计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191000060

基于复杂攻击的脆弱水印图像完整性认证算法

郑秋梅　刘　楠　王风华
中国石油大学(华东)计算机科学与技术学院　山东 青岛 266580
(zhengqm@upc.edu.cn)

摘　要　将图像用于司法、医学等重要领域时,往往需要对图像进行完整性认证,判断图像是否被恶意篡改。基于脆弱水印的图像完整性认证方法可以用来实现图像的篡改检测和篡改定位。针对脆弱水印在图像篡改检测中定位精度和抗复杂攻击不能同时满足的问题,提出了一种基于复杂攻击的脆弱水印图像完整性认证算法。向彩色图像 R,G,B 3个通道中嵌入相同的脆弱水印,目的是检测出图像任意通道的篡改。将图像以2×2大小分块来提高定位精度,采用块验证和组验证的认证方式来检测复杂攻击的图像篡改,使用非等长图像置乱变换提高算法的普适性和抗复杂攻击的能力。仿真结果表明,所提算法具有较好的不可见性,对简单攻击、复杂攻击和混合攻击类型下的图像篡改具有更高的定位精度。

关键词:脆弱水印;图像篡改;篡改检测;定位精度;复杂攻击
中图法分类号　TP309

Complex Attack Based Fragile Watermarking for Image Integrity Authentication Algorithm

ZHENG Qiu-mei,LIU Nan and WANG Feng-hua
College of Computer Science and Technology,China University of Petroleum Huadong,Qingdao,Shandong 266580,China

Abstract　When the image is used in judicial,medical and other importantfields,it is often necessary to authenticate the integrity of the image to determine whether the image have been tampered with maliciously. The authentication method of image integrity based on fragile watermarking can be used to detect and locate the image tampering. In order to solve the problem that the localization accuracy and anti-complex attacks of fragile watermarking in image tamper detection can not be satisfied simultaneously,a complex attack based fragile watermarking for image integrity authentication algorithm is proposed in this paper. The fragile watermarking is embedded into the color image's R,G and B channels to detect any channel tampering. In order to improve the localization accuracy,2 × 2 image blocks are divided. Block authentication and group authentication are used to detect the image tampering of complex attacks,and non-equilateral image scrambling transformation is used to improve the universality and anti-complex attacks ability of the algorithm. The simulation results show that the proposed algorithm has better invisibility and higher localization accuracy for image tampering under common attacks,complex attacks and multi attacks.

Keywords　Fragile watermarking,Imagetampering,Tamper detection,Localization accuracy,Complex attack

1　引言

互联网技术的发展促进了图像、视频等数字作品的传播,而不法分子对图像内容的恶意篡改会导致严重的后果。将图像用于司法、医学、电子商务或工程设计等领域时,需要对图像进行完整性认证,以判断图像是否被篡改。脆弱水印[1-2]可用来实现图像的篡改检测和定位[3-4]。基于脆弱水印的图像完整性认证方法是当前研究的热点。

脆弱水印算法根据篡改定位的最小单位分为像素级和分块级[5]。像素级水印算法定位精度高,可以精确定位简单攻击下的篡改,但对于拼贴攻击等复杂攻击具有一定的局限性[6]。分块级水印算法能够准确检测复杂攻击下的篡改,但定位精度与分块大小有关,有较大的虚警率。国内外学者在像素级和分块级篡改定位上进行了大量研究,以提高脆弱水印篡改定位精度和抗复杂攻击的能力。Yeung等[7]首先提出像素级脆弱水印算法,将水印嵌入图像最低位,实现简单且定位精度高,但无法抵抗复杂攻击。Chen等[8]通过比较像素邻域中不一致像素的数目来判定像素的有效性,以抵抗拼贴攻击,但无法抵抗均值攻击等复杂攻击。Munir等[9]对水印周期延拓后替换LSB平面,来检测简单攻击的篡改,该算法不受图像尺寸的限制。Tong等[10]将图像按2×2大小分块,结合最高有效位和最低有效位检测图像内容篡改,但无法抵抗

到稿日期:2019-10-12　返修日期:2020-03-08　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(51274232);山东省自然科学基金(ZR2018MEE004);中央高校基本科研业务费专项资金(19CX02030A)
This work was supported by the National Natural Science Foundation of China (51274232),Natural Science Foundation of Shandong Province,China (ZR2018MEE004) and Fundamental Research Funds for the Central Universities of Ministry of Education of China (19CX02030A).
通信作者:刘楠(liunan0919@126.com)

拼贴攻击。Kang 等[11]将图像按 4×4 大小分块，根据奇异值的变化检测图像篡改，能够抵抗多种复杂攻击，且漏检率较低，但其定位的最小篡改区域为 16 像素，存在较大的虚警率，由于采用 Arnold 变换置乱图像块，要求水印和宿主图像为等长图像。

