基于公钥和脆弱水印的图像认证算法*)

张鸿宾　杨　成
(北京工业大学计算机学院　北京100022)

摘　要　本文提出一种基于公钥和小波域脆弱水印的图像完整性认证的算法。该算法不但能够检测和定位窜改，而且能同时验证图像的所有权。该水印方案为盲水印，检测过程不需要原始图像和附加信息。由于该认证算法建立在密码学方法的基础上，因此它是单向、不可逆的，未授权者很难伪造或修改原有的水印。文中分析了针对逐块独立的一类水印算法的"矢量量化攻击"，提出了认证链等对抗措施。理论分析和实验结果表明，本文算法具有较好的性质和较高的安全性。

关键词　多媒体认证，数字水印，小波变换，哈希函数，RSA 公钥算法

An Image Authentication Scheme Based on Public Key and Fragile Watermarking

ZHANG Hong-Bin　YAHG Cheng
(Computer Institute of Beijing University of Technology, Beijing 100022)

Abstract　In this paper, we propose an image authentication scheme based on public key and fragile watermarking in wavelet domain for ownership verification and authentication on integrity of image. The scheme is not only able to detect minus modification made to the image and meanwhile indicate the specific location that have been modified, but also can check for the owner of the image according to the user key used in the verification process. The scheme is a blind watermark, and the original image or any other information is not necessary in the process of verification. Since the algorithm is based on the security of the cryptographic function, and requires a user key during both the insertion and the extraction procedures, it is not possible for an unauthorized user to forge a watermark or alter the existing watermark so that the resulting will pass the test.

Keywords　Multimedia authentication, Digital watermarking, Wavelet transform, Hash function, RSA public key encryption

1　引言

近年来数字多媒体的应用取得了很大的进展。数字媒体易于编辑、合成、复制和传播等优点在推动信息化社会前进的同时，也使它的知识产权保护和真实性、完整性的认证等问题成为人们关注的焦点。推动了以知识产权保护和完整性认证等为目标的数字水印技术的研究。

数字水印利用人的视听觉系统的特性，有控制地将一些标识嵌入到多媒体数据之中。这些标识信息以后可以用作版权证明、完整性认证、拷贝控制或内容注释等目的。数字水印的早期研究主要集中在版权保护的鲁棒(robust)水印上。随着研究的深入，人们发现，数字水印在多媒体完整性的认证方面，同样有很好的应用前景。这一点目前人们重视得还不够。

所谓多媒体数据的认证就是要确认数据是否完整(integrity)、有无窜改，以及真实(reality)和来源可靠。多媒体数据的认证主要有两个功能：一是确认多媒体数据的完整、真实和可靠，二是可以作为电子证据(electronic evidence)。多媒体认证在电子商务和政务、法庭证据、新闻传媒、金融、保险、公安和医疗文档的认证以及军事情报等领域都有广泛的应用。

传统的密码学在多媒体完整性的认证上有安全性较高的优点，但也存在一定的局限。一方面，多媒体数据往往允许一定程度的改变(如有损压缩等)，而密码学的认证方法不容许任何变化，也很难确定窜改的程度和位置。另一方面，密码学的认证方法需要另外的文件保存和传送认证信息，这使得它的安全性存在一定的漏洞[1]。

近年来人们开始把脆弱(fragile)和半脆弱(semi-fragile)水印用于多媒体完整性的认证上[2~4]。利用脆弱水印的多媒体认证问题可以粗略地描述如下：

给出一个多媒体数据 f 和数字水印 w，通过有控制地修改 f 将 w 不可感知地嵌入到 f 中，产生一个加水印后的数据 g，使得：

1. 水印 w 可以从 g 中盲抽取出来；

2. 如果 g 的内容没有变化，则从 g 中抽取出的水印 w' 将和 w 完全相同；

3. 如果 g 的内容被修改了，则 w' 将和 w 不同；

4. 根据 w' 和 w 之间的差别可以判定 g 的内容被窜改的位置和程度；

5. 嵌入的水印应能抵抗各种旨在去除、破坏或使认证歧义的攻击。

在上述的定义中，如果"g 的内容"是指 g 的每一位原始数据，那么这种水印称为脆弱水印，它不允许对数据有一点修改，这时的认证一般称为完全认证。如果"g 的内容"是指 g 的语义层次上的内容，则这种水印称为半脆弱水印。在数据的内容语义不变的前提下，它允许对原始数据作一些修改，这时的认证成为内容认证。医疗和军事情报等场合往往需要完全认证，而在其它的场合，数据往往要经过压缩的处理，这时需要的一般是内容认证。

