

一种像素级的图像篡改认证算法

谢建全^{1,2}, 阳春华¹

(1. 中南大学 信息科学与工程学院, 湖南 长沙 410083;

2. 湖南财经高等专科学校 信息管理系, 湖南 长沙 410205)

(xiejianquan@sina.com)

摘 要:提出了一种用于图像内容像素级篡改认证的脆弱水印算法,能准确识别图像中被篡改的像素点,并且能容忍图像传输过程中出现的个别认证信息位的传输错误。这种算法以向左右和上下各扩展若干个像素点来产生认证信息,结合周围像素来确定嵌入的水印比特而引入基于图像的不确定性的某些算法,常常会出现被篡改的单个像素点不能被准确定位的问题,本算法可有效解决这一问题,认证信息进行加密后再进行嵌入,解决了多数像素级认证算法所出现的安全漏洞问题。

关键词:脆弱水印;图像认证;篡改检测

中图分类号: TP393.08 **文献标识码:** A

Pixel-level image modification authentication algorithm

XIE Jian-quan^{1, 2}, YANG Chun-hua¹

(1. School of Information Science and Engineering, Central South University, Changsha Hunan 410083, China;

2. Department of Information Management, Hunan Finance and Economics College, Changsha Hunan 410205, China)

Abstract: A fragile watermarking algorithm used to authenticate the pixel-level modification to the image content was proposed, which can distinguish pixel points modified in image and tolerate transmission errors of individual authentication information bit in the process of image transmission. In this algorithm, authentication information was produced by extending more and more pixel points toward both left-right and up-down, which ensured the accurate location of individual juggled pixels resulting from some algorithms based on image uncertainty by combining surrounding pixels to determine the embedded watermarking bits. Authentication information was embedded after encryption to avoid most problems of security bugs caused by most pixel-based authentication algorithms.

Key words: fragile watermarking; image authentication; modification detection

0 引言

随着多媒体技术的飞速发展和日趋完善,人们可以方便地对数字媒体,如音乐、视频或图像等进行篡改。然而各类无意或有意的篡改攻击可能会导致严重的后果,比如医学图像中的任意一点变化都可能使医生对病情做出错误的判断;作为法庭举证的照片如果被恶意修改后可能扭曲事件的真实情况,现在多媒体照片一般不能作为法庭的证据,就是因为多媒体照片可以很容易地被各种软件修改而不被发现;如果篡改涉及到国家安全、新闻图片、历史文献等重要内容时,可能会造成不良的社会影响或重大政治经济损失。因此,人们在使用数字媒体(数字图像、数字音频、数字视频)时,常常对其完整性、内容的真实性产生质疑。为解决这一问题,需要一种新的技术来保护数字媒体。作为数字水印技术的一个分支,认证水印(authentic watermarking)通过嵌入水印来达到这一目的。

基于数字水印的认证为图像数据的真实性与完整性认证,以及篡改检测提供了一个简便的工具。与传统的基于数字签名的数据认证相比,基于数字水印的数据认证的主要优

点在于不需要额外的附加认证信号,水印信息离散地分布到数字媒体的各个部分,不是附加在数字图像的后面,从而提高了攻击难度,增加了安全性。到目前为止已经提出了很多种认证水印技术,大体上分为易损水印和半易损水印两类。易损水印是一种在数字图像作品发生任何形式的改变时都会损坏的水印;半易损水印一般指能承受图像进行诸如 JPEG 压缩、加少量噪声的偶然修改,但会被图像内容的恶意篡改损坏的水印。由于一些图像处理操作,例如 JPEG 压缩并不破坏图像的内容,因此需要半易损水印对破坏图像内容的一些蓄意操作进行检测。但是在某些应用中,例如法律证据图像,医疗图像等细节敏感图像,甚至 JPEG 压缩也会破坏图像的细节。易损水印正是针对这一类应用而设计的,在没有附加额外信息的情况下,易损水印能够检测出篡改位置或区域。

1 像素级的篡改定位算法

迄今为止,图像篡改认证根据其篡改的定位能力可以分为三种^[1]: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[2]等;另一种是分块级的定位能力,即未被篡改的最小单元

收稿日期:2006-12-06;修订日期:2007-02-12

基金项目:国家自然科学基金资助项目(60574030);湖南省教育科学“十一五”规划课题(XJK06CXJ012)

作者简介:谢建全(1964-),男,湖南双峰人,教授,博士研究生,主要研究方向:信息安全技术; 阳春华(1965-),女,湖南双峰人,教授,博士生导师,主要研究方向:优化理论。