由此可见，在宿主图像中仅嵌入脆弱水印无法实现复杂攻击下的图像篡改检测和定位，图像块的大小在一定程度上决定着篡改定位的精度。针对该问题，本文提出了一种基于复杂攻击的脆弱水印图像完整性认证算法。该算法将图像按 2×2 大小分块，以保证篡改定位精度，认证过程分为块认证和组认证来检测不同攻击类型的篡改，同时使用非等长图像置乱变换提高了算法抗复杂攻击的能力，也提高了算法对任意尺寸图像的适用性。

2 相关理论知识

本文算法使用非等长图像置乱变换置乱图像块，利用奇异值表征图像块信息，通过 Logistics 映射决定水印位的嵌入。

2.1 非等长图像置乱变换

Arnold 变换有良好的周期性，在数字水印方面得到了广泛的应用。Arnold 变换一般考虑等长图像，对于一幅 $N\times N$ 图像，通常 Arnold[12] 变换的定义如下：

$$\begin{bmatrix}x'\\y'\end{bmatrix}=\begin{bmatrix}1&1\\1&2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}(\mathrm{mod}\ N),x,y\in(0,1,\cdots,N-1) \tag{1}$$

在实际应用中，图像并非都是等长的。本文使用由 Arnold 变换推广的非等长图像置乱变换。对于一幅 $M\times N$ 图像，非等长图像置乱变换[13]的定义如下：

$$\begin{bmatrix}x'\\y'\end{bmatrix}=\begin{bmatrix}a&b\\c&d\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}\mathrm{mod}\begin{bmatrix}x\\y\end{bmatrix}$$
$$x\in\{0,1,\cdots,M-1\},y\in\{0,1,\cdots,N-1\} \tag{2}$$

其中，$\boldsymbol{A}=\begin{bmatrix}a&b\\c&d\end{bmatrix}$ 为置乱变换矩阵，a,b,c,d 为非负整数，且 $|\boldsymbol{A}|>0$。当 $\boldsymbol{A}$ 的元素满足 $a=1,b>0,c=nq,d=bc+1$ 时，对任意非等长图像进行置乱，其变化周期都存在。其中，q 为 $N/\gcd(M,N)$，b 和 n 为正整数。

使用非等长图像置乱变换可以提高算法的普适性，置乱后分组使组内图像块在空间上不连续，有利于通过组验证判断图像是否被篡改。

2.2 奇异值分解

奇异值分解[14-15]是一种将对称图像对角化的线性代数技术，对于大小为 $M\times N$ 的矩阵 $\boldsymbol{A}$，$\boldsymbol{A}\in R^{m\times n}$ 的奇异值分解的定义如下：

$$\boldsymbol{A}=USV^T \tag{3}$$

其中，$\boldsymbol{U}\in R^{m\times m}$ 为左奇异值矩阵，$\boldsymbol{V}\in R^{n\times n}$ 为右奇异值矩阵，$\boldsymbol{S}\in R^{m\times n}$ 为对角矩阵且满足：

$$\sigma_1\geqslant\sigma_2\geqslant\sigma_3\geqslant\cdots\geqslant\sigma_n\geqslant0 \tag{4}$$

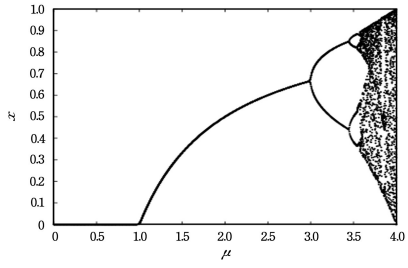
本文算法利用了奇异值的敏感性，根据图像块奇异值的变化检测图像篡改。

2.3 Logistic 映射

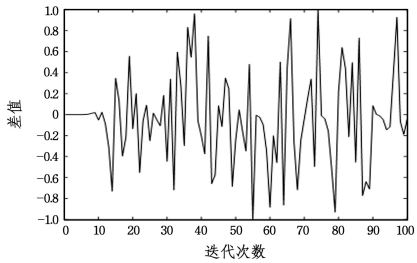
在数字水印技术中，使用混沌序列实现水印加密、图像置乱等功能。Logistic 映射[16]是经典的一维混沌序列，其定义如下：

$$x_{n+1}=\mu x_n(1-x_n),\mu\in[0,4],x_n\in(0,1),n=1,2,3,\cdots \tag{5}$$

图 1(a)为 Logistic 映射形态图，Logistic 映射对初值极其敏感，取 $\mu=4$，初始值分别为 0.3980 和 0.3981，两个序列的差值如图 1(b)所示。本文算法通过比较 Logistic 序列提取的水印与嵌入水印的异同，来检测图像是否被篡改。



(a) Logistic mapping diagram



(b) Sequences difference

图 1 Logistic 映射

Fig. 1 Logistic mapping

3 基于复杂攻击的脆弱水印图像完整性认证算法

本文宿主图像均是深度为 24 的彩色图像，在彩色图像 R、G、B 3 个通道中嵌入相同的脆弱水印，目的是检测出图像任意通道的篡改。为提高算法的普适性和抗复杂攻击的能力，采用非等长图像置乱变换；将图像按 2×2 大小分块，以保证较高的定位精度。

