

鲁棒的区域复制图像篡改检测技术

骆伟祺^{1),2)} 黄继武^{1),2)} 丘国平³⁾

¹⁾(中山大学信息科学与技术学院 广州 510275)

²⁾(广东省信息安全重点实验室 广州 510275)

³⁾(诺丁汉大学计算机学院 英国)

摘 要 区域复制把数字图像中一部分区域进行复制并粘贴到同一幅图像的另一个区域中,以达到去除图像中某一重要内容的目的,是一种简单而有效的图像篡改技术.现有检测算法对区域复制后处理的鲁棒性较差.文中针对此篡改技术,提出了一种有效的检测与定位篡改区域算法.该算法首先将图像分解为小块并比较各小块间的相似性,最后利用“主转移向量”方法去除错误的相似块对得到篡改的区域.实验数据说明该算法能有效地对抗多种区域复制的后处理操作,包括高斯模糊、加性白高斯噪声、JPEG压缩及它们的混合操作.

关键词 图像盲认证;图像篡改;篡改检测;区域复制;鲁棒性

中图法分类号 TP391

Robust Detection of Region-Duplication Forgery in Digital Image

LUO Wei-Qi^{1),2)} HUANG Ji-Wu^{1),2)} QIU Guo-Ping³⁾

¹⁾(School of Information Science and Technology, Sun Yat-Sen University, Guangzhou 510275)

²⁾(Guangdong Key Laboratory of Information Security Technology, Guangzhou 510275)

³⁾(School of Computer Science, University of Nottingham, NG8 1BB, UK)

Abstract Region duplication forgery, in which a part of a digital image is copied and then pasted to another portion of the same image in order to conceal an important object in the scene, is one of the common image forgery techniques. This paper describes an efficient and robust algorithm for detecting and localizing this type of malicious tampering. The algorithm first divides an image into small overlapped blocks and then compares the similarity of these blocks, finally identifies possible duplicated regions using the main shift vector. The experimental results show that above method is more robust comparing with other existing algorithms and can successfully detect this type of tampering for images that have been subjected to various forms of post region duplication image processing, including, blurring, noise contamination, severe lossy compression, and a mixture of these processing operations.

Keywords blind image authentication; image forgery; detection of tampering; region duplication; robustness

1 引 言

各种高级图像处理算法及相应图像处理软件的

出现使得人们很容易对数字图像进行修改而不留下明显的痕迹.这必然会带来一些涉及到诸如法律取证中的图像真实性、图像媒体的版权、个人的隐私保护等相关的问题.在2004年美国总统大选中曾广为

收稿日期:2006-07-30;最终修改稿收到日期:2007-08-14.本课题得到国家杰出青年科学基金(60325208)、国家自然科学基金(60633030, 90604008)、国家“九七三”重点基础研究发展规划项目基金(2006CB303104)和广东省自然科学基金团队项目(04205407)资助. 骆伟祺,男,1980年生,博士研究生,研究方向包括数字图像鉴证、处理及模式识别. 黄继武,男,1962年生,博士,教授,博士生导师,目前研究领域为多媒体信息安全和信息隐藏. E-mail: issbjw@mail.sysu.edu.cn. 丘国平,男,1964年生,博士,副教授,目前研究兴趣为视觉信息处理的计算方法及工具.

流传的一张图片:总统候选人 Kerry 和著名反战女艺人 Jane Fonda 的合照^[1],最后被证实是由两幅其它图像拼接起来的伪造图像.其实这仅仅是图像造假问题的一个例子.如今流传于网络上各式各样的图片,人眼是很难判定其真伪性的.如果虚假的图片被滥用,这将对我们的社会、个人生活等带来极大的负面影响.随着图像处理算法和相应处理工具的发展,这些伪造的图片将会越来越多地出现在人们面前.图像认证成为了一个十分重要的研究课题^[2-3].

可以利用数字签名^[4-5]或脆弱/半脆弱水印技术^[6-8]实现图像认证.但基于签名或水印认证技术的一个限制是:必须在媒体数据建立的同时人为地进行预处理,如计算图像 Hash 值或水印嵌入操作.这使得其应用范围受到了很大的约束.而且目前基于签名和水印图像认证方法还存在一些问题,如安全性尚未能得到保证,水印抗恶意攻击的能力有待进一步加强等.因此需要有一种新的认证技术.最近,基于图像本身性质的盲认证技术^[9-20]引起了国际上学者们的重视.此技术假设:在自然图像中存在着某些统计上的性质,倘若我们对图像数据进行修改则会改变其潜在的统计上的规律.这一假设也是我们判定一个图像是否被修改和进行篡改定位的依据.盲认证技术的最大特点是它不需要事先给媒体数据添加任何的签名或水印信息,仅仅利用媒体数据本身的特性就可达到认证的目的,因此有学者称之为被动式认证(passive-blind authentication)^①.

近年来,国际上对盲认证已有一些初步的研究. Farid 等人提出了一些统计特征检测区域复制(region duplication)^[11]、彩色滤波插值图像(color filter interpolation)^[12]、重采样(re-sampling)^[13]、光源方向检测^[14]等算法. Fridrich 等人提出检测“复制-粘贴”篡改检测^[15],利用数码相机传感器固有的噪声对图像数据进行认证等^[16-17]. Ng 和 Chang 建立了图像拼接库和计算机图形(CG)库,并提出了检测图像拼接、区分 CG 和真实图像的算法^[18-19]. 尽管如此,盲认证在国内外仍处于起步阶段,很多问题还有待解决.

