基于数字水印的图像认证技术

吴金海 林福宗
(清华大学计算机科学与技术系智能技术与系统国家重点实验室 北京 100084)

摘 要 伴随着数字水印技术的发展,用来解决数字图像的真实性问题的图像认证技术在近年来发展迅速.它主要包括两大部分:篡改检测与篡改定位.有两种技术手段可供它使用:数字签名和数字水印.该文详细讨论了在设计基于数字水印的图像认证算法时常见的若干关键问题,阐述了基于数字水印的精确认证和模糊认证算法各自的发展过程及其国内外现状,并指出了将来继续努力的方向.

关键词 图像认证;篡改检测;篡改定位;数字水印;多媒体安全
中图法分类号 TP309

Image Authentication Based on Digital Watermarking

WU Jin-Hai LIN Fu-Zong
(*State Key Laboratory of Intelligent Technology & Systems, Department of Computer Science & Technology, Tsinghua University, Beijing* 100084)

Abstract With the development of digital watermarking technology, image authentication technology, which can be used to verify the integrity of a digital image, has been under active study in recent years. It mainly contains two parts: tamper detection and tamper localization, where the former aims at verifying the integrity of an image, while the latter is dedicated to localize maliciously tampered pixels if any. Both digital signature and digital watermarking can be utilized by image authentication. This paper focuses on image authentication based on digital watermarking technology. Several general key problems when designing such an algorithm are discussed first, including the allowable manipulations, the feature extraction, the precision of tamper localization, the restorableness of tampering, the influence of watermark embedding and the security of the system. Then, the literature and the state of the art of two classes of watermarking based image authentication algorithms (exact authentication and selective authentication) are introduced, respectively. As for exact authentication, there is no scheme which is secure enough and can localize the tampering to a single-pixel level up to now. The most important problem is to make a good tradeoff between the fine tamper localization and the system security. As to selective authentication, there are no satisfying schemes in the literature. The future direction is pointed out finally. The goal of exact authentication is to achieve finer tamper localization when keeping enough security, while the urgent affairs of selective authentication are to extract one kind of image feature which can describe the content of an image adequately.

Keywords image authentication; tamper detection; tamper localization; digital watermarking; multimedia security

1 引 言

现代数字技术的发展把人们带进了一个崭新的世界.对于多媒体数据而言,数字化后的媒体数据显然具有传统模拟时代无法比拟的许多优点,比如易于创作、易于存储、易于发布,同时也提高了媒体的欣赏质量.然而,任何事物都有正反两方面.数字媒体在带给人们方便的同时,也引入了一些潜在的风险,它们很容易被修改.虽然大多数情况下人们修改

收稿日期:2003-06-05;修改稿收到日期:2004-06-15.本课题得到国家自然科学基金(60135010)资助.**吴金海**,男,1979 年生,硕士,研究方向包括数字水印、图像认证、媒体安全等.E-mail: wu-jh01@mails.tsinghua.edu.cn.**林福宗**,男,教授,研究方向包括远程教育、机器学习、数字水印等.E-mail: linfz@mails.tsinghua.edu.cn.

文件都有合法的目的,但有些时候也有人不注意甚至是怀着恶意改变原来作品的内容,并造成严重的后果.以数字图像为例,对 X 光医学图像的一个不经意的修改可能会造成误诊;作为法庭证据的照片如果被恶意地修改后就可能扭曲法律事件的真实面貌,使好人蒙冤而坏人却逍遥法外;新闻发布网站需要确认所发布的图像是否属实,等等.在这些场合下,我们需要明确地知道图像是否被修改过,也就是说,我们需要对图像的真实性(也称完整性)进行验证.在下文中,如果不作特殊说明,所提到的图像都是指以二进制形式保存的数字图像,而不包括传统意义下的印刷型图像.

传统的数据认证技术是数据安全领域的一种重要技术,用来验证数据的真实性,已经发展得非常成熟.它把所有的数据当作二进制比特流,然后计算该比特流的哈希散列值并产生消息认证码,或者用非对称密码加密哈希值而产生数字签名,最后把它附加在原来消息的末尾并一起传送出去.在传输过程中,任何比特的改变都可能导致认证失败.

数字图像同样是以二进制形式保存的数据,因此也能用传统的数据认证技术来解决其真实性问题.然而,把图像当作二进制数据来处理仅仅是计算机的特长,除了那些进行图像压缩等类研究的人员之外,普通人更关心的是一幅图像所表达的意义.对同一幅图像而言,不同的人从不同的角度去理解就会得出不同的解释.人们往往只关心自己感兴趣的那部分内容.因此,人们也经常使用各种图像处理方法来获取自己所需要的信息.比如,未经过压缩的图像通常具有巨大的数据冗余,人们在传输或保存之前一般都会选取适当的压缩方法进行压缩.又如,人们可能需要在不同的图像文件格式之间进行转换.这些操作无疑会改变图像的二进制比特流,却不影响人们对图像内容的认识.

传统的数据认证技术在适应这些图像处理需求方面显得无能为力.一旦为图像的二进制比特流产生了消息认证码(或数字签名),任何图像处理操作所引起的二进制比特流的变化都可能导致认证失败,这是人们不希望看到的.而且,在认证失败的情况下,这些数据只是被当作"没用"的东西而被抛弃,但实际上导致认证失败的可能只有一小部分数据.如果能找到被修改的数据的位置,那么无疑保留了那些未被修改的数据的价值.此外,传统的数据认证技术只能把所产生的消息认证码(或数字签名)附加在数据的末尾,增加了数据量.而蓬勃发展起来的数字水印技术充分利用图像的冗余和人类的视觉感知特性,能把信息在不被知觉的情况下隐藏到图像中,无疑为解决图像的真实性问题注入了新鲜的血液.数字图像认证技术就是为了完成这些使命而诞生的,下文中把它简称为图像认证.