是一个图像块,这种认证也称为分块认证,比如文献[3]等,为了尽可能定位准确,分块应尽可能小;还有一种是没有任何篡改定位能力,只能简单确定图像是否被篡改,这种认证称为无分割认证。像素级篡改定位能确定到每一个像素是否被篡改,因此定位能力最强,并且在受到篡改时能容易区分是有意篡改还是在传输过程中受到了干扰,但相对来说要实现鲁棒的像素级定位也是最困难的。

1997年,文献[2]提出一种基于易碎数字水印的单像素认证算法,该算法使用二值函数把灰度值0到255映射为0或1,并以一个二维标识作为水印,修改图像的像素灰度使其映射值与标识中相应的比特相同从而嵌入易损水印,从而可以将篡改定位精确到一个像素点。此后,文献[4,5]均发现文献[2]的这种技术具有致命的缺点,即每一个像素中嵌入的水印比特具有确定性。由于嵌入的水印比特具有确定性的原因,这时总可以发起攻击来伪造一幅能通过认证的图像^[6],也就是说攻击者可以轻易地伪造加水印图像来实施攻击。针对这一缺点,文献[7]提出了一种易损水印方案,通过结合周围像素来确定嵌入的水印比特,从而引入基于图像的不确定性。这种方法的优点是可以抵制文献[4,5]中提出的攻击,但是其局部检测性能会有所下降。本文在文献[7]的基础上进行改进,提出一种新的认证算法,它既可以抵抗文献[4,5]中提出的攻击,又能对被篡改的单个像素点进行准确定位。

设图像为 $I_{M \times N}$,其中 M 和 N 分别表示图像的高和宽,其中的各像素点表示为 $x_{i,j}$,其中 i,j 为相应的像素点的坐标, $i \in \{1,2,\dots,M\}$, $j \in \{1,2,\dots,N\}$ 。本文算法的关键是一个像素点的认证信息 $w_{i,j}$ 是与该像素点上下和左右各 k 个点共计 $2k+1$ 个像素的函数,即:

$$w_{i,j} = f(x_{i-k,j}, x_{i-k+1,j}, \dots, x_{i-1,j}, x_{i,j}, x_{i+1,j}, \dots, x_{i+k,j}, x_{i,j-k}, \dots, x_{i,j-1}, x_{i,j+1}, \dots, x_{i,j+k}) \quad (1)$$

式中 $f(\cdot)$ 为二值映射函数,其结果为0或1, $k \geq 1$, k 的具体值可根据需要引入的图像的不确定性程度而定,一般取 $k=2$ 就可满足要求。当(1)式中像素点的第一个坐标 $i-y \leq 0$ 时($1 \leq y \leq k$),将相应的坐标值加 M ;当 $i+y > M$ 时($1 \leq y \leq k$),将相应的坐标值减 M ,即保证相应的坐标值在 $\{1,2,\dots,M\}$ 范围之内;对第二个坐标也进行同样的处理,只是将 M 改为 N ,保证第二个坐标相应的值在 $\{1,2,\dots,N\}$ 范围之内。这实际是以 (i,j) 为中心,向左右和上下各扩展 k 个像素点来产生认证信息,从而引入基于图像的不确定性,达到抵抗文献[4,5]中提出的攻击。

由算法可知,当一个像素点被篡改时,会导致多个像素点的认证信息的改变,表面上会导致不能准确定位被篡改的像素点,实际上这种一个像素点的改变会导致多个像素点的改变是有规律的,向左右和上下各扩展的 k 个像素点都会发生认证信息的改变。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。即一个点是否被篡改可以用下式表示:

$$T_{i,j} = \prod_{y=-k}^k A_{i,j+y} \cdot \prod_{z=-k}^k A_{i+z,j} \quad (2)$$

其中 $A_{i,j}$ 表示接收到的图像的像素点 (i,j) 的认证信息与

重新计算的认证信息的比较结果,不一致为1,一致为0,式中的坐标值超出范围时,按前面说明的方法处理。当 $T_{i,j}$ 为1时,表示坐标为 (i,j) 的像素点受到篡改,否则未被篡改,从而能将篡改准确定位到被篡改的像素点。

2 水印信息的嵌入与认证检测

在进行水印嵌入时,可能会对图像造成影响。对于版权认证等方面的水印只要满足人类视觉不可见性要求即可,但对篡改定位就不一定适用。特别对于精确认证而言,应该能够消除水印嵌入给需认证的内容带来的影响。为解决这个问题,通常的做法是将图像分为两部分,一部分用于认证,一部分则被修改以承载水印。LSB算法是一种非常成熟的算法,它算法简单、可嵌入容量固定和嵌入速度快,所修改的是人类视觉系统无法感知的最低位(LSB位),目前几乎全部的隐写算法中都可以找到LSB算法的影子。虽然LSB算法容易受到噪声干扰和滤波等影响,鲁棒水印和半脆弱水印使用LSB算法隐藏信息可能遇到问题,但作为脆弱水印使用LSB算法是完全可以的,因此本算法采用LSB算法嵌入认证水印信息,将图像的高7位用于认证,而用最低位承载水印。但为了保证认证系统的安全性,在水印的嵌入时还需进行一定的处理。

