

图像复制-粘贴篡改快速取证

王忠美, 龙永红

(湖南工业大学 计算机与通信学院, 湖南 株洲 412008)

摘要: 提出了一种新的篡改检测算法, 将图像分为大小相同的重叠块, 然后提取每块的特征向量, 并对所求出的特征向量通过基数排序来减少搜索空间, 减少运算量。对排序后的特性向量求相邻的特征向量的差异(偏移向量), 并对偏移向量进行统计, 如果存在大量的相同偏移向量可认定为篡改区域。最后利用数学形态学操作获得最终的篡改定位结果。

关键词: 复制-粘贴篡改; 篡改取证; 基数排序; 数学形态学

中图分类号: TP391.4

文献标志码: A

文章编号: 1673-9833(2010)06-0058-05

Fast Forensics for Image Copy-Move Forgery

Wang Zhongmei, Long Yonghong

(School of Computer and Communication, Hunan University of Technology, Zhuzhou Hunan 412008, China)

Abstract: Proposes a novel method for detecting copy-move forgery: the given image is divided into overlapping blocks of equal size, feature of each block is then extracted and represented as a vector. All the extracted feature vectors are then sorted using the radix sort. The difference (shift vector) of the positions of every pair of adjacent feature vectors in the sorting list is computed. The accumulated number of each of the shift vectors is evaluated. Finally, mathematical morphology is performed to obtain the final result.

Keywords: copy-move forgery; forensics of tampering; radix sort; mathematical morphology

0 引言

数码相机的普及和图像处理软件的广泛应用, 使得图像篡改操作越来越容易, 并且人眼难以分辨。针对数字图像篡改现象, 提出了用于检测图像真实性的图像取证技术。国内外针对数字图像鉴别的核心技术研究主要有2大类: 主动图像取证和被动图像取证。主动取证技术主要是指基于脆弱数字水印^[1]和基于数字签名的图像取证技术。主动取证技术需要对图像进行预处理, 如提取摘要或插入水印。然而, 在较多实际情况中, 是要求在不依赖任何预签名提取或预嵌入水印信息的前提下对图像的真伪进行检测, 这就是图像的被动取证。

近年来, 被动数字取证技术受到了学者们的广泛关注, 并已经取得了初步的研究成果。Fridrich等^[2]提出了利用数码相机传感器的固有噪声对图像进行取证。Popescu等^[3]提出了利用图像的统计特征来对图像的真实性进行检测。彩色滤波图像(CFA), 重采样^[4], 光源方向检测^[5]等方法用来进行图像真实性取证, 施云庆等人提出采用自然图像的统计特征的盲取证技术对图像的真实性进行判断, 其中主要有Ng提出图像的双相干系数和边缘百分比特征方法和图像质量度量等。国内, 数字图像被动取证技术刚刚开始, Luo等^[6]对JPG图像剪切重压缩进行检测, 提出了新的图像复制篡改区域的检测。朱秀明等^[7]在图像重采样和图像模

收稿日期: 2010-09-19

通信作者: 王忠美(1984-), 男, 湖北荆州人, 湖南工业大学硕士研究生, 主要研究方向为数字图像处理与模式识别,

E-mail: ldwangzm2008@163.com

糊取证方面取得一定进展。但被动取证技术还处在起步阶段, 还有很多亟待解决的问题。

图像复制-粘贴篡改是在一幅图像上进行的篡改操作, 是指把数字图像中的一部分区域进行复制并粘贴到同一幅图像中的另一个区域中, 它是一种简单而有效的图像伪造手段, 通常篡改者会对复制-粘贴后的图片进行后处理操作, 如添加噪声, 压缩等以使篡改图像难以辨认与检测。对图像的复制-粘贴篡改检测 Fridrich 等提出了离散余弦逆变换 (DCT) 系数模糊匹配法, Popescu 等提出了主成分分析法, 但都有缺陷, 如 Fridrich 仅考虑了 JPEG 后处理情况。骆伟祺^[8]提出了复制-篡改的一些合理性假设, 得到一种新的区域复制-粘贴篡改检测方法, 能有效对抗篡改的后处理操作, 但不能对有旋转的复制-粘贴篡改操作进行检测, 且检测时间复杂度高。

