

研 究 生 文 献 综 述 报 告

**题目：基于数字水印的图像认证算法研究**

|  |  |  |
| --- | --- | --- |
| **学号** | **：** | **M201470343** |
| **姓名** | **：** | **沈嘉琪** |
| **专业** | **：** | **机械电子工程** |
| **指导教师** | **：** | **陈永府** |
| **院（系、所）** | **：** | **机械科学与工程学院** |

**华中科技大学研究生院制**

根据导师在数字水印方面的研究，结合武汉天喻软件公司数据安全防护的技术研究，综合国内外相关领域的发展趋势，确定本人的论文题目为“基于数字水印的图像认证算法研究”，具体研究面向图像认证领域的数字水印技术。

下面，分别从一般的数字水印和认证领域的数字水印两个角度对国内外研究现状进行综述，并进行总结，进一步指出目前存在的问题和可能的发展方向。

# 一、国内外研究现状

由于信息技术的飞速发展，越来越多的资料以数字的形式存在，然而数字媒体存在着易扩散和易修改的不足，数字媒体的内容是否安全、完整、可信问题日益凸显。传统的加密技术能够在数据传输过程中起到保护作用，但数据被接受或者解密之后将不再受到保护，同时加密的信息内容容易引起好事者的注意，激发他们的破译热情。另外，密码学中的完整性是通过数字签名方式实现的，它并不是直接嵌入到数字信息当中，因此无法察觉信息在经过加密之后的再次传播中内容的改变；最后，传统的加密手段无法对数字信息的修改进行局部定位，也无法恢复。因此，为克服传统的加密技术的管理不便、不能定位篡改、不能区分篡改强度等缺点，迫切需要寻求另外一种途径。数字水印作为信息安全领域的一个新方向，甚至被认为是多媒体内容保护的最后一道防线，目前得到了学术界的广泛的研究，其发展也相当迅速。

## 1.1 数字水印简介

### 1.1.1数字水印概念

数字水印是一种有效的数字产品版权保护和数据安全维护的技术，是信息隐藏技术研究领域的一个重要分支[1,2]。它将具有特定意义的标记（水印），利用数字嵌入的方法隐藏在数字图像、音频、文档、图书、视频等数字产品中，用以证明创作者对其作品的所有权，并作为鉴定、起诉非法侵权的证据；同时，通过对水印的检测和分析保证数字信息的完整可靠性，从而成为知识产权保护和数字多媒体防伪的有效手段。

数字水印是一种十分贴近实际应用的数据隐藏技术，虽然其具有一定的共有特性，如不易察觉性、安全可靠性（不易被破解、伪造），而更多的特性要求往往来自特定的应用需求。例如，从信息安全的保密角度而言，隐藏的信息如果被破坏掉，系统可以视为安全的，因为秘密信息并未泄露；但是，在数字水印系统中，隐藏信息的丢失即意味着版权信息的丢失，从而失去了版权保护的功能，这一系统就是失败的。

数字水印不同于传统的加密技术，因为其目的不在于限制正常的信息交流，而是保证信息的安全或对信息加以说明等，但信息在传播过程中必然会遭到各种人为或自然地破坏，这就要求数字水印具备如下的特征[3,4]：

（1）透明性（Invisibility）：利用人类视觉系统或人类听觉系统属性，经过一系列隐藏处理，使目标数据没有明显的降质现象，而隐藏的数据却无法人为地看见或听见。

（2）鲁棒性（Robustness）：鲁棒性又称为稳健性、健壮性。鲁棒性是指数字水印在经过有意或无意的处理操作之后，依然能够保持水印的完整性和鉴别的准确性。

（3）安全性（Security）：数字水印不能被非法攻击者检测或即使检测到也不能被破坏。

（4）不可检测性：指隐蔽载体与原始载体具有一致的特性。如具有一致的统计噪声分布等，以便使非法拦截者无法判断是否有隐蔽信息。

（5）自恢复性：由于经过一些操作或变换后，可能会使原图产生较大的破坏，如果只从留下的片段数据仍能恢复隐藏信号，而且恢复过程不需要宿主信号，这就是所谓的自恢复性。

其中水印的透明性和鲁棒性有一定的对立，因为要保证水印的鲁棒性，必须嵌入尽可能多的水印容量，而这恰恰是透明性所反对的。所以在数字水印系统中必须在鲁棒性和透明性之间寻求最佳结合点。

### 1.1.2数字水印基本框架

通用的数字水印算法包含两个基本方面：水印的嵌入和水印的提取或检测。水印可由多种模型构成，如随机数字序列、数字标识、文本或图像等。从鲁棒性和安全性考虑，常常需要对水印进行随机化以及加密处理。

设为水印信号，为密钥（通常是一个随机发生器的种子），那么处理后的水印由函数定义如下：

（2.1）

 （2.1）

如果水印所有者不希望水印被其他人知道，那么函数应该是非可逆的，如经典的DES加密算法等。这是将水印技术与加密技术结合起来的一个通用方法，目的是提高水印的可靠性、安全性和通用性。

水印的嵌入过程，见图2-1，其中虚线标识该条件并非必要条件。设有编码函数，水印和原始图像，那么水印图像可表示如下：

 (2.2)

其中由式（2.1）定义。



图1-1 水印的嵌入流程

水印的提取或检测时水印算法中最重要的步骤



图1-2 水印的检测流程

若将这一过程定义为解码函数，那么输出的可以是一个判定水印存在与否的0-1决策，也可以是包含各种信息的数据流，如文本、图像等。如果已知原始图像和版权疑问的图像，则有：