下文首先简要描述图像认证要解决的问题,接着介绍两种常用的图像认证方法,然后详细讨论设计基于数字水印的图像认证算法时经常出现的若干关键问题,随后介绍基于数字水印的精确认证和模糊认证算法的发展过程及其国内外现状.最后,针对图像认证技术目前仍然存在的问题,阐述将来进一步努力的方向.

2 图像认证问题描述

图像认证技术是对数字图像的感知内容进行认证的一门技术.它要回答以下两个问题:

(1) 图像是否真实,也就是图像是否被恶意篡改?

(2) 如果图像不真实,那么哪些地方不真实?

回答问题(1)的方法常被称为"篡改检测",而回答问题(2)的方法常被称为"篡改定位".关于图像认证的概括性描述,可以参考文献[1~4].

以一个例子来说明图像认证所做的事情,如图 1 所示.其中,图 1(a)表示现实世界中的一幅图

(a) 真实图像 (b) 篡改后的图像 (c) 篡改定位结果

图 1 图像认证示例

像，它是清华大学校园内的一个标志性建筑物，镌刻着被许多清华学子铭记于心的“行胜于言”四个大字. 图 1(b)描绘了与图 1(a)类似的场景，只是“行胜于言”已经变为“言胜于行”. 如果图 1(b)没有和图 1(a)放在一起，而是单独置于某个网页上，那么未曾到过清华大学的人无法根据肉眼辨认出图像的真假，尤其当小孩子看到这幅图时，或许会把“言胜于行”记在心上，这无疑将对社会风气造成不良的影响. 图像认证就是要在图像发布之前对图像进行适当的处理，使得将来在没有原始真实图像的情况下能够判断一幅新图像的真实性. 而且如果发现图像不真实，还能把被篡改的位置找出来. 图 1(c)中的白色部分表示在没有图 1(a)的情况下找出的由图 1(a)变成图 1(b)所修改的位置.

3　图像认证方法

要解决数字图像的真实性问题，一般有两种方法：基于数字签名的方法和基于数字水印的方法. 它们的主要差别在于：基于数字签名的方法把认证信息与原始图像隔离开，而基于数字水印的方法把认证信息嵌入原始图像.

3.1　基于数字签名的图像认证

Friedman 最早提出“可靠的照相机”的概念，它使用传统的数据认证技术实现图像认证[5]. 然而，传统的数据认证技术只能进行比特流的认证，无法考虑到图像的感知内容. 因此需要把传统数据认证的消息摘要模块进行适当的修改，才能用来实现图像认证. Schneider 和 Chang 最早提出一种可行的基于数字签名的图像认证技术的框架[6].

在发送方签名的过程如下：

(1)从需要签名的图像中提取表达图像内容的特征.

(2)根据提取出的特征数据产生特征摘要.

(3)用密钥对特征摘要进行加密，得到签名信息.

(4)保存签名信息.

在接收方验证签名的过程如下：

(1)用与发送方签名时一样的方法，从待验证图像中产生特征摘要.

(2)提取出发送方保存的签名信息，并用发送方提供的密钥进行解密.

(3)用一定的度量方法，比较两个特征摘要，并给出认证结果.

到目前为止，有许多基于数字签名的图像认证系统. 它们的主要区别就是提取出来的用来表达图像内容的特征不同. 这些特征包括灰度直方图[6]、边缘[7]、特征点[8]、图像灰度低阶矩[9]、块灰度均值[10]，DCT 变换系数[11]、DWT 变换系数等等.

产生签名信息时既可以用对称加密算法，也可以用非对称加密算法.

签名信息一般有两种保存方式：(1)直接把签名信息放在图像文件的开头部分，验证时直接从文件头提取签名信息. 然而这样保存的签名信息如果经过文件格式转换就容易被去除；(2)把签名信息独立保存在第三方数据库里，验证时从数据库中提取出签名信息. 相比起来，第 2 种方法更加可靠一点，但是实现起来稍微复杂.

3.2　基于数字水印的图像认证

数字水印作为一种信息隐藏技术，在近几年来得到广泛的研究和发展[12,13]. 目前对数字水印算法有各种各样的分类标准. 如果把鲁棒性作为分类标准，目前有 3 种水印类型：(1)鲁棒水印；(2)易碎水印；(3)半易碎水印. 它们的主要区别在于水印信号经过各种信号处理之后能否被检测出来. 顾名思义，鲁棒水印的目标就是使水印信号在经过各种各样的信号处理之后仍能被清楚地检测出来；易碎水印的目标正好与之相反，那就是只要图像信号被处理过，水印信号就无法恢复原样；而半易碎水印的目标正好介于两者之间，只有少数允许的处理操作不会改变水印信号的原貌.

数字水印技术的应用前景非常广阔，包括版权保护、完整性认证(篡改证明)、数字指纹、广播监测、拷贝控制等等. 到目前为止，版权保护和完整性认证是水印技术的两种主流应用. 但是不管从设计者的角度还是从攻击者的角度来看，这两种水印算法都是不同的. 从设计者的角度来说，版权保护的主要目的就是使版权信息不能被盗版者轻易去除，而这正是完整性认证所不需要的. 在用于完整性认证的水印技术中，更需要的是从检测出的被修改过的水印信号中，判断出被修改的地方以及被修改的严重程度. 从攻击者的角度来说，攻击版权保护的水印系统的主要目标是去掉原始的版权信息，甚至再加上自己的伪装的版权信息. 而完整性认证水印系统的攻击者的主要目的显然不在于此，而是在保证原来的水印信号不受影响的情况下，尽可能地修改图像的内容，达到欺骗认证系统的目的. 所以，对设计者而言，只能用易碎水印或半易碎水印来实现完整性认证；对攻击者而言，就只好使用伪造攻击.