本文提出了一种有效的图像复制-粘贴篡改检测的方法, 对图像进行分块处理, 对每块提取相关的特征, 利用这些特征进行图像篡改检测, 并且该方法对复制-篡改操作的后处理 (如有损压缩, 高斯噪声等操作) 具有鲁棒性。

1 复制-粘贴篡改模型

复制-粘贴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。复制-粘贴篡改后的图像中通常存在着两处完全相同 (或相似) 的区域, 而且一般情况下对图像内容进行复制-粘贴篡改, 选择的复制区域通常是连通的、面积较大的区域; 同时图像复制区域和粘贴区域之间存在着相关性, 即对应的相似图像块对具有相同的偏移向量。由此可知, 复制区域和粘贴区域应该是由许多相邻的、具有相同偏移向量的图像块对组成的连通区域对, 而不是由零散相似块对组成的集合。因此如果图像中存在复制-粘贴区域, 则对应的偏移向量的统计值应该较大, 且远远大于其余偏移向量的统计值。综上所述, 本文给出了一个基本图像复制-粘贴篡改模型, 如图 1 所示。

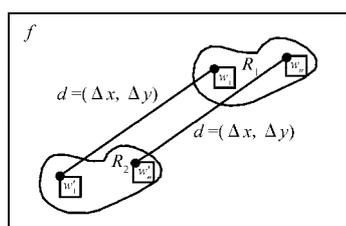


图 1 图像复制-粘贴篡改模型

Fig. 1 The model of copy-move image forgery

该模型满足以下合理的假设:

1) 被复制的区域 R_1 是一个连通的、面积较大的区域, 而不是若干零散像素的集合;

2) 复制区域 R_1 和粘贴区域 R_2 不相交, 即

$$R_1 \cap R_2 = \emptyset;$$

3) 设定的图像块的大小应小于待检测的复制区域, 即 $w_1 < R_1$;

4) R_1, R_2 中对应的相似图像块对具有相同的偏移向量 $d = (\Delta x, \Delta y)$ 。

复制-粘贴篡改模型可描述为: 在 1 幅复制-粘贴篡改后的图像中, 存在 2 个区域 R_1, R_2 和 1 个偏移向量 $d = (\Delta x, \Delta y)$, 如 $\forall (x_1, y_1) \in R_1, (x'_1, y'_1) \in R_2$, 则有 $f(x_1, y_1) = f(x'_1, y'_1)$ 。其中 $x'_1 = x_1 + \Delta x, y'_1 = y_1 + \Delta y$, $f(x_1, y_1)$ 和 $f(x'_1, y'_1)$ 分别表示图像在点 (x_1, y_1) 和点 (x'_1, y'_1) 处的灰度值, $f(x_1, y_1)$ 为复制区域, $f(x'_1, y'_1)$ 为粘贴区域。

复制-粘贴篡改图像中的复制区域和相对应的粘贴区域基本上相似, 而通过大量实验发现, 在自然图像中 (除有大片平坦区域的图像外) 存在这样较大面积的相似区域的可能性是较小的。基于这个观点, 可以将一幅图像中存在较大面积的相似区域对作为判断图像被篡改过的依据, 即复制-粘贴伪造图像盲取证算法的关键, 就是判断一幅给定的图像中是否存在一个或多个如图 1 中所示的区域对 R_1 和 R_2 。若存在, 则表明图像很有可能被复制-粘贴篡改过。但一幅图像经过复制-粘贴篡改后常常会进行后处理操作如加噪声, 进行 JPEG 压缩等操作, 已达到逼真的效果, 这样被复制的区域 R_1 与被粘贴的区域 R_2 不是精确相等, 如果仅从像素值考虑图像取证将会失败, 笔者采用基于块的特征进行复制-粘贴篡改检测。

