文章编号: 1006 - 4729(2012)04 - 0369 - 06

数字图像取证技术的发展

魏为民[1], 胡胜斌[2], 赵 琰[1]

(1. 上海电力学院,上海 200090; 2. 上海宝信软件股份有限公司,上海 201203)

摘 要: 随着图像处理技术的快速发展,数字图像被广泛地应用于互联网. 与此同时,视觉上难以觉察的篡改图像也日益泛滥. 数字图像被动取证作为一种不依赖任何预签名提取或预嵌入信息来鉴别图像真伪和来源的技术,正成为多媒体安全领域新兴的研究热点,有着广泛的应用前景.

关键词: 数字取证; 图像篡改; 篡改检测; 数字水印; 数字图像被动取证
中图分类号: TP393.08; TP751 **文献标志码:** A

Survey on the Study of Image Forensics

WEI Wei-min[1], HU Sheng-bin[2], ZHAO Yan[1]

(1. *Shanghai University of Electric Power, Shanghai* 200090, *China*; 2. *Shanghai Baosight Software Co., Ltd., Shanghai* 201203, *China*)

Abstract: Digital images are widely used in the Internet and many other applications. As image processing techniques are developing at a rapid speed, tampering digital images without leaving any obvious traces becomes easier and easier. Passive image forensics, a technology of detecting image authenticity and source without relying on any pre-extraction or pre-embedded information, has become a hot topic with a promising prospect in multimedia security.

Key words: digital forensics; image forgery; tampering identification; digital watermark; passive image forensics

利用 Photoshop 等图像处理软件可非常容易地修改数字图像的内容,用肉眼无法辨别图像的真假. 在胶片时代,修改照片需要非常有经验的专家在暗房里费时费力地操作,而在数字时代,数字图片取代了胶片,图像处理软件,如 Photoshop 等软件功能日益强大,使得任何人都可篡改图片,修改后的图片即便是专家亦很难辨别其真假. 在某些情况下对数字图像的恶意篡改和传播,会给社会和人们的生活带来巨大危害,因此数字图像的内容保护和完整性认证成为国内外研究的热点.

1 数字图像的篡改

美国 Dartmouth 大学的 FARID Hany 教授将数字图像篡改手段分为如下 6 类[1].

(1) 合成 是指同一幅图像内或不同图像之间的复制、粘贴操作,以此造成某种假象或者隐藏

收稿日期: 2012 - 03 - 16

通讯作者简介: 魏为民(1970 -),男,副教授,博士,湖北蒲圻人. 主要研究方向为图像处理,数字取证,信息隐藏,数据挖掘. E-mail: weiweimin@hotmail.com.

基金项目: 上海市自然科学基金(11ZR1414300); 上海市教育委员会科研创新项目(11YZ194); 上海电力学院人才引进基金(K2010-020).

图像中某个对象. 为消除伪造图像中的篡改痕迹，往往会对篡改部分进行缩放、旋转和润饰等处理.

(2) 变体　通过找出源图像和目标图像对应的特征点，以不同的权重叠加两幅图像，在把一幅图像渐变成另一幅图像的同时，兼有两幅图像的特征.

(3) 润饰　一种图像修补技术，最主要的操作是在同一幅图像中对局部区域进行复制粘贴，并用模糊操作扫除边缘拼接的痕迹.

(4) 增强　通过改变图像特定部位的颜色、对比度等来着重突出某部分内容，这种操作虽然不明显改变图像内容，但可弱化或突出某些细节.

(5) 计算机生成　是艺术家或者程序员利用计算机软件，如 3Ds Max，SoftImage XSI，Maya，Terragen 等生成图像，可以分为 Photorealistic Computer Graphics (PRCG) 和 Non-Photorealistic Computer Graphics (NPRCG) 两类，其中大量的 PRCG 与自然图像从视觉上已经做到很难区分的程度. Computer Graphics (简称 CG) 的生成过程是: 首先构建一个 3D 多边模型模拟期望的形状，然后为模型赋予颜色和纹理，将修饰好的模型用模拟光源照射并送到一个虚拟的照相机前生成图像. 从产生机理来看，真实照片与 CG 有很多不同之处，前者是从真实世界投影到图像设备传感器上形成的，形成过程非常复杂; CG 则是通过许多基本的图像处理步骤来粗略模拟真实照片的形成过程而得来的，不管有多逼真，总是机器而非自然的产物，它与自然的真实照片在图像统计特征(如直方图的连续性、色彩数量、细小纹理的复杂程度等) 方面会有一些差异，这些差异被很多技术用作区分两种图像的重要信息.