 （2.3）

或者：

 （2.4）

式（2.4）中为提取出的水印，为密钥，为决策阀值函数，做相关检测。这种形式的检测函数是创建水印框架的一种最简便方法，如假设检验或水印相似性实验。

检测器的输出结果如充分可信则可在法庭上作为版权保护的潜在证据，那么这实际上要求水印的检测过程和算法应该完全公开。对于假设检验的理论框架，可能的错误有如下两类：

第一类错误（纳伪）：检测到水印但实际上不存在。该类错误用虚警率（Probability Of False Alarm）衡量。

第二类错误（弃真）：没有检测到水印而水印实际存在，用漏报概率表示。

总错误率为，且当变小时检测性能变好，但是检测的可靠性只与误识率有关。

为了能提高水印算法的鲁棒性，可以在水印信息中包含纠错码和检测码，可以保证包含水印的图像在受到干扰的情况下仍能够可靠和完整地提取出水印信息。

众多的水印算法，每种都有自己的特点，嵌入或检测的过程可能在某些细节上有所不同，但总体来看，都与上述的一般原理是一样的。

### 1.1.3数字水印分类

本节将从不同的角度，对现有的数字水印算法进行分类，以便更加全面地了解和掌握数字水印算法的理论与技术[5]。

**1. 按特性划分**

分成鲁棒型水印和脆弱型水印。鲁棒型水印主要用于数字产品的版权保护，它必须保证对原始版权的准确无误的标识。脆弱型水印主要用于数据的真伪鉴别和完整性鉴定，又称为认证。

**2. 按水印所附载的媒体划分**

分成图像水印、音频水印、视频水印、文本水印等。随着数字技术的发展，会有更多种类的数字媒体出现，同时也会产生相应的水印技术。

**3. 按检测过程划分**

分成明文水印和盲水印。明文水印在检测过程中需要原始数据，而盲水印的检测不需要原始数据。一般来说，明文水印的鲁棒性比较强，但其应用受到存储成本的限制。

**4. 按水印嵌入位置划分**

分成时空域数字水印和变换域数字水印。时空域数字水印是直接在信号空间上叠加水印信息，具有较大的信息嵌入量，但鲁棒性较弱。变换域数字水印则分别是在DCT、DFT等频域、时/频变换域和小波变换域上隐藏水印，具有较强的鲁棒性。

**5. 按用途划分**

可以分为证件防伪水印、版权标识水印、篡改提示水印等。

**6. 按内容划分**

分成有意义水印和无意义水印。有意义水印是指水印本身也是某个数字图像或数字音频片段的编码；无意义水印则只对应于一个序列号。

**7. 从外观上划分**

分成可见水印和不可见水印。可见水印主要目的在于明确标识版权，防止非法的使用。不可见水印目的是为了将来起诉非法使用者，作为起诉的证据，以增加起诉非法使用者的成功率，保护原创者和所有者的版权。

**8. 根据所采用的用户密钥的不同分类**

分为私钥水印和公钥水印。私钥水印在加载和检测水印过程中采用同一密钥。公钥水印则在水印的加载和检测过程中采用不同的密钥。

### 1.1.4数字水印典型算法

自数字水印提出以来，已经得到了迅速的发展。目前，已经提出了众多的数字水印算法，也有许多公司推出了数字水印的产品。但总的来说，由于数字水印的研究是基于计算机科学、密码学、通信理论、算法设计和信号处理等领域的思想和概念，一个数字水印方案一般总是综合利用这些领域的最新进展。以图像为载体的数字水印技术是水印技术研究的重点之一。在该领域发表的论文数目要远大于以音频、视频等信号为载体的水印方面的论文。我们按数字水印嵌入的位置对算法进行划分。一类是空间域算法：通过直接对宿主信息做变换来嵌入水印。另一类是变换域算法：通常首先对宿主信息做特殊变换（DCT、DFT、DWT），然后在变换的基础上嵌入水印信息。在实际应用中，往往将空间域算法或变换域算法与其它相关理论相结合以得到不同特性的水印算法，如感知模型算法、扩频算法等。

**1. 空间域算法**

该算法主要是指在图像的空间域中嵌入水印的技术。最简单和有代表性的方案就是用水印信息代替图像的最低有效位（LSB）或者多个位平面的所有比特的算法。这里的水印信息指的是二值比特序列。图像的最低有效位也称为最不显著位，它是指数字图像的像素用二进制表示时的最低位。LSB算法是由Schyndel[6,7]提出的,该算法首先把一个密钥输入到一个m-序列（Maximum-Length Random Sequence）发生器来产生水印信号，然后将此m-序列重新排列成2维水印信号，并按像素点逐一插入到原始图像像素值的最低位。由于水印信号被安排在最低位，因此它是不可见的，但基于同样的原因，水印也可以被轻易的去掉、破坏，因此该方法鲁棒性较差。Wolfgang等人给出了一种改进的方法，将水印嵌入到随机选取的像素最低位，这种改进的方法虽然增强了水印的安全性，但是其鲁棒性却没有多大的改变。Bender等人提出的基于统计的拼凑水印算法以及纹理块映射编码方法则是空间域水印技术的典型设计。Patchwork算法根据密码伪随机地将图像的像素分成两部分，一部分的像素值加1，另外一部分减1。这样在检测时，将这两部分相减，如果有水印，则相减结果应该为2N(N为像素对的个数)，否则为0。纹理块映射编码方法则是将数字信息隐藏于数字图像的任意纹理部分，该算法对于滤波、压缩和扭转等操作具有抵抗能力，但仅适用于具有大量任意纹理区域的图像，而且尚不能完全自适应。

