部分，其中一部分用于提取图像特征，即组成等待潜入的数字水印；另外一部分用于水印嵌入。子块划分算法如下：

(1)对于设定的初始数值，首先进行下列(1)式的一定次数的迭代，以去除系统的暂态：

$$x_0=4x_0(1-x_0) \tag{1}$$

(2)取编号

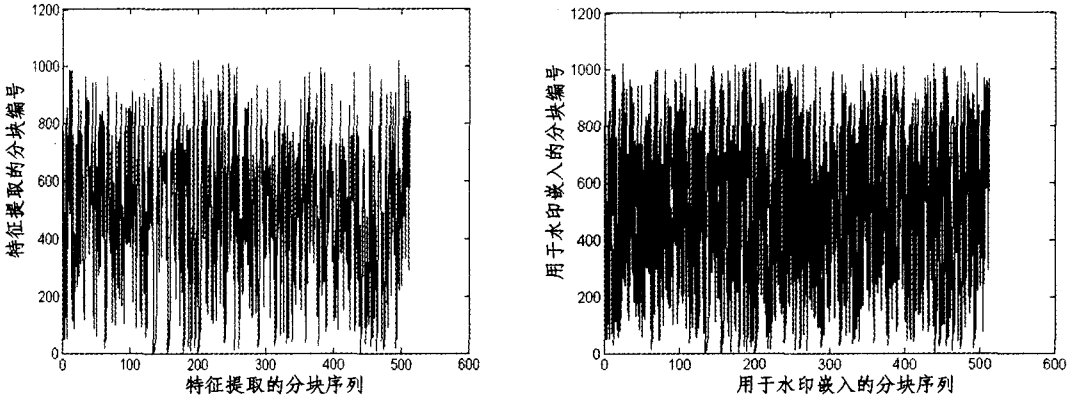
$$l=\mathrm{mod}((Abs(x_i)-Floor(abs(x_i))\times 10^{13},1024)+1 \tag{2}$$

显然，$l\in[1,1024]$。

(3)反复执行式(1)和(2)，直至找到 1024 个互不相同的编号。

(4)将前面的 512 个编号代表的图像子块依次排列，用于图像的特征提取。

由于混沌系统对系统初始数值的敏感以及 Logistic 的遍历性，因此随机提取的子块分布也呈随机性。在系统初始值为 0.7132 时，得到的用于特征提取和水印嵌入的子块编号结果如图 1。



(a) 用于特征提取的图像子块　　(b) 用于水印嵌入的图像子块

图 1　用于特征提取和嵌入水印的图像分块序列比较

对于用于特征提取的部分的每一个图像子块，用下列的算法进行特征提取：

(1)计算小的子块内像素值和的平均值，并除以 16，得到量化值 Δ：

$$\Delta=\frac{1/64\sum_{i=1}^{64}I(i)}{16} \tag{3}$$

(2)根据下面公式计算提取的四比特数字 w：

$$w=\begin{cases}0000 & Min<\Delta\leqslant Min+\dfrac{Max-Min}{16}\\ 0010 & Min+\dfrac{Max-Min}{16}<\Delta\leqslant Min+\dfrac{2(Max-Min)}{16}\\ \vdots & \\ 1111 & Min+\dfrac{15(Max-Min)}{16}<\Delta\leqslant Max\end{cases} \tag{4}$$

其中，Min 和 Max 表示 64 个数值中的最大和最小值。对选择的 512 个子块进行循序提取，将组合成 2048 个比特的数字水印序列 $W=\{w(i)\mid w(i)\in\{0,1\},i=1,2,\cdots,2048\}$。

2.2 水印嵌入

离散小波变换(Discrete Wavelet Transform，DWT)是变换域通常采用的方法。小波变换具有时域和频域的良好局部化性质，图像的小波分解就是将图像分解到频域中，同时保留图像在空间上的分布，便于选择理想的水印信号频域添加范围，并使得水印具有多分辨率检测能力。图像经小波分解后，逼近子带集中了原始图像的绝大部分信息，是对原始图像的最佳逼近。这里聚集着图像的大部分能量，具有很强的抗干扰能力。而高频子带系只是图像的细节信息，因此水印应该首先嵌入小波图像低频系数[11]。

嵌入方法将采用分块嵌入规则。将上述剩余的 512 个子块按照序列生成的先后顺序进行逐个嵌入。嵌入算法采用二级二维小波分解技术。由于小波分解后的低频子带是对原始图像的最佳逼近，而水印应该放在视觉系统感觉上最重要的分量上，因此本文选择了小波逼近子带，即 LL_2 作为水印嵌入区域。每个小的子块嵌入四个比特，嵌入算法描述如下：