到目前为止，人们已经提出了许多利用数字水印的认证算法。Walton 和 van Schyndel 等人分别提出了空间域的脆弱水印算法[2,5]。他们将水印信号嵌在了图像位平面的最低位

*)本课题得到国家自然科学基金、北京市自然科学基金、863计划和北京市教委科技发展计划的资助。

上。这些方法的主要问题是安全性上有漏洞，而且不能经受有损压缩。Wolfgang 和 Delp 提出的 VW2D 脆弱水印算法将一个双极性的二维 m 序列嵌入数据的空间域中[6]。该算法利用 m 序列和原图像的内积给出了窜改程度的相对度量。但 Delp 算法的检测过程需要原图像。它的另一个主要问题是和许多利用最低位嵌入的算法一样，不能抵抗只修改高位而不影响低位的攻击。Yeung 和 Mintzer 的方法使用了查找表来控制像素的改变，利用二维标志图像作为脆弱水印[3]。这种方法的安全性取决于推断查找表的难度，而且它也不能经受有损压缩。Wong 提出了一种将图像块的数据、大小和密钥加密成摘要后嵌入最低位平面的方法[7,8]。Wong 的方法有较强的安全性，但在抵抗针对分块算法的“矢量量化攻击”上存在着漏洞。虽然文[9]对算法作了改进，在哈希函数的输入中加入了图像序列号和块编号，但图像序号的发布，管理和传递又产生了新的问题，而且使用图像序列号的水印检测使得这种算法成为半盲的水印算法。

由于在各种有损压缩方法中广泛使用了各种变换如 DCT、DWT 等，而且由于变换系数的修改和量化有较好的感知模型研究成果可以参考，因此变换域的水印引起了人们更大的兴趣。Wu 和 Liu 提出了一种 DCT 频域的查找表方法[10]。水印是通过修改量化后的 DCT 系数实现的，是文[3]的方法在 DCT 域的一种推广。Kunder 和 Hatzinakos[11]以及 Xie 和 Arce[12]分别提出了小波域的脆弱水印算法。Kunder 的算法通过量化 Haar 小波变换系数的过程来嵌入水印，而 Xie 的方法则是在 SPIHT 压缩域有选择地嵌入水印位。由于图像的小波分解同时具有空间和频率的信息，因此小波域的脆弱水印便于定位和刻画窜改。

综上所述，尽管有不少图像认证的算法发表，但在性能和安全性上仍有许多问题需要进一步研究。作者认为，一个安全性高的认证算法必须和密码学的方法相结合，而且完整性认证最好和所有权的验证同时进行。本文提出一种密码学和脆弱水印相结合的图像认证算法，它能检测并定位对图像的任何窜改，并通过密钥来验证图像的所有权。该算法为盲水印算法。为了抵抗对基于分块的水印算法的“矢量量化”攻击，本文提出了一种认证链的水印方法。这些方法保证了认证算法具有较高的安全性。

下面的第2节详细描述本文的认证算法，第3节是实验结果和分析，最后是小结。

2 基于公钥和脆弱水印的图像认证和所有权验证算法

本文算法的基本思想是，将图像的小波变换系数经哈希函数变换后，用私钥对其加密，将加密后的信息作为水印嵌入图像小波变换重要系数的最低位(LSB:Least Significant Bit)上。检测时，从相应的嵌入位置提取出水印信息并用公钥进行解密，然后将其和从待测图像中得到的哈希函数值进行比较，以此来进行图像的窜改检测和所有权验证。为了确定被窜改的位置，算法将图像分块，水印的嵌入和检测是逐块进行的。为了抵抗对分块水印算法的“矢量量化攻击”，算法采用了将相邻块链接起来一起认证的认证链方法。和传统的数字签名相比，本文算法的认证信息直接嵌入了图像之中，不需要单独的文件来存放。和一般的脆弱水印方法相比，本文算法的长处是同时验证了数据的完整性和所有权人，伪造者很难伪造水印和窜改已有的水印。下面分别介绍本文算法的主要步骤。