(6) 绘画　专业人员或艺术家利用 Photoshop 等图像处理软件进行图像制作. 这类图像往往跟真实场景的照片有较大差别，不会引起混淆. 但由于当前扫描仪的扫描精度越来越高，某些绘画或打印作品经过扫描所得到的数字图像与自然图像或由计算机生成的图像难以区分.

图 1 是一些影响重大的图像篡改例子，图 1a 为 2003 年《洛杉矶时报》刊登的伊拉克战争现场照片，被普遍认为具有角逐普利策最佳新闻图片奖的实力，却被人指出是由图 1b 和图 1c 合成的. 2004 年美国参议员 KERRY John 争取民主党总统提名初选时，网络上出现如图 1d 所示与美国反越战女影星 FONDA Jane 同台的照片，事实上，图 1e 的 KERRY 照片摄于 1971 年，而图 1f 的 FONDA Jane 照片则是摄于 1972 年. 2008 年 7 月世界各大媒体纷纷转载伊朗革命卫队网站上刊发的图 1g 所示“导弹齐射”照片，《纽约时报》却发现，这张 4 枚导弹腾空而起的照片存在人为修改的痕迹: 其中左起第 3 枚导弹很可能是照片上其他两枚导弹克隆的产物，伊朗方面随后撤掉这幅照片，将其变成如图 1h 所示的只有 3 枚导弹同时升空的图片，但未就此事作任何解释. 图 1i，图 1j，图 1k 都被确认为是合成的照片，分别被网友戏称为广场鸽、藏羚羊和周老虎，堪称新的“吉祥 3 宝”. 接连出现的虚假新闻照片事件，打碎了不少人长期以来对影像真实性的信心，正如 FARID 所说，“我们生活在一个不再能够相信自己所看到或者听到的世界中”[2].

上述 6 种篡改方式可归并称之为数字图像真实性篡改. 如果结合基于数字图像完整性的隐秘分析取证和数字图像版权认证，还可增加如下 3 种篡改类型[3].

(1) 数字图像完整性篡改　即数字图像水印和隐写术，利用数字图像中的冗余空间携带秘密信息. 数字图像的冗余空间主要在图像的最低比特有效位(Least Signed Bits，LSB)，由于冗余信息对人类视觉感知贡献较小，携带有秘密信息的图像与原始图像在视觉上虽然并无差异，但它破坏了数字图像的完整性.

(2) 数字图像原始性篡改　指的是 2 次获取图像，即原始的“现场”图像经过数字处理后形成的新数字图像. “现场”图像是经过一次图像获取设备获得的. “2 次获取”图像经过两次以上图像获取设备获得，如照片的扫描图、照片的照片等. 例如广受关注的华南虎照片事件就属于数字图像原始性篡改取证范畴，其取证的主要焦点在于数字华南虎照片是否经过了 2 次获取，即老虎照片的原始现场是平面虎还是立体虎的问题.

(3) 数字图像版权篡改　主要改变的是图像版权等一些额外附加信息，并不改变图像内容和像素信息. 对于数字图像作品的版权篡改主要集中在对数字图像作品的作者或所有者的版权篡改、对数字图像作品的购买者的篡改，以及对数字图像作品防打印或复印功能的攻击篡改. 例如通过修改照片的 EXIF (Exchangeable Image File Format) 信息从而篡取所有者版权.

a 伊拉克战场合成照片　b 现场照片1　c 现场照片2

d Kerry和Jane合成照片　e Kerry原始照片　f Jane原始照片

g 伊拉克“导弹齐射”照片　h 导弹发射原始照片

i 广场鸽接种禽流感疫苗　j 青藏铁路为野生动物开辟生命通道　k 周老虎

图1　图像篡改例子

借助各种先进的图像处理技术,伪造者可使伪造图像更为逼真,已有多种技术可以从图像中准确分割出感兴趣对象(Region Of Interest, ROI)以便制作合成图像[4,5].