将子带的系数被 Δ 整除。如果此时结果为偶数，那么如果水印值为 0，子带系数不变，否则子带系数采用下列公式进

行修改：

$$I'(i,j)=I(i,j)+\alpha\Delta \tag{5}$$

其中，量化值$\Delta=2^2$，α取大于零的偶数；反之，如果子带系数被整除的结果为奇数，当水印值为1时，子带系数不变，为0时则执行(5)式进行系数的修改。

2.3 水印检测和提取

从水印的提取和嵌入过程可以看出，嵌入水印的图像中既可以提取图像特征水印，又可以进行水印的提取，因此本算法在水印提取过程中不需要原始载体图像。为了明晰原始水印和水银图像中的图像特征水印以及提取的水印三者之间的关系，我们设原始嵌入的水印为W，从嵌入水印的图像中进行特征提取的水印为W_0，从嵌入水印的图像中提取的水印为W_t，则W_t的提取过程为：

采用和嵌入过程一样的混沌映射、初始参数和分块规则，对于用于进行水印提取的512个子块的每一个小块进行小波变换，将逼近子带的系数被Δ整除。此时结果为偶数，那么如果水印的位提取为0，否则提取水印的位为1。记原始水印和提取的水印的相关度为

$$\lambda(W,W_t)=\frac{\sum_{i=1}^{2048}w(i)w_t(i)}{\sqrt{\sum_{i=1}^{2048}w_t^2(i)}\sqrt{\sum_{i=1}^{2048}w^2(i)}} \tag{6}$$

原始水印和从含水印的图像中进行特征提取的水印的相关度为

$$\lambda(W,W_0)=\frac{\sum_{i=1}^{2048}w(i)w_0(i)}{\sqrt{\sum_{i=1}^{2048}w_0^2(i)}\sqrt{\sum_{i=1}^{2048}w^2(i)}} \tag{7}$$

通过对这两个相关度的判断，我们可以得到关于图像内容的认证和篡改情况。混沌系统的初始值对提取的过程至关重要，因为混沌系统对系统的初始状态极为敏感，所以混沌系统不同的初始值对上述两个相关度也有重要影响。例如，当系统初值为$x_0=0.7132$时，提取的图像特征水印信息和初值为$x_0=0.71320000000001$时提取的图像特征水印的相关度为$\lambda=0.5003$。对于下列的同样为256灰度颜色的“Moon surface”、“Aerial”、“Airplane”、“Clock”和“Lenna”图像(图2)进行同样的测试，得到混沌初始状态具有微小差别时的提取特征的相关度见表1。

表1 不同图像对不同混沌初值形成水印的相关度

图像	相关度
Moon surface	0.5336
Aerial	0.5017
Airplane	0.5611
Clock	0.5329
Lena	0.5027

3 版权信息的认证及篡改的识别

在本文提出的数字图像水印实现方案中，版权信息的认证与恶意篡改的识别是通过比较式(6)和(7)得到的。由于水印的嵌入和提取采用的是分块置乱算法，因此可以推断，对于恶意的剪切攻击，两个相关度将呈现出比较大的差别。而对于一般的常规图像处理，如有损压缩、叠加噪声等，两个相关度则不会呈现太大的差别。因此，我们可以将$\lambda(W,W_0)$作为一种辅助手段进行版权信息的认证和篡改识别。而文[12]则是借助于水印差图中稀疏点与稠密点的比值来判断攻击类别，在某一阈值范围内判断偶然攻击和恶意攻击，而水印差图即原始水印和提取水印之间的差异。