2.1 水印的产生和嵌入

对于任意一幅尺寸为 $m\times n$ 的灰度图像 $x_{m,n}$，嵌入不可见的水印 w 后的图像记为 $y_{m,n}$。将图像 $x_{m,n}$ 划分为尺寸为 $i\times j$ 的块，并依次编号为 $b_1,\cdots,b_r$。水印的嵌入是按块进行的，其过程如图1所示。

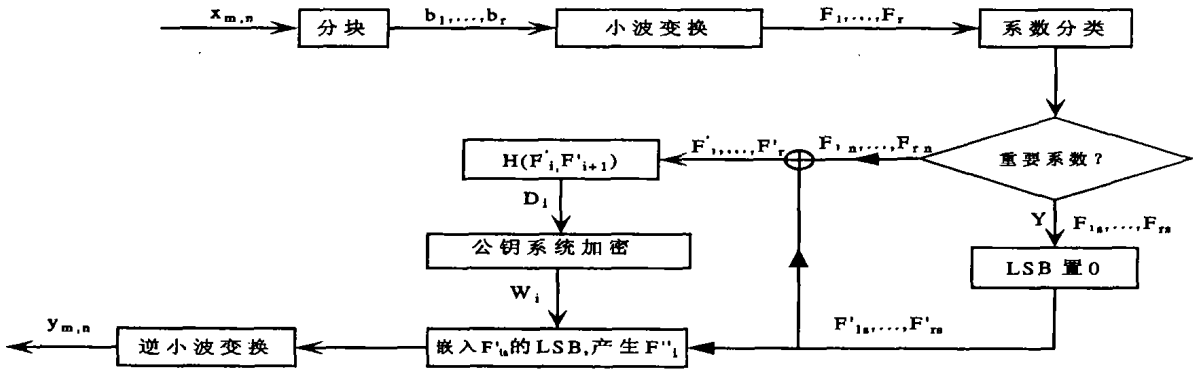


图1 水印的产生和嵌入流程

采用 Calderbank 和 Daubechies 等提出的整数小波[13]，对每个图像块作3级整数小波变换后，可以得到各块的小波变换系数组 $F_1,\cdots,F_r$。从每个 $F_i(1\leqslant i\leqslant r)$ 的最低频子带中选择绝对值较大的系数作为重要系数。根据小波零树编码的思想[14]，从这些重要系数及其在较高分辨率上对应的子块系数中选出128个重要系数，记为 F_{is}。其余非重要系数记为 F_{in}。将 F_{is} 的最低位分别置为零，记为 F'_{is}。以后水印将嵌在 F'_{is} 的最低位上。由 F'_{is} 和 F_{in} 一起组成新的系数组 $F'_i,1\leqslant i\leqslant r$。

为了能够抵抗“矢量量化攻击”，本文把相邻两个图像块的系数链接起来组成一个认证链。认证链的定义如下：

$$G_i=\begin{cases}(F'_i,F'_{i+1}),1\leqslant i\leqslant r-1\\(F'_i,F'_1),i=r\end{cases}$$

将 G_i 和相应的图像块号 i 经哈希函数 H 加密后形成摘要(digest)，再经过公钥算法加密后产生水印，嵌在图像块 b_i 的小波系数 F'_{is} 的最低位上。如果某个图像块 $b_i(i=1,\cdots,r)$ 发生了变化，则验证过程中含有 $b_{(i-1)\bmod r}$、b_i 和 $b_{(i+1)\bmod r}$ 三个图像块将会出现不匹配的情况。而图像块 $b_{(i-1)\bmod r}$ 和 $b_{(i+1)\bmod r}$ 的完整性可以分别由 $b_{(i-2)\bmod r}$ 和 $b_{(i+2)\bmod r}$ 来确定，因此可以定位窜改是否发生在 b_i 上。