3.1 水印嵌入算法

水印嵌入过程如图 2 所示，嵌入算法主要分为 5 个步骤，具体如下。

步骤 1(图像分块) 将大小为 $M\times N$ 的彩色图像 I 的 R、G、B 通道分离，分离后图像分别为 I_r、I_g、I_b，向 3 个通道嵌入相同的脆弱水印，使算法既能检测多通道篡改，也能检测单一通道篡改。以 R 通道为例，将图像 I_r 划分为 2×2 的非重叠图像子块，记为 $Block$。对于分块级篡改定位算法，图像分块大小决定了篡改检测的最小单位，对图像进行 2×2 分块可以在一定程度上减小虚警率，从而提高篡改定位精度。

步骤 2(块置乱变换) 分块后图像块数为 $\frac{M}{2}\times\frac{N}{2}$，根据 2.1 节求非等长图像置乱变换的变换矩阵，取 $a=1,b=1,n=1$，$q=\dfrac{\frac{N}{2}}{\gcd\left(\frac{M}{2},\frac{N}{2}\right)}$，那么 $c=\dfrac{\frac{N}{2}}{\gcd\left(\frac{M}{2},\frac{N}{2}\right)}$，$d=1+$

$\frac{\frac{N}{2}}{\gcd\left(\frac{M}{2},\frac{N}{2}\right)}$，即分块图像 $Block$ 的变换矩阵为 $\boldsymbol{A}=\begin{bmatrix}1 & 1\\ \frac{\frac{N}{2}}{\gcd\left(\frac{M}{2},\frac{N}{2}\right)} & 1+\frac{\frac{N}{2}}{\gcd\left(\frac{M}{2},\frac{N}{2}\right)}\end{bmatrix}$。对图像 $Block$ 进行 k 次置乱变换，将置乱后的图像记为 $ScrBlock$。非等长图像置乱变换根据宿主图像尺寸和分块大小选定参数后求变换矩阵，该算法不使用唯一的变换矩阵，可自适应地计算变换矩阵，提高了算法的灵活性；另外，非等长图像置乱变换使得算法适用于任意尺寸的图像，提高了算法的普适性。

步骤 3(块 SVD 分解) 为更好地表达图像块信息，仅用一个 LSB 平面记录认证信息是不够的，故本文将水印信息嵌入图像的 2 个 LSB 平面，每个图像块信息可由 8 位二进制数表示。将置乱图像 $ScrBlock$ 的 2 个 LSB 平面像素值置 0，得到图像 $ScrZerolsb$，如式(6)所示：

$$ScrZerolsb(i,j)=ScrBlock(i,j)-(ScrBlock(i,j)\ \mathrm{mod}\ 4) \tag{6}$$

对 $ScrZerolsb$ 图像块进行 SVD 分解，将图像块奇异值矩阵 S 的迹 $trace(\boldsymbol{S})$ 的值映射至[0,31]范围内，用 5 位二进制数表示，作为水印块验证位，记为 BAB。将相邻 6 个图像块划分为一组，将组内图像块迹的平均值映射至[0,7]范围内，用 3 位二进制数表示，该值作为水印组验证位，记为 GAB。块验证位记录图像内容信息，若图像内容被篡改，则提取水印的 BAB 改变，表示图像被篡改；分组打破了图像块的相对独立性，若组内图像块不是来自于同一幅图像，则分组验证失败，表示图像被篡改。

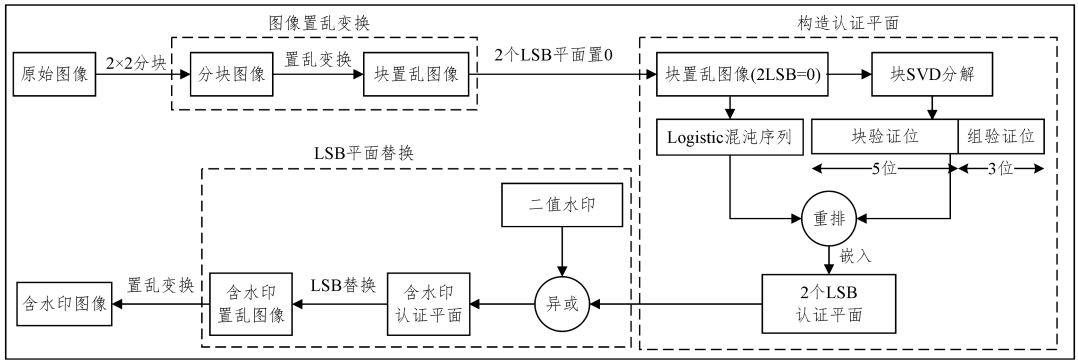


图 2 水印嵌入的流程

Fig. 2 Process of watermarking embedding

步骤 4(构造认证平面) 由步骤 3 得到 8 位认证序列，按照 Logistic 序列重新排列后嵌入认证平面。Logistic 混沌序列初始值的计算方法如下：

$$x_1=\frac{mean+1}{257} \tag{7}$$

$$\mu=3.569\,945\,66+(std-\lfloor std\rfloor)\times 0.43 \tag{8}$$

其中，$mean$ 和 std 分别为图像块的均值和标准差。认证序列根据图像块信息迭代产生的 Logistic 序列自适应地嵌入，得到认证平面 AP。当图像块篡改时，提取的认证序列改变，图像块无法通过认证，表示图像块被篡改。认证序列的嵌入过程如图 3 所示。