**2. 变换域算法**

基于变换域的技术可以嵌入水印数据而不会引起感观上的察觉[8,9]，这类技术一般基于常用的变换，如离散傅立叶变换（DFT）、离散余弦变换（DTC）、离散小波变换（DWT）、Fourier-Mellin变换或其它变换等。从目前的情况看，变换域方法日益普遍。因为变换域方法通常都具有很好的稳健性，对数据压缩、常用的滤波处理以及噪声等均有一定的抵抗能力。并且一些水印算法还结合了当前的图像和视频压缩标准（如JPEG、MPEG等），因而有很大的意义。

在变换域水印算法中，数字图像首先进行一种特定的正交变换，该变换可以针对整幅图像或图像的各部分，比如对图像进行分块，块大小一般为8×8或者16×16嵌入空间是图像的某个频带或某些频带，这些频带对应的变换系数遵循一定的规则被修改、替换或交换。

变换域水印算法与空间域水印算法比较具有一下优点：

■在变换域中嵌入的水印信号能量可以散布到空间域的所有位置上，有利于保证水印的不可察觉性。

■在变换域，人类视觉系统和听觉系统的某些特性可以方便地结合到水印编码过程中。

■变换域的方法可与数据压缩标准相兼容，从而实现在压缩域内的水印算法，同时，也能抵抗相应的有损压缩。

**3. 感知模型算法**

为了得到水印的鲁棒性和不可见性的最佳折中[10,11]，一些基于人类生理特性的水印方法被提出。人的生理模型包括人类视觉系统HVS和人类听觉系统HAS。利用视觉模型的基本思想均是从视觉模型导出的JND（Just Noticeable Difference）描述来确定在图像的各个部分所能容忍的数字水印信号的最大强度，从而能避免破坏视觉质量。

随着图像数字水印算法研究的深入，研究人员提出了其它类型的算法，这其中包括基于独立分量分析（ICA）的数字水印技术以及基于神经网络的图像水印技术。

**4. 压缩域水印算法**

基于 JPEG、MPEG 标准的压缩域数字水印系统不仅节省了大量的完全解码和重新编码过程[12]，而且在数字电视广播及 VCD 中有很大的使用价值。相应地，水印检测与提取也直接在压缩域数据中进行。

### 1.1.5数字水印算法的评价指标

对水印系统或水印方案的评估是多方面的，不仅需要对鲁棒性进行评估，而且包括对由水印处理而引起的变形进行主观的或定量的评估，过去，只有很少一部分研究者对水印处理带来的图像降质进行定量的分析。一般来说，在水印鲁棒性和可见性之间需要进行折衷。因此，为了能够进行公平的基准测试和性能评估，必须确保各种水印系统是在可比较的条件下进行测试与研究。到目前为止，还没有统一的水印算法衡量标准，下面只介绍一些常用的检测参数。

**1. 不可察觉性（Imperceptibility）**

用户应该无法从含水印图像中直接看出水印，它要求水印信号不能比宿主信号强，这是检测不可见水印的前提要求。对含水印图像最直接的检测方法是通过多个观察者对含水印图像做观察，得出各自的判断。

**2. 峰值信噪比（PSNR：Peak Signal-to-Noise Ration）**

在水印嵌入过程中，由于水印能量要远远小于原始信号，所以可以把水印作为噪声处理。观察其峰值信噪比，尽管这一引用不是很精确，但在某种程度上，它还是能够很好地比较水印的稳健性。其计算公式如下：为宿主信号，为水印信号，为含水印信号，为像素点，为像素总个数。

 （2.5）

**3. 归一化汉明距离**

若水印信号是二进制序列，则可以计算提取水印信号与嵌入水印信号之间的归一化汉明距离来检测其相似性。计算公式如下：其中与分别表示嵌入水印信号和提取水印信号，代表水印长度，为异或运算。

 （2.6）

**4. 归一化相关系数**

为了检测提取水印信号和嵌入的水印信号之间的相似性，一般可以通过它们的归一化相关系数（NC：Normalized Cross-Correlation），其计算公式如下：其中与分别表示嵌入水印信号和提取水印信号，代表水印长度，对于鲁棒型水印系统，一般在有信号失真时，值越大越好，而对于易碎水印系统，值越小越好，这样有助于提高检测的可信度。

 （2.7）

水印存在与否的判断标准为：若则可以判定被测图像中有水印存在，否则没有水印。的选择要同时考虑虚警概率和漏警概率，减小，漏警概率降低而虚警概率提高，反之，同样成立。

**5. 相似系数**

相似系数也是用来评价提取水印信号与嵌入水印信号的相似程度，其计算公式如下：其中与分别表示嵌入水印信号和提取水印信号。

 （2.8）

### 1.1.6 数字水印的主要应用领域

**1.版权保护**

即数字作品的所有者可用密钥产生一个水印，并将其嵌入原始数据，然后公开发布他的水印版本作品。当该作品被盗版或出现版权纠纷时，所有者即可利用图2-3或图2-4的方法从盗版作品或水印版本作品中获取水印信号作为依据，从而保护所有者的权益。对这种应用领域来说，水印技术必须有较好的鲁棒性、安全性、透明性和水印嵌入的不可逆性

