黄惠芬 孙占全 著

数字图像
司法取证技术

Shuzi Tuxiang Sifa Quzheng Jishu

山东大学出版社

数字图像司法取证技术

黄惠芬　孙占全　著

山东大学出版社

图书在版编目(CIP)数据

数字图像司法取证技术/黄惠芬,孙占全著.—济南:山东大学出版社,2015.4

ISBN 978-7-5607-5267-9

Ⅰ.①数… Ⅱ.①黄… ②孙… Ⅲ.①数字图象处理—计算机犯罪—证据—调查—研究 Ⅳ.①D918

中国版本图书馆 CIP 数据核字(2015)第 082835 号

责任策划:陈 珊
责任编辑:李云霄
封面设计:张 荔

出版发行:山东大学出版社
社 址 山东省济南市山大南路 20 号
邮 编 250100
电 话 市场部(0531)88364466

经 销:山东省新华书店经销
印 刷:济南铁路印刷厂
规 格:880 毫米×1230 毫米 1/32
7 印张 187 千字
版 次:2015 年 4 月第 1 版
印 次:2015 年 4 月第 1 次印刷
定 价:16.00 元

前　言

随着图像编辑和处理工具的迅速发展和数字图像在生活和工作中的大量应用，修改图像内容也变得相对容易，数字图像正面临着被随意篡改和伪造的威胁，负面影响不断增多。不真实数字图像的不断出现，最终会让人失去对照片的信任，迫切需要研究实用的图像资料取证技术。

图像取证属于计算机取证范畴，有别于普通计算机取证，图像取证还需对图像内容的真实性和原始性等进行审查鉴定。该技术可应用于新闻出版、法院、公安、国家安全、档案管理及防伪等各个领域，它的广泛使用将具有深远的社会和经济效益。

本书全面介绍了数字图像司法取证的意义、原理、技术基础和主要方法。全书共分 10 章。第 1 章，绪论。概要介绍了数字图像司法取证的研究背景、意义，并对国内外在这一领域的研究现状及发展动态进行了分析。第 2 章，电子证据司法鉴定基础。简述了电子证据的基本概念、特征以及电子证据与数字证据、计算机证据、科学证据之间的区别，介绍了电子证据的证明力标准，总结归纳了电子证据司法鉴定的发展状态及常用的技术基础和基本步骤。第 3 章，数字影像司法鉴定。介绍了数字影像司法鉴

定的基本程序和相关法律，对数字影像司法鉴定的常用技术进行了详细描述。第4章，数字图像处理基础。在分析数字图像成像原理及表示的基础上，给出了数字图像的常用统计模型。第5章，数字图像取证基础。对数字图像取证的两个方面——主动取证和盲取证进行了介绍。第6章，基于水印的内容主动认证技术。本章立足于数字水印系统的相关流程框架、功能要求等的分析在多媒体认证中的应用，重点分析了现行的通用经典算法，在此基础上提出了一种脆弱性水印算法。第7章，数字图像内容篡改方法及取证技术。分析了复制粘贴、重采样、模糊润饰等常用篡改手法，介绍了针对这些常用篡改的取证技术。第8章，模糊图像清晰化还原技术。在分析数字图像退化成因的基础上，对数字图像去模糊复原技术——图像去噪、图像增强、直方图修正、图像锐化、同态增晰及其他技术进行了介绍。第9章，基于大数据的专用数字影像司法取证图像库。以与数字影像司法鉴定相关的图像自然属性特征为准则，给出了专用影像司法取证图像库的建库准则，介绍了图像库的存储、图像获取及图像库分类的技术方法。第10章，总结与展望。对本书的工作进行了综述，并对今后的研究工作和方向提出了展望。

本书由黄惠芬制定编写大纲并负责全书大部分内容的编写工作。孙占全、常玉红参加了第5、7、9章部分内容的编写，王志红、贺永会参加了第3、4、6章部分内容的编写。本书最后由黄惠芬统稿。

本书的编写工作得到了国家自然科学基金“面向司法鉴定的

数字影像综合性取证方法研究”(基金号:61402271)、山东省自然科学基金“不依赖传统密码学的外包数据库安全关键技术研究”(基金号:ZR2012FQ006)和山东省高等学校科研计划项目“分布式网页图像数据信息隐藏检测研究”(项目编号:J14LN51)的支持,在此特表感谢。

