用于图像认证的半脆弱数字水印技术综述

汤文明,李海华
(杭州电子科技大学 图形图像研究所,浙江 杭州 310018)

摘要:伴随着计算机网络技术及多媒体信息技术的快速发展,为了更好地保护数字图像,用于图像认证的数字水印技术已成为当今研究的热点。主要阐述了当前用于图像认证的半脆弱水印技术研究现状及其攻击行为分析,并对几种典型的图像认证算法进行了介绍。通过对典型认证算法的研究,提出了认证数字水印技术未来的发展方向。

关键词:图像认证;半脆弱水印;篡改检测
中图分类号:TP391.41 **文献标识码**:A **文章编号**:1001 - 4551(2010)04 - 0115 - 04

Overview on semi-fragile watermarking for image authentication

TANG Wen-ming, LI Hai-hua
(Institute of Graphics and Image, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: With the development of internet and digital of the multimedia information, and aiming at protecting the digital image, digital watermarking used for image authentication has been a hot research focus. The image authentication's research status and analysis of aggressive behavior were proposed. Some typical image authentication algorithm was introduced. Finally, the future development of digital watermarking for image authentication was also analyzed and conjectured.

Key words: image authentication; semi-fragile watermarking; tamper detection

0 引 言

数字水印作为数字媒体作品知识产权保护的一种有效手段,目前得到了广泛地研究和发展[1-2],并已成为国际学术界研究的一个热点[3]。对数字产品的保护技术一般可以分为两类:①用于版权保护的鲁棒数字水印技术;②用于内容完整性、真实性认证的脆弱性数字水印和半脆弱性数字水印。

图像认证系统在医学、商业、军事、法律和新闻等方面均有很广泛地应用。认证水印是利用人类视觉系统的冗余,在不影响数字媒体感官质量的前提下,将与媒体内容相关或不相关的标志信息作为水印嵌入到媒体内容中,当媒体内容需要认证时,则可以根据提取的信息来判断其是否真实完整。计算机网络传输中的图像认证一般分为两类[4]:精确认证和模糊认证。精确认证是把图像作为一个整体,对图像的任何篡改均是不允许的,即使图像有 1 byte 的改动,图像就不能通过认证系统的认证,它使用脆弱水印技术。而模糊认证则是指在图像作品内容基本不变的情况下,允许作品有一定程度的失真,但对图像内容有明显改变的恶意篡改则不允许通过认证,它使用的是半脆弱水印技术。半脆弱水印对保持图像内容基本不变的操作(如图像压缩、图像滤波和图像增强等)是鲁棒的,而对恶意篡改是脆弱的。它作为一种新的数字媒体认证技术,在近几年得到了迅速的发展,并在数字音频和视频领域也有了很大的发展。

本研究首先对图像认证半脆弱水印技术的研究目标和研究现状进行叙述和分析,并指出当前典型的半脆弱水印算法的不足之处,然后对半脆弱认证系统的攻击行为进行分析。最后展望图像水印认证技术未来的发展方向。

1 半脆弱水印技术发展和研究现状

在实际应用中,数字图像因其数量较大,通常以压缩方式存储或传输,同时图像处理软件各异,图像格式众多,最终用户所要认证的图像通常是原始图像经有损

收稿日期:2009 - 07 - 23
作者简介:汤文明(1981 -),男,福建漳州人,主要从事信息处理、计算机安全方面的研究. E-mail: tangwm@hundsun.com

压缩或其它常规操作处理后的图像,因此,对图像进行内容认证的半脆弱水印技术在现实生活中更为实用。

基于半脆弱水印的认证系统应满足以下基本要求[5]:①鲁棒和脆弱的兼备性;②感知的透明性;③篡改的敏感性和可定位性;④盲检测性;⑤安全性。

半脆弱数字水印技术按照其实现方法可以分为空域法和变换域法。

1.1 空域法研究现状

空域水印算法是直接通过修改数字图像中的像素位来嵌入水印,如直接修改图像像素的最低位[6],该方法嵌入的水印信息量极少,算法鲁棒性差。水印信息很容易被滤波、图像压缩、图像增强等图像常规操作所破坏。Lin 等人通过改进空域像素之间的相关性来提高检测效果[7]。通过密钥在原始图像 DCT 域提取伪随机的白噪声序列,并将此序列作为认证序列,把序列叠加到每个 8 ×8 的 DCT 块上的三角矩阵中,对 DCT 块反变换,并结合水印强度因子合成含水印图像。通过提取水印和原伪随机序列求相关来完成认证。此算法的优点是对于 JPEG 有损压缩后被篡改的图像检测准确率很高,但是对于图像边缘和纹理较多的情况,算法的检测率较低,可靠性差。