一个数字水印系统一般包括两大模块：水印嵌入模块和水印检测模块. 水印嵌入模块把水印信号嵌入到原始图像中，得到嵌入水印后的图像；水印检测模块从待检测图像中提取出原先嵌入的水印信号

并做进一步判断.它们的具体实现过程根据水印技术应用场合的不同而有所区别.用于图像完整性认证的数字水印系统中,水印嵌入过程和水印检测过程分别如图 2 和图 3 所示.

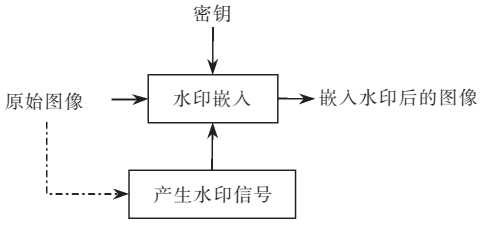

图 2 用于图像认证的水印嵌入过程

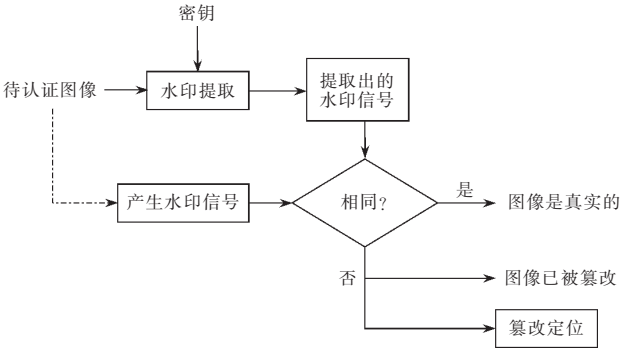

图 3 用于图像认证的水印检测过程

需要说明的是,图 2 和图 3 中的一条虚线表示所产生的水印信号不一定和图像内容有关.因此,在用于完整性认证的水印系统中,有两种不同的水印信号:一种是表明发送方身份的标识,比如一个图标[14]或者一个只有发送方才知道的伪随机序列码[15];另一种就是从图像中提取出的表达图像内容的特征,这些特征和上述的基于数字签名的认证系统中的特征集合类似,比如边缘信息[16]、模式块映射[17]、块均值[18]、DCT 变换系数[19]、DWT 变换系数[20]、校验和[21]等等.水印信号的不同也决定了签名的验证方式的不同.一般说来,用发送方身份标识作为水印信号,在验证时必须掌握水印信号的构成的先验知识;而用表达内容的特征作为水印信号时,只需要知道水印的提取方法,经过计算比较两者是否相同就能得出判断结果.

3.3 两种认证方法比较

以上讨论的两种认证方法各有优缺点.对基于数字水印的认证技术而言,主要有 3 个好处:(1)把认证信息作为水印信号嵌入原始图像信号中,不需要额外的数据管理;(2)在对图像信号进行修改时,水印信号会经历和原始图像信号一样的变化过程,有助于利用检测出的水印信号判断篡改的类型;(3)可以比较精细地定位篡改.当然,基于数字水印的认证技术也有它的不足之处,主要表现在:(1)水印的嵌入会对原始图像信号产生一定的影响,而且许多时候这种影响是无法恢复的;(2)原始图像必须有足够的冗余空间用来嵌入水印信号,因此对于压缩过的数据源来说,使用数字水印技术进行认证显得有些困难;(3)在获得较为精细的篡改定位的同时,可能牺牲了系统的安全.基于数字水印的图像认证技术的这些不足恰恰是基于数字签名的认证技术中不可能有的.数字水印的好处是否胜过它的不足,应根据不同的应用环境做出判断.

4 基于数字水印的图像认证技术研究的关键问题

设计基于数字水印的图像认证算法需要考虑各种各样的因素,其中最关键的包括 6 个方面:对图像的哪些修改是允许的? 该提取图像的哪些特征作为认证的判断依据? 怎么找到图像中被修改的位置? 如果图像已经被修改了,能否恢复出被修改部分的原来面目? 如何减轻水印信号的嵌入对图像质量的影响? 系统是否存在安全漏洞?

4.1 允许的修改

一般说来,对图像的操作可以分为两类:第 1 类是允许的操作,常见的图像处理方法比如高质量压缩、文件格式转换、数/模或模/数转换等都可以归入这一类;第 2 类是不允许的操作,也称恶意篡改,比如剪切-替换操作等.对图像认证而言,第 1 类操作一般不会产生认证失败,而第 2 类操作是图像认证系统需要检测的主要目标.但是上述对图像处理方法的归类并不适用于所有的图像认证系统.一般说来,哪种操作可以被系统所接受依赖于实际的应用环境.比如,对高质量的 JPEG 压缩而言,在一般情况下可以归入第 1 类操作,是系统能够容忍的修改.然而对医疗图像而言,任何轻微的变化(包括 JPEG 压缩在内)都可能引起对图像内容的不同解释,这时 JPEG 压缩就可能被归入第 2 类操作,而成为认证系统需要检测的目标之一.