2 检测算法

本文提出的算法是基于块特征匹配的, 首先将图像分成有重叠区域的块, 提取每块的特征向量, 利用特征向量将块进行排序, 减少了相似块匹配的搜索空间; 然后利用匹配阈值得到相似块对; 最后进行篡改区域的定位操作。

2.1 特征向量的提取

对待检测的图像进行分块操作, 将图像分为重叠的块, 重叠块的划分为相邻的图像块之间移动 1 个像素形成, 每个块的大小为 16×16 像素, 对每个块用 1 个 9 维的特征向量 $v = (x_1, x_2, \dots, x_9)$ 来表示。对图像块 A 形成的重叠块 B 特征向量的定义如下: 首先将每个块分成 4 等分小块 S_1, S_2, S_3 和 S_4 , 假设 $f_{Ave}(\cdot)$ 表示灰度的平均值, 如式 (1) 中描述, f_1 表示块 B 的灰度平均值, f_2, f_3, f_4, f_5 表示子块 S_2, S_3, S_4, S_5 平均灰度的比值, $f_6,$

f_7, f_8, f_9 表示块 S_1, S_2, S_3, S_4 与块 B 的差异。

$$f_i = \begin{cases} f_{Ave}(B), & i=1; \\ f_{Ave}(S_{i-1}) / (4f_{Ave}(B) + \varepsilon_1), & 2 \leq i \leq 5; \\ f_{Ave}(S_{i-1}) - f_{Ave}(B), & 6 \leq i \leq 9. \end{cases} \quad (1)$$

将由式(1)求得的 f_i 的值转化为整数 x_i , 其中 x_i 的范围为 $0 \sim 255$, 转化式(2)如下:

$$x_i = \begin{cases} \lfloor f_i \rfloor, & i=1; \\ \lfloor 255 * f_i \rfloor, & 2 \leq i \leq 5; \\ \lfloor 255 * (f_i - m_2) / (m_1 - m_2 + \varepsilon_2) \rfloor, & 6 \leq i \leq 9. \end{cases} \quad (2)$$

式中: $m_1 = \max_{6 \leq i \leq 9} \{f_i\}$ 且 $m_2 = \min_{6 \leq i \leq 9} \{f_i\}$; $\lfloor \cdot \rfloor$ 表示向下取整操作, 将 f_i 转化为整数。通过将数据转化为整数, 可以减少下一步中特征向量排序的时间。通过这些包含 9 个特征的特征向量, 能表示出复制区域和粘贴区域的信息。

2.2 特征向量排序

假设给定的原始图像的大小为 $N \times N$ 像素, 分块操作块大小为 $b \times b$ 像素, 则总块数 k 为: $k = (N-b+1)^2$, 块对应的特征向量为 $v_1, v_2, v_3, \dots, v_k$ 。为减少特征向量比较搜索空间, 对特征向量进行基数排序操作。

设 $v_1, v_2, v_3, \dots, v_k$ 表示块 B_1, B_2, \dots, B_k 排序后的特征向量, $P(B_i)$ 用来记录每个块 B_i 左上角的位置, 偏移向量 U 用来表示排序后的 2 个特征向量之间的差异, 偏移向量定义如下:

$$U_i = P(B_{i+1}) - P(B_i), \quad (3)$$

对排序后的特征向量按式(3)计算, 得到偏移向量为 U , 并统计相同偏移向量个数 N 。如果图像中存在复制-粘贴篡改区域, 则被篡改与被复制的部分对应的特征向量比较接近, 因此可以用相同偏移向量的总数 N 来判断图像是否为复制-粘贴篡改区域。笔者通过设置一个阈值 θ 和偏移向量总数 N 的比较, 确定图像是否为篡改图像, 如果 $N > \theta$, 则可能为复制-粘贴篡改图像。由于自然图像的连续性, 相邻的块具有较强的相似性, 在篡改检测中设置图像块的距离阈值 T 来排除邻近块的干扰, 当偏移向量个数 $N > \theta$ 且块之间距离 $D > T$ 时, 确定匹配块是否属于复制-粘贴部分。