在书稿的编写过程中,得到了北京电子技术应用研究所周琳娜研究员,山东省计算中心郑晓势研究员、赵彦玲副研究员的热情帮助。他们提出了很多宝贵的意见,在此深表感谢。特别感谢山东英才学院帅相志院长、迟萍萍处长的大力相助,他们对本书编写工作的支持和关注是本书的动力源泉。

由于作者水平有限,加之编写时间紧张,书中难免有不足之处,恳请广大读者提出宝贵意见,以便我们再版时修改和完善。

编 者

2015 年 3 月

目　录

第1章 绪 论

1.1 研究背景及意义

数字影像资料司法鉴定技术是指在诉讼活动中，具备相应执业资格的司法鉴定人运用物理学和计算机学的原理和技术，对录像带、磁盘、照片、手机等载体上记录的数字影像信息的真实性、完整性以及所反映的情况过程进行鉴定，对记录影像中的人体、物体作出同一认定，并提供司法鉴定意见的技术。

因监控设备、光照条件等因素影响，犯罪现场监控录像的影像质量往往不能清楚地记录犯罪的物证信息，出现视频分辨率低、散焦模糊、运动模糊、图像压缩、色彩层次不够丰富、噪声掩盖真实图像等问题，这些问题严重影响影像作为证据使用。公安等司法机关迫切需要便捷、有效的专业设备和技术，有效改善影像质量，为诉讼和审讯等提供可靠证据，更有效地打击各类违法犯罪行为。另一方面，随着图像编辑和图像处理工具的迅速发展以及数字图像在工作生活中的大量应用，修改图像内容也变得相对容易，数字图像正面临着被随意篡改和伪造的威胁，负面影响不断增多。不真实数字图像的不断出现，最终会让人失去对照片的信任。遗憾的是，上述问题一直没有得到妥善解决。首要的问题是数字图像的完整性、原始性和真实性鉴别在技术上尚不成熟，因此迫切需要研究实用的图像资料取证技术。

图像取证属于计算机取证范畴，有别于普通计算机取证，图像取

证需对图像内容的真实性和原始性等进行审查鉴定。该技术可应用于新闻出版、法院、公安、国家安全、档案管理及防伪等各个领域，它的广泛使用将具有深远的社会和经济效益。

1.2 国内外研究现状及发展动态分析

数字影像司法鉴定技术具有高度针对性，它针对司法鉴定各个环节、不同来源的数字影像证据的缺陷进行技术研究，是图像数据造假的对抗手段和监控图像数据有效的保障手段，是司法的关键环节，是物证鉴定的拓展。它涉及图像处理、模式识别、数据信息恢复等关键技术。

1.2.1 司法鉴定中的数字影像取证方法研究现状

数字影像资料司法鉴定作为司法鉴定的一个分支，是随着计算机技术的发展而发展的。计算机取证是对存在于计算机和相关外围设备中(包括网络介质)的潜在的、有法律效力的电子证据的确定与获取，其研究起步比较晚。而图像资料司法鉴定技术作为与计算机取证相关联的分支，也是近几年刚刚开始，主要研究模糊影像还原取证和影像资料真实性鉴定。

(1)模糊影像还原取证技术

数字影像去模糊复原是通过具体的图像退化模型来估计原始图像，如散焦模糊、运动模糊、大气湍流、成像角度引起的图像变形等。有一些图像的退化模型无法确定，对这一类的图像的复原称为“盲目复原”。这些技术各有不同的复原取证目的：①图像增强：增强影像中的细节或特征。②图像去噪：去除影响观感的干扰。③图像去压缩：修复影像中降质的成分。④图像去模糊：复原影像中失真的内容。⑤图像修复、去遮挡：修复损毁或遮挡部分。⑥图像分析与理解：分析以得到客观的信息、数据。

实际案情中很少有图像只包含单一或者理想的模糊原因，这就

决定了在实际工作中处理过程没有单一的处理方法，都是各种方法的结合。推荐模糊图像复原取证的过程如下：①几何校正。②色彩空间转换、校正与滤光。③减少噪声：噪声通指图像中与真实场景不一致的像素值，分为三类：随机噪声（光点变化）、散弹噪声、准周期噪声（干涉）。④调整光照问题：去除非一致的光照，如夜间或逆光的影像。⑤对比度调整：通常情况下，提高对比度是图像处理的第一步，但事实上将其放在色彩校正、光照一致化、去噪之后比较恰当，这样一方面可以保持前面方法处理的结果，另一方面又可使对比度的范围尽可能地调大。⑥对焦问题：噪声图像模糊，需要去模糊处理。⑦插值与放大。