一种简单而有效去除图像中某重要物体的篡改方法是:挑选图像中某一区域进行复制并粘贴到欲去除的区域中.由于在同一幅图像中有着一致的噪声、纹理、颜色等信息,同时造假者往往会在“复制-粘贴”后做模糊、加噪、JPEG 压缩等操作,使得篡改图像更难于被人辨别与检测.一些学者已对这一篡改形式提出了算法, Fridrich 在文献^[15]中分析并

利用 DCT 系数的性质, Popescu 和 Farid 则利用了主分量分析的方法^[11].但他们都有各自的缺陷,如 Fridrich 仅讨论了 JPEG 后处理的情况. Farid 算法的时间复杂度较高,且在一定的后处理下(如质量因子小于 50 的 JPEG 压缩,添加高斯白噪声后 SNR 小于 24dB 等)算法会出现较高的误判率.

本文分析了经过各种区域复制后处理(如高斯模糊、加高斯白噪声、有损 JPEG 压缩等),对应篡改图像块的特征变化,选取了一组对各种常用后处理鲁棒的空域统计特征,提出了区域篡改的判别准则,实现了对区域复制的识别.实验证明,与文献^[11, 15]相比,本文提出的算法能更有效地抵抗更多,更强的后处理操作.

本文在第 2 节首先给出区域复制篡改的模型.相应的检测算法与算法复杂性分析在第 3 节进行详细地描述;第 4 节是实验的结果和讨论;最后,在第 5 节给出全文的总结.

2 区域复制篡改模型

区域复制是一种图像的局部篡改技术,它把图像中的某一区域进行复制,粘贴到同一图像的不相交区域上,并进行一定的后处理操作,以达到去除图像中某一重要特征的目的.

从其篡改的操作过程可知,篡改后的图像中至少存在两个较大面积的相似区域.基于大量实验,我们得出如下结论:在自然图像中(除有大片平坦区域的图像外)存在相似(包括颜色,形状,纹理等信息)大面积区域的可能性是很少的(基于对图像库的统计结果,“大面积区域”假定为不小于原始图像尺寸的 0.85%).若我们检测到在一幅图像中存在大面积的相似区域,则很有可能是被区域复制篡改过的.本文所提出的算法也以此作为判断图像是否被篡改的准则.

一般对图像内容的修改,都是针对图像中某个连通的区域,并且往往是利用离篡改块较远的区域进行复制粘贴,这样得到的图像会更难于被人察觉.并且要实现有意义的篡改,其篡改的区域一般较大,我们假设篡改块的大小不小于原始图像尺寸的 0.85%.据此,首先给出区域复制篡改模型的一些合理假设(如图 1 所示):

(1) 被复制的区域 D_2 是一个连通“无洞”的

① <http://advent.ctr.columbia.edu/advent/>

- 区域;
- (2) 仅有一个区域 D_1 被篡改改为 D_2 , 且 $D_1 \cap D_2 = \emptyset$;
 - (3) 对应篡改块的转移向量距离大于 L ;
 - (4) D_2 的面积大于 0.85% 的图像大小.

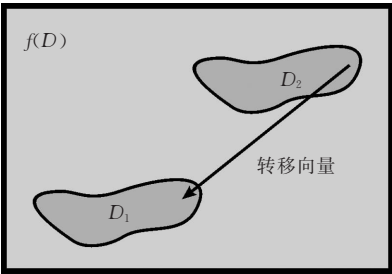


图 1 区域复制篡改模型

这可描述为

在一幅篡改图像 $f'(D)$ 中, \exists 两个区域 $D_1, D_2 \subset D$ 和一个转移向量 $\mathbf{d} = (d_1, d_2) (|D_1| = |D_2| > |D| \times 0.85\%, |\mathbf{d}| > L)$

s. t. $\forall (x_1, y_1) \in D_1, f'(x_1, y_1) = f(x_2, y_2),$

$x_2 = x_1 + d_1, y_2 = y_1 + d_2, (x_2, y_2) \in D_2,$

其中, $f(x, y)$ 是原始图像的灰度值 (对于彩色图像是一个表示 RGB 各颜色通道灰度值组成的三维向量), $D = \{(x, y) | 1 \leq x \leq M, 1 \leq y \leq N\}, D_2$ 是被复制的区域, D_1 是被篡改的区域. 因此篡改后的图像 f' 可表示为

$$f'(x,y)=\begin{cases} f(x,y), & (x,y) \notin D_1 \\ f(x+d_1,y+d_2), & (x,y) \in D_1 \end{cases}.$$

检测算法所要做的就是判断一幅给定的图像中是否存在这样不知其形状与位置的区域 D_1, D_2 . 如果存在则定位出其区域.

倘若篡改后的图像不做任何的后续处理, 则图像 D_1, D_2 中的值是精确相等, 检测是一个简单的区域匹配的问题. 但是, 这样篡改后的图像会在 D_1 区域产生不一致的边界信息. 为了消除其边界效应, 同时为了增加检测的难度, 篡改者往往会对 $f'(x, y)$ 做后续的处理操作, 如加噪、模糊、有损的 JPEG 压

缩等. 经过后处理操作的篡改图像, 其边界效应会明显减少, 视觉上更难以发现, 并且 D_1, D_2 区域上的对应像素的值一般都变得不相等, 检测的复杂性将大大增加.

3 检测算法及时间复杂度分析

我们的检测算法是基于块匹配的. 算法首先将待测图像分解成有重叠区域的小块, 从中提取出每块的特征, 然后选择恰当的阈值度量各分块的相似性得到相似块对, 最后去除错误的匹配块对以定位出篡改区域.

其中选择怎样的特征能度量篡改后图像块的相似性以抵抗不同后处理操作攻击是算法的关键之一. 另外, 由于图像内容复杂性及临近区域相似性的影响, 根据所选择特征找到的匹配块对可能不同时出现在区域 D_1 和 D_2 中 (我们称之为错误的匹配块对). 这些块对的多少与所选择的特征和判别阈值有关. 因此, 如何设定恰当的阈值使得在对后处理操作具有较强鲁棒性的同时错误块对尽量少是算法的另一个关键. 最后, 如何消除错误的匹配块对并定位出篡改区域, 如何提高算法的执行效率等也是算法必须解决的问题. 检测算法的具体步骤如下.