如果待分割的对象具有丰富的细节边缘(如毛发等)并有一定的透明性,可用 Bayesian 方法[6]、Poisson 方法[7]解决这一类修边(Matting)问题,即提取这些具有丰富细节边缘的对象并使之与新背景融合在一起. 而当伪造者去除图像中的某些对象后,可以采用图像修复(Inpainting)技术根据周边内容情况自然填补空白位置[8]. 这都可用于增进伪造图像的乱真效果.

2　数字取证

随着计算机技术的成熟与广泛使用,利用计算机和其他数字产品进行犯罪的诸多证据都以数字形式通过计算机或网络进行存储和传输,因此出现了电子证据(Digital Evidences)这一概念. 由于电子证据的特殊性,其获取、存储、传输和分析都需要特殊的技术手段和严格的程序,否则难以确保证据的客观性、关联性和合法性. 数字取证(Digital Forensics)作为法学和计算机科学的交叉科学,能揭示与数字产品相关的犯罪、过失行为,利用一切合法的科学方法和工具,从不同学科的

角度及其相互关系等方面进行研究. 简而言之,数字取证就是对以 0/1 二进制表示的数据电文进行识别、保存、收集、检查、分析和呈堂的活动过程[9].

国外从 20 世纪 80 年代就开始研究数字取证,在取证思想、理论、技术和方法等方面取得不少成果. 在 90 年代后期,提出了 5 种较为典型的计算机取证过程模型,即: 基本过程模型(Basic Process Model),事件响应过程模型(Incident Response Process Model),法律执行过程模型(Law Enforcement Process Model),过程抽象模型(Abstract Process Model),其他过程模型. 从取证技术使用的角度看,根据 DFRWS(Digital Forensic Research Workshop) 框架[10],取证技术可以分为以下 6 类:

(1) 识别类 判定可能与指控或突发事件相关的项目、属性和数据;

(2) 保存类 保证证据状态的完整性;

(3) 收集类 提取或捕获突发事件的项目及其属性或特征;

(4) 检查类 对突发事件的项目及其属性或特征进行检查;

(5) 分析类 为了获得结论而对数字证据进行融合、关联和同化;

(6) 呈堂类 客观、有条不紊、清晰、准确地报告事实.

目前,美国至少 70% 的法律部门拥有自己的计算机取证实验室,经过资格认定的取证专家使用专门技术,通过网络或从犯罪现场获取的计算机和外部设备进行证据的提取和分析,并将这些证据提交法庭作为裁决的依据. 近年来,我国研究机构和有关部门意识到数字取证的重要性,已经开始进行理论探讨和技术开发.

3 数字图像取证

作为计算机取证的一个重要分支,数字图像取证技术(Digital Image Forensics) 是对源于数字图像资源的数字证据进行确定、收集、识别、分析,以及出示法庭的过程[11]. 不同于先前的计算机取证,数字图像取证主要是针对数字图像内容的完整性和原始性,而不是对计算机文件或磁盘格式的取证. 从现有的数字图像取证类型来看,数字图像认证方法可分为 3 类: 一是基于数字水印图像认证的主动方法[12],在被保护的图像中预先嵌入脆弱水印,篡改图像将破坏水印而暴露篡改行为,其局限性在于水印嵌入会对载体图像造成轻微变化且无法保护大量未嵌入水印的图像; 二是基于数字签名图像认证的半主动方法[13],利用图像内容生成长度很短的认证码、数字签名或视觉摘要(Visual Hash),认证码和数字签名对任何改动都很敏感,但视觉摘要仅对恶意篡改比较敏感,对压缩、滤波等合法处理不敏感,认证时可通过确认图像内容和认证码、签名或摘要是否匹配即可,该方法虽然没有改动图像,但需预先产生辅助数据; 三是被动取证,这类方法既不需要事先在图像中嵌入水印,也不需要依赖辅助数据,仅根据待认证的图像本身判断其是否经过篡改、合成、润饰等伪造处理. 实际应用中待认证图像往往既未被嵌入脆弱水印,也没有辅助信息可以利用,因此被动取证是更具现实意义的图像认证方法. 正是由于认证条件的苛刻,使得被动取证成为更具挑战性的学术课题,对多媒体信息安全、刑侦、甄别虚假新闻等方面具有重要意义.