以上各种模糊图像复原处理方法之间存在一定的互补性，其处理过程之间也会互相影响和干扰，如去噪声的过程是对图像平滑的过程。该过程中，将会对图像的边缘或轮廓产生不利的影响，而图像去模糊、复原、锐化的过程会增强图像的高频分量，也会增加噪声。所以，模糊图像复原的任何处理都是在一定的主观要求下对这一对矛盾的折中，其处理过程具有顺序性，处理方法具有多次重复性。

目前，国内该技术的主要研究单位有中科院自动化所、公安部二所的物证鉴定中心、北京电子技术应用研究所、华夏物证鉴定中心等。其主要研究成果有：图像增强、图像滤波、正交变换、形态学操作、几何变换、图像运算、图像特征、图像转换等常规图像处理技术，图像去噪、图像复原、图像融合、图像超分辨率（如人脸超分辨率、车牌恢复/增强）等模糊图像高级复原技术及软件，视频图像序列的增强、放大、多帧处理超分辨率还原等技术手段和软件等。

国际主流警用图像和视频分析系统和常规图像处理软件（也就是各影像鉴定机构常用的图像处理软件）主要有：①PhotoShop、PhotoImpact、Image-ProPlus 等通用图像处理软件；②美国 Ocean Systems（dTective，Claer ID）；③美国 SALIENT STILLS（MIT Media Lab），VideoFOCUS；④荷兰 IMIX 公司的“影博士”（IMIX Vision Support Systems）；⑤美国 Cognitech 公司研发的“识慧”模糊

图像处理系统。

模糊图像复原所面临的难题是：①噪声类型不可知；②模糊模型不可知；③随着硬盘录像系统越来越多的使用，记录的视频流文件绝大多数都经过压缩，尤其以 MP4 格式的压缩量最大，其压缩比达到 10∶1 到 100∶1 或更高，造成图像分辨率低，难以进行处理。目前还没有专门针对硬盘压缩图像进行处理的系统。

(2)数字影像篡改取证技术

数字影像篡改取证技术近几年才刚刚兴起，目前仍然处于起步阶段，现有的学术研究成果主要是针对某一种具体操作的检测，发展方向为：一方面，检测并判断出联合处理操作；另一方面，从数码相机拍摄的图像所独有的统计特性入手，来判断图像在数码相机拍摄以后有没有受到过篡改。

该技术相关的研究队伍以美国的大学和研究机构居多。其中，美国 Dartmouth 大学的 Hany Farid 教授领导的科研团队，用对数字图像进行多尺度小波分解和高阶统计建模的方法对拍摄的照片图像、扫描图像以及计算机生成的图像进行盲来源鉴别和取证。美国 Columbia 大学的 Shih-Fu Chang 领导的科研团队，在图像的来源取证和拼接图像取证上取得了一些成绩，通过用数码相机响应函数定性描述 CCD 镜头失真校正、色彩插值、白平衡、非线性伽马校正以及传感器噪声等，鉴别和确认这些过程在图像中引入的独有特征，对不同相机拍摄的照片进行取证。Shih-Fu Chang 和 Tian-Tsong 还提出了检测图像拼接伪造的数学模型，并利用双一致性即归一化后的双谱对拼接的伪造图像进行分析取证。美国 Binghamton 大学的 J. Fridrich 研究团队的主要贡献在于将分析隐蔽通信存在的隐密分析技术和数字图像盲取证技术研究结合，研究数字图像完整性研究的隐密分析取证技术，还提出基于量化 DCT 的方法来解决复制粘贴检测中的运算量过大和鲁棒性差问题。在数字图像来源取证方面，美国纽约大学的 Mehdi Kharrazi 等人提出利用平均像素值、RGB 色彩通道相关系数、邻域像素分布重心、色彩能量比、小波统计特征以

及图像质量矩阵等组成的特征向量来对相机来源进行鉴别和取证。美国 Purdue 大学的 E. J. Delp 教授领导的科研团队利用打印机光电硒鼓转动时存在轻微的不均匀性、传动齿轮的离心速率的不均匀性以及后座等装置的轻微晃动所导致的打印品缺陷，来对不同的打印机进行鉴别和取证。而 F. Cutzu 等人也开展了相关工作鉴别照片和扫描的油画。美国 Maryland 大学的吴旻从统计学的角度，对数字图像空域滤波、JPEG 重压缩、重采样、亮度调整等操作进行了分析和检测。