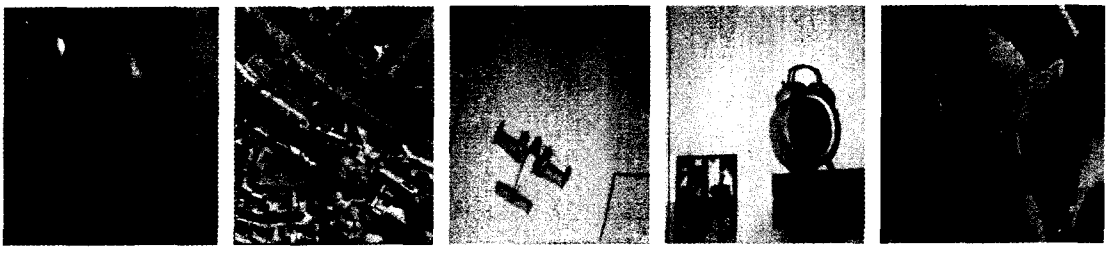


图2 不同的测试图像

按照本文提供的特征提取水印与嵌入算法，对于一幅图像，可以提取两种水印。而这两种水印的提取则是利用了水印图像的全部，因此从对图像进行的常规处理与一般的恶意攻击考虑，常规处理一般对图像的影响是全局性的，所以提取的两种差别较小，而恶意攻击则一般是局部小面积攻击。从嵌入算法可以看出，可能影响的只是两种水印中的一方。因此，通过适当选取阈值，我们提出如下的版权信息判断与恶意篡改认证的认证规则(见表2)：

表2 版权信息判断与恶意篡改认证的认证规则

条件	判断结果
$\lambda(W,W_0)=\lambda(W,W_t)=0$	图像可靠
$\lambda(W,W_0)>T$ 同时 $\lambda(W,W_t)>T$	已进行常规图像处理但内容可靠
$\lambda(W,W_0)>T$ 同时 $\lambda(W,W_t)<T$	图像受到恶意剪切攻击
$\lambda(W,W_0)>T$ 同时 $\lambda(W,W_t)<T$	图像已经不可用

4 仿真实验

图3 嵌入水印后的图像

本实验采用的是256×256×8bit的Woman图像，混沌系统的初始数值是0.7123，初始迭代次数为3000次，拉伸因子$\alpha=4$。图像采用“db1”小波进行二级分解，检测阈值T定

为 0.8。在无篡改的情况下，内嵌水印后图像如图 3 所示，此时得峰值信噪比 $PSNR=42.1060$。可以看出，该嵌入算法具有较高的峰值信噪比，而且良好的主观视觉效果也证实了本算法实现的水印具有不可感知性，隐藏效果好。

4.1 恶意剪切攻击实验

首先，对嵌入水印的图像进行恶意的剪切攻击，剪切比例和两个相关度的数值的测试结果见图 4。

从图 4 可以看出，对于恶意的剪切攻击，当剪切比例在以内时，两个相关度呈现比较大的差别。因此，当两个相关度值的差大于 0.5 时，我们可以认定这种攻击为恶意攻击。实际上，从水印的嵌入过程可以看到，由于用于嵌入水印的图像占有图像的一半，而另外一半是用于表征图像特征的水印提取，所以当剪切比例不大时，导致两个相关度有较大的差异。需要说明的是，在当剪切比例大于 50%，即使能够得到良好的水印检测效果，我们也将认为检测效果无效。

4.2 常规图像变换攻击实验

对于图像的有损压缩以及添加噪声，有的学者将其归并到恶意攻击一类[6]，这里我们将其看作是对图像进行的一些常规处理。通过进行的有损小波压缩实验，得到的实验压缩比例和得到的两个相关度的比较图见图 5(a)。

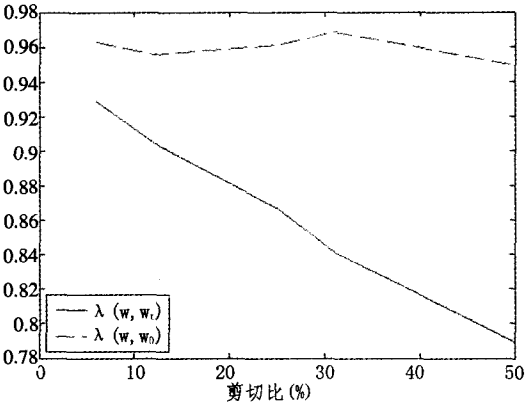
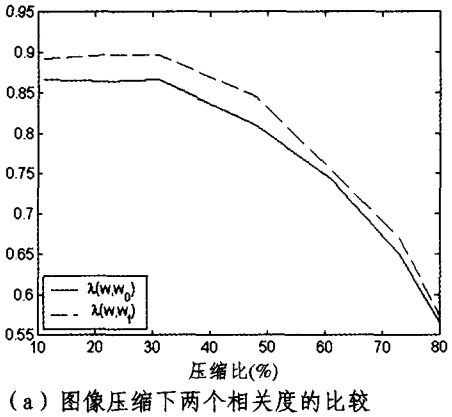
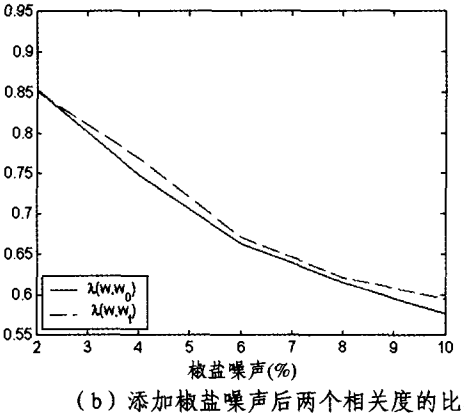