3.1 特征的提取与阈值的选择

设待检测的图像为 $M \times N$ 大小的彩色图像, 算法首先将其分解为 $b \times b$ 的小块, 相邻小块之间只有一行 (或一列) 不相交, 共得到 $S = (M - b + 1)(N - b + 1)$ 个图像块. 按行优先顺序记录每个图像块的 7 个特征, 并用向量表示成 $\mathbf{V}_i = (c_1, c_2, c_3, c_4, c_5, c_6, c_7)$, 其中:

- (i) c_1, c_2, c_3 分别是彩色图像块 R, G, B 三个通道的平均值 (对于灰度图像仅需记录亮度分量的平均值).
- (ii) 计算彩色图像块的亮度分量 $Y = 0.299R + 0.587G + 0.114B$. 用 c_4, c_5, c_6, c_7 分别记录 Y 分量如下 4 个方向上的特征, 如图 2 所示.

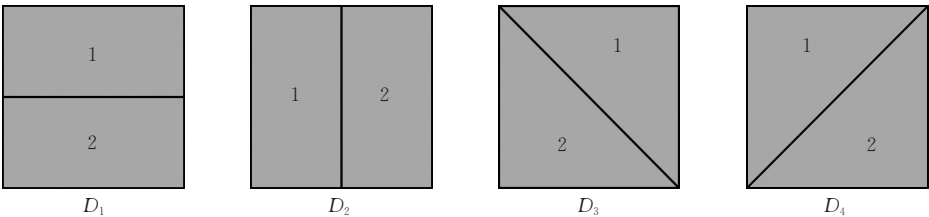


图 2 4 种分解模式

$$c_i = \text{sum}(\text{part}(1)) / \text{sum}(\text{part}(1) + \text{part}(2)), \\ i = 4, 5, 6, 7.$$

特征合理性分析与阈值的选择. $\mathbf{V}_i = (c_1, c_2, c_3, c_4, c_5, c_6, c_7)$ 体现了图像块的“平均值”信息, 相当于直流或低频分量数据, 对于区域复制篡改的各种后续处理都具有很好的鲁棒性.

以对加性白高斯噪声 ϵ (设均值为 0, 方差为 δ^2) 操作为例. 假设所加的噪声对图像中每个像素都是独立、同分布的. 设加噪后的第 i 个图像块为 B'_i , 我们得到

$$B'_i = B_i + \epsilon_{b \times b} = (f + \epsilon)_{b \times b},$$

$$c'_1 = \text{mean}(\text{red}(B'_i)) = \frac{\sum \text{red}(f + \epsilon)}{b^2} = c_1 + \epsilon',$$

其中, $\epsilon' = \frac{\sum \epsilon}{b^2}$, $E(\epsilon') = 0$, $D(\epsilon') = \frac{\delta^2}{b^2}$. 从而加噪后的

的特征与原始没加噪声前的特征相近, 即

$$c'_1(c'_2, c'_3) \approx c_1(c_2, c_3).$$

类似地

$$c'_4 = \frac{\text{sum}(\text{part}(1)) + \epsilon_1}{\text{sum}(\text{part}(1) + \text{part}(2)) + \epsilon_2},$$

其中

$$E(\epsilon_1) = 0, D(\epsilon_1) = \frac{b^2}{2} \delta^2,$$

$$E(\epsilon_2) = 0, D(\epsilon_2) = b^2 \delta^2.$$

在所加噪声不太强的情况下 (实验表明当 SNR 值大于 20dB 时), 有下面不等式成立:

$\text{sum}(\text{part}(1)) \gg \epsilon_1$, $\text{sum}(\text{part}(1) + \text{part}(2)) \gg \epsilon_2$. 从而我们可以得到

$$c'_4(c'_5, c'_6, c'_7) = \frac{\text{sum}(\text{part}(1)) + \epsilon_1}{\text{sum}(\text{part}(1) + \text{part}(2)) + \epsilon_2} \\ \approx \frac{\text{sum}(\text{part}(1))}{\text{sum}(\text{part}(1) + \text{part}(2))} \\ = c_4(c_5, c_6, c_7).$$

因此, 特征(i), (ii)对加高斯白噪声的后处理具有很好的鲁棒性. 同样, 对于有损的 JPEG 压缩处理和高斯模糊, 这些操作相当于一个低通滤波器, 它丢弃了部分高频信息, 但不会对图像的低频和直流信息有太大的改变. 因此所选的特征对这些操作也有较好的鲁棒性.

区域复制检测算法的关键在于如何在一幅图像中寻找并定位出篡改区域. 这个问题类似于视频压缩编码中可变宏块的运动估计. 但在运动估计中, 宏块间的比较都是基于像素级的比较, 如 MAE, MSE 等, 若用这些误差评判准则确定经后处理的相似块对, 块对中对对应像素间的噪声将会累加, 不利于阈值

的选取和对后处理操作鲁棒性要求. 因此, 视频编码中的块匹配准则 MAE, MSE 等不太适用于检测区域复制篡改. 为此我们做了如下实验: 先随机选取 100 幅图像做区域复制篡改, 篡改块大小为 16×16 . 然后对篡改图像做不同的后处理操作, 包括添加高斯白噪声 (SNR 为 20~40dB)、有损的 JPEG 压缩 (质量因子为 40~90). 对于每幅经过篡改后的图像, 比较原始块与篡改块间的变化. (i) 用类似传统视频压缩编码中的宏块误差比较方法: 先把块转换到 YC_bC_r 空间, 分别比较 Y, C_b, C_r 3 个通道的 MAE 值变化; (ii) 用我们提出的 7 个特征比较. 把上述各种情况下绝对差的平均值作为图像中衡量两个小块 (16×16) 间是否相似的一个阈值.