国内研究数字图像篡改盲取证的机构以北京电子技术应用研究所、北京邮电大学、大连理工大学信息安全研究中心、中山大学及同济大学计算机系为主，有相关文献发表。大连理工大学信息安全研究中心刘文锋、孔祥维等在 2005 年的 PCM 会议上发表文章，利用色彩变化提出检测图像篡改和检测隐藏信息的方法。王波、孙璐璐和孔祥维等人在 2006 年提出了利用异常色调率来对相机拍摄的照片图像进行取证的方法，可以对合成后经过模糊处理的篡改伪造图像进行鉴别。周琳娜、郭云彪和杨义先等利用同态滤波、移动平均滤波和数学形态学的方法检测经过模糊处理的数字图像篡改，针对模糊操作的取证正确率在 90%左右。

数字图像篡改盲取证技术的最新进展有：数字图像真实性分析可以用于判断一幅图像是照片图像还是计算机生成图像，以及对图像的复制粘贴、旋转缩放、尺寸变换、模糊锐化、亮度调整、双重压缩、镜像翻转、二次拍摄等操作的分析。

(3)专用数字影像司法取证图像库

国际互联网上有海量的图像数据，以这些数据的采样作为数字影像司法取证的测评样本并不是合适方案。现实可行的方案是根据数字取证与图像的可能相关因素，对可信赖的图像进行属性划分，作为测试图像样本的分类依据和图像库的建立准则。针对数字取证相关性因素对互联网上海量数据进行分类和建模，有效地利用国际互联网上海量图像数据对标准图像库进行扩充，是建立和完善标准图

像库的下一个方向。

国内外已有大量公开或者收费的数字图像数据库。荷兰阿姆斯特丹 ALOI 图像库用于研究不同物体在不同环境、状态以及方位条件下记录时的人体感官差别。该图像库由数以千计的小物体(如小球、小杯等)图像组成,每个物体又以不同视角、明亮程度、色彩照明光条件提供上百幅样本图像,图像库数据量高达 11 万。英国 AT&T 人脸图像库提供了 400 多幅人脸测试图像,主要用于人脸识别等方法的研究。Bologha 大学的指纹数据库保存了 3500 多个指纹,用于与指纹技术相关的研究。CCITT 的标准传真图像库,仅仅由 8 幅图像组成,堪称规模最小的图像数据库。国内也有大量的医学病例图像库、商业徽标图像库、公司产品图像库以及艺术图像库等。除此之外,还有大量商业或者个人网站提供的许多用于制作网页和网络多媒体作品的图像素材。

1.2.2 专用数字影像司法取证图像库研究现状

随着图像处理技术的不断发展,图像数据资源越来越丰富,图像的类型、内容也越来越丰富,为了有效进行司法取证的研究,提高数字影像司法取证的正确率,对图像库的要求也越来越高,需要图像库规模也越来越大。如何获取丰富的图像资源、如何对海量图像数据进行有效的存储、如何建立有效的专用图像库,都是数字影像司法取证图像库所要解决的关键问题。传统的基于 Oracle 等关系数据的存储、基于串行的数据挖掘分析等方法已经不能很好地处理这些大规模数据,需要结合大数据的技术实现。

现有数据中心很难满足大数据存储、处理、分析、挖掘的需求,存储能力的增长远远赶不上数据的增长,大数据也导致高可扩展性成为信息系统最本质的需求。Hadoop 的出现为上述问题的解决提供了框架平台。基于分布式的存储技术近年来得到迅速发展,是目前对于非结构化、半结构化大规模数据存储最有效的存储方式,形成了很多相关的技术产品和解决方案。如基于分布式文件存储的 HDFS

系统，适合一次写入多次读取的大规模非结构化数据的存储，在网络搜索引擎领域得到大规模的应用。对于结构化数据的分布式存储，HBase是目前应用最广泛的NoSQL数据库形式。对于大规模文件的管理，MongoDB提供了高效的解决方案。对于实时性要求较高的业务系统数据，Storm、S4提供了基于流数据的解决方案。不同形式的数据类型和数据需求，需要不同的大数据存储解决方案，通常需要多种技术的结合来实现实际业务需求。