Dittmann[8]把提取的图像边缘特征作为水印信息嵌入到空间域中,并采用 Canny 边缘检测器,通过比较被测图像边缘和提取的水印信息是否一致来判断图像的真伪。该算法对图像常规操作和恶意篡改识别能力强,但对各种压缩操作表现敏感。此外,算法的缺点还有误检率比较高,因为一旦图像被篡改,它的特征会随之改变,因而造成比较时的错误。

1.2 变换域法研究现状

变换域半脆弱水印算法首先对图像进行可逆变换,然后修改变换域系数来实现水印的嵌入。变换域数字水印嵌入的信息量大、安全性高。目前大多数半脆弱水印算法采用 DCT 和 DWT 变换。

1.2.1 DCT 域半脆弱水印算法

通常基于 DCT 域的半脆弱水印算法是为了抵抗 JPEG 压缩而提出的。Tewfik[9]提出一种基于 JPEG 编码方法的半脆弱水印技术。该方法首先对原始图像的每个 8 ×8 图像块进行 DCT 变换,然后把各个图像块的信号排序,用 Hilbert 对照 JPEG 量化表把向量分解成更小向量,再把子向量纵排列成 Hadamard 矩阵,采用 Zig-Zag 扫描法选取 DCT 系数进行奇偶性量化,将调制后的 DCT 系数嵌入水印的图像块,结合图像块形成含水印图像。最后通过比较待测图像的量化系数与原图量化系数的奇偶性相符情况来完成认证。此算法的优点是脆弱性强,但对常规处理反应敏感,因为系数的奇偶性在图像经过一般信号处理后是很容易改变的。Lin 和 Chang[10]给出了一种可以在一定程度上抵抗 JPEG 压缩和剪裁与替换操作的半脆弱水印技术,该技术可以识别被篡改的块的位置,并且可以利用来自原图的一个粗糙图像来对篡改块进行恢复。所提出的算法基于 JPEG 压缩前后 DCT 系数的两个不变特性:①如果 DCT 系数被修改为 JPEG 量化步长的倍数,那么在未来的 JPEG 压缩中,该系数可以被确切重建;②JPEG 变换前后两个 8 ×8 子块相同位置的系数关系保持不变。算法利用第 2 个特性来形成认证信号,而利用第 1 个特性来将其嵌入到 DCT 系数中。此算法的优点是虚警率近似为 0,并且对大多数攻击,比如去噪、剪切、直方图均衡等检测效果好,抗 JPEG 压缩能力强。但算法的篡改定位精度不够准确,只能定位到以块为单位的地方,而对于图像部分像素的改变无法准确定位,且恢复后的图像质量较差。Yang S Y[11]等人对原始图像进行每个 8 ×8 块 DCT 变换,然后选择嵌入水印区域和嵌入系数,计算差别敏感门限值、修正系数,以实现水印的生成和嵌入。认证时根据 Zig-Zag 系数序列和 DCT 系数之间的关系对 DCT 反变换以复原原始图像。此算法对常规操作和 JPEG 有损压缩有良好特性。Lin 和 Podilchuk[12]等人在原始图像的每个 8 ×8 块的 DCT 中频系数上叠加不同的伪随机序列,由于自然图像中一般平滑区较多,边缘区较少,认为在没有边缘存在的情况下,图像相邻像素差值信号的能量主要是由水印引起的,通过一个改进的运算来进行图像认证。算法的优点是对于 JPEG 有损压缩后被篡改的图像检测准确率很高,但算法的检测率较低,可靠性也差。

1.2.2 DWT 域半脆弱水印算法

由于小波变换在时域和频域具有良好的局部定位性质,同时与现有的图像压缩标准 JPEG2000 相融合,故小波域的数字水印认证技术有更高的实用意义。Walton[13]和 Kundur、Hatzinakos[14]通过量化小波系数来嵌入水印,利用小波空间域和频率域定位出篡改位置,并估计当前图像被篡改的程度。其中 Walton 对图像的 Haar 小波系数进行量化,根据量化步长的大小来控制水印的鲁棒性,最后利用攻击估计函数将图像遭受的恶意篡改与非恶意篡改区分开来。该算法的缺点是很容易让攻击者得知量化步长而改变图像内容使提取的水印信息仍然不变。Lu 把小波系数分成掩蔽门限位(MTUS)[15]。在频域中选择同尺度、同方位、绝对值大于 JND 阈值的小波系数,再用 CWS(Cocktail Wa-