本文采用了 Rivest 提出的 MD5哈希函数[15]，它将任意长度的输入数据散列为128位长的输出位串，即：

$$H(s)=(d_1,d_2,\cdots,d_{128})$$

其中 s 为任意长度的数据串，$d_i,i=1,\cdots,128$，为哈希函数的二值输出。MD5哈希函数的性质是，当输入位串 s 有一点变化时，输出位串 $(d_1,d_2,\cdots,d_{128})$ 会有显著的变化。而在已知一个输入位串 s 的输出结果时，寻找另一个任意长度的输入位串

s'使得 $H(s')=H(s)$在计算上不可行的。

上述的水印嵌入过程可以归纳如下(见图1)：

1)令 b_i 表示图像 $x_{m,n}$的第 i 个数据块。将 b_i 的小波变换系数 F_i 中的128个重要系数 F_{in}的最低位置为零后和其余的系数一起形成新的系数组 F'_i；

2)将 F'_i 和下一图像块的系数组链接起来形成认证链 G_i，$1\leqslant i\leqslant r$；

3)用 MD5哈希函数对 G_i 加密，形成摘要 D_i：

$D_i=H(i,G_i)=(d'_1,d'_2,\cdots,d'_{128})$

4)用公钥算法的私钥 K'对 D_i 加密后形成水印 W_i：

$W_i=E_{k'}(D_i)$

式中 $E_{k'}(\cdot)$是公钥加密函数，K'是私钥。后面的实验中采用了RSA 公钥算法[16]；

5)将 W_i 嵌入 F'_{in}的各个最低位后，和非重要系数 F_{im}一起产生加水印后的小波系数组 F''_i。然后进行逆向小波变换，产生加水印后的图像 $y_{m,n}$的图像块。

2.2 水印的提取和图像认证

水印提取和图像认证的过程如下(图2)：

1)对任意一幅待认证图像 $z_{m,n}$，将其分为 $i\times j$ 大小的块，记为 $C_1,\cdots,C_r$；

2)对每一个图像块 $C_i(1\leqslant i\leqslant r)$分别进行小波变换，得到小波系数组 E_i。从 E_i 中选择出128个重要系数 E_{in}；