根据是否容忍图像在一定程度下被修改,可以把现有的图像认证系统分为两类:第 1 类是不允许有任何的修改,称为精确认证,有时可能称为硬认证[1,2]或完全认证[3];第 2 类是允许一定程度的修改,比如 JPEG 压缩等,称为模糊认证,有时可能称为软认证[1,2]或选择认证[12].这两类认证都可以用基于数字签名的方法和基于数字水印的方法.当使用基于数字水印的方法时,精确认证常使用易碎水印,而模糊认证常使用半易碎水印.

4.2 特征提取

确定了认证系统允许的操作以后,必须以一定

的方式来表达不同的修改造成的不同的结果，达到区分可允许操作和恶意篡改的目的. 这就是特征提取的作用. 所选取的特征要满足两个要求：(1)如果对图像的操作在允许的范围之内，提取的特征应该保持不变或者变化量是可以接受的；(2)这些特征对恶意篡改足够敏感.

当不允许对图像有任何修改也就是需要精确认证时，往往不需要特征提取这一步，或者说图像的原始像素灰度值就是最好的特征. 当允许对图像有一定程度的修改也就是说需要模糊认证的时候，许多学者试验了不同的图像低层表征，比如边缘信息、灰度直方图、图像灰度低阶矩、特征点、DCT 变换系数、DWT 变换系数等等. 这些特征都只能从不同的侧面表达图像的内容，并不能适用于所有的应用环境. 只有当特定的应用环境所关心的图像内容恰恰是所提取出的图像低层特征时，这些特征对认证系统才有意义. 另外，当这些特征的数据量不大时，通常也可以用于基于数字签名的认证系统.

4.3　篡改定位能力

目前的许多基于数字水印的图像认证系统都能够找到图像被修改的地方，这个特性常被称为篡改定位特性. 之所以要使图像认证系统具有篡改定位特性，是因为这些信息可以用来推断：(1)篡改的动机；(2)可能的真凶；(3)篡改的严重程度. 因此，图像认证系统的篡改定位能力也是设计者经常需要考虑的一个问题.

实现篡改定位的一个重要手段就是把图像分割为各个大小相同的独立的基本处理单元，然后对每个单元进行相同或类似的处理. 迄今为止，图像认证系统的定位能力可以分为 3 种水平：一种是像素级的定位能力，即可以确定每一个像素是否被篡改过，这种认证也称为单像素认证，比如文献[14]等；另一种是分块级的定位能力，即未被篡改的最小单元是一个图像块，这种认证也称为分块认证，比如文献[22]等；还有一种是没有任何篡改定位能力，这种认证称为无分割认证. 随着定位能力的不同，可能使用的认证技术也不同. 一般来说，单像素认证系统不使用数字签名技术.

4.4　篡改可恢复性

一些设计者并不满足于仅仅知道图像是否被修改和哪里被修改，他们追求的另一个目标是如果知道图像被修改的位置，如何去恢复它们被修改前的真正面目. 这个特性被称为篡改可恢复性. 一般说来，有两种恢复策略：一种是精确恢复，也就是能够恢复成和原来一模一样的效果；另一种是模糊恢复，也就是能把被修改的地方恢复成和原来的内容相差不多. 对于图像认证来说，精确恢复并不是它的目标，取而代之的是模糊恢复，它容忍恢复后的图像和原来图像之间存在一定程度的差别. 只要这种差别不影响对图像的重要内容的解释，这种恢复就是有意义的.

在现有的文献中，共有 3 种模糊恢复方法：第 1 种方法是在原来的图像中嵌入冗余信息，比如嵌入错误纠正码(Error Correction Code, ECC)[23]，在一定程度上可以达到恢复的功能；第 2 种方法是在图像中嵌入原图像的低分辨率版本，达到自嵌入的效果[24]；第 3 种方法是盲恢复[25]，也就是先判断出对图像的修改类型，并估计适当的参数，然后进行逆向操作. 当然，只有对图像的修改操作是可逆的，盲恢复才能起作用.

4.5　水印嵌入的影响

由于常用的水印方法会使原始图像产生或多或少的失真，通常情况下假设对图像内容的这种影响可以忽略不计. 然而，对于精确认证而言，如果能够消除水印嵌入带来的失真，无疑更有意义. 因此，有些学者提出可擦除水印技术(也称可逆水印)，使得接收方在判断出图像未被修改的情况下可以去掉水印信号，恢复出图像的原来面目. Honsinger 等提出用模运算法则实现水印可逆嵌入与提取[26]，Fridrich 等分别提出两种可逆水印技术[27]：一种是用无损压缩的方法，获得冗余空间，以供嵌入签名信息；另一种方法是基于 JPEG 压缩量化表的方法，把量化表系数分别减小一半，使得量化系数都变成偶数，留下的全 0 最低有效位作为水印嵌入空间. 此外，还有使用整数小波变换技术实现可逆水印[28].

另外，水印嵌入还要考虑和图像文件格式的兼容问题：一是重新保存为需要的图像文件格式是否会破坏已经嵌入的水印信号；另一是如果水印信号不会被破坏，那么对图像质量的影响是否不可忽略，尤其是压缩后的图像.

4.6　安全性

一个认证系统能否投入实际的使用，最关键的因素就是系统的安全性. 对图像认证系统而言，除了传统密码学中所存在的安全隐患之外，影响系统的安全性还有两个方面：(1)篡改定位特性带来的安全漏洞；(2)模糊认证中提取出的表达图像内容的低层特征不能充分表达图像内容引起的安全漏洞. 恶意的攻击者可能利用这些漏洞来修改或伪造真实图像而不被认证算法发现，达到欺骗认证系统的目的.