**2.加指纹**

为避免未经授权的拷贝制作和发行，出品人可以将不同用户的ID或序列号作为不同的水印（指纹）嵌入做平的合法拷贝中。一旦发现未经授权的拷贝，就可以根据此拷贝所恢复出的指纹来确定它的来源。对这种应用领域来说，水印技术除了具有版权保护的应用中的特性外，还必须具有防止串谋攻击（多拷贝攻击）等功能。

**3.标题和注释**

即将作品的标题、注释等内容（如，一副照片的拍摄时间和地点等）以水印形式嵌入做平中，这种隐式注释不需要额外带宽，而且不易丢失

**4.篡改提示**

当数字作品用于法庭、医学、新闻和商业时，常需要确定它们的内容是否被修改、伪造或者特殊处理过。为实现该目的，通常可以将原始图像分成多个独立块，再将每个块加入不同的水印。同时可通过检测每个数据块中的水印信号，来确定作品的完整性。与其他水印不同的是，这类水印必须是脆弱的，并且检测水印信号时，不需要原始数据。

**5.使用控制**

这种应用的一个典型的例子是DVD防拷贝系统，即将水印信息加入DVD数据中，这样DVD播放机可通过检测DVD数据中的水印信息而判断其合法性和可拷贝性。从而保护制造商的商业利益。

**6.数字水印技术应用与版权保护的实例**

ImageBridge是Digamarc公司的一个用于保护万维网上数字图像版权的软件。当你购买了该软件后，Digamarc会给你一个全球唯一的ID，供你嵌入数字图像使用，Digamarc的MarcSpider会不停的在网络上搜索，当发现有嵌入图像经修改后的版本后，它会将改图片的相关信息通过email的方式告诉版权所有者。该系统被包括大不列颠图书馆、华盛顿邮报社等诸多知名机构、公司采用，取得了良好的实际效果。

## 2.1 图像认证水印技术概述

认证水印技术是传统密码学与隐蔽通信相结合的产物，它涉及到密码学、通信理论以及数字图像处理等方面的技术。认证水印是把标志信息作为水印嵌入到数字多媒体内容中（保证嵌入的信息不影响感官质量为前提），当需要对多媒体内容进行认证时，提取出水印信息以鉴定数字多媒体内容是否真实完整。基于数字水印的图像认证系统的一般框架实现包括：认证水印的产生、嵌入和检测认证。

### 2.1.1 认证水印的特殊要求

作为数字水印技术的一种，用于数字图像完整性保护和内容认证的脆弱水印的嵌入原理与一般水印的嵌入原理基本相同，从数字信号处理的角度上看都是对载体图像的调制过程。脆弱水印除了具有一般水印系统的基本特征如不可感知性、安全性等特点外，还应能够完成认证功能。具体来说，脆弱水印应满足以下特殊要求：

（1）检测篡改。脆弱水印的最基本的要求就是具有认证功能，即载体图像遭到篡改时，认证水印系统能有效地检测篡改并指出数字图像被篡改的程度及定位篡改发生的位置，甚至有些应用场合要求系统能够恢复出被篡改的图像内容。

（2）脆弱性和鲁棒性。为了鉴别数字图像的内容是否发生篡改，则要求数字水印为脆弱性的水印。即一旦检测不出水印的存在，则说明载体图像内容已经被正常的信号处理操作改变或被篡改。实际的应用场合，大部分数字图像要经过一定的压缩处理后才在网络上传输，以便提高信息传输量，这就要求脆弱性的数字水印也具有一定的鲁棒性。

（3）不可感知性。同鲁棒性水印相同，脆弱性水印也要求人眼很难觉察到嵌入水印前后载体图像的变化，以便保证数字图像的使用价值。

（4）盲检测。水印检查时，使用原始的宿主信号，更有利于检测和提取信息，如鲁棒性水印用于数字媒体的版权保护，检测时允许使用原始数据，但是检测时用到的原始宿主信号容易暴露给恶意的攻击者，而且在某些应用中，并不能获得原始载体图像信息。脆弱水印的功能就是用来验证待检测图像的完整性、真实性，如果接收方有原始图像，那对图像真实性进行验证就显得多余。对于认证系统来说，盲检测这一点是必须的。因此，除了极个别的方案外，目前主要研究的是盲检测水印技术。

（5）更高的安全性。脆弱性水印与鲁棒性水印的最大差别就在于脆弱性水印对各种攻击保持着高度的敏感性，脆弱性水印主要应用于对各种重要数据的精确认证和内容的模糊认证，例如医疗过程中的影像，或者是用于法庭举证的图片、视频等等。在这种场合下，脆弱水印的安全性就成为非常重要的问题。为了增加水印的安全性，很多研究者采用对水印信息加密的方式，利用一个控制水印信息的密钥来保证水印不被攻击者获取。密钥的选择方式有很多种，根据载体图像的特征选择最合适的控制密钥能大大提高水印的安全性。