表 1 各特征变化的平均值

特征	均值	特征	均值
Y	8.1	C_3	3.0
C_b	5.6	C_4	0.006
C_r	6.5	C_5	0.005
C_1	2.5	C_6	0.005
C_2	1.5	C_7	0.005

从实验结果可看到本文选取的特征, 经过各种后处理操作后, 变化不大. 有利于篡改块对的寻找与减少匹配的运算量. 对得到的均值, 随机挑选 30 幅没有经过任何篡改的图像作测试, 比较在自然图像中两种方法找到的相似块对数 (错误的匹配块对). 我们的算法平均有 0.2188 对/块, 而采用 MAE 方法平均有 2.88 对/块以上. 因此利用我们所构造的特征要明显优于传统的视频压缩中宏块匹配方法所做的结果. 用 MSE 分析也会得到类似的结论.

我们把得到的均值作为算法衡量块与块间是否相似的阈值.

3.2 相似块对寻找及错误块对去除

从 3.1 节的分析和实验可知, 用我们所提取的特征与阈值找到的相似块对中同样会出现错误匹配块对. 如何消除错误块对影响及定位出篡改区域是本文算法的另一关键. 我们采取了“主转移向量”方法, 具体方法如下.

(i) 首先利用 3.1 节得到的 7 个特征和阈值寻找出图像中的所有相似块对, 设有 n 对. 对每一块对我们计算它们的转移向量 \mathbf{d}_i (从一个图像块到另一图像块的变化向量, 它类似于视频帧间预测编码的运动向量. 在本文提出的算法中认为 $\mathbf{d}_i, -\mathbf{d}_i$ 是同一个向量), $i = 1, 2, \dots, n$.

(ii) 统计这 n 个转移向量, 从中挑选一个出现

频率最大的作为主转移向量. 然后把这 n 个块对中转移向量不等于主转移向量的块对认为是错误的块对给去掉, 剩下的块对则认为是经过区域复制篡改的区域. 因为从区域复制篡改过程看, 相当于图像中的某个区域 D_2 做了一个平移操作, 将 D_1 中的信息进行了覆盖, 而 D_1, D_2 中各对应小块的移动方向和距离是一致的, 如图 3 所示.

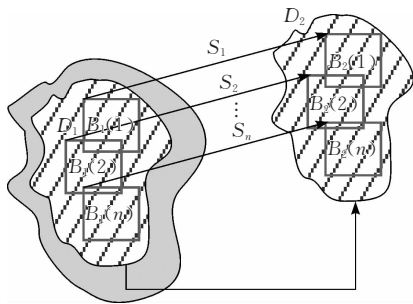


图 3 对应块有一致的转移向量

当然, 在没有经过篡改的图像中也会出现相似的区域, 尤其是图像中那些平坦的区域, 如天空、地板等. 但从大量实验中我们注意到, 对于一般的自然图像, 其相似区域一般较小, 所以当篡改区域的面积较大时, 对应的篡改图像块对所形成的转移向量往往是最多的.

3.3 判定篡改图像和定位篡改区域

在 3.2 节中, 我们把具有与主转移向量一致的块对放入一个与待测图像同样大小的二进制图像对应位置上. 由假设, 利用数学形态学方法从中提取出二值图像最大的两个连通分量, 并将其连通分量的“空洞”填补上. 其中数学形态学是一种基于集合关系运算的图像处理方法, 它为表达和提取图像形状特征等方面提供了一个有效的工具^[21]. 膨胀和腐蚀是形态学中两类基本的操作, 它们可以有效地用于图像边界、连通分量、凸壳等的提取及区域填充等处理.

设经过以上步骤得到的两个区域为 R_1, R_2 , 若 $\min(|R_1|, |R_2|) > \alpha M \times N \times 0.85\%$, 且 $||R_1| - |R_2|| / \max(|R_1|, |R_2|) < Tr$, 则认为图像是经过区域复制篡改过, 此时以 $tag = 1$ 表示; 否则, 认为没有被篡改, $tag = 0$. 由于篡改图像往往会经过一些后处理操作, 检测到的区域 R_1, R_2 面积可能与原始篡改区域 D_1, D_2 会有不同, 我们利用参数 α 度量其减少的程度, 利用 Tr 度量检测到的 R_1, R_2 之间的面积差异的大小, 并取 $\alpha = 61\%$, $Tr = 6\%$. 此阈值是利用我们算法对 100 幅图像进行不同大小的区域篡改, 再经过不同的后处理操作, 包括模糊、加噪、JPEG 压缩等所得到的篡改图像

(共 5600 幅, 如第 4 节中所述) 进行检测, 得到的两个篡改区域之间变化的最大值.

下面给出以上步骤的时间复杂度分析:

(1) 将图像 (尺寸为 $M \times N$) 分解为 $b \times b$ 大小且相邻块间只有一行 (列) 不相交的小图像块. 算法时间复杂度为 $O(MN)$. (3.1 节步骤)

(2) 对于 (1) 中得到的每个图像块提取 7 个特征. 复杂度为 $O(b^2)O(MN)$. (3.1 节步骤)

(3) 两两比较 (2) 中所有分块的 7 个特征, 利用选定的阈值找到所有的相似块并记录块对的转移向量. 统计转移向量出现的频率, 设出现最多的为主转移向量. 复杂度为 $O(M^2 N^2)$. (3.1 节, 3.2 节步骤)

(4) 把具有与主转移向量一致的块对放入一个与待测图像同样大小的二进制图像对应的位置上. 利用数学形态学方法提取出最大的两个连通分量并进行区域填充. 复杂度为 $O(MN)$. (3.3 节步骤)

(5) 对图像是否被篡改做判断. 若出现区域复制则定位出篡改位置. 复杂度为 $O(1)$. (3.3 节步骤)

算法总的时间复杂度为

$$\max(O(MN), O(b^2)O(MN), O(M^2 N^2), O(MN), O(1)) = O(M^2 N^2).$$

实验中我们采用测试图像的大小为 300×400 , 分块大小设为 16×16 . 机器的配置是: 主频为 2.8GHz 的 Pentium(R) CPU, 512MB 内存. 利用 Matlab 7.0 实现算法, 平均处理一幅图像的时间在 50s 左右.