3)从 E_{in}最低位上提取出用私钥加密后的水印 U'_i，并用相应的公钥进行解密，有

$U_i=E_k(U'_i)$，

式中 $E_k(\cdot)$是公钥系统的解密函数，k 是相应私钥 k'的公钥；

4)将重要系数 E_{in}的最低位置为0后，与非重要系数 E_{im}一起产生新的系数组 E'_i；

5)和嵌入时一样，构造认证链 D_i 并用哈希函数求出其摘要 V_i；

6)比较 U_i 和 V_i 是否相同，并由此和前后块的情况确定图像块 C_i 是否被窜改。

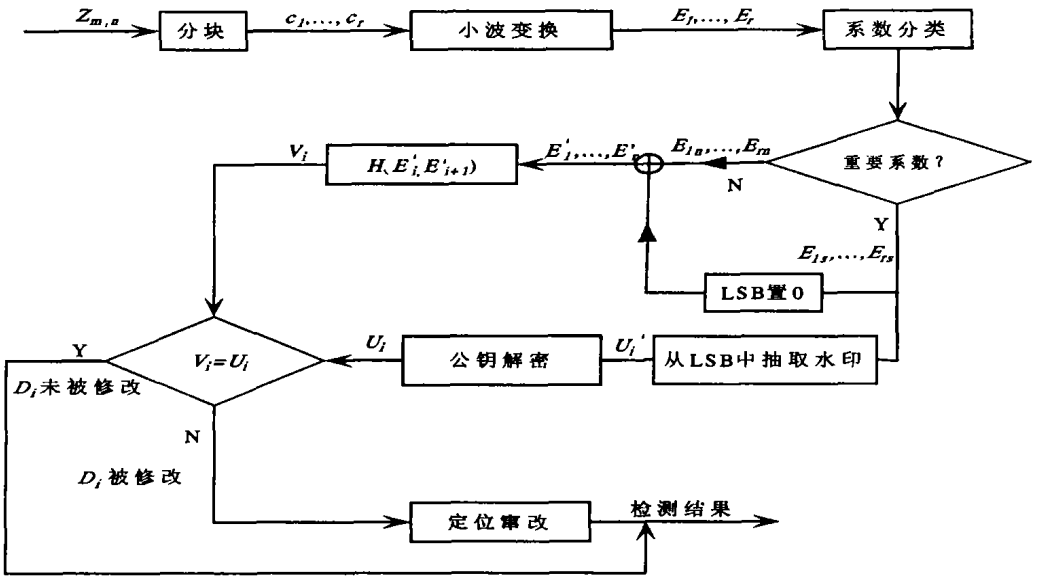


图2 水印提取和图像认证流程

2.3 水印算法的性能分析

由上述水印嵌入和提取的过程可以看出，本文的认证算法具有以下的性质：

1)如果加水印后的图像没有被修改，即必有 $U_i=V_i$，$1\leqslant i\leqslant r$。反之，如果图像被修改了，这时由哈希函数的性质可知，U_i 和 V_i 将有很大的不同，以此可以推断和定位窜改；

2)图像所有权的验证通过水印嵌入和提取时的密钥来进行；

3)由于水印嵌在了小波系数的最低位平面上，因此对图像质量的影响很小。这一点也为后面的实验所证实；

4)水印的提取不需要原图像；

5)由于本文水印的目的是检测对图像的窜改，因此将水印放在小波变换系数的最低位上不会影响算法的安全性。任何企图去除水印的处理都将被检测出来。另外，试图修改小波变换系数的高位而使它们有相同的哈希输出值是很困难的。攻击128位输出的 MD5的最好方法要大约$2^{128/2}$次搜索[15]，这在计算上是困难的。这一点使本文算法的安全性要高于引言中提到的一般认证算法。

2.4 图像分块、认证链和矢量量化攻击

1. 图像分块的原则 一般的数字签名的方法不能定位窜改的位置。为了能够定位窜改，本文采用了将图像分块和逐块进行水印嵌入和提取的方法。块尺寸的选择考虑了以下的因素：

a. 定位精度。块的尺寸越大，对窜改的定位精度越低。

b. 安全性。如果块的尺寸过小，可能嵌入不下全部的哈希函数的输出值(MD5是128位)。而只嵌入一部分哈希值将为使用穷尽搜索的伪造水印提供方便，降低算法的安全性。

c. 图像质量。每块中水印的生成和嵌入应该不影响图像的质量。

综合考虑这些因素，本文算法中采用了16×16的图像块。这样可以嵌入 MD5的全部128位，同时又有较好的定位精度和安全性。

2. 矢量量化攻击和认证链 如前所述，采用分块算法便于定位窜改的位置，同时也具有便于处理、计算开销小等优点。但这种逐块独立地进行水印嵌入和检测的算法很容易受到“矢量量化攻击”。矢量量化攻击是由 Holliman 和 Memon 提出的针对逐块独立一类水印算法的一种有效攻击方法[17]。下面先简单介绍一下这种攻击的方法，然后分析本文提出的对抗措施。

对于某个密钥 k，若从两个图像块 Y_i 和 Y_j 中用密钥 k 抽取出的水印相同，则称 Y_i 和 Y_j 是密钥 k 等价的。对于一幅或多幅加入水印的图像，密钥 k 可以把所有的图像块划分为密