（6）更强的敏感性。图像认证系统要求能够检测出被保护图像上的篡改操作。目前针对数字图像的各种攻击和处理可分为两类：

①无意的攻击。采用压缩编码标准如 JPEG、JPEG2000 等对图像进行编码；数字图像处理中的图像增强等操作。

②有意的攻击。保护数字图像的真实性、完整性是图像认证系统的主要功能，因此存在某些设法篡改图像内容但不会破坏水印信息以求通过认证的“伪认证”攻击。

### 2.1.2 认证水印所受攻击类型

用于数字多媒体内容完整性、真实性认证的水印信息本身具有一定脆弱性，常见的对鲁棒水印的攻击方法，如简单攻击，删除攻击，混淆攻击对认证水印系统的效果微乎其微。图像认证系统最需要抵抗的是试图篡改图像的内容却不损坏水印信息的“伪认证”攻击，为保证认证水印嵌入时的安全性，通常使用密码算法或密钥，使攻击者无法提取水印信息，但当同一密钥嵌入多幅图像后，攻击者会对这些数字图像进行大量分析，破译密钥，进而对图像进行恶意篡改操作，却仍可通过认证。目前，关于数字水印认证技术的攻击算法报道还较少，根据已有的文献来看，对图像水印认证系统可能实施的有效攻击有以下几种：

1．恒值攻击（Constant-average Attack）：恒值攻击是指在不改变水印基础上任意改变像素的值保证图像块的像素均值不变，例如一个2×2的图像块的四个像素的像素值为[50 60; 60 60]，将图像块像素信息修改为[20 90; 90 20]，图像块的均值不变。

2．内容攻击（Content Attack）：内容攻击和恒值攻击有些类似。也是在保持不改变水印的前提下，对非水印部分任意修改。有些算法的水印产生与非水印层的无关，因此，只要不去修改水印，被篡改的图片就可以通过验证，不会发现图片已经被篡改了。

3．拼贴攻击（Collage Attack）主要常见于基于分块的脆弱水印算法中，它利用水印算法在多幅图像中嵌入同一水印信息时存在的安全漏洞，对于嵌入相同水印信息的多幅图像，在保持图像分块相对位置保持不变的情况下，将属于不同图像的图像块进行拼贴得到篡改后的图像。如果攻击者在进行图像拼贴时，精心地选择图像块，保证拼贴后图像一定的视觉质量，那么几乎所有的基于独立分块的脆弱水印算法都将无法检测到篡改。

4．量化攻击（Vector Quantization Attack）也是一种针对基于独立分块的脆弱水印算法的攻击方法。这种攻击的前提条件是每个图像块中嵌入的水印信息与其他图像块的内容无关，这样如果两个图像块中嵌入的水印信息相同，就可以将它们的内容互换而不会被检测到篡改。预防量化攻击最有效的方法是设计脆弱水印算法时保证每个图像块生成的水印信息依赖于其他图像块的内容。

5．统计分析攻击：当采用同一密朗在不同图像中嵌入同样认证信息时，攻击者通过对大量取证图像统计分析，有可能分析出水印嵌入规律。

### 2.1.3 认证水印技术的发展

脆弱数字水印从1994年被提出至今，已有近20年的发展历史，在这期间，脆弱水印凭借其表现出的巨大的社会经济潜力，越来越受到人们的重视。IHW（国际信息隐藏学术研讨会）、IWDW（国际数字水印研讨会）、ICIP（IEEE 图像处理国际会议）等国际重要学术会议定期报道数字水印技术的最新发展，1999 年国内召开了第一届 CIHW 信息隐藏学术会议，之后每年召开一次，对国内数字水印技术的发展进行报道。英国剑桥大学、美国麻省理工学院、日本NEC研究所等著名研究机构最早开始数字水印相关领域的研究，并取得不少的成果。国内对于脆弱水印的研究是从2003年开始的，中科院自动化研究所、北京大学、清华大学等一批国内科研机构和院校也逐步投入到脆弱水印的研究领域。

由于脆弱水印的类型错综复杂，很难一类一类地分清，下面从脆弱水印算法的功能角度，对脆弱水印算法的国内外发展进行综述。

#### 1．像素级篡改定位的脆弱水印技术发展

像素级篡改定位的脆弱水印方法，可以将篡改定位到像素的级别。1995年Walton[10]率先提出的LSB方法是最早出现的像素级篡改定位的空域脆弱水印方法，其主要思想是计算载体图像像素点的最高七个比特位的校验和，这种计算校验和的方法本质上是一种特殊的 Hash 函数，算法中的每个字由八个部分组成，每个部分又由像素的最高七个比特位组成。根据校验和随机选取一定数目的像素点，将这些像素点的LSB位用校验和中相应的比特位进行替换，从而实现水印信息的嵌入。认证时重新计算像素的最高七个比特位的校验和，并与提取的水印信息进行比较，从而实现图像的精确认证。这种方法简单且很容易实现，但是无法检测像素交换的伪认证篡改，篡改定位能力差，而且由于水印信息位于LSB平面，很容易被移除掉，因此安全性也较差。文献[11]使用二值图像作为水印内容，将图像尺寸和最高七个比特位进行Hash运算，对运算结果与二值水印进行异或操作，操作结果经公钥加密后嵌入到LSB位。认证时，首先对图像尺寸和最高七个比特位进行Hash运算，将运算结果与解密后图像的LSB位进行异或操作，从得到的水印模式中可以直观地看出图像被篡改的区域。该算法安全性较高，并且具有篡改定位的能力，但是容易遭受量化攻击，算法虚警率较高。文献[12]也是以二值图像为水印，通过将水印的某些特征映射为随机数0或者1，形成一个查找表LUT（Look Up Table），通过利用LUT对像素进行量化实现水印信息的嵌入。这种算法的安全性依赖于LUT表的推断难度[13]，并且可能遭受拼贴攻击[14]。文献[15]将文献[12]的算法从空间域推广到变换域，使得算法可以与JPEG压缩标准相兼容。该算法采用JPEG压缩标准中的量化矩阵，对载体图像的 DCT 变换系数进行量化，在LUT中，将量化结果映射为随机数 0 或者 1。水印嵌入时，判断水印比特值与LUT表中DCT系数的映射值是否相同，如果相同，则保持该系数不变，如果不同，则将DCT系数量化为与其距离最近的系数。像素级篡改定位的脆弱水印的优点是能够将篡改定位到像素上，缺点是很容易被破坏掉，对噪声非常敏感，容易遭受精心设计的伪认证攻击。文献[16]将某一点的像素与其周围的上下和左右各*k*个点，共计2*k+*1个像素点结合起来生成认证信息，最后嵌入到图像最低位。如果一个像素点的认证信息发生改变*,*但其向左右和上下各扩展*k*个像素点的认证信息有些没有发生改变，则该像素点未被篡改，认证信息的改变是由于左右和上下各扩展*k*个像素点中的一个或多个被篡改所致，或者是该像素点的认证信息在传输中出现差错所致。只有该点以及其左右和上下各扩展*k*个像素点都发生认证信息的改变才是相应的像素点被篡改。但是基于像素级的算法都不能抵御拼贴攻击，因此，现在基于像素级的算法的研究已经很少见了。