4 实验结果

我们随机地从网上下载 100 幅 300×400 大小的图像做实验. 实验中各参数设置如下: $L = 50$, $b = 16$, 各特征相似阈值由 3.1 节中实验方法得到 $[2.5, 1.5, 3.0, 0.006, 0.005, 0.005, 0.005]$. 如果图像是被篡改过的, 但算法给出 $tag = 0$; 或 $tag = 1$, 但待测图像没有被篡改过, 此时算法出现误判, 令 $J = 1$ 表示, 否则 $J = 0$. 当输入图像被篡改过, 算法没有出现误判时, 我们定义检测率 r 和错误率 w :

$$r = \frac{|R_1 \cap D_1| + |R_2 \cap D_2|}{|D_1| + |D_2|},$$

$$w = \frac{|R_1 \cup D_1| + |R_2 \cup D_2|}{|D_1| + |D_2|} - r.$$

以下是对两个测试图像的实验结果.

在没有任何后处理的情况下, 本文算法的检测率达到 0.99123, 错误率为 0.1045. 经过各种后处理的检测结果见图 5~图 7 (图像下方的向量表示 (r, w)).

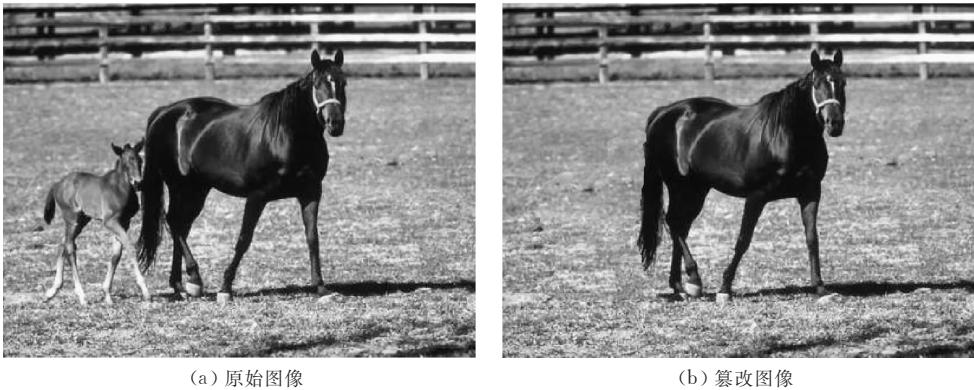


图 4 测试例子 1



图 5 对图 4 图像添加不同强度的高斯噪声后的检测结果



图 6 对图 4 图像进行不同质量因子 JPEG 压缩后的检测结果

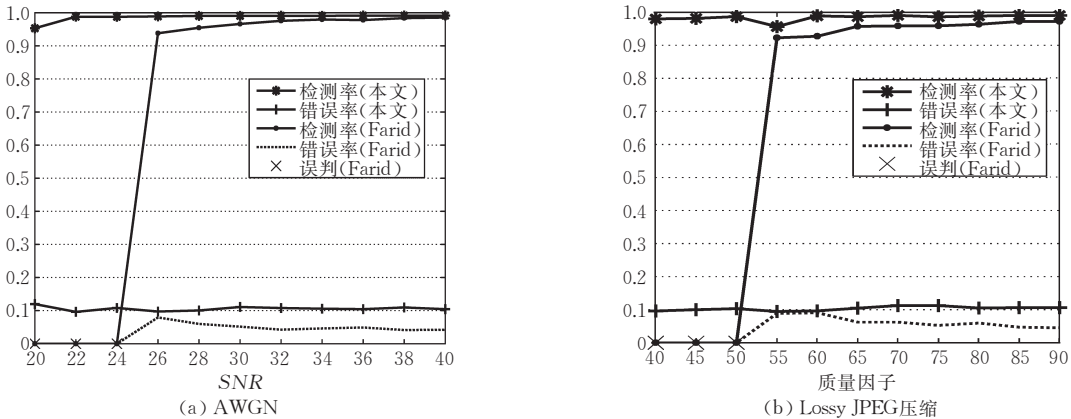


图 7 与 Farid 算法^[11](默认参数应用于 Green 通道)的比较

从上面的对比我们可以看到,本文的算法可以抵抗更强的攻击,即使 SNR 下降到 20dB, 仍有 $(r, \omega) = (0.9530, 0.1198)$. JPEG 质量因子下降到

40 时, $(r, \omega) = (0.9799, 0.0958)$. 而 Farid 算法则在 SNR 小于 24dB 或在质量因子小于 50 时就产生了误判.

另一例子如图 8 所示(文献[11]所用的例子). 在没有任何后处理时 $(r, w) = (0.9926, 0.0622)$.

图 9~图 11 是在不同后处理下的检测结果. 实验结果与例子 1 相似.