钥 k 等价的一些等价类 $\{C_1, C_2, \cdots, C_m\}$，其中 m 是所有可能的等价水印数。Holliman 和 Memon 的攻击方法正是利用了密钥 k 等价类的图像块间的替代仍然会抽出相同水印的性质。对于一个加印的图像块 X_i，攻击者可以用一个密码 k 等价的图像块 Y_i 去替换它。另外，攻击者也可以利用密钥 k 等价的性质，把一幅含有水印的图像 X' 中的水印 W，拷贝到另一幅未加水印的图像 Y 中，形成有伪造水印 W 的图像 Y'。为了改善伪造水印后的图像质量，可以采用"矢量量化"的方法，即挑选或构造一个图像块 Y_i，它属于所需要的等价类 C_i，而和原图像块 Y_i 最接近。Y_i 可以从 X' 的等价类图像块 C_i 即码本(codebook)中挑选。这就是这种伪造水印的方法称为"矢量量化攻击"的原因。

由于本文的算法是一种分块的公钥水印算法，攻击者可以利用公钥找到等价类，因此必须研究对抗矢量量化攻击的方法。

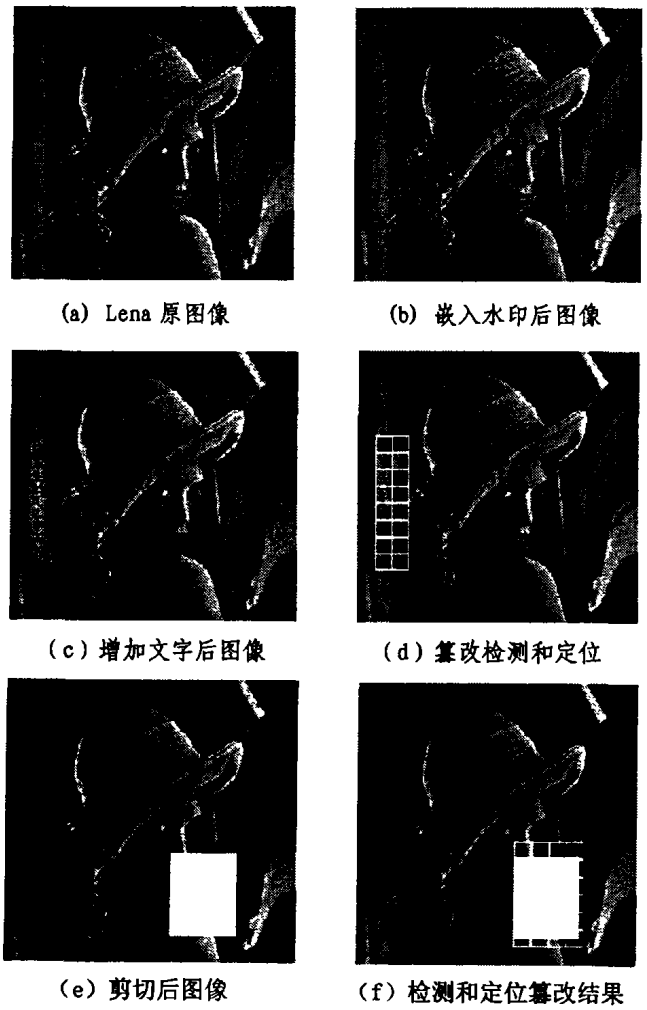
对抗矢量量化攻击的措施可以从以下两方面来考虑：

1)取消逐块算法中各块间的独立性，使得每块中水印的产生不仅和本块有关，而且依赖于其它的图像块。

2)尽量增加等价类的数目以及减少每个等价类"码本"中图像的数目，使攻击者很难在码本中找到相匹配的图像块。

本文采用的认证链方法在产生水印时，不仅和正在处理的块有关，而且依赖于下一个图像块。这使得图像块的"剪切和替换"变得更加困难。另外，本文算法的水印是基于各块图像内容的，而且图像块的尺寸较大(16×16)，这使得每一等价类中码本的数量会大大减少，因此很难找到足够匹配的图像块去作替换而不降低图像质量和不产生痕迹。