#### 2．块级篡改定位的脆弱水印技术发展

基于分块的脆弱水印方法，可以将篡改精确地定位到分块级别。1998 年，Wong 等在文献[17,18]提出一种基于独立分块的脆弱水印方法。该算法将图像划分为互不重叠的独立小块，每个小块都会生成相应的水印并嵌入到图像块自身中，从而实现了篡改的分块定位。但是，由于图像分块的独立性，导致算法容易遭受伪造真实图像数据的量化攻击[19]。为了消除分块的独立性，文献[20,21]使用分层的分块认证方法，文献[22]使用图像索引与分块编号的方法，取得显著的效果并进一步提高了篡改定位的能力。1999年，Fridrich在文献[23]中首次提出一种DCT变换域的分块自嵌入水印。该算法首先将载体图像划分为互不重叠的大小为 8×8 的图像块，将一个子块最高七位的DCT系数按照特定的码长进行量化编码，并采用固定偏移方式嵌入到另一个子块的LSB平面位。该算法不但能够定位篡改位置，而且可以利用提取的水印信息近似地恢复被篡改的内容，但是算法可能遭受伪造攻击[19]。文献[24,25]为了克服固定偏移方式的缺点，提出一种基于混沌的 DCT域分块自嵌入水印方法。该算法利用混沌系统的极端敏感性和伪随机特性，实现水印信息的随机偏移嵌入，进一步扩大了密钥空间，提高了算法的安全性。该算法可以有效地定位被篡改的图像块，并且可以近似地恢复被篡改的图像块。但是由于算法采用8×8的分块方式，篡改定位的精度较低，采用DCT域嵌入的恢复信息非常有限，篡改恢复的效果也较差。为了提高定位精度，文献[26]利用 SHA512 算法和单向函数生成水印，结合层次结构与滑动窗口技术实现水印的嵌入，使得算法能够将篡改定位到 2×2 分块，并能有效抵抗量化攻击。基于分块的脆弱水印最大的特点是，可以将篡改定位到分块级别上，安全性较高，缺点是精确篡改定位能力较差。文献[27] 提出了一种针对彩色图片检测的算法。算法将彩色图片分成不重叠的4×4小块，每一个小块的最重要的6位用来产生验证数据，采用块截断编码的方式来生成恢复数据。为了抵御量化攻击，采用一个2维变换公式来生成块映射。但是由于水印未加密，很容易被他人攻击，此外该算法也不能抵挡拼贴攻击。在文献[28]中，Zhang根据像素将图像分为不重叠的2×2的小块，之后对每一个小块进行DCT变换。同样的，Zhang的算法也将水印分成两种数据，恢复数据和验证数据。恢复数据由DCT变换后的DC系数生成，验证数据则是根据这些恢复数据生成。并采用logistic映射生成块映射。但是，zhang的算法并不能抵御恒值攻击，并且恢复效果也不佳。

#### 3．篡改可恢复的脆弱水印技术发展

为了提高图像的利用价值，有些图像认证的应用场合要求脆弱水印具备篡改恢复能力。篡改恢复可以分为精确恢复和近似恢复。精确恢复要求将被篡改的数据完全恢复到原始状态。由于图像认证只要求恢复后的篡改内容与原始内容大体一致，允许恢复后的数据与原始数据存在一定的偏差，因此对于图像认证只需实现近似恢复即可。脆弱水印的近似恢复方法一般可以分为自嵌入、纠错码恢复和盲恢复三种。自嵌入是将图像自身特征内容作为水印信息，纠错码恢复是在原始图像中嵌入错误纠正码，盲恢复是首先判断篡改类型，然后对篡改进行逆向操作。在三种方法中自嵌入水印是最为重要的一种篡改恢复方法。