目前，针对图像认证系统的攻击主要有 4 种方法：

(1)量化攻击[29,30]；

(2)密码分析攻击[30]；

(3)黑盒攻击[31]；

(4)特征选取攻击[32].

量化攻击的前提条件是图像中每一个认证单元(比如一个图像块或图像的一个像素)所嵌入的水印信号与其它认证单元的内容无关.于是,只要在两个认证单元中嵌入的水印信号相同,就可以把它们互相替换来修改图像内容而不会导致认证失败,这就是量化攻击.为了使攻击后得到的图像具有一定的意义和较好的质量,有时需要多幅图像,尤其是攻击彩色图像时.防止这种攻击的最有效的方法就是使每个认证单元的认证信息依赖于其它认证单元的内容,这种改进方法的一个缺点就是篡改定位能力将受到影响.

密码分析攻击的前提条件是攻击者拥有几幅用同样密钥嵌入相同水印信号的图像.其目的是找出图像认证算法中使用的秘密信息,比如密钥等.一种可以用来防止密码分析攻击的方法就是每幅图像使用不同的密钥嵌入水印;另一种方法是使图像内的每一个认证单元所嵌入的水印信号依赖于其它认证单元的内容或者全局信息;有时可以把这两种方法结合起来.

黑盒攻击的前提条件是攻击者拥有一个"黑盒"验证器.黑盒有时也称为 Oracle,黑盒攻击也称为 Oracle 攻击.黑盒的输入是一幅待认证图像,输出是篡改检测的结果,即"成功"或"失败",有时还输出篡改定位结果.发动攻击时,需要不断地有计划地修改图像,然后把修改后的图像作为黑盒的一个输入,并观察认证结果.只要验证器的输出是"成功",就表明对图像的这次修改不会被检测到.经过多次成功修改的组合,就可能推断出水印嵌入所用的密钥或其它秘密信息,从而可以随心所欲地修改图像而不引起认证失败.一种可以有效阻止黑盒攻击的方法就是限制对验证器的访问频率,比如一分钟内相同的人只能验证一次,验证一定次数之后自动阻止相同人员对验证器的访问等等.

特征选取攻击的前提条件是设计者所选取的用来表达图像内容的特征并不能充分表达图像的全部内容,以致无法根据这些特征来区分可接受操作和恶意篡改.比如,当把图像边缘信息作为特征,就可以发动颜色攻击,保持图像的边缘不变.再比如,当把图像的灰度直方图作为特征,那么构造一个与原有图像具有相同的灰度直方图而内容不同的图像就是攻击者的目标之一.对于特征选取攻击,目前尚无很有效的方法.当然,选取什么特征作为认证信息,这也依赖于系统设计者在不同的应用环境下对图像内容的定义.如果说设计者所关心的就是图像的边缘信息不被修改,那么边缘信息就完全是唯一可以作为认证信息的特征,而不存在特征选取攻击的可能.

5 基于数字水印的图像认证技术的发展与研究现状

精确认证与模糊认证是图像认证的两大分支,前者不允许对图像内容进行任何操作,后者允许不改变图像感知内容的处理.这两种认证分别可以使用数字签名技术和数字水印技术,本节着重介绍基于数字水印的精确认证和基于数字水印的模糊认证的发展和研究现状.

5.1 基于数字水印的精确认证的发展过程

1995 年,Walton 首次提出用易碎数字水印的方法实现图像认证[21],其主要思想是首先随机选择一些像素,然后计算它们的灰度值中除了最低有效位(Least Significant Bit,LSB)之外的其它有效位的校验和,并作为水印信号嵌入到其 LSB 中.但是,使用校验和的方法不够安全,而且该方法的篡改定位能力比较差.

1997 年,Yeung 和 Minzter 提出一种基于易碎数字水印的单像素认证算法来实现精确认证[14,33].该算法通过适当修改每个像素的灰度值而把 1 比特水印信号嵌入其中,从而可以把篡改定位精确到一个像素.1998 年,Wu 等人把 Yeung-Mintzer 的方法从空间域推广到频率域[34],使算法能和 JPEG 图像文件格式相兼容.此后,Memon 和 Fridrich 等人发现 Yeung-Mintzer 方法存在安全漏洞,并不断提出改进方法[29,30,35~37].Li, Zhong 和 Lu 等人也相继提出各自的改进算法[38~40].所有这些在 Yeung-Mintzer 方法的基础上发展起来的算法都使用易碎水印技术,而且图像的每个像素都携带一个比特的水印信号.然而,Fridrich 等指出这类水印算法天生缺乏安全的保证,而且不管如何改进,只要仍然保持这种顺序处理每个像素的特性,即依次在每个像素嵌入一个比特水印信号,那么总可以发动 Oracle 攻击来伪造一幅真实的图像[41].Wu 等人最近提出一种基于易碎水印的把篡改检测与篡改定位相分离的单像素认证,可以抵抗已知的各种攻击[42],包括 Oracle 攻击.

1998 年,Wong 等人提出两个类似的基于易碎水印的分块认证算法[22,43~45].唯一的区别就是其中之一的水印信号使用加密的消息认证码,而另一个使用数字签名.算法的主要思想是把图像分割为各个独立的小块,然后分别在各小块上嵌入各自的水印.Holliman 和 Memon 指出这类分块独立算法存在致命的缺陷,并提出伪造真实图像的量化攻击,他

们同时指出消除分块独立性可以阻止量化攻击[29]. Celik 等人提出一种分层的分块认证算法，消除了 Wong 算法的分块独立性[46,47]. Fridrich 等人采用分块编号和图像唯一索引来消除分块独立性，其效果比 Celik 的方法更好，篡改定位能力更强一些[41]. 然而，所有这些基于易碎水印的分块认证算法的共同特点是只能把篡改定位精确到图像分块上. 它和基于易碎数字水印的单像素认证算法相比，安全性较高，但篡改定位能力被削弱了许多.