图 4 不同剪切比例下相关度的比较



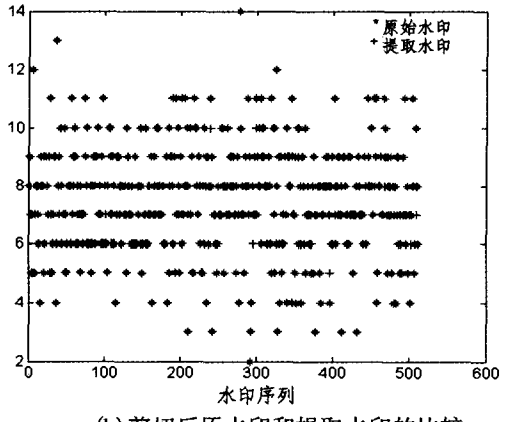
(a) 图像压缩下两个相关度的比较



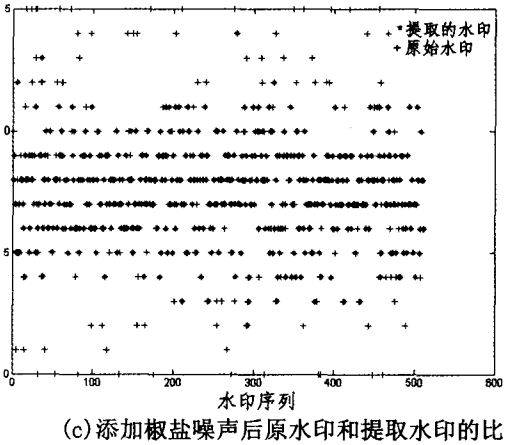
(b) 添加椒盐噪声后两个相关度的比较

图 5 常规攻击实验的相关度比较

(a) 水印图像的剪切攻击



(b) 剪切后原水印和提取水印的比较



(c) 添加椒盐噪声后原水印和提取水印的比

图 6 剪切攻击实验结果

从图中可以看出，本方法对图像的压缩具有较强的鲁棒性能，而且两个相关度的差值并不大，和我们提出的版权信息判断与恶意篡改认证的认证规则相符合。

同样进行的测试是在含有水印图像中添加椒盐噪声，测试结果如图 5(b)。显而易见，水印的相关度对椒盐噪声比较敏感，但是这两个相关度仍然呈现出相近的特性，这也验证了上述的版权信息认证规则。

4.3 攻击识别实验

对水印图像进行局部的修改，修改效果如图 6(a)，对提取的水印和原始水印每四位组成一个数。图 6(b)是提取的水印和原水印的效果比较图，此时 $\lambda(W,W_t)=0.9648$。通过比较，容易确定修改的图像块。同样，图 6(c)是对含有水印的图像添加椒盐噪声后原水印和提取水印的效果图，此时有 $\lambda(W,W_t)=0.9250$。

4.4 水印同步攻击实验

假定水印嵌入算法和提取算法被窃取，但是密钥没有公开，这时图像可能受到其它一些水印嵌入的攻击。这种情形下，我们随机选择了 1000 组合原始水印同样长度的水印进行相关度测试(其中有一个和原始水印相同)，得到的结果如图 7。

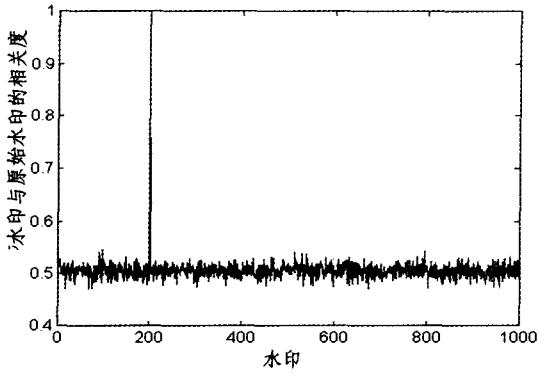


图 7 水印同步实验结果

从上面的图形可以看出，即使受到假冒水印的攻击，算法和相关度也能够对水印进行识别，只要适当选取阈值即可。本文我们选择 $T=0.85$，即可较好地区别对水印图像的各种篡改和完成对图像的认证。