自嵌入水印最早由Fridrich[23]于1999年提出，之后出现了许多改进的自嵌入水印算法[29-32]。文献[31]在文献[22]的基础上，重新设计了DCT量化系数的编码位长，进一步提高了算法的安全性。但是，该算法用水印数据完全替换LSB位平面，对载体图像修改较大，且利用DCT系数恢复的图像质量也不高。文献[32]进一步改进自嵌入方法，每个图像块中嵌入的水印信息包含了子块的认证信息及其偏移子块的压缩信息，水印信息同样嵌入到子块的LSB位平面，但最多修改LSB位平面中一半的数据，因此含水印图像的质量较好。文献[33,34]对文献[20]中分层的分块认证方法进行了改进，利用分层检测的方法对水印进行篡改检测与定位。该算法具备一定的篡改恢复能力，但篡改恢复质量不高，尤其是在篡改比例较大的情况下。文献[35]在分析文献[33]的基础上提出一种图像篡改检测与恢复的双水印算法，该算法中每个图像块的生成的水印信息都有一个拷贝版本，通过将两份相同的水印嵌入到不同的图像块中，大大提高了篡改恢复的质量，在篡改量较大的情况下仍然能够保持较高的恢复质量。不过，该算法存在较为严重的安全性问题，该算法不能抵御多种常见的“伪认证”攻击，比如拼贴攻击，恒值攻击等。文献[36]改进了文献[18]的算法，通过量化的方式为原始图像创建了一个量化表，用来恢复使用，并且与文献[35]，拥有更好的恢复效果。但是算法同样存在安全漏洞。

# 二、目前存在的问题

从上述脆弱水印算法、及安全性综述可看出，篡改恢复有助于推断取证图像遭到的攻击方式及攻击者的目的和意图，是数字图像真实性认证动技术的特点，也是评价图像算法优劣的重要技术指标。导致现有图像真实性认证算法恢复质量不高的因素主要包括以下几个方面。

1、伪认证攻击：无论是脆弱水印或是半脆弱水印，伪认证攻击都是对其威胁最大的攻击方法。由前面综述可知，现有部分（甚至最新的）图像认证算法在遭到伪认证攻击，不能检测篡改区域或篡改检测性能低，这为攻击者伪造能通过认证的伪造“取证图像”提供了可能的途径。能通过认证伪造“认证图像”的存在是图像真实性主动取证技术面临的最大安全隐患。同时，篡改不能被检测出来也直接影响图像恢复质量，这是由于篡改恢复仅对被判定为篡改的图像块实施。因此，如何提高算法抵抗伪认证攻击的能力，是数字图像真实性认证走向实际应用必须要解决的关键难题。

2、定位精度：由于目前的算法不能完全抵抗伪认证攻击，以及算法自身的缺陷问题，比如存在较大的虚警率或者漏警率，所以导致算法的定位精度依旧还有提高的空间。因此，如何设计有效的算法，尽可能的提高算法的定位精度。

3、恢复效果：目前关于水印的恢复算法有很多，目前水印对于小面积的篡改有较好的恢复效果，但是随着篡改面积的增大，算法的恢复效果就会显著性下降。

# 三、可能发展的方向

脆弱水印技术作为水印技术的一个重要分支，目前还是一个未成熟的研究领域，尚有许多问题有待于进一步深入地研究，未来的图像认证水印技术在如下几个方面需要作深入地探讨：

(1)基于图像内容的水印认证算法的研究

基于图像内容的水印信息既可以增强系统抵御统计攻击能力，又可避免在认证检测端额外提供原始水印信息，同时也可以区分恶意操作与非恶意操作，因此更适用于图像认证。目前虽然也有一部分算法采用与图像内容相关的水印信息，但针对不同的应用提取图像的何种特征，生成的水印长度如何与原始图像容量相匹配以及如何使水印的嵌入不引起图像特征的变化等问题还有待于更深入的分析和研究。

(2)水印安全性问题

实际应用对水印的保密安全有不同程度的要求。现有的脆弱水印技术大多采用基于私钥的加密方案，大量私钥信息通常很难管理，如何将半脆弱水印认证系统与密码学中的公开密钥算法结合，设计安全可靠的公钥水印认证算法，同时建立相应的标准或协议也将是一个重要的研究方向。

( 4)视、音频水印认证技术的研究

大量消费类数字视频产品的推出，使得以半脆弱水印为重要组成部分的视、音频真伪鉴别技术的市场需求更加迫切，然而，由于包括时间域掩蔽效应等特性的更为精确的理论模型尚未完全建立，使得目前视频、音频半脆弱水印技术的性能不太理想，同时现有的音、视频的编码格式也在一定程度上限制了水印技术的引入。因此，未来音、视频半脆弱水印技术也将会成为一个研究热点。

(5)评测标准的建立

目前在将多媒体认证水印技术推向标准化方面还有很多工作要做，比如水印嵌入算法和检测算法的理论研究、水印的构造模型、水印能量和容量的理论估计、对水印系统进行公正的比较和评价方法等，此外数字水印技术还面临着许多社会和法律问题需要解决。

# 四、参考文献

[1]王炳锡,彭天强.信息隐藏技术－北京：国防工业出版社，2007，9:53-54.

[2]Christine I, Edward J, Digital Watermarking: Algorithms and applications,IEEE

[3]张春田，苏育挺，管晓康.多媒体数字水印技术[J]. 通讯学报，2000，(9): 46-52.

[4]沈晓峰. 基于奇异值分解的数字图像水印技术研究[D].苏州大学，2008， 12(12):10-11.

[5]杨义先，钮心忻.数字水印理论与技术－北京：高等教育出版社，2006:20-22，29-32.