5.2　基于数字水印的模糊认证的发展过程

van Schyndel 等人首次提出数字水印的概念[48]，并把一个扩展的 m-随机序列加到图像的 LSBs 上以隐藏信息. 该 m-序列由某个密钥控制产生，最后利用相关检测方法检查水印信号是否存在. Wolfgang 和 Delp 通过嵌入一个双向 m-序列推广了 Van Schyndel 的算法，增加算法抗修改的鲁棒性，提高篡改定位能力[49,50]. 然而他们的算法在认证时需要原始的未被篡改的真实图像，显得不实用.

Zhu，Swanson 和 Tewfik 三人首次提出利用在某个向量空间进行量化可以实现水印嵌入，且能用于模糊认证[15,51]. 他们的量化算法既可以用在空间域，也可以用在 DCT 频率域. 量化步长由空间域(或频率域)的掩蔽数值决定，从而使水印的嵌入不会导致图像质量的下降. Kundar 等人提出一种在小波变换域(DWT)利用量化进行水印嵌入的算法[20,52]. 该算法可以在空间域和频率域找出篡改的位置，并估计当前图像被篡改的程度. 然而，利用量化实现水印嵌入具有天生的弱点，只要知道量化步长就能很容易地改变图像内容而使提取出的水印信号仍然不变. Fridrich 等人提出一种利用向某些随机向量投影的方法产生图像特征并作为水印信号且嵌入在图像中的认证算法[17]. Li 等人提出用图像边缘信息作为图像特征并嵌入到图像中[53]. Bassali 等人提出用量化后的块均值作为图像特征并嵌入到图像中[18].

Lin 和 Chang 等人把基于数字签名的 SARI 系统扩展成基于数字水印的系统[19]，提取的图像特征仍然是 DCT 系数之间的大小关系，水印信号是根据 JPEG 压缩中用不同步长进行量化所产生的不变性而嵌入的. Sun 和 Chang 等人把基于数字水印的 SARI 系统的设计方法扩展到 DWT 域，以适应 JPEG2000 图像压缩标准[54,55].

5.3　国内外现状小结

对于精确认证而言，算法有两个目标：(1)篡改检测过程希望能检测到对图像的任何修改；(2)篡改定位过程希望能把篡改定位精细到一个像素. 由于精细的篡改定位与系统的安全性之间存在矛盾，精确认证算法的设计思路主要有两种：一种是在保证系统具有合理安全性的前提下尽量提高篡改定位的精度；另一种是在保持精细的篡改定位的前提下试图提高系统的安全性. 迄今为止，Fridrich 等人提出的分块认证是比较安全可靠的一种精确认证算法，它可以把篡改定位到大小为 128 个像素的子块上[41]. Wu 等人最近提出的把篡改检测与篡改定位相分离的易碎水印算法可以抵抗目前已知的各种攻击，且可以把篡改定位精细到一个像素，在迄今为止已出现的所有单像素认证算法中是最为安全的一种[42].

与精确认证不同，模糊认证的重点不再是精细的篡改定位能力. 许多模糊认证算法的主要区别在于各自选取的表达图像内容的低层特征的不同. 由于这些特征无法充分表达图像内容，修改图像的内容而保持这些特征不变是常有的事，因此模糊认证算法到目前为止还不够成熟，还有很长的路要走.

就图像认证能处理的图像文件格式而言，目前多数的精确认证算法还只能处理无压缩的原始图像，而模糊认证则能处理多种文件格式的图像.

实际上，还有一些非主流的认证算法，比如可逆水印认证、可恢复认证等等. 目前的可逆水印认证算法是无分割认证，不具有篡改定位能力. 可恢复认证在近年来不是很受重视.

6　总结与展望

数字图像认证技术用来解决数字图像的真实性问题，主要包括两个部分：篡改检测与篡改定位. 根据篡改检测类型分，图像认证技术可以分为精确认证与模糊认证；根据篡改定位能力分，可以分为单像素认证、分块认证和无分割认证. 图像认证主要有两种方法：基于数字签名的认证和基于数字水印的认证. 本文主要讨论了基于数字水印的图像认证技术.

精确认证要解决的主要问题是篡改定位与系统安全之间的矛盾，理想的目标是提出一种既安全可靠又能把篡改定位精细到一个像素的算法，这个目标尚未完全实现. 此外，还应该扩展精确认证算法，使它与各种图像文件格式兼容. 此时由于文件格式本身的原因，可能无法把篡改定位精细到一个像素，但是也要尽量提高篡改定位的精度，同时减少图像质量的失真.

对于模糊认证而言，当务之急是寻找一种或多种能充分表达图像内容的特征. 另一种思路是分析各种图像处理操作的本质，从而达到把某种操作与其它操作区别出来的目的. 尤其是随着 JPEG2000 图像压缩标准的成熟，研究能够抵抗 JPEG 及 JPEG2000 压缩

的模糊认证算法具有重大的现实意义和广阔的市场前景.提高模糊认证的篡改定位能力不是当务之急.