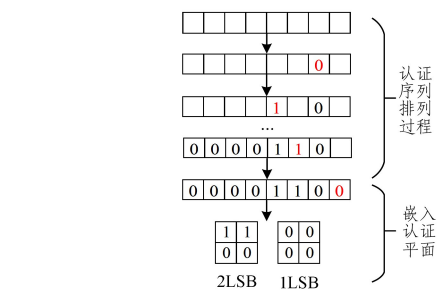


图 3 构造认证平面流程图

Fig. 3 Process of authentication plane construction

步骤 5(LSB 平面替换) 将与宿主图像同样大小的二值水印与认证平面异或：

$$APWM=AP\oplus WM \tag{9}$$

其中，AP 为认证平面，WM 为水印图像，$APWM$ 为含水印的认证平面。用含水印的认证平面替换置乱图像 $ScrBlock$ 的 2 个 LSB 平面后，进行($T-k$)次置乱变换(T 为置乱变换的周期)，得到含水印图像 I_{wm}。

3.2 水印提取算法

水印提取过程如图 4 所示，该算法主要分成 3 个步骤，具体如下。

步骤 1 提取含水印图像的 2 个 LSB 平面，获得图像块的块验证位和组验证位，分别记为 $OriginalBAB$ 和 $Original\text{-}GAB$。根据图像 6 个 MSB 平面的块信息，计算块验证位和组验证位，分别记为 BAB 和 GAB。

步骤 2 提取二值水印 WM'，原二值水印记为 WM。首先进行块匹配，如式(10)所示：

$$WM'=\begin{cases}WM(i,j), & BAB(i,j)=OriginalBAB(i,j)\\ \overline{WM(i,j)}, & BAB(i,j)\neq OriginalBAB(i,j)\end{cases} \tag{10}$$

每组最多的组验证位 GAB 作为该组验证位 $GABnum$，然后进行组匹配，如式(11)所示：

$$WM'=\begin{cases}WM(i,j), & GAB(i,j)=GABnum(i,j)\\ \overline{WM(i,j)}, & GAB(i,j)\neq GABnum(i,j)\end{cases} \tag{11}$$

步骤 3 比较提取的二值水印 WM' 与原二值水印 WM 的区别，以定位图像的篡改区域。

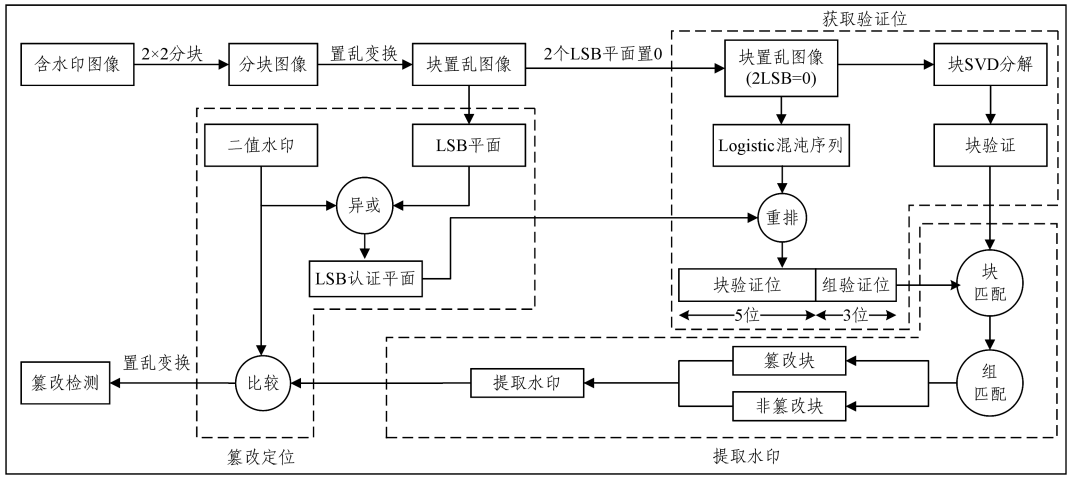


图 4 水印提取的流程

Fig. 4 Process of watermarking extracting

4 实验结果分析

本文在 MATLABR2017b 平台上测试算法的不可见性及篡改检测和定位能力。评价脆弱水印算法不可见性和篡改检测性能的权威指标如下。

1)峰值信噪比(PSNR)[17]:衡量水印嵌入后图像的质量,若图像大小为 $M\times N$,其 $PSNR$ 值的计算如式(12)所示:

$$PSNR(I,I')=10\log_{10}\frac{255^2}{MSE}(dB) \tag{12}$$

$$MSE=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-I'(i,j)]^2 \tag{13}$$