[6]E.Wolfgang，R.Podilchuk，C.Delp.Perpetual watermarks for digital images and vide，Proceedings of the IEEE，1999.87(7):1108-1126.

[7]Pei-Min Chen.A Robust Digital Watermarking Based on a Statistic Approch. The International Conference on Information Technology: Coding March 27-29，2000 p.116.

[8]Pereira Shelby，VoloshynovskiySviatoslav，Pun Thierry.Optimized wavelet domain watermark embedding strategy using linear programming.Proceedings of SPIE-The International Society for Optical Engineering 4056 Apr 26-Apr 28 2000 SPIE:490-498.

[9]Quan Liu，Qinsong Ai. A Combination of DCT-based and SVD-based Watermarking Scheme. International Conference on Signal Processing Proceedings，2004.

[10] Walton S. Information Authentication for a Slippery New Age. Dr.Dobbs Journal, 1995,20(4):18-26.

[11] Wong P W. A Public Key Watermark for Image Verification and Authentication. Proceedings of the IEEE International Conference on Image Processing[C], Chicago, Oct.1998:455-459.

[12] Yeung M, Mintzer F. Invisible watermarking for image verification[J]. Journal of

Electronic Imaging, 1998,7(3):578-591.

[13] Memon N, Shende S, Wong P. On the security of the Yeung-Mintzer authentication watermark[C] // Proceedings of the IS & TPICS Symposium, Savannah, Georgia, 1999: 301-306.

[14] Fridrich J, Goljan M, Memon N. Further attacks on Yeung-Mintzer fragile watermarking scheme[C]//Proceedings of the SPIE, Security and Watermarking of Multimedia Contents Ⅱ, San Jose,CA. 2000,3971:428-437.

[15] Wu M, Liu B. Watermarking for image authentication[A]. In: Proceedings of the IEEE International Conference on Image Processing [C], Chicago, USA, 1998, 2: 437-441.

[16]谢建全, 阳春华. 一种像素级的图像篡改认证算法[J]. 计算机应用, 2007, 27(6):1337-1342.

[17] Wong P W. A watermark for image intergrity and ownership verification[C]//Proceeding of the IS & T PIC Conference, Oregon, Portland, 1998.

[18] Memon N, Wong P. Secret and public key authentication watermarking schemes that resist quantization attack[C]//Proceeding of the SPIE International Conference on Security and Watermarking of Multimedia Contents Ⅱ. San Jose, USA, 2000, 3971:417-427.

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

[20] Celik M, Sharma G, Saber E, et al. A hierachical image authenticcation watermark with improved localization and security[C]//Proceeding of the IEEE International Conference on Image Processing, Thessaloniki, Greece. 2001,2:502-505.

[21] Celik M, Sharma G, Saber E, et.al. Hierachical watermarking for secure image authentication with localization[J]. IEEE Transactions on Image Processing, 2002,11(6):585-595.

[22] Fridrich J. Security of fragile authentication watermarks with localization[C] // Proceedings of the SPIE, Security and Watermarking of Multimedia Contents Ⅳ. San Jose, Alifornia, 2002,4675:691-700.

[23] Fridrich J, Goljan M. Images with self-correcting capabilities[A].In Proceedings of International Conference on Image Processing[C], Kobe, Japan 1999, 3: 25-28.

[24] 和红杰，张家树. 基于混沌置乱的分块自嵌入水印算法[J]. 通信学报，2006, 27(7): 80-85.

[25] 和红杰，张家树. 基于混沌的自嵌入安全水印算法[J]. 物理学报，2007，56(6): 3092-3100.

[26] 张宪海, 杨永田. 基于脆弱水印的图像认证算法研究[J]. 电子学报, 2007, 35(1):34-39.

[27]Huang S C, Jiang C F. A color image authentication and recovery method using block truncation code embedding [J]. Journal of Marine Science and Technology, 2012, 20(1): 49-55.

[28]Zhang J, Zhang Q, Lv H. A novel image tamper localization and recovery algorithm based on watermarking technology[J]. Optik-International Journal for Light and Electron Optics, 2013, 124(23): 6367-6371.

[29] 和红杰，张家树. 区分不同篡改的脆弱水印算法[A]. 第十二届全国图像图形学术会议[C], 2005, 140-144.

[30] 和红杰，张家树，田蕾. 能区分图像或水印篡改的脆弱水印方案[J]. 电子学报，2005,

33 (9): 1557-1561.

[31] 张鸿宾, 杨成. 图像的自嵌入及窜改的检测和恢复算法[J]. 电子学报, 2004, 32(2): 196-199.

[32] 钱振兴, 程义民, 王以孝等. 一种图像自嵌入方法[J]. 电子学报, 2006, 34(7): 1347-1350.

[33] Lin P L, Hsieh C K, Huang P W. A hierarchical digital watermarking method for image tamper detection and recovery[J]. Pattern Recognition, 2005, 38(12):2519-2529.

[34] 刘泉, 江雪梅. 用于图像篡改定位和恢复的分层半脆弱数字水印算法[J]. 通信学报, 2007, 28(7):104-110.

[35] Lee T Y, Lin S D. Dual watermark for image tamper detection and recovery[J]. Pattern Recognition, 2008, 41(11):3497-3506.

[36] Yang C W, Shen J J. Recover the tampered image based on VQ indexing[J]. Signal Processing, 2010, 90(1): 331-343.

# 五、导师评语