2.3 检测算法完整描述

对完整的检测算法描述为:

1) 对图像进行分块操作, 设图像大小为 $N \times N$ 像素, 将图像分解为块大小为 $b \times b$ 像素的重叠的图像块, 则共得到 $(N-b+1) \times (N-b+1)$ 个图像块。

2) 计算每个图像块的特征向 $v=(x_1, x_2, \dots, x_9)$, 并对每个特征向量求得特征值进行取整操作, 便于特征向量进行排序操作, 并对特征向量进行排序操作。

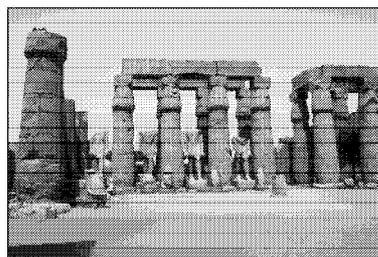
3) 对排序后的特征向量对应的图像块, 按照偏移向量的计算公式, 计算出 2 个图像块对的偏移向量, 统计相同的偏移向量的个数 N , 与阈值 θ 进行比较, 并对块间距离 D 和 T 比较, 当满足条件 $N > \theta$ 且 $D > T$ 时确定存在篡改操作。

4) 利用数学形态学的方法去掉一些满足条件的单独块, 用区域填充的方法填补检测到的复制区域与粘贴区域的“空洞”, 得到最终的篡改区域定位。

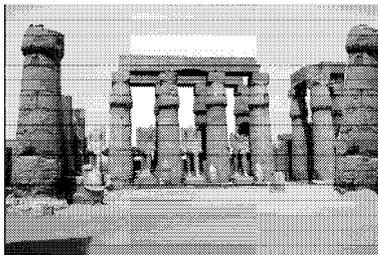
3 实验结果及分析

从中国科学院自动化所的 CASIA1 图像库中, 选取部分复制和粘贴部分没有旋转的复制-粘贴篡改图像, 进行篡改检测实验, 所选取的篡改图像中可能存在一个或多个篡改区域。分块的块大小为 16×16 像素, 选取用来与偏移向量数目进行比较的阈值 $\theta=100$ 和块之间距离阈值 $T=10$, 进行图像复制-粘贴篡改检测, 如果偏移向量数目 $N > \theta$ 且块之间距离 $D > T$ 时确定为复制-篡改区域, 并对篡改区域定位。

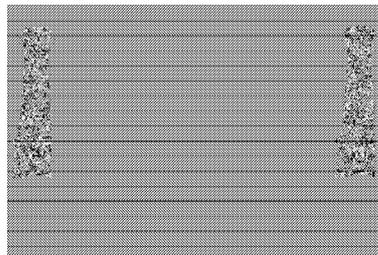
对篡改后的图像检测结果如图 2 所示:



a) 原始图像



b) 篡改图像



c) 检测结果

图 2 原始图像检测结果

Fig. 2 The result of the original image

为了验证算法的鲁棒性, 对篡改后的图像进行不同的后处理操作, 常见处理操作为对图像采用不同的

质量因子进行 JPEG 压缩, 添加不同程度的高斯噪声, 用不同等级的模糊因子进行高斯模糊处理等。下面给出了对图 2 b) 中篡改图像进行后续处理后, 对篡改图像进行篡改检测结果如图 3 和 4 所示:

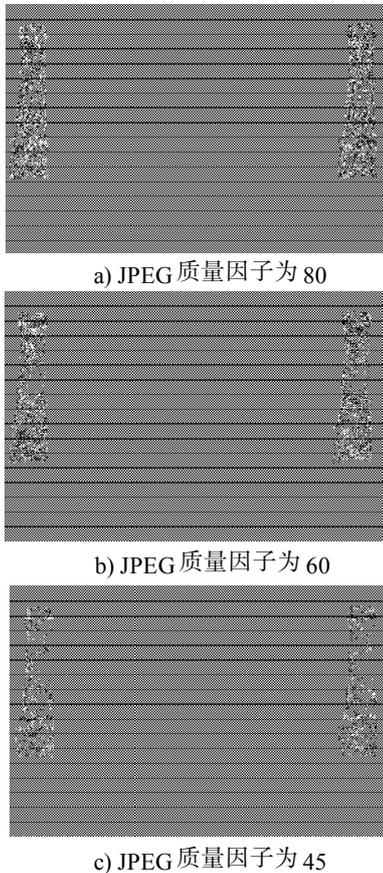
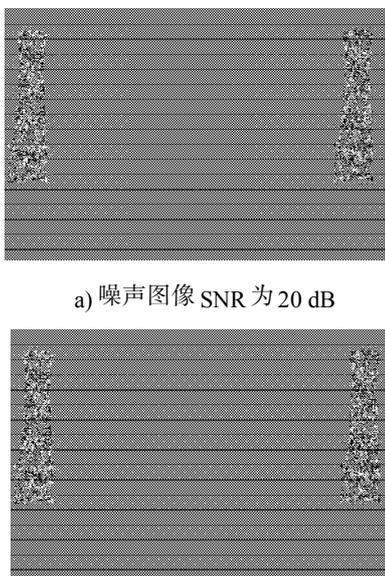


图 3 对图 2 中的篡改图像进行不同质量因子 JPEG 压缩后的检测结果
Fig. 3 The result of compressed image with different JPEG qualities



a) 噪声图像 SNR 为 20 dB
b) 噪声图像 SNR 为 30 dB



c) 噪声图像 SNR 为 40 dB

图 4 对图 2 中的篡改图像添加不同强度噪声后的检测结果

Fig. 4 The result of the images with different noise

对多幅复制-粘贴篡改的图像进行检测, 得到的检测结果如表 1, 对篡改图像采用不同后处理操作后, 得到对应操作下的检测率。

表 1 不同后处理操作后的检测率

Table 1 The detection rate with different operation

篡改图像后处理方式	图像数 / 幅	检测率 / %
没有后处理	50	98
JPEG 压缩	QF 为 80	96
	QF 为 60	92
	QF 为 45	86
高斯噪声	SNR 为 20	94
	SNR 为 30	78
	SNR 为 40	54

通过实验验证, 该算法对复制-粘贴篡改图像的检测具有较高的检测率, 对经过压缩, 添加噪声等后处理后的篡改图像也具有较好的检测效果。但当图像的复制-粘贴部分存在旋转时效果不理想, 不能有效地检测到图像的篡改部分。

4 结语

本文提出了一种简单有效的图像复制-粘贴篡改检测方法, 对篡改图像进行分块操作, 提取图像块的特征向量, 并通过基数对特征向量排序减少了块匹配的搜索空间, 减少了时间复杂度。本文所采用的图像特征能较好地抵抗篡改图像后处理操作。实验发现, 不能对存在任意角度旋转的篡改进行检测。下一步工作是寻求旋转不变特征篡改检测和其它类型的篡改操作的检测, 如图像的拼接操作等, 同时把图像篡改检测方法用于视频篡改检测。

参考文献:

[1] 宋玉杰, 谭铁牛. 基于脆弱水印的图像完整性验证研究[J]. 中国图像图形学报, 2003, 8(1): 1-7.
Song Yujie, Tan Tieniu. The Research of Image Integrity