数字图像主动取证是指事先向待取证的图像中嵌入信息,在取证的过程对嵌入信息进行认证的技术[3-13]. 现有的主动取证技术包括鲁棒性数字水印防伪技术、脆弱性数字水印防篡改技术,以及数字指纹、数字签名认证技术. 这些技术所采用的基本思路都是通过嵌入或添加附加信息对数字图像进行真实性和完整性的鉴别.

(1) 鲁棒性数字水印 主要应用于数字图像作品著作权保护,如标识图像作者、作品序号、完成时间等作品信息. 目前有大量的图像处理软件可以非常方便地对数字图像进行格式转换或其他处理,如有损压缩、滤波、平滑、信号裁剪、图像增强、重采样和几何变形等,这些操作有些是恶意的,有些是无意识的. 鲁棒性数字水印要求含水印图像在经过这些处理后,只要载体图像没有被破坏到不可使用的程度,都应该能够正确提取其中嵌入的版权信息.

(2) 脆弱性数字水印 在保证正常的人类感知质量的前提下,将数字、序列号、文字、图像标志等作为数字水印嵌入到多媒体数据中,当媒体受到篡改攻击并引起怀疑时,根据脆弱水印的状态就可判断载体图像是否受到过篡改,并能确定篡改区域. 它主要用于数字图像作品的完整性取证.

与鲁棒性数字水印的要求相反,对载体图像的微小处理就应该能改变或毁掉其中嵌入的脆弱水印. 按照实现方法的不同,脆弱性水印可分为空间域方法和变换域方法两类.

(3) 数字指纹　是指一个客体所具有的、能够把自己和其他相似客体区分开的数字特征. 它主要应用于对数字图像的使用目标(如图像的版权信息或购买使用者的个人信息)进行取证,其目的是为了防止数字图像产品被非法复制或者追踪非法散布数据的授权用户. 常用的指纹方案有合谋安全指纹、叛逆者追踪指纹、非对称指纹和匿名指纹等.

不同成像设备会在图像中产生不同的内在特征. 尽管大多数篡改并不会引起人们视觉上的怀疑,但会不可避免地改变原始图像的固有特征,或引起图像某种统计特性的变化. 数字图像被动取证是指在不依赖任何预签名提取或预嵌入信息的前提下,对图像的真伪和来源进行鉴别和取证. 数字图像被动取证可分为以下 3 类[14].

(1) 图像真实性鉴别　判断数字图像在最初获取之后是否经历任何形式的修改或处理,亦称之为防伪检测. 根据图像鉴别的取证特征,这类取证技术可分为基于图像伪造过程遗留痕迹的检测方法[15-21]、基于成像设备一致性的检测方法[22-25]和基于自然图像统计特性的检测方法[26-29] 3 类.

(2) 图像来源鉴别　判断生成图像的数据获取设备,包括数字相机、扫描仪、可拍照手机,以及计算机等. 由于各种图像生成设备的特征不同,其生成的图像也会具有不同的内在特征,图像来源认证就是通过分析提取这些能够区别图像来源的特征并建立特征数据库,从而对数字图像的来源进行认证[30-34]. 许多图像真实性鉴别技术亦可应用于图像来源鉴别.

(3) 图像隐写分析取证　不仅要判断数字图像中是否嵌入了秘密信息,而且还需要提取秘密信息作为呈堂证据. 目前的隐写分析研究基本集中在检测图像中是否隐藏有无法提取的秘密信息方面[35]. 隐写分析的未来研究是进一步解决如何确定隐写所用的方法、嵌入软件、密钥等,从而实现对秘密信息的正确提取.