2)篡改检测性能[16]:本文采用虚警率(R_{fa})、漏警率(R_{md})、误检率(R_{fd})来衡量算法篡改检测的性能,各指标定义如下:

$$R_{fa}=\frac{N_{fd}}{N-N_t} \tag{14}$$

$$R_{md}=\frac{N_t-N_{td}}{N_t} \tag{15}$$

$$R_{fd}=\left(1-\frac{N_t}{N}\right)R_{fa}+\frac{N_t}{N}R_{md} \tag{16}$$

其中,N 为图像 I 中像素个数,N_t 为实际篡改像素个数,N_{fd} 为未篡改像素识别错误的个数,N_{td} 为篡改像素正确识别的个数。本文选用峰值信噪比评价算法的不可见性,选用虚警率、漏警率和误检率评价算法对图像篡改的检测和定位性能。

4.1 水印的不可见性

实验使用 10 张不同大小的彩色图像进行算法不可见性测试,嵌入水印与宿主图像尺寸相同,测试结果如表 1 所列。由表 1 可以看出,本文算法的 $PSNR$ 值在 44 dB 左右,根据人眼视觉特性,当 $PSNR$ 值大于 30 dB 时,两幅图像在视觉上没有差异,因此本文算法具有较好的不可见性。

表 1 水印不可见性的测试结果

Table 1 Results of imperceptibility of watermarking

Image	baboon	barbara	girl	cablecar	cornfield
Size	480×500	576×720	576×720	480×512	480×512
PSNR	44.0174	44.0870	44.0765	44.1595	44.1029
Image	lena	airplane	pepper	sailboat	tiffany
Size	512×512	512×512	512×512	512×512	512×512
PSNR	44.9951	44.0260	44.1434	44.0438	43.7507

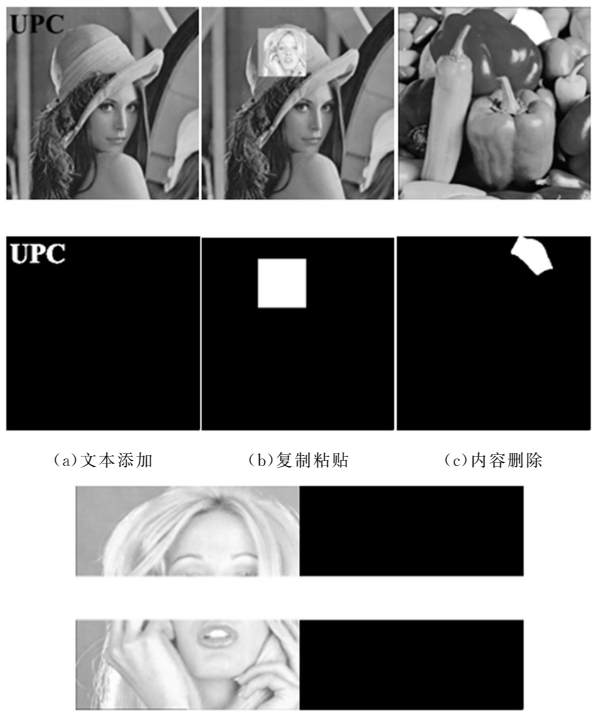
4.2 攻击实验

实验选用 512×512 的经典彩色图像作为测试图像,测试算法在简单攻击、复杂攻击和混合攻击[18]下的性能。为了验证算法的篡改检测性能和定位精度,将其分别与文献[9]和文献[11]的算法进行对比。

4.2.1 简单攻击实验

脆弱水印简单攻击包括文本添加攻击、复制粘贴攻击、内容删除攻击和剪切攻击[19]。实验分别对水印图像进行这 4 种简单攻击,图 5 为篡改检测的效果。

简单攻击下 3 种算法篡改检测性能的对比如表 2 所列。由表 2 可以看出,分块级篡改定位算法在误检率上的表现明显优于像素级篡改定位,本文算法在保证不漏检的前提下,降低了虚警率,提高了篡改定位精度。



(a)文本添加 (b)复制粘贴 (c)内容删除

(d)剪切

图 5 简单攻击测试的结果

Fig. 5 Results of common attacks

表 2 简单攻击下篡改检测性能的对比

Table 2 Comparison of tamper detection under common attacks

Attack		Text addition	Copy and paste	Content remove	Crop
R_{fa}	Proposed	0.0020	0.0011	0.0010	0
	Ref. [11]	0.0055	0.0021	0.0027	0.0097
	Ref. [9]	0	0	0	0
R_{md}	Proposed	0	0	0	0
	Ref. [11]	0	0	0	0
	Ref. [9]	0.3984	0.1290	0.4055	0.4958
R_{fd}	Proposed	**0.0019**	**0.0010**	**0.0010**	**0**
	Ref. [11]	0.0054	0.0020	0.0027	0.0078
	Ref. [9]	0.0043	0.0081	0.0095	0.0968

本文测试了算法在简单攻击下对不同篡改比例的检测性能，测试结果如表 3 所列。实验表明，文献[9]的算法几乎不存在虚警率，但算法漏警率在 50%左右，随着篡改比例增加，误检率近似呈线性增加；文献[11]的算法对篡改区域可正确检测，但当篡改比例超过 20%时，误检率明显升高；本文算法能够检测出篡改区域，在篡改比例不高于 35%时，检测率在 99%以上，当篡改比例达到 50%时，篡改检测率能够保持在 90%以上。综上，本文算法对简单攻击的检测性能优于文献[9]和文献[11]的算法。