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,如果嵌入方法存在漏洞,恶意的攻击者可能利用这些漏洞来修改或伪造图像而不被认证算法发现,达到欺骗认证系统的目的。在上面表述的算法中,如果映射函数 $f(\cdot)$ 被攻击者所掌握,攻击者就可能在篡改图像后再通过重新计算认证信息来伪造图像,但却能通过认证。为保证安全性,本算法对需嵌入的水印信息使用混沌密码等序列密码进行加密后再嵌入。因此水印信息的嵌入主要分三步进行:

- 1) 利用映射函数 $f(\cdot)$ 计算每个像素点高7位的认证信息 $w_{i,j}$,得到长度为 $M \times N$ 的二进制认证序列 W ;
- 2) 选用某种混沌加密算法 E 在密钥 key 的控制下对二进制认证序列 W 进行加密,得到加密的长度仍为 $M \times N$ 的嵌入信息 EW ;
- 3) 将加密后的二进制序列 EW 按顺序嵌入到图像 $I_{M \times N}$ 的最低位(LSB)。

在进行认证检测时,基本上是嵌入的逆过程,但是否被篡改需要按(2)式进行计算和判断,它可以分为如下4个步骤:

- 1) 利用 $f(\cdot)$ 计算接收到的图像 I^* 每个像素点高7位的认证值 $w_{i,j}^*$,形成一个长度为 $M \times N$ 的二进制认证序列 W^* ;
- 2) 使用与嵌入认证信息时相同的密钥 key 和加密算法 E 对 W^* 进行加密计算,得到二进制序列 EW^* ;
- 3) 将二进制序列 EW^* 按顺序与图像 I^* 的最低位进行比较,如果相同则 $A_{i,j} = 0$,否则 $A_{i,j} = 1$;
- 4) 按(2)式计算各像素点对应的 $T_{i,j}$,若 $T_{i,j} = 1$,则认为坐标为 (i,j) 的像素点被篡改,否则相应的像素点没有被篡改。

3 仿真试验

为验证本算法的性能,以 $256 \times 256 \times 8$ 的原始Lena灰度图像进行试验,对嵌入认证水印信息的图像的部分像素点进行篡改,然后进行认证处理。嵌入认证信息的图像如图1(b)所示,可见嵌入的认证信息具有很好的视觉不可见性,然后对嵌入认证信息的图像中进行篡改,本试验采用在图像中加入

(下转第1342页)

class SVM 模型在纯净数据集训练下都具有较高的识别准确率。传统 SVM 对噪声十分敏感,在添加噪声之后识别率大大降低。而 RSVM 和 One-class SVM 在有噪声和无噪声情况下检测准确率变化不大,并且在经过 Online training 算法的改进后,训练样本数据集处于实时更新的状态,因此对于检测准确率都有一定的提升。

3)除了上面讨论的虚警率和识别正确率之外,另一个值得关注的问题是训练时间。理想的入侵检测系统应该在尽可能短的训练时间内保持尽可能高的检测准确率。当进行新样本分类时,SVM 的计算复杂性与支持向量个数成线性比,因为 RSVM 的支持向量个数比标准 SVM 和 One-class SVM 少,所以 RSVM 只需更少的运行时间。此外通过引入 Online training 算法,在本实验中入侵检测系统能够实时处理不断更新的训练样本,节省了训练时间。在表 3 中,我们列出了实验中三种 SVM 的向量个数和训练时间百分比,其中以 Online training 算法改进前的 SVM 训练时间为基准,设为 100%。由表 3 可知,经过 Online 算法的改进,SVM、RSVM 和 One-class SVM 在训练时间方面都有显著的改善。尤其在有噪声数据训练时,RSVM 和 One-class SVM 时间开销比 SVM 算法小。

4 结语

本文将 Online training 算法引入 SVM、RSVM 和 One-class SVM 中进行异常检测。实验数据集采自 1999DARPA 数据集,实验表明在经过 Online training 算法改进的三种 SVM 在所需的支持向量数目和训练时间上有一定幅度的下降,而且检测正确率有所提高。这说明 Online SVM 在有效检测和推广能力等性能上有着良好的表现,并且由于样本训练过程的实时性,使得基于 SVM 的入侵检测系统成为一个实时系统,