termarking Scheme)调整小波系数来完成水印嵌入,用原量化信息作为密钥储存来恢复原图像。两种水印的检测分别进行以完成认证。该方法考虑了两种水印的特性,但是用作图像认证时,必须存储原始水印信息,并且鲁棒性和脆弱性的均衡也是问题。Yu[16]等人通过量化小波系数的加权平均值来嵌入水印,认为小波系数的变化服从高斯分布,对图像进行恶意攻击导致小波系数的变化往往具有较大的方差,而由偶然因素造成图像失真引起的系数变化往往具有较小的方差,从而将恶意篡改与非恶意篡改区分开来。这种方法较直接量化小波系数有更好的鲁棒性。Paquet[17]等人则结合人类视觉系统(HVS)量化小波包系数。算法首先生成 ID 序列,并用此 ID 序列选择小波变换函数及分解层数;再运用此 ID 选择所要嵌入水印的区域及系数,利用人类视觉掩蔽特性,对不同的小波系数选择合适的嵌入强度,完成小波包系数的量化,经小波包重构后得到含水印的图像;最后用密钥提取水印,结合小波包系数区域分别进行带内频域及带间空域比较以得出认证结果。不足的是很容易遭遇搜索攻击。Eggers[18]先用 Scalar 编码器对水印信息进行编码得到二值码书,再用对应的 Scalar 量化函数对所选 8 ×8 块中的系数量化,即可把码书的各个元素嵌入相应宿主信息内。同时为了使嵌入随机化,嵌入过程引入二值伪随机序列。认证时,把待测信息与相应的量化函数、步长因子、二值伪随机序列混合运算得到验证值,若无水印则验证值近似为 0,若水印存在,其验证值绝对值应接近步长因子值的一半,通过这种方法完成认证。该算法的优点是虚警率为 0,且对 JPEG 压缩稳健,能容忍一般的图像处理操作,缺点是对直方图均衡及锐化敏感。Hu[19]提出由低频率领域图像特征产生二值水印进行水印生成及嵌入,由小波变换逆变换进行认证。此算法定位能力强,并具有篡改恢复能力,其安全性也较高,但对图像旋转等变形失真非常敏感。

1.2.3　其它算法

Lu[20]提出了一种基于矢量量化的半脆弱水印技术。先把原图编码分块,用图像尺寸和图像块大小进行相关运算并结合特定二值图像产生水印,把图像块视为矢量,用 VQ 编码器在码书中找出最佳匹配码字代替该图像块;用码字下标代替输入矢量作为索引值反向到解码器,对应每个水印比特在码书里寻找合适的匹配码字,并把水印嵌入图像块相应位置中。检测时,把待测图像分块,用解码器根据索引值用同一码书重构相应的码字,根据嵌入规则提取水印。此法对 JPEG 与 VQ 压缩具有较好性能。但对于拼贴攻击等的抗攻击能力相对比较薄弱。

2　半脆弱水印攻击行为分析

半脆弱水印主要是用来保证数字媒体内容的真实性,其本身具有一定脆弱性的。因此施加于鲁棒水印的攻击方法(如简单攻击、同步攻击、削去攻击、混沌攻击)对认证水印系统的影响是不明显的,而试图篡改图像内容却不损坏水印信息的“伪认证”攻击对认证水印系统的影响就较为明显了。目前,关于数字水印认证技术的攻击算法文献较少,就已有的文献来看,主要有以下几种:

(1) VQ 攻击。该类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。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对篡改定位能力有一定影响。

(2) Holliman-Memon 攻击。密钥资源有限且密钥与图像无关致使当密钥和所嵌入的水印完全相同的情况下,同一位置可能会隐藏相同的水印信息,所以攻击者可以交换两个可信图像同一位置的图像块而不会影响提取的水印信息。

(3) 密码分析攻击。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。

(4) 特征选取攻击。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。

3　半脆弱水印的发展方向

半脆弱水印技术作为数字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,尚有许多亟待解决和进一步深入研究的问题,笔者认为未来半脆弱水印技术应在以下几个方面展开探讨和研究:

(1) 与图像压缩编码算法相结合。数字图像的网络传输发布都需要经过压缩编码(JPEG 或 JPEG2000 等)。目前针对 JPEG 编、解码器设计的水印认证算法已有文献报道,但针对 JPEG2000 及其它图像压缩标准的认证水印算法还很少。随着 JPEG2000 编码标准的不断成熟,研究能够抵抗 JPEG 及 JPEG2000 压缩的半脆弱水印图像认证算法具有深远的意义。