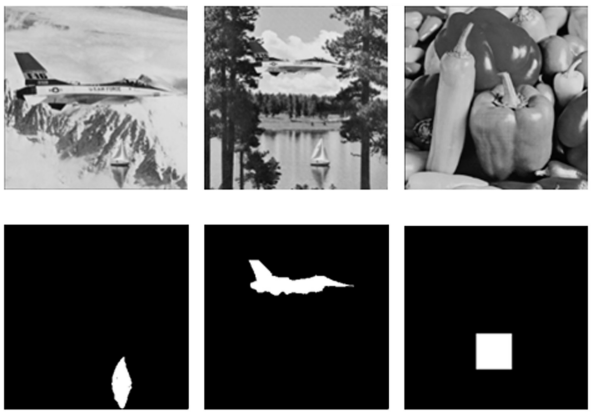
表 3 简单攻击下各算法对不同篡改比例的检测性能对比

Table 3 Comparison of each algorithm with different tamper ratio under common attacks

TamperRate		1%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
R_{fa}	Proposed	0.0004	0.0005	0.0007	0.0009	0.0011	0.0015	0.0020	0.0043	0.0491	0.0723	0.1582
	Ref. [11]	0.0008	0.0013	0.0020	0.0033	0.0054	0.0211	0.0540	0.1061	0.1883	0.2927	0.3768
	Ref. [9]	0	0	0	0	0	0	0	0	0	0	0
R_{md}	Proposed	0	0	0	0	0	0	0	0	0	0	0
	Ref. [11]	0	0	0	0	0	0	0	0	0	0	0
	Ref. [9]	0.4996	0.5031	0.4942	0.4944	0.4951	0.4948	0.4948	0.4946	0.4946	0.4938	0.4944
R_{fd}	Proposed	**0.0004**	**0.0004**	**0.0006**	**0.0008**	**0.0009**	**0.0011**	**0.0014**	**0.0028**	**0.0295**	**0.0404**	**0.0800**
	Ref. [11]	0.0008	0.0013	0.0018	0.0028	0.0043	0.0160	0.0380	0.0695	0.1138	0.1636	0.1905
	Ref. [9]	0.0048	0.0252	0.0483	0.0728	0.0986	0.1194	0.1464	0.1698	0.1956	0.2178	0.2444
Detection Rate	Proposed	**0.9996**	**0.9996**	**0.9994**	**0.9992**	**0.9991**	**0.9989**	**0.9986**	**0.9972**	**0.9705**	**0.9596**	**0.9200**
	Ref. [11]	0.9992	0.9987	0.9982	0.9972	0.9957	0.9840	0.9620	0.9305	0.8862	0.8364	0.8095
	Ref. [9]	0.9952	0.9748	0.9517	0.9272	0.9014	0.8806	0.8536	0.8302	0.8044	0.7822	0.7556

4.2.2 复杂攻击实验

脆弱水印复杂攻击包括拼贴攻击、仅内容篡改攻击和均值攻击[20]，实验分别对水印图像进行 3 种复杂攻击，本文算法的篡改检测结果如图 6 所示。



(a)拼贴攻击 (b)仅内容篡改 (c)均值攻击

图 6 复杂攻击测试

Fig. 6 Results of complex attacks

1)拼贴攻击：使用相同的水印嵌入算法得到两幅含水印的图像，将一幅图像的部分复制到另一幅图像中，保证其空间位置和大小不变。图像块间具有相对独立性的水印算法无法抵抗拼贴攻击。本文将水印图像“sailboat”中的部分复制到水印图像“airplane”中。

2)仅内容篡改攻击：在不改变水印位的前提下，修改未嵌入水印部分。水印位未与图像内容建立联系的水印算法无法抵抗仅内容篡改攻击。本文向水印图像“sailboat”的非水印位添加对象。

3)均值攻击：在不改变水印位的前提下，修改图像块像素值，但保证图像块均值不变，例如一个 2×2 图像块原像素值为[60 60;60 60]，均值攻击后图像块像素值为[20 100;20 100]。水印位未与图像内容建立联系或仅考虑单一均值因素的水印算法无法抵抗均值攻击。本文对水印图像“pepper”进行均值攻击。

复杂攻击下 3 种算法篡改检测性能的对比如表 4 所列。由表 4 可以看出，分块级篡改定位算法能够检测复杂攻击下的篡改，但像素级篡改定位算法对复杂攻击下的篡改检测效果不理想甚至无法检测，而本文算法不仅能检测出图像篡改，而且定位精度也优于其他两种算法。