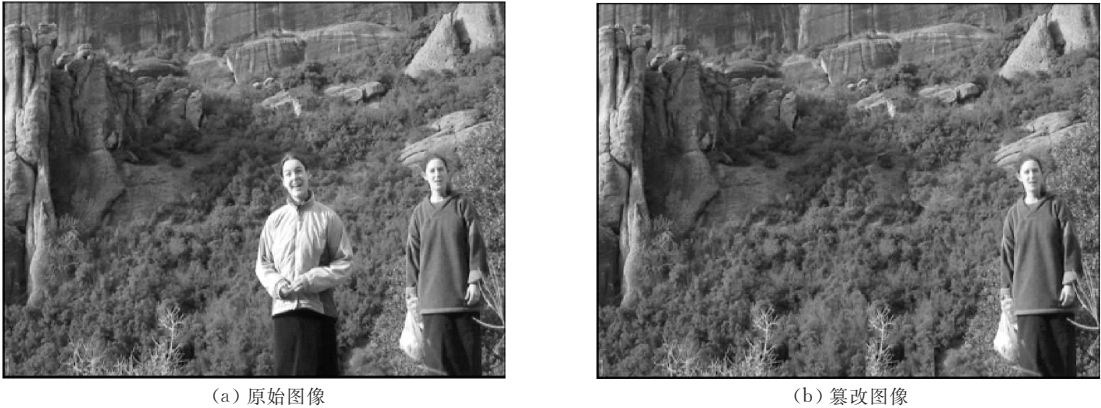


图 8 测试例子 2

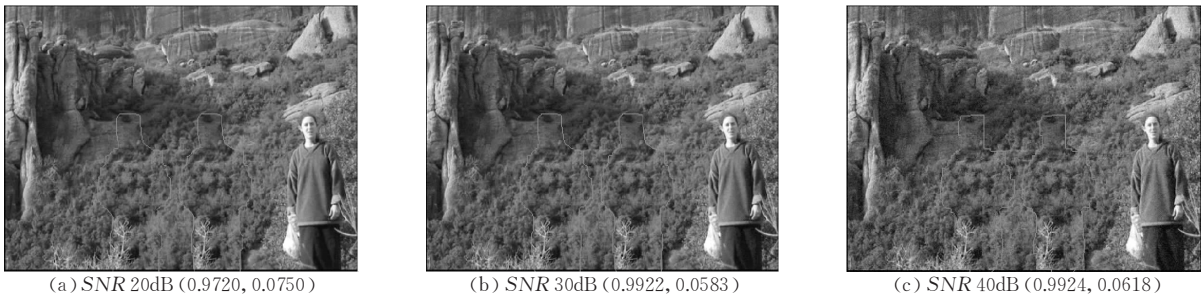


图 9 图 8 图像中添加不同强度的高斯噪声后的检测结果

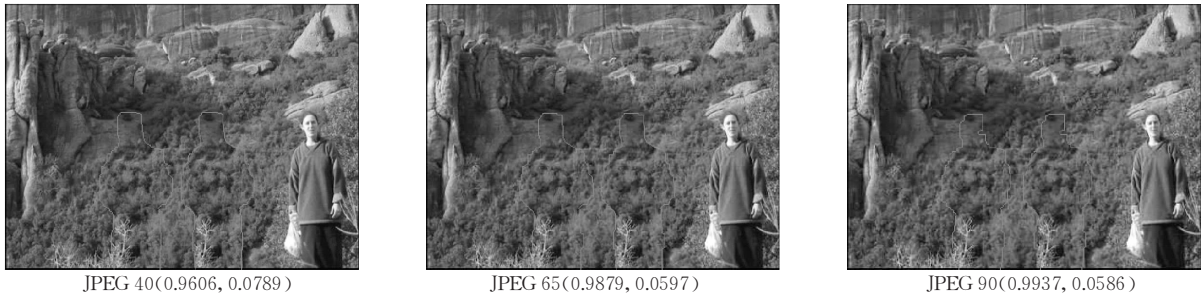


图 10 对图 8 图像进行不同质量因子 JPEG 压缩后的检测结果

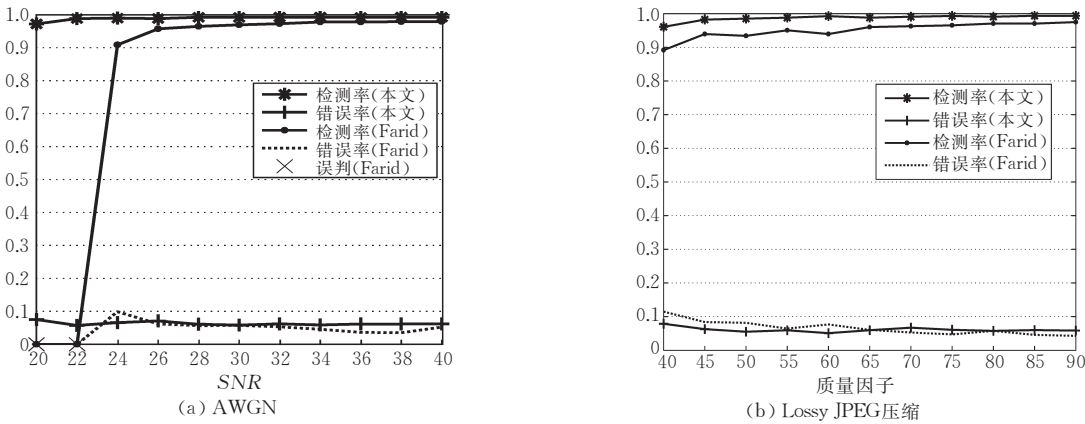


图 11 与 Farid 算法^[11]的比较

此外,本文算法还能抵抗下面的攻击类型:高斯模糊($n_1 = n_2 = 5$, 方差 $\delta^2 = 1$)及混合操作(先进行高

斯模糊,然后加高斯白噪声($SNR = 24\text{dB}$),最后做质量因子为 60 的 JPEG 压缩). 上述两个例子的检

测结果如表 2 所示.

表 2 抗高斯模糊与混合操作

	检测率/错误率	
	高斯模糊	混合操作
例子 1	(0.9891, 0.1093)	(0.9845, 0.0984)
例子 2	(0.9892, 0.0663)	(0.9772, 0.0676)

从上面的两个测试例子我们可以看到,与 Farid 方法相比,本文提出的算法检测效果很好:在没有出现误判情况下达到高检测率和较低的错误率,并且能抵抗更强的攻击和更多类型的攻击. 为了进一步测试算法的有效性和对各种后续操作的鲁棒性,我们随机选取了 100 幅图像(大小均为 300×400)进行测试,对于每一幅图像,我们随机地选取一个方块进行复制,并把它粘贴到同一图像中的不相交的区域中,然后再对这些篡改后的图像进行不同的操作:高斯模糊、加高斯白噪声、有损 JPEG 压缩及其它们的混合操作. 在测试中我们选取方块大小分别是 32×32 , 48×48 , 64×64 , 80×80 ,它们相当于图像大小的 0.853%, 1.920%, 3.413%, 5.333%, 我们算法的精度是 16×16 ,相当于图像大小的 0.213%. 最后统计它们在不同后处理操作下 100 幅图像的检测率均值、错误率均值及误判率. 表 3 给出了在没有进行后操作下的检测结果. 从表中数据可以看到,所有图像的检测率高达 99.9%,而错误率均低于 5%. 但在篡改块比较小时,由于受到图像临近相似区域等因素的影响出现了误判: 32×32 有 4 幅图像, 48×48 有 1 幅图像.