3 实验结果



(a) Lena 原图像　(b) 嵌入水印后图像

(c) 增加文字后图像　(d) 篡改检测和定位

(e) 剪切后图像　(f) 检测和定位篡改结果

图3 实验结果

我们用各种不同性质的图像进行了实验，下面是部分实验结果。

图3(a)是256×256的8比特的 Lena 原图像，图3(b)是嵌入水印后的图像。从中很难看出和原图有什么变化。图3(c)是添加文字窜改后的图像，图3(d)是对窜改的检测和定位结果，图3(e)和(f)分别是对原图像的剪切以及检测和定位窜改的结果。这些实验表明，本文的水印算法有较好的水印不可见性和检测及定位窜改的能力。

小结 本文提出了一种基于公钥和脆弱水印的图像完整性认证的方法。该方法不但能检测和定位窜改，而且能同时验证图像的所有权。由于该算法建立在密码学的基础上，因此它是单向、不可逆的，未授权者很难伪造水印或修改原有的水印。它的安全性只依赖密码的秘密性而不是水印算法的不公开性。文中讨论了图像分块的原则，分析了针对逐块独立一类水印算法的"矢量量化攻击"，提出了认证链等对抗措施，极大地增加了这种攻击成功的难度。本文算法的思想也可以用对称密钥系统来实现。和公钥系统相比，对称密钥系统的计算量较小，但需要传递密钥，这在某些应用中是不可能或是不安全的。

图像完整性的认证是当前迫切需要解决的一个问题，本文的算法为实用的图像认证系统提供了一种可能的方法。

参 考 文 献

1 Lin E T, Delp E J. A review of fragile image watermarks. In: Proc. of the Multimedia and Security Workshop(ACM Multimedia'99), Orlando, 1999. 25～29

2 Walton S. Information authentication for a slippery new age. Dr. Dobbs Journal, 1995, 20(4): 18～26

3 Yeung M, Mintzer F. Invisible watermarking for image verification. Journal of Electronic Imaging, 1998, 7(3): 578～591

4 Wolfgang R B, Delp E J. A watermark for digital images. In: Proc. IEEE Int. Conf. on Image Processing, 1996, 3: 219～222

5 van Schynael R, Tirkel A, Osborce C. A digital watermark. In: Proc. of the IEEE ICIP, Austin, Texas, Nov. 1994, 2: 86～90

6 Wolfgang R B, Delp E J. Fragile watermarking using the VW2D watermark. Proc. SPIE, Security and Watermarking of Multimedia Contents, pp. 204-213, San Jose, California, Jan 25-27, 1999

7 Wong P W. A watermark for image integrity and ownership verification. In: Proc. IS&T PIC Conf. Porland, OR, May 1998

8 Wong P W. A public key watermark for image verification and authentication. In: Proc. ICIP, Chicago, IL, Oct. 1998

9 Wong P W, Memon N. Secret and public key image watermarking schemes for image authentication and ownership verification. IEEE Trans on Image Processing, 2001, 10(10): 1593～1601

10 Wu M, Liu B. Watermarking for image authentication. In: Proc. ICIP, Chicago, IL, Oct. 1998

11 Kundur D, Hatzinakos D. Towards a telltale watermarking technique for tamper-proolfing. In: Proc. ICIP, Chicago, Illinois, Oct. 1998, 2: 409～413

12 Xie L, Arce G. Joint wavelet compression and authentication watermarking. In: Proc. IEEE ICIP, Chicago, Illinois, Oct. 1998, 2: 427～432

13 Calderbank R, Daubechies I, Sweldens W, Yeo B L. Wavelet transforms that map integers to integers. Appl. Comput. Harmon. Anal, 1998, 5(3): 332～369

14 Shapiro J M. Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans. on Signal Processing, 1993, 41(12)

15 Rivest R L. The MD5 message digest algorithm. Tech. Rep., 1992

16 Rivest R L, Shamir A, Adleman L. A method for obtaining digital signature and public-key cryptosystems. Commun. ACM, 1978, 21: 120～126

17 Holliman M, Memon N. Conterfeiting attacks on oblivious blockwise independent invisible watermarking schemes. IEEE Trans. Image Processing, 2000, 6(3): 432～441