网络通信技术的迅速发展和多媒体数字产品的爆炸式增长,对数字图像进行真实性和完整性认证变得日益紧迫和重要,其应用涉及国家安全、司法、新闻出版、电子政务、医疗、网络通信、科学研究、电子商务和工程设计等各个领域. 采用数字水印和数字指纹等主动取证技术进行图像认证是一个方兴未艾的前沿课题,其迫切的市场需求和广泛的应用前景已吸引众多的研究者投入到这一行列. 其中脆弱性水印和半脆弱性水印技术的研究和应用尚处于起步阶段,在理论和实际成果方面还远不如鲁棒性水印技术那么成熟,还存在许多有待深入研究的问题.

4 结　语

随着图像处理软件功能的日益强大和简便,使得图像篡改技术得以迅速发展,主动取证技术由于受到诸多应用条件的限制,已无法从根本上遏制图像篡改现象,现在数字图像取证技术更倾向于被动取证研究. 实施被动取证不需要有关图像来源的先验知识,也不要求图像中含有事先嵌入的附加信息,因而在实际应用中具有重要意义. 数字图像被动取证技术可应用于网络图像的真实性过滤、电子政务文书和电子商务证书图像的鉴别、数字图像来源鉴定、法律证据图像的完整性和可信性认证、军事图像信息的鉴别等方面. 目前,数字图像被动取证是一个前沿的研究领域,国内外的研究基本上还处于探索阶段,其挑战性高、创新空间大,因而吸引了众多院校、研究机构人力、财力的投入. 预计在不久的将来,被动取证在理论研究和实际应用方面都会得到迅猛发展.

参考文献:

[1] FARID H. Creating and detecting doctored and virtual images: Implications to the child pornography prevention act [EB/OL]. [2004-09-01] http://www.ists.dartmouth.edu/library/100.pdf.

[2] FARID H. Image forgery detection [J]. IEEE Signal Processing Magazine, 2009, 26(2): 16-25.

[3] 周琳娜, 王东明. 数字图像取证技术 [M]. 北京: 北京邮电大学出版社, 2008: 35-40.

[4] LUO Q, KHOSHGOFTAAR T M. Unsupervised multiscale color image segmentation based on MDL principle [J]. IEEE Trans. on Image Processing, 2006, 15(9): 2 751-2 761.

[5] LI Y, SUN J, TANG C K, *et al*. Lazy snapping [J]. ACM Trans. on Graphics, 2004, 23(3): 303-308.

[6] CHUANG Y Y, CURLESS B, SALESIN D, *et al*. A bayesian approach to digital matting [C] // IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001: 264-271.

[7] SUN J, JIA J, TANG C K, *et al*. Poisson matting [J]. ACM Trans. on Graphics, 2004, 23(3): 315-321.

[8] 王朔中，克达尔，秦川，等. 应用热传导模型的偏微分方程图像修复 [J]. 上海大学学报，2007，13(4)：331-336.

[9] 蒋平，黄淑华，杨莉莉. 数字取证 [M]. 北京：清华大学出版社，2007：50-53.

[10] PALME G. A road map for digital forensic research [J]. Technical Report DTRT0010-01, DFRWS, 2001, 1(1): 15-20.

[11] NG T, CHANG S, SUN Q. Blind detection of photomontage using higher order statistics [C] // IEEE Int. Symposium on Circuits and Systems, Canada, 2004: 688-691.

[12] PETITCOLAS F, ANDERSON R J, KUHN M G. Information hiding—a survey [C] // Proc. of the IEEE, 1999, 87(7): 1 062-1 078.

[13] SWAMINATHAN A, MAO Y, WU M. Robust and secure image hashing [J]. IEEE Trans. on Information Forensics and Security, 2006, 1(2): 215-230.

[14] 吴琼，李国辉，涂丹，等. 面向真实性鉴别的数字图像盲取证技术综述 [J]. 自动化学报，2008，34(12)：1 458-1 466.