表 4 复杂攻击下篡改检测性能的对比

Table 4 Comparison of tamper detection under complex attacks

Attack		Collage	Contentonly	Mean
R_{fa}	Proposed	0.0008	0.0015	0
	Ref. [11]	0.0022	0.0055	0
	Ref. [9]	0	0	0
R_{md}	Proposed	0.0017	0	0
	Ref. [11]	0	0	0
	Ref. [9]	0.9398	1	0.9999
R_{fd}	Proposed	**0.0008**	**0.0014**	**0**
	Ref. [11]	0.0022	0.0052	0
	Ref. [9]	0.0170	0.0410	0.0381

本文测试了复杂攻击下算法对不同篡改比例的检测性能，测试结果如表 5 所列(数据保留 4 位小数，计算结果数量级为 10^{-5}时，忽略不计)。

表 5 复杂攻击下算法对不同篡改比例的检测性能对比

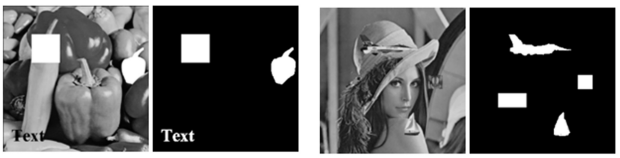
Table 5 Comparison of each algorithm with different tamper ratios under complex attacks

Tamper Rate		1%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
R_{fa}	Proposed	0	0	0.0001	0.0001	0.0002	0.0002	0.0005	0.0024	0.0288	0.0684	0.1561
	Ref.[11]	0	0	0.0001	0.0006	0.0009	0.0156	0.0497	0.1066	0.1810	0.2799	0.3728
	Ref.[9]	0	0	0	0	0	0	0	0	0	0	0
R_{md}	Proposed	0	0	0	0	0	0	0	0	0	0	0
	Ref.[11]	0	0	0	0	0	0	0	0	0	0	0
	Ref.[9]	0.9347	0.9675	0.9762	0.9805	0.9832	0.9850	0.9868	0.9873	0.9887	0.9887	0.9893
R_{fd}	Proposed	0	0	0.0001	0.0001	0.0002	0.0002	0.0004	0.0016	0.0174	0.0382	0.0781
	Ref.[11]	0	0	0.0001	0.0005	0.0007	0.0117	0.0349	0.0690	0.1085	0.1536	0.1884
	Ref.[9]	0.0094	0.0491	0.0961	0.1477	0.1948	0.2434	0.2937	0.3444	0.3951	0.4453	0.4940
Detection Rate	Proposed	1	1	0.9999	0.9999	0.9998	0.9998	0.9996	0.9984	0.9826	0.9618	0.9219
	Ref.[11]	1	1	0.9999	0.9995	0.9993	0.9883	0.9651	0.9310	0.8915	0.8464	0.8116
	Ref.[9]	0.9906	0.9509	0.9039	0.8523	0.8052	0.7566	0.7063	0.6556	0.6049	0.5547	0.5060

实验表明，文献[9]的算法对仅内容篡改攻击和均值攻击几乎没有篡改检测的能力，对拼贴攻击仅能检测出篡改区域的轮廓；文献[11]的算法对篡改区域可正确检测，但检测区域的轮廓不够圆滑，呈现锯齿状，误检率较高；本文算法能够检测出篡改区域，篡改比例不高于 35%时，检测率在 99%以上，篡改比例达到 50%时，篡改检测率在 90%以上。综上，本文算法对复杂攻击的检测性能优于文献[9]和文献[11]的算法。

4.2.3 混合攻击实验

为测试算法对混合攻击的检测能力，本节进行了两组实验，本文算法的篡改检测效果如图 7 所示。



(a)剪切＋文本添加＋内容删除 (b)拼贴＋均值＋复制粘贴＋仅内容篡改

图 7 混合攻击测试的结果

Fig. 7 Results of multi attacks

混合攻击下 3 种算法篡改检测性能的对比如表 6 所列。由表 6 可以看出，本文算法能较好地抵抗混合攻击，篡改检测定位的性能优于文献[9]和文献[11]的算法。

表 6 混合攻击篡改检测性能对比

Table 6 Comparison of tamper detection under multi-attacks

Attack		Multiattacks(a)	Multiattacks(b)
R_{fa}	Proposed	0.0025	0.0029
	Ref.[11]	0.0081	0.0073
	Ref.[9]	0.4848	0.8410
R_{md}	Proposed	0	0.0006
	Ref.[11]	0	0
	Ref.[9]	0.0341	0.0457
R_{fd}	Proposed	0.0023	0.0028
	Ref.[11]	0.0075	0.0069
	Ref.[9]	0.5152	0.7978

当篡改区域的轮廓不规则时，文献[11]的算法检测的篡改区域轮廓呈现明显的锯齿状，弧度较小的轮廓检测出的边界为直线；本文算法检测的篡改区域轮廓虽不如像素级篡改定位算法准确，但是比文献[11]的算法检测的轮廓效果要好。由于文献[11]的算法的分块大小为 4×4，篡改定位的最小单位为 16 像素，本文算法的分块大小为 2×2，篡改定位的最小单位为 4 像素，因此在最差的情况下，当图像只有 1 个像素被篡改时，文献[11]的算法检测结果为 16 个像素被篡改，而本文算法检测结果为 4 个像素被篡改，因此本文算法的虚警率低于文献[11]的算法，篡改定位精度优于文献[11]的算法。