更加符合现实工作的要求,因此它在网络安全中的应用将会越来越广泛。

参考文献:

- [1] 李辉,官晓宏.基于支持向量机的网络入侵检测[J].计算机研究与发展,2003,40(6):799-807.
- [2] ESKIN E, ARNOLD A. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data [EB/OL]. <http://www1.cs.columbia.edu/ids/publications/>, 2006.
- [3] FRIEDMAN J. Another approach to polychotomous classification [EB/OL]. <http://www.stat.stanford.edu/reports/friedman/>, 2006.
- [4] LAU KW, WU H. Online training of support vector classifier[J]. Pattern Recognition, 2003, 36(8): 1913-1920.
- [5] SONG Q, HU WJ, XIE WF. Robust support vector machine for bullet hole image classification[J]. IEEE Transactions on Systems, Man and Cybernetics, 2002, 32(4): 440-448.
- [6] SCHOLKOPF B, PLATTZ JC. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443-1471.
- [7] HU WJ, LIAO YH, VEMURI VR. Robust support vector machines for anomaly detection in computer security[J]. The 2003 International Conference on Machine Learning and Application. Los Angeles, California, 2003, 14(6): 1449-1459.
- [8] WANG YX, JOHNNY W, ANDREW M. Anomaly intrusion detection using one class SVM[A]. Proceeding of the Fifth Annual IEEE System, Man and Cybernetics Information Assurance Workshop[C]. New York, USA, 2004. 358-364.
- [9] MIT Lincoln Laboratory [EB/OL]. <http://www.ll.mit.edu/IST/ideval/>, 2006.

(上接第 1338 页)

文字“lena”,篡改后的图像如图 1(c)所示,再进行认证检测,检测结果如图 1(d)所示,其中用白色表示被篡改的像素点,可见本算法能准确发现被篡改的像素点,而不是多数算法那样只能定位一个一定大小的被篡改的像素块。为试验本算法的抗干扰能力,对嵌入认证水印信息的图像的 LSB 位加入 1% 的均衡噪声后进行认证处理,图像能通过认证,未报告有被篡改的像素点。由此可见,该算法能够完成像素级的篡改检测,并且对最低有效位的噪声干扰具有一定的鲁棒性。



图 1 篡改认证检测结果

4 结语

本文提出了一种用于图像内容像素级篡改认证的脆弱水印算法,能准确识别图像的被篡改的像素点,同时解决了文献[4]、[5]所提到的多数像素级认证算法所出现的安全漏洞问题;另外算法还能容忍图像传输过程中出现的个别认证信息位的传输错误,即不会将认证信息位出现传输错误的像素点错误判定为被篡改。该算法的主要特点包括以下几个方面:1)一个像素点的认证信息与该像素点上下和左右各 k 个点相关;2)当一个像素点及其上下和左右各 k 个像素点的认证值都出现错误时,相应的像素点被认为受到篡改,否则没有被篡改;3)认证信息经过加密后再进行嵌入。在图像的篡改认证

中,有时还需要对被篡改的数据进行恢复,本算法暂时还没有涉及篡改的恢复问题,这是下一步要解决的问题。

参考文献:

- [1] 吴金海,林福宗.基于数字水印的图像认证技术[J].计算机学报,2004,27(9):1153-1160.
- [2] YEUNG M, MINTZER F. An invisible watermarking technique for image verification[A]. Proc. IEEE International Conference on Image Processing[C]. Santa Barbara, USA, 1997, 2: 680-683.
- [3] WONG P, MEMON N. Secret and public key image watermarking schemes for image authentication and ownership verification[J]. IEEE Transactions on Image Processing, 2001, 10(10): 1593-1601.
- [4] HOLLIMAN M, MEMON N. Counterfeiting attacks for block-wise independent watermarking techniques[J]. IEEE Trans. on Image Processing, 2000, 9(3): 432-441.
- [5] FRIDRICH J, GOLJAN M, MEMON N. Further attacks on Yeung-Mintzer fragile watermarking scheme[A]. In: Proc. SPIE Photonic West, Electronic Imaging 2000, Security and Watermarking of Multimedia Contents[C]. San Jose, California, January 24-26, 2000. 428-437.
- [6] ALBANESI MG, FERRETTI M, GUERRINI F. A taxonomy for image authentication techniques and its application to the current state of the art[A]. In: Proceedings of the 11th International Conference of Image Analysis[C]. Palermo, Italy, 2001. 535-540.
- [7] FRIDRICH J, GOLJAN M, BALDOZA AC. New fragile authentication watermark for images[A]. In: Proc ICIP[C]. Vancouver, Canada, September 10-13, 2000. 446-449.