(2) 安全性问题。安全性一直是水印认证系统的关键。实际应用对水印的保密安全有不同程度的要求。现有的半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。对于一个完善的半脆弱水印认证系统,系统的各个环节,如密码的产生、发布和管理都是必不可少的考虑因素。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。

(3) 完善音频、视频水印认证体系。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。

(4) 水印认证系统评测标准的建立。水印认证技术要得到广泛的应用,就必须建立一套完善的标准,以适应各种不同水印认证系统的质量评测。同时需要解决许多的社会和法律问题。

4 结束语

图像认证的数字水印技术是当今研究的热点。本研究分析了图像认证半脆弱水印技术的研究目标和研究现状,介绍了几种典型的图像认证算法,重点指出了当前典型的半脆弱水印算法的不足之处,并对半脆弱认证系统的攻击行为进行分析,最后提出了认证数字水印技术未来的发展方向。

参考文献(References):

[1] 孙圣和,陆哲明. 数字水印技术及应用[M]. 北京:科学出版社,2004.

[2] COX I J, MILLER M L, BLOOM J A. 数字水印[M]. 王 颖,黄志蓓,译. 北京:电子工业出版社,2003.

[3] LU CS, LIAO HYM. Multipurpose watermarking for image authentication and protection[J]. **IEEE Transactions on Image Processing**,2001,10(10):1579 – 1592.

[4] 秦 川,单承赣. 一种基于半脆弱水印的图像内容认证算法[J]. 信息安全与通信保密,2005,5(6):66 – 69.

[5] 朱晓冬. 数字水印技术的研究[D]. 长春:吉林大学计算机科学与技术学院,2004.

[6] 张鸿宾,杨 成. 基于 JPEG 压缩不变量和半脆弱水印的图像认证[J]. 北京工业大学学报,2004,30(11):214 – 218.

[7] LIN E T, PLD ILCHUK C I, DELP E J. Detection of image alterations using semi-fragile watermarks[C]//Proceedings of SPIE Conference on Security and Watermarking of Multi-media Contents Ⅱ. San Jose, California, USA,2000:152 – 163.

[8] DITTMANN J. Contentfragile watermarking for image authentication[C]//Proceedings of SPIE Conference on Security and Watermarking of Multimedia Contents Ⅲ,2001:175 – 184.

[9] MANSOUR M F, TEWFIK A H. Robust High Capacity Data Embedding[C]. ICASSP 2001, Utab, 2001.

[10] LIN C Y, CHANG S F. Semi-fragile watermarking for authenticating JPEG visual content[C]//Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents II, San Jose, USA,2000:140 – 151.

[11] YANG S Y, LU Z D, ZOU F H. A Novel Semi-fragile Watermarking Technique for Image Authentication[C]//ICSP Proceedings,2005:2282 – 2285.

[12] LIN E T, PODILCHUK C I, DELP E J. Detection of image alterations using semi-fragile watermarks[C]//Proc. of SPIE Security and Watermarking of Multimedia Contents II, San Jose,2000:152 – 163.

[13] WALTON S. Information authentication for a slippery new age[J]. **Dr. Dobb's Journal**,1995,20(4):18 – 26.

[14] KUNDUR D, HATZINAKOS D. Digital watermarking for telltale tamperproofing and authentication[C]//Proceedings of the IEEE Special Issue on Identification and Protection of Multimedia Information,1999:1167 – 1180.

[15] LU C S, HYM L I. Multipurpose watermarking for image authentication and protection[J]. **IEEE Transactions on Image Processing**,2001,10(10):1579 – 1592.

[16] YU G J, LU C S, LIAO H Y, et al. Mean quantization blind watermarking for image authentication[C]//IEEE International Conference on Image Processing, Vancouver BC, Canada,2000:706 – 709.

[17] PAQUET A H, WARD R K. Wavelet-based Digital Watermarking for Image Authentication[C]. Proceedings of 2002 IEEE Canadian Conference on Electrical and Computer Engineering,2002.

[18] EGGERS J, GIROD B. Blind watermaking applied to image authentication[C]//Proceedings of IEEE ICASSP, Salt lake city, UT,2001:7 – 11.

[19] HU Y P, HAN D Z. Using Two Semi-fragile Watermark for Image Authentication[C]//Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, IEEE2005,2005:5481 – 5489.

[20] LU Z M, LIU C H, XU D G, et al. Semi-fragile image watermarking method based on index constrained vector quantization[J]. **Electronics Letters**,2003,39(1):35 – 36.

[编辑:张 翔]