精确认证与模糊认证有各自适合的市场前景和需求,不能简单地说孰优孰劣.而且,在精确认证研究中遇到的问题对于将来提高模糊认证的篡改定位能力也是有帮助的.

参 考 文 献

1 Zhu B. B., Swanson M. D., Tewfik A. H.. When Seeing Isn't Believing. IEEE Signal Processing Magazine, 2004, 21(2): 40～49

2 Zhu B. B., Swanson M. D.. Multimedia authentication and watermarking. In: Feng D., Siu W. C., Zhang H. eds. Multimedia Information Retrieval and Management, Springer-Verlag, 2003, Ch. 7, 148～177

3 Lin Ching-Yung. Watermarking and digital signature techniques for multimedia authentication and copyright protection [Ph. D. dissertation]. Columbia University, New York, 2000

4 Albanesi M. G., Ferretti M., Guerrini F.. A taxonomy for image authentication techniques and its application to the current state of the art. In: Proceedings of the 11th International Conference of Image Analysis, Palermo, Italy, 2001, 535～540

5 Friedman G. L.. The trustworthy digital camera: Restoring credibility to the photographic image. IEEE Transactions on Consumer Electronics, 1993, 39(4): 905～910

6 Schneider M., Chang S.-F.. A robust content based digital signature for image authentication. In: Proceedings of the IEEE International Conference on Image Processing, Lausane, Switzerland, 1996, 3: 227～230

7 Queluz M. P.. Towards robust, content based techniques for image authentication. In: Proceedings of the IEEE 2nd Workshop on Multimedia Signal Processing, Los Angeles, 1998, 297～302

8 Bhattacharjee S., Kutter M.. Compression tolerant image authentication. In: Proceedings of the IEEE International Conference on Image Processing, Chicago, USA, 1998, 1: 435～439

9 Queluz M. P.. Content based integrity protection of digital images. In: Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents, San Jose, USA, 1999, 3657: 85～93

10 Lou D. C., Liu J. L.. Fault resilient and compression tolerant digital signature for image authentication. IEEE Transactions on Consumer Electronics, 2000, 46(1): 31～39

11 Lin C.-Y., Chang S.-F.. A robust image authentication method distinguishing JPEG compression from malicious manipulation. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(2): 153～168

12 Cox I. J., Miller M. L., Bloom J. A.. Digital Watermarking. New York: Academic Press, 2002

13 Cox I. J., Miller M. L., Bloom J. A. 著. 王 颖, 黄志蓓等译. 数字水印. 北京:电子工业出版社, 2003

14 Yeung M., Mintzer F.. An invisible watermarking technique for image verification. In: Proceedings of the IEEE International Conference on Image Processing, Santa Barbara, USA, 1997, 2: 680～683

15 Zhu B., Swanson M. D., Tewfik A. H.. Transparent robust authentication and distortion measurement technique for images. In: Proceedings of the 7th IEEE Digital Signal Processing Workshop, Loen, Norway, 1996, 45～48

16 Dittmann J., Steinmetz A., Steinmetz R.. Content-based digital signature for motion pictures authentication and content-fragile watermarking. In: Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, 1999, 2: 209～213

17 Fridrich J.. Image watermarking for tamper detection. In: Proceedings of the IEEE International Conference on Image Processing, Chicago, USA, 1998, 2: 404～408

18 Bassali H., Chhugani J., Agarwal S., Aggarwal A., Dubey P.. Compression tolerant watermarking for image verification. In: Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, 2000, 1: 430～433

19 Lin C.-Y., Chang S.-F.. Semi-fragile watermarking for authenticating JPEG visual content. In: Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents II, San Jose, USA, 2000, 3971: 140～151

20 Kundur D., Hatzinakos D.. Towards a telltale watermarking technique for tamper proofing. In: Proceedings of the IEEE International Conference on Image Processing, Chicago, USA, 1998, 2: 409～413

21 Walton S.. Image authentication for a slippery new age. Dr. Dobb's Journal, 1995, 20(4): 18～26

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

23 Lee J., Won C. S.. A watermarking sequence using parities of error control coding for image authentication and correction. IEEE Transactions on Consumer Electronics, 2000, 46(2): 313～317

24 Fridrich J., Goljan M.. Images with self-correcting capabilities. In: Proceedings of the IEEE International Conference on Image Processing, Kobe, Japan, 1999, 3: 792～796

25 Kundur D., Hatzinakos D.. Semi-blind image restoration based on telltale watermarking. In: Proceedings of the 32nd Asilomar Conference on Signals, Systems and, Computers, Pacific Grove, California, 1998, 2: 933～937

26 Honsinger C. W., Jones P., Rabbani M., Stoffel J. C.. Lossless recovery of an original image containing embedded data. U. S. Patent No. 6,278,791, 2001

27 Fridrich J., Goljan M., Du R.. Lossless data embedding for all image formats. In: Proceedings of the SPIE Photonics West, Vol. 4675, Electronic Imaging 2002, Security and Watermarking of Multimedia Contents IV, San Jose, California, 2002, 572～583

28 Tian J.. Wavelet-based reversible watermarking for authentication. In: Proceedings of the SPIE Photonics West, Vol. 4675, Electronic Imaging 2002, Security and Watermarking of Multimedia Contents IV, San Jose, California, 2002, 679～690

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

30 Fridrich J., Goljan M., Memon N.. Cryptanalysis of the Yeung-Mintzer fragile watermarking technique. Journal of Electronic Imaging, 2002, 11(2): 262～274

31 Wu J., Zhu B., Li S., Lin F.. Efficient oracle attacks on Yeung-Mintzer and variant authentication schemes. In: Proceedings of the IEEE International Conference on Multimedia & Expo (ICME'04), Taiwan, 2004