人工对图像进行分类的方式已不可取，一是效率非常低，需要花费大量的人力和物力；二是很难快速建立大规模的数据库。因此，基于人工智能、机器学习的方法，利用数据挖掘、分析模型自动对图像采集、分类，从而高效准确地建立大规模专用图像数据库是最有效的方式。数据挖掘是从大量数据集中发现新模式的过程，结合了人工智能、机器学习、统计和数据库，是目前分析数据的最有效手段。国内外很多学者从事这方面的研究，很多数据挖掘方法已被应用到实际当中。随着数据规模的扩大，很多传统的数据挖掘方法已不实用，针对大规模数据密集型的并行数据挖掘方法研究是近年来信息领域的研究重点。有效的并行算法和实现技术是实现大规模数据挖掘的关键。很多并行挖掘算法以不同技术实现，如多线程、MPI技术、MapReduce技术、工作流技术等。不同的实现技术有不同的性能和使用特性，MPI模式适用于计算密集型问题，特别适用于仿真，但编程复杂度较高，对运行环境的时延要求高，容错性较差。MapReduce是信息检索领域提出的一种适于数据分析的云技术，适合于数据密集型的并行数据挖掘。传统的MapReduce架构只是单向的Map和Reduce过程，不支持迭代，不适合复杂的数据挖掘算法。最新由美国印第安那大学教授提出的Twister软件，是一种迭代MapReduce模型，支持算法的迭代，大大提高了MapReduce算法的实用性。当前国内外的厂商开始尝试使用Hadoop来存储和处理所产生的大数据。Facebook借助机群运行Hadoop，支持其数据分析和机器学习。诺基亚研究中心在Hadoop的分布式计算模型MapReduce的基础

上开发了分布式 co-clustering(DisCo)框架,可以高效处理海量数据;通过巧妙划分数据、安排计算流程来最大化数据的局部性,提高并行度,利用 MapReduce 进行非负矩阵分解。Google 的研究学者提出了对大数据的分布式优化问题的 MapReduce 结合 BigTable 的解决方法。Yahoo 的 Hadoop 应用包含搜索、日志处理、用户建模、内容优化、垃圾邮件过滤器以及广告计算等。其中所用的许多算法都是由 MapReduce 来实现的,比如大规模矩阵分解等。

参考文献

[1]张春田,苏育挺,张静.数字图像压缩编码.北京:清华大学出版社,2006.

[2]王永全,施少培,周琳娜.声像资料司法鉴定实务.北京:法律出版社,2013.

[3]周琳娜.数字影像认证技术的应用及典型案例分析.北京:计算机学会前沿学术讲座,2010.

[4]周琳娜,张茹,郭云彪.数字图像内容取证.北京:高等教育出版社,2011.

[5]崔夏荣,苏光大.基于模式噪声的数字图像来源鉴别.光电子·激光,2007,18(11):1239-1243.

[6]Farid H, Lyu S. Higher-order wavelet statistics and their application to digital forensics. Madison: IEEE Workshop on Statistical Analysis in Computer Vision, 2003: 1-8.

[7]Farid H. Creating and detecting doctored and virtual images: implications to the child pornography prevention. New Hampshire: Dartmouth College, 2004.

[8]Lyu S, Farid H. How realistic is photorealistic? IEEE Transactions on Signal Processing, 2005, 53(2): 845-850.

[9]Ng T-T, Chang S-F, Tsui M-P. Camera response function

estimation from a single-channel image using differential invariants. ADVENT Technical Report #216-2006-2. Columbia University, 2006.

[10]Ng T-T, Chang S-F. A model for image splicing. Singapore: IEEE International Conference on Image Processing, 2004: 1169-1172.

[11]Ng T-T, Chang S-F, Sun Qibin. Blind detection of photomontage using higher order statistics. Canada: IEEE International Symposium on Circuits and Systems, 2004: 688-691.

[12]Kharrazi M, Sencar H T, Memon N. Blind source camera identification. Singapore: IEEE International Conference on Image Processing, 2004: 709-712.

[13]Martone A F, Mikkilineni A K, Delp E J. Forensics of things, image analysis and interpretation. IEEE Southwest Symposium, 2006: 149-152.

[14]Mikkilineni A K, Chiang P J, Ali G N, et al. Printer identification based on texture features. International Conference on Digital Printing Technologies, 2004: 306-312.

[15]Cutzu F, Hammoud R, Leykin A. Estimating the photorealism of images: Distinguishing paintings from photographs. IEEE Conference on Computer Vision and Pattern, Recognition, 2003, 2: 18-20.

[16]Swaminathan A, Wu Min, Ray Liu K J. Nonintrusive component forensics of visual sensors using output images. IEEE Transactions on Information Forensics and Security, 2007.

[17]Swaminathan A, Wu Min, Ray Liu K J. Image tampering identification using blind deconvolution. Proceedings of the IEEE International Conference on Image Processing. Atlanta, GA, 2006: 2311-2314.