- Verification Based on Fragile Watermarking[J]. Journal of Image and Graphics, 2003, 8(1): 1-7.
- [2] Fridrich J, Soukal D. Detection of Copy-Move Forgery in Digital Images[R]. Cleveland: [s.n.], 2003.
- [3] Popescu A C, Farid H. Exposing Digital Forgeries by Detecting Duplicated Image Regions[R]. Dartmouth: Dartmouth College, 2004.
- [4] Poesscu A C, Farid H. Exposing Digital Forgeries in Color Filter Array Interpolated Image[J]. IEEE Transactions on Signal Processing, 2005, 53(10): 3948-3959.
- [5] Poesscu A C, Farid H. Exposing Digital Forgeries by Detecting Traces of Re-Sampling[J]. IEEE Transactions on Signal Processing, 2005, 53(2): 758-767.
- [6] Johnson M K, Farid H. Exposing Digital Forgeries by Detection Inconsistencies[C]//Proceedings of the 7th Workshop on Multimedia and Security. New York: ACM, 2005: 1-10.
- [7] 朱秀明, 宣国荣, 姚秋明, 等. 基于EM算法的数字取证中图像重采样检测[C]//六届全国信息隐藏暨多媒体信息安全学术研讨会. 哈尔滨: 哈尔滨工业大学学报, 2006: 868-871.
- Zhu Xiuming, Xuan Guorong, Yao Qiuming, et al. The Detection of Resampling in Digital Image Forensics Based on EM Algorithm[C]//The 6th National Information Hiding and Multimedia Security Conference. Harbin: Journal of Harbin Institute of Technology, 2006: 868-871.
- [8] 骆伟祺, 黄继武, 丘国平. 鲁棒的区域复制图像篡改检测技术[J]. 计算机学报, 2007, 30(11): 1998-2007.
- Luo Weiqi, Huan Jiwu, Qiu Guoping. The Technology of Robust Regional Copy-Move Image Forgery Detection[J]. Journal of Computer, 2007, 30(11): 1998-2007.

(责任编辑: 罗立宇)

(上接第40页)

便携性、资源丰富性和教学辅助性, 注重人(学生)一机(智能移动终端)交互, 其本质上是移动学习终端作为教师角色对学生进行施教, 教与学互动过程在移动学习终端和学生之间发生, 而对作为教学主体的教师、学生、家长和学校间的互动重视不足。本文提出的多方互动式的移动学习系统使得教师和家长可以对学生进行离线的现场或在线的远程多媒体教学、指导和评价, 并进而实现学生之间的协同学习、教师、家长和学生之间的信息沟通, 从而发挥教师在教学过程中的主导性以及学生在学习过程中的能动性。

参考文献:

- [1] 刘豫钧, 鬲淑芳. 移动学习—国外研究现状之综述[J]. 教育技术, 2004, 14(3): 12-16.
- Liu Yujun, Ge Shufang. An Overall View of the Mobile Learning Domain[J]. Modern Educational Technology, 2004, 14(3): 12-16.
- [2] 天极 yesky. 以学为本—诺基亚本土化软件行学一族[EB/OL]. (2007-09-16), [2010-06-15]. <http://digi.cfan.com.cn/mobile/labs/2007-09-16/1189945581d106466.shtml>.
- Celestial Pole Yesky. Study First-Nokia Localization Software Group[EB/OL]. (2007-09-16), [2010-06-15]. <http://digi.cfan.com.cn/mo-bile/labs/2007-09-16/1189945581d106466.shtml>.
- [3] 傅健, 杨雪. 国内移动学习理论研究与实践十年瞰览[J]. 中国电化教育研究, 2009 (7): 36-41.
- Fu Jian, Yang Xue. Ten Years Scan of Theory and Practice of Domestic Movable Study[J]. China Educational Technology, 2009 (7): 36-41.
- [4] 侯伟, 朱东鸣. 移动学习技术开发实践的研究[J]. 现代教育技术, 2010, 20(1): 115-119.
- Hou Wei, Zhu Dongming. Research on the Practice of Developing Technologies on Mobile Learning[J]. Modern Educational Technology, 2010, 20(1): 115-119.
- [5] Schreurs J. Mobile E-Learning Course Scenario Model on PDA [J]. International Journal of Mobile Learning and Organization, 2008, 2(4): 358-368.

(责任编辑: 罗立宇)