结束语 本文提出了一种基于混沌系统的图像分块算法，该算法将图像分成两个大小相等的部分，其中一部分用来提取基于图像特征的水印，另外一部分则用来嵌入提取的特征水印。水印嵌入在小波变换的逼近子带中。水印提取算法不需要原始图像，而且通过比较原始水印和水印图像中的特征水印与提取的水印的两个相似度，能够区分图像受到攻击的种类。算法对恶意的篡改攻击能够较好地进行定位。实验结果表明，该算法在保持对常见的小波压缩稳健的同时，能有效地区分偶然失真与恶意篡改。

参 考 文 献

1 Chao H M, Hsu C M, Miaou S G. A Data Hiding Technique with Authentication, Integration and Confidentiality for Electronic Patient Records. IEEE Trans Inf Technol Biomed, 2003, 6 (1): 46～53

2 Kundur D, Hatzinakos D. Digital watermarking for telltale tamper-proofing and authentication. Proc. IEEE, 1999, 87(7): 1167～1180

3 Lin E T, Podilchuk C I, Delp E J, Detection of image alterations using semi-fragile watermarks. Proc. SPIE, 2000(3971): 152～163

4 张静，张春田. 用于 JPEG2000 图像认证的半脆弱性数字水印算法. 电子学报，2004，32(1)：157～160

5 李春，黄继武. 一种抗 JPEG 压缩的半脆弱图像水印算法. 软件学报，2006，17(2)：315～32

6 陈生潭，侯振华，王虹现. 双重认证的变换域图像半脆弱数字水印算法. 计算机辅助设计与图形学学报，2005，17(5)：1114～1119

7 王兴元，石其江. 基于图像特征和超混沌迭代的图像认证算法. 计算机研究与发展，2005，42(11)：1896～1902

8 Zhao D W, Chen G R, Liu W B. A chaos-based robust wavelet-domain watermarking algorithm. Chaos, Solitons and Fractals, 2004, 22: 47～54

9 Hassan M H, Gilani S A M. A Semi-fragile signature based scheme for ownership identification and color image authentication. Transactions on Engineering. Computer and Technology, 2006, 13: 308～311

10 Celik M U, Sharma G, Saber E, et al. Hierarchical Watermarking for Secure Image Authentication with Localization. IEEE Trans Image Process, 2002, 11 (6): 585～595

11 黄达人，刘九芬，黄继武. 小波变换域图像水印嵌入对策和算法. 软件学报，2002，13(7)：1290～1296

12 Hu J Q, Huang J W, Huang D R, et al. Image fragile watermarking based on fusion of multi-resolution tamper detection. Electronics Letters, 2002, 38(24):1512～1513

(上接第 218 页)

们形成的聚类之间的相关性，而且无需事先给出无从知晓的聚类的数量以及聚类的距离阈值，因此易于实现，并取得了良好的实验效果。当然，本方法对于内存的需求量较大，待消解项过多的测试语料可能难以一次性评测，所以，我们下一步拟采用 LP Chunking 的方法来评测待消解项过多的测试语料。

参 考 文 献

1 王厚峰，梅铮. 鲁棒性的汉语人称代词消解. 软件学报，2005，16(5)：700～707

2 Soon W M, Ng H T, et al. A machine learning approach to coreference resolution of noun phrases. Computational Linguistics, 2001, 27 (4): 521～544

3 Ng V, Cardie C. Improving machine learning approaches to coreference resolution. In: Proc. of the ACL, Philadelphia, 2002

4 李国臣，罗云飞. 采用优先选择策略的中文人称代词的指代消解. 中文信息学报，2005，19 (4)：24～30

5 Luo X, Ittycheriah A, et al. A mention～synchronous coreference resolution algorithm based on the Bell tree. In: Proc. of the ACL, Barcelona, 2004

6 Demaine E D, Emanuel D, et al. Correlation Clustering in General Weighted Graphs. Theoretical Computer Science, 2006, 361 (2-3): 172～187

7 Quinlan R J. C4. 5: Programs for Machine Learning. San Francisco, CA: Morgan Kaufmann, 1993

8 Wang H F, Mei Z. An empirical study on pronoun resolution in Chinese. In: Gelbnkh A, ed. Proc. of the 5th CiCLing Conf., Heidelberg, 2004