以上实验结果表明，本文算法具有很好的不可见性，对简单攻击和复杂攻击下的图像篡改，具有良好的检测性能。该算法不仅在单一攻击类型下检测性能良好，在混合攻击下仍有较好的检测效果。

结束语 本文将图像以 2×2 大小分块来提高定位精度，使用非等长图像置乱变换提高抗复杂攻击能力和算法的普适性。经实验测试，所提算法具有较好的不可见性，在简单攻击、复杂攻击和混合攻击下进行图像篡改检测，具有良好的检测效果。当篡改区域为不规则形状时，该算法在篡改定位精度上的优势高于其他分块算法。本文采用漏警率和虚警率对图像篡改性能进行评价，其值越低，表示算法的篡改检测能力越高。该指标只能评价算法检测篡改与实际篡改的差别，但是无法评价算法篡改检测是否成功，这是下一步需要研究的内容。

参 考 文 献

[1] LIU X L, LIN C C, YUAN S M. Blind dual watermarking for color images' authentication and copyright protection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(5): 1047-1055.

[2] TIWARI A, SHARMA M, TAMRAKAR R K. Watermarking based image authentication and tamper detection algorithm using vector quantization approach[J]. International Journal of Electronics and Communications (AEU), 2017(78): 114-123.

[3] ZHAN X, LI Y, LEI Y R, Chen C R. A grayscale images fragile watermark authentication system based on the singular value decomposition[J]. Journal of Inner Mongolia University (Natural Science Edition), 2015, 44(3): 385-388.

[4] SHI J P, WU Y Q. Fragile watermarking for fingerprint images based on QR decomposition in NSCT-Domain[J]. Journal of Applied Science — Electronics and Information Engineering, 2017, 35(6): 735-744.

[5] JING X Z, JIANG W Z. Fragile watermarking capable of locating tampered blocks in JPEG images[J]. ACTA ELECTRONICA SINICA, 2010, 38(7): 1585-1589.

[6] ANSARI I A, PANT M, AHN C W. SVD based fragile watermarking scheme for tamper localization and self-recovery[J]. International Journal of Machine Learning and Cybernetics, 2016, 7(6): 1225-1239.

[7] YEUNG M M, MINTZER F. An invisible watermarking technique for image verification[C]// Proceedings of theICIP'97. Santa Barbara, California: IEEE, 1997: 680-683.

[8] CHEN F, WANG H X. Secure fragile watermarking algorithm with tampered-pixels localization[J]. Journal of the China Railway Society, 2011, 33(1): 63-68.

[9] MUNIR R. A chaos-based fragile watermarking method in spatial domain for image authentication[C]// 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA). IEEE, 2015.

[10] TONG X, LIU Y, ZHANG M, et al. A novel chaos-based fragile watermarking for image tampering detection and self-recovery [J]. Signal Processing: Image Communication, 2013, 28(3): 301-308.

[11] KANG Q, KE L, HU C. An SVD-based fragile watermarking scheme with grouped blocks[C]// 2nd International Conference on Information Technology and Electronic Commerce (ICITEC). IEEE, 2014.

[12] SHAO L P, QIN Z, GAO H J, et al. 2-Dimension non equilateral image scrambling transformation[J]. Acta Electronica Sinica, 2007, 35(7): 1290-1294.

[13] LI Y K, FENG Q S, ZHOU F, et al. 2-D Arnold transformation and non-equilateral image scrambling transformation[J]. Computer Engineering and Design, 2009, 30(13): 3133-3135.

[14] ANSARI I A, PANT M. Multipurpose image watermarking in the domain of DWT based on SVD and ABC[J]. Pattern Recognition Letters, 2016, 16(94): 228-236.

[15] SHEHAB A, ELHOSENY M, MUHAMMAD K, et al. Secure and robust fragile watermarking scheme for medical images[J]. IEEE Access, 2018, 6: 10269-10278.

[16] ZHOU W, YU L, WANG Z, et al. Binocular visual characteristics based fragile watermarking scheme for tamper detection in stereoscopic images[J]. International Journal of Electronics and Communications(AEU), 2015, 70(1): 77-84.

[17] SHI H, LI M C, GUO C, et al. A region-adaptive semi-fragile dual watermarking scheme[J]. Multimedia Tools and Applications, 2016, 75(1): 465-495.

[18] TRIVEDY S, PAL A K. A Logistic map-based fragile watermarking scheme of digital images with tamper detection[J]. Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2017, 41: 103-113.

[19] ZHANG H. Research on fragile watermarking schemes for image content authentication and self-recovery[D]. Weihai: Shandong University, 2018.

[20] XIE G X. Self-recovery fragile watermarking algorithm for image content authentication[D]. Chengdu: Southwest Jiaotong University, 2017.

ZHENG Qiu-mei, born in 1964, postgraduate, professor, postgraduate supervisor. Her main research interests include digital watermarking and image processing.

LIU Nan, born in 1996, postgraduate. Her main research interests include image processing and digital image watermarking.