32 Wu J., Zhu B., Li S., Lin F.. New attacks on SARI image authentication system. In: Proceedings of the SPIE, Vol. 5306, Security and Watermarking of Multimedia Contents VI, San

Jose, California, USA, 2004, 602～609
33 Yeung M., Mintzer F.. Invisible watermarking for image verification. Journal of Electronic Imaging, 1998, 7(3): 578～591
34 Wu M., Liu B.. Watermarking for image authentication. In: Proceedings of the IEEE International Conference on Image Processing, Chicago, USA, 1998, 2: 437～441
35 Memon N., Shende S., Wong P.. On the security of the Yeung-Mintzer authentication watermark. In: Proceedings of the IS&T PICS Symposium, Savannah, Georgia, 1999, 301～306
36 Fridrich J., Goljan M., Memon N.. Further attacks on Yeung-Mintzer fragile watermarking scheme. In: Proceedings of the SPIE, Vol. 3971, Security and Watermarking of Multimedia Contents II, San Jose, CA, 2000, 428～437
37 Fridrich J., Goljan M., Baldoza A. C.. New fragile authentication watermark for images. In: Proceedings of the IEEE International Conference Image Processing, Vancouver, Canada, 2000, 1: 446～449
38 Li C. T., Yang F. M., Lee C. S.. Oblivious fragile watermarking scheme for image authentication. In: Proceedings of the IEEE International Conference Acoustics, Speech, & Signal Processing, Orlando, FL, USA, 2002, VI: 3445～3448
39 Zhong H., Liu F., Jiao L. C.. A new fragile watermarking technique for image authentication. In: Proceedings of the International Conference Signal Processing, Beijing, China, 2002, 1: 792～795
40 Lu H., Shen R., Chung F.. Fragile watermarking scheme for image authentication. Electronics Letters, 2003, 39(12): 898～900
41 Fridrich J.. Security of fragile authentication watermarks with localization. In: Proceedings of the SPIE, Vol. 4675, Security and Watermarking of Multimedia Contents IV, San Jose, California, 2002, 691～700
42 Wu J., Zhu B., Li S., Lin S.. A secure image authentication algorithm with pixel-level tampering localization. In: Proceedings of the IEEE International Conference on Image Processing (ICIP'04), Singapore, 2004
43 Wong P. W.. A watermark for image integrity and ownership verification. In: Proceedings of the IS&T PIC Conference, Oregon, Portland, 1998
44 Wong P. W.. A public key watermark for image verification and authentication. In: Proceedings of the IEEE International Conference on Image Processing, Chicago, USA, 1998, 1: 455～459
45 Memon N., Wong P.. Secret and public key authentication watermarking schemes that resist vector quantization attack. In: Proceedings of the SPIE, Vol. 3971, International Conference on Security and Watermarking of Multimedia Contents II, San Jose, USA, 2000, 417～427
46 Celik M., Sharma G., Saber E., Tekalp A. M.. A hierarchical image authentication watermark with improved localization and security. In: Proceedings of the IEEE International Conference on Image Processing, Thessaloniki, Greece, 2001, 2: 502～505
47 Celik M., Sharma G., Saber E., Tekalp A. M.. Hierarchical watermarking for secure image authentication with localization. IEEE Transactions on Image Processing, 2002, 11(6): 585～595
48 van Schyndel R. G., Trikel A. Z., Osborne C. F.. A digital watermark. In: Proceedings of the IEEE International Conference on Image Processing, Austin, Texas, 1994, 2: 86～90
49 Wolfgang R. B., Delp E. J.. A watermark for digital images. In: Proceedings of the IEEE International Conference on Image Processing, Laussane, Switzerland, 1996, 3: 219～222
50 Wolfgang R. B., Delp E. J.. Fragile watermarking using the VW2D watermark. In: Proceedings of the SPIE, Vol. 3657, Security and Watermarking of Multimedia Contents I, San Jose, USA, 1999, 204～213
51 Swanson M. D., Zhu B., Tewfik A. H.. Robust data hiding for images. In: Proceedings of the 7th IEEE Digital Signal Processing Workshop, Loen, Norway, 1996, 37～40
52 Kundur D., Hatzinakos D.. Digital watermarking for telltale tamper-proofing and authentication. In: Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information, 1999, 87(7): 1167～1180
53 Li C. T., Lou D. C., Chen T. H.. Image authentication and integrity verification via content-based watermarks and a public key cryptosystem. In: Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, 2000, 3: 694～697
54 Sun Qibin, Chang Shih-Fu, Kurato Maeno, Suto Masayuki. A quantitative semi-fragile JPEG2000 image authentication system. In: Proceedings of the IEEE International Conference on Image Processing, Rochester, USA, 2002, 2: 921～924
55 Sun Qibin, Chang Shih-Fu. Semi-fragile image authentication using generic wavelet domain features and ECC. In: Proceedings of the IEEE International Conference on Image Processing, Rochester, USA, 2002, 2: 901～904

WU Jin-Hai, born in 1979, M. S.. His research interests include digital watermarking, image authentication and multimedia security.

LIN Fu-Zong, professor. His research interests include distance education, machine learning, digital watermarking, etc..

Background

This paper focuses on the research of multimedia security, especially image authentication based on digital watermarking. The project is partially supported by the National Nature Science Foundation of China under grant No. 60135010. Research goal is to design secure image authentication algorithms which can be put into practice in the future.

This paper is an overview of the image authentication technology based on digital watermarking.