[15] 魏为民，王朔中，唐振军，一类数字图像篡改的被动认证 [J]. 东南大学学报：自然科学版，2007，37(1)：58-61.

[16] POPESCU A C, FARID H. Exposing digital forgeries by detecting traces of resampling [J]. IEEE Trans. on Signal Processing, 2005, 53(2): 758-767.

[17] GALLAGHER A C. Detection of linear and cubic interpolation in JPEG compressed images [C] // Proc. 2nd Canadian Conf. on Computer and Robot Vision (CRV'05), Washington, DC, USA, 2005: 65-72.

[18] MAHDIAN B, SAIC S. Blind authentication using periodic properties of interpolation [J]. IEEE Trans. on Information Forensics and Security, 2008, 3(3): 529-538.

[19] WEI Wei-min, WANG Shuo-zhong. Estimation of image rotation angle using interpolation-related spectral signatures with application to blind detection of image forgery [J]. IEEE Trans. on Information Forensics and Security, 2010, 5(3): 507-517.

[20] FARID H. Exposing digital forgeries from JPEG ghosts [J]. IEEE Trans. on Information Forensics and Security, 2009, 4(1): 154-160.

[21] JOHNSON M K, FARID H. Exposing digital forgeries in complex lighting environments [J]. IEEE Trans. on Information Forensics Security, 2007, 3(2): 450-461.

[22] POPESCU A C, FARID H. Exposing digital forgeries in color filter array interpolated images [J]. IEEE Trans. on Signal Processing, 2005, 53(10): 3 948-3 959.

[23] FARID H. Blind inverse gamma correction [J]. IEEE Trans. on Image Processing, 2001, 10(10): 1 428-1 433.

[24] FARID H, POPESCU A. Blind removal of lens distortions [J]. Journal of the Optical Society of America, 2001, 18(9): 2 072-2 078.

[25] JOHNSON M K, FARID H. Exposing digital forgeries through chromatic aberration [C] // Proc. 8th Workshop on Multimedia and Security, Geneva, Switzerland, 2006: 48-55.

[26] LYU S, FARID H. How realistic is photorealistic [J]. IEEE Trans. on Signal Processing, 2005, 53(2): 845-850.

[27] NG T, CHANG S. A model for image splicing [C] //IEEE Int. Conf. on Image Processing, 2004(2): 1 169-1 172.

[28] AVCIBAS I, MEMON N, SANKUR B. Steganalysis using image quality metrics [J]. IEEE Trans. on Image Processing, 2003, 12(2): 221-229.

[29] BAYRAM S, AVCIBAS I, SANKUR B, *et al*. Image manipulation detection [J]. Journal of Electronic Imaging, 2006, 15(4): 041102.

[30] DIRIK A E, SENCAR H T, MEMON N. Digital single lens reflex camera identification from traces of sensor dust [J]. IEEE Trans. on Information Forensics and Security, 2008, 3(3): 539-552.

[31] DEHNIE S, SENCAR H T, MEMON N. Digital image forensics for identifying computer generated and digital camera images [C] //Proc. 2006 IEEE Int. Conf. on Image Processing, Atlanta, USA, 2006: 2 313-2 316.

[32] KHARRAZI M, SENCUR H T, MEMON N. Blind source camera identification [C] //Proc. 2004 IEEE Int. Conf. on Image Processing, Singapore, 2004: 709-712.

[33] TSAI M J, WU G H. Using image features to identify camera sources [C] // Proc. 2006 IEEE Int. Conf. Acoustics, Speech and Signal Processing, Toulouse, France, 2006: 297-300.

[34] LUKÁŠ J, FRIDRICH J, GOLJAN M. Digital camera identification from sensor pattern noise [J]. IEEE Trans. on Information Forensics and Security, 2006, 1(2): 205-214.

[35] FRIDRICH J, GOLJAN M, SOUKAL D, *et al*. Forensic steganalysis: determining the stego key in spatial domain steganography [C] //Proc. SPIE, Security, Steganography, and Watermarking of Multimedia Contents VII, San Jose, USA, 2005: 631-642.

（编辑 苏 娟）