表 3 无后处理的检测结果

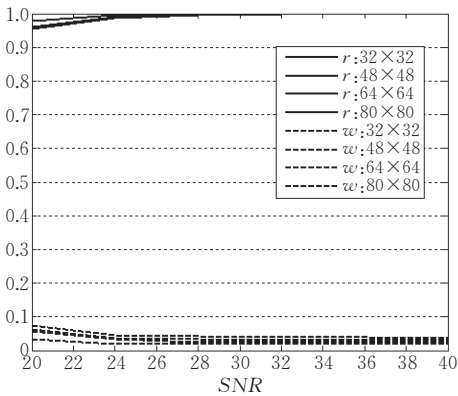
无后续操作	检测率均值	错误率均值	误判率
32×32	0.9998	0.0491	0.04
48×48	0.9999	0.0254	0.01
64×64	0.9998	0.0219	0.00
80×80	1.0000	0.0191	0.00

表 4 给出了篡改图像经过高斯模糊($n_1=n_2=5$, 方差 $\delta^2=1$)操作后算法的检测结果.

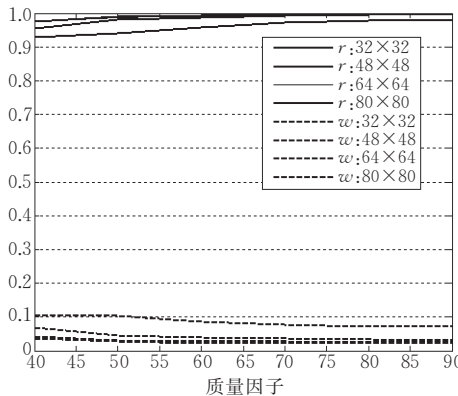
表 4 高斯模糊检测结果

高斯模糊	检测率均值	错误率均值	误判率
32×32	0.9464	0.0926	0.07
48×48	0.9677	0.0613	0.02
64×64	0.9766	0.0439	0.00
80×80	0.9797	0.0371	0.00

图 12 给出了 100 幅篡改图像经过不同强度的噪声干扰、JPEG 压缩后的检测率和错误率平均值. 对应的误判率见表 5 和表 6. 最后,表 7 给出在混合后处理的操作下的检测结果.



(a) AWGN



(b) Lossy JPEG 压缩

图 12 不同强度噪声(a)和压缩因子(b)下的检测率和错误率

表 5 添加不同强度噪声后的误判率

AWGN 参数/dB	误判率			
	32×32	48×48	64×64	80×80
20	0.20	0.04	0	0.01
24	0.08	0.01	0	0.01
28	0.06	0.01	0	0
32	0.06	0.01	0	0
36	0.06	0.01	0	0
40	0.05	0.01	0	0

表 6 不同质量因子 JPEG 压缩下的误判率

JPEG 参数	误判率			
	32×32	48×48	64×64	80×80
40	0.18	0.03	0	0
50	0.13	0.02	0	0
60	0.12	0.01	0	0
70	0.12	0.01	0	0
80	0.13	0.01	0	0
90	0.12	0.01	0	0

表 7 混合操作下的检测结果

混合操作	检测率均值	错误率均值	误判率
32×32	0.9037	0.1295	0.13
48×48	0.9326	0.0869	0.02
64×64	0.9521	0.0646	0.00
80×80	0.9505	0.0618	0.00

从以上的实验结果可以看到在篡改图像块大于

2%图像大小时(大约 48×48),算法的检测效果都比较好.但在篡改图像块比较小或篡改图像经过了比较大的后处理修改,如添加高斯噪声使得 SNR 小于 24dB,或在 JPEG 压缩其质量因子小于 50 情况下,会出现相对较高的错误率.

5 结 论

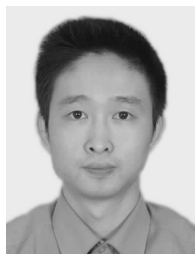
本文提出了一个区域复制篡改的自动检测与定位算法,主要的贡献如下:(1)对区域复制的图像篡改过程进行了模型化和数学描述;(2)把篡改检测转化为相似块的匹配问题.通过选择一组对篡改后处理鲁棒的特征,有效地实现了复制区域的检测;(3)所提出的区域复制检测算法,具有很强的抗后处理能力.算法仅需要在空域上进行,而无需变换到其它空间,因而算法实现上也较文献[11,15]简单.

区域复制利用了同一图像中有着相似的颜色,纹理等特性,使得被篡改后的图像在视觉上很难被发觉,而且能十分简单地实现篡改,但对比其它更高级的篡改方法(如文献[22-23]),其主要缺点在于篡改后图像特征相对明显.如今,一些结合计算机视觉和计算机图形学的更高级的篡改方法正逐步发展,相应的检测技术还有待研究.

致 谢 感谢 Farid 博士给我们提供其算法的源代码作比较实验.感谢 Fridrich 博士对我们提出的相关问题的讨论!

参 考 文 献

- [1] Light K. Fonda, Kerry and Photo Fakery. The Washington Post, Saturday, Feb. 28, 2004; A21
- [2] Farid H. A picture tells a thousand lies. New Scientist, 2003, 179(2411): 38-41
- [3] Zhu B B, Swanson M D, Tewfik A H. When seeing isn't believing. IEEE Signal Processing Magazine, 2004, 21(2): 40-49
- [4] Schneider M, Chang S F. A robust content based digital signature for image authentication//Proceedings of the Image Processing, Lausanne, Switzerland, 1996, 3: 227-230
- [5] Swaminathan A, Mao Y, Wu M. Robust and secure image hashing. IEEE Transactions on Information Forensics and Security, 2006, 1(2): 215-230
- [6] Cox I J, Miller M, Bloom J. Digital Watermarking. San Francisco, USA: Morgan Kaufmann Publishers, 2002
- [7] Katzenbeisser S, Petitcolas F. Information Techniques for Steganography and Digital Watermarking. Boston, MA: Artec House, 2000
- [8] Craver S A, Wu M, Liu B et al. Reading between the lines: Lessons from the SDMI challenge//Proceedings of the 10th Usenix Security Symposium. Washington DC, 2001
- [9] Lyu S. Natural image statistics for digital image forensics [Ph. D. dissertation]. Dartmouth College, Hanover, New Hampshire, USA, 2005
- [10] Popescu A C. Statistical tools for digital image forensics [Ph. D. dissertation]. Dartmouth College, Hanover, New Hampshire, USA, 2004
- [11] Popescu A C, Farid H. Exposing digital forgeries by detecting duplicated image regions. Dartmouth College, Hanover, New Hampshire, USA: TR2004-515, 2004
- [12] Popescu A C, Farid H. Exposing digital forgeries in color filter array interpolated images. IEEE Transactions on Signal Processing, 2005, 53(10): 3948-3959
- [13] Popescu A C, Farid H. Exposing digital forgeries by detecting traces of re-sampling. IEEE Transactions on Signal Processing, 2005, 53(2): 758-767
- [14] Johnson M K, Farid H. Exposing digital forgeries by detecting inconsistencies in Lighting//Proceedings of the ACM Multimedia and Security Workshop. New York, 2005: 1-9
- [15] Fridrich J, Soukal D, Lukas J. Detection of copy-move forgery in digital images//Proceedings of the Digital Forensic Research Workshop. Cleveland OH, USA, 2003
- [16] Lucas J, Fridrich J, Goljan M. Digital "bullet scratches" for images//Proceedings of the IEEE International Conference on Image processing. Genova, Italy, 2005: III-65-8
- [17] Lukcas J, Fridrich J, Goljan M. Detecting digital image forgeries using sensor pattern noise//Proceedings of the SPIE: Security, Steganography, and Watermarking of Multimedia Contents. San Jose, California, USA, 2006, VIII6072(1): 362-372
- [18] Ng T T, Chang S F. A model for image splicing//Proceedings of the IEEE International Conference on Image Processing. Singapore, 2004: 1169-1172
- [19] Ng T T, Chang S F, Hsu J et al. Physics-motivated features for distinguishing photographic images and computer graphics//Proceedings of the ACM Multimedia. Singapore, 2005: 239-248
- [20] Geradts Z J, Bijhold J, Kieft M et al. Methods for identification of images acquired with digital cameras//Proceedings of the SPIE: Enabling Technologies for Law Enforcement and Security, 2001, 4232(1): 505-512
- [21] Gonzalez R C, Woods R E. Digital Image Processing. New Jersey: Pearson Education, 2002
- [22] Criminisi A, P'erez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212
- [23] Wei L Y. Texture synthesis by fixed neighborhood searching [Ph. D. dissertation]. Stanford University, Palo Alto, California, USA, 2001



LUO Wei-Qi, born in 1980, Ph. D. candidate. His research interests include digital image forensics, image processing and pattern recognition.

HUANG Ji-Wu, born in 1962, Ph. D., professor, Ph. D. supervisor. His current research interests include multimedia security and data hiding.

QIU Guo-Ping, born in 1964, Ph. D., associate professor. His current research interests include computational methods and tools for visual information processing.

Background

Detection of Region-Duplication in digital image is in the field of passive/blind forensics, which is a new research area in information forensics.

With the proliferation of digital cameras and computers, as well as software for image editing, the problem of digital image forgery is potentially very serious. Authentication of digital images becomes an important issue. Unlike the signature-based and watermark-based methods, which need to do some operations such as signature generated or watermark embedded in advance, the new forensic is passive and blind. It needs not any extra side information in detection. The method assumes that different imaging devices or processing etc. would introduce different inherent patterns into the output images. These underlying patterns are consistent in the original un-tampered images and would be altered after some kinds of manipulations. Therefore the patterns can be used as evidences for image source identification and alteration detection, i. e. the two main issues in passive forensics. Several research groups, e. g. Farid H et al. in Dartmouth College, Fridrich J et al. in University of Binghamton, Wu Min et al. in University of Maryland, Chang S F et al. in University of Columbia etc., have started to investigate the new technology and some works have been reported. However passive technology for image forensics is still in its infancy. There

are many open issues.

Region-Duplication is one of the common image forgery techniques. The attacker may perform some post processing attack after Region-Duplication operation, which makes the task of detecting forgery significantly harder. The key of the detection algorithm is the robustness against the post image processing, such as noise contamination, Lossy JPEG compression, blurring etc. Several researchers have developed methods for detecting this form of forgery. Fridrich analyzed the DCT coefficients for each block, while Farid employed principal component analysis (PCA) to capture the image blocks' features.

The authors' method extracts the 7 low-frequency components for each overlapping block. All of the methods are based on block matching. The main difference of these methods is the choice of the features. From the experiments, the features of the authors' methods are more robust to various post region duplication image processing operations comparing with the prior methods.

This work is supported by the National Natural Science Foundation of China under grant No. 90604008, the National Natural Science Funds for Distinguished Young Scholar under grant No. 60325208 and the National Natural Science Foundation of Guangdong under grant No. 04205407.