伪造数字图像盲检测技术研究

李 杭◎著

吉林大学出版社

伪造数字图像盲检测技术研究

李 杭 著

吉林大学出版社

图书在版编目(CIP)数据

伪造数字图像盲检测技术研究 / 李杭著. -- 长春 : 吉林大学出版社, 2016.1

ISBN 978-7-5677-5724-0

Ⅰ. ①伪… Ⅱ. ①李… Ⅲ. ①数字图象处理-研究 Ⅳ. ①TN919.8

中国版本图书馆 CIP 数据核字(2016)第 026284 号

书　　名：伪造数字图像盲检测技术研究
作　　者：李　杭　著

责任编辑：张宏亮　　责任校对：王瑞金　　封面设计：中图时代
出版发行：吉林大学出版社　　印刷：河南新华印刷集团有限公司
开　　本：880mm×1230mm　1/32　　版次：2016 年 1 月第 1 版
印　　张：4.5　　字数：120 千字　　印次：2016 年 1 月第 1 次印刷
书　　号：ISBN 978-7-5677-5724-0　　定价：29.80 元

社　　址：长春市明德路 501 号　　邮编：130021
发 行 部：0431-89580028/29
网　　址：http://www.jlup.com.cn
E - mail：jlup@mail.jlu.edu.cn

目 录

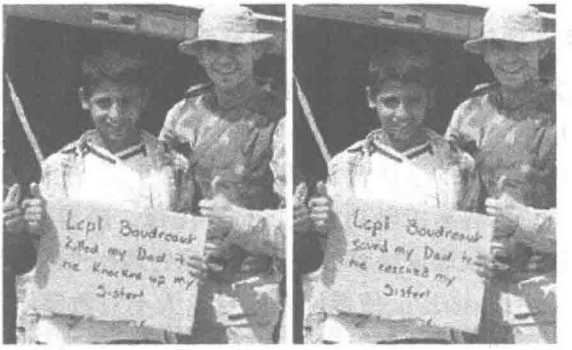
第一章 绪 论

一、数字图像被动盲取证技术的研究背景及意义

当今社会的迅速发展，使得数字媒体在我们的日常生活、学习和工作中扮演的角色越来越重要。近些年来由于数字时代的来临以及低成本高分辨率数码相机的迅速普及，我们在日常生活、工作和学习中比以往任何时候都更加容易接触到大量的数字图像。随着互联网的大众化以及那些操作简单获取途径容易的数字图像处理软件的大量使用，如最为常见的 Adobe Photoshop，使得许多非专业人士对数字图像的修改、编辑和存储已不再是什么难题。而且这些篡改图像的内容的伪造痕迹单靠人眼是难以辨别的。虽然这些操作一方面可以增强数字图像的视觉效果，使得我们能够更好地享受数字化带给我们的视觉享受。但是，同时也滋生了许多恶意的念头，不乏非专业人士甚至是专业人士会出于不同的目的来恶意传播一些篡改伪造图像，并以此来达到不可告人的动机和目的。这些篡改伪造图像一旦被正式媒体、科学研究甚至法庭等应用，不仅会影响我们的正常生活秩序，而且还会影响我们人身安全甚至严重扰乱社会稳定性。近一些年来，无论是在政界还是新闻界等领域，国内外已经出现了许多令人备感震惊的图像篡改伪造案例。

2003 年夏天，美国海军陆战队一等兵特德·博顿瑞尔克斯曾在一张硬纸板上随手写下了“欢迎海军陆战队”的字样，并把它交到一名伊拉克小孩的手中，然后两个人一起微笑着合影留念。然而一年多后，当这张照片最终通过互联网被再次传到他的电脑中时，他被惊得目瞪口呆。此时纸板上原先的文字“欢迎海军陆战队”已经不翼而飞，取而代之的竟是：“一等兵博顿瑞尔克斯杀死了我爸爸，然后又强奸了我姐姐”！一时间，很多人对博顿瑞尔克斯的行为展开了激烈的批判。世人的强烈呼声使得五角大楼下令对此展开彻底调查，美国白宫并为此事件成立了专门调查小组来调查该事件的真实性，然而由于找不出证明该照

片被篡改过的证据,因此就连美国海军犯罪调查办公室的专家也对此束手无策。此后,在另一个倾向于布什政府的网站上,竟出现了第3个版本——卡片上的字又变成"一等兵博顿瑞尔克斯救了我爸爸,然后又救了我姐姐"!



(a)篡改图像1　　(b)篡改图像2

图1-1　"硬纸板"事件篡改图片

2004年美国总统大选竞争最为激烈的时期,因特网上疯狂流传出年轻时的约翰·特里与反越战歌星简·方达同台的图片。虽然后来经多方验证,这是一张人为的合成恶意篡改的图片。但是此次照片事件对约翰·特里的政治生涯造成的影响却是远远无法估量的。图1-2是2004年美国总统大选期间广为流传的篡改伪造图片。

在科学研究中也发生了不少严重违背科学研究真实性原则的数字图像篡改伪造案例。2005年,韩国首尔大学教授黄禹锡论文造假事件,在整个学术界引起了空前的轰动。后经韩国首尔大学调查委员会及其本人证实,黄禹锡的论文中所涉及的十一个病人的干细胞株上所衍生出来的干细胞,只有两个是真的来源于病人的干细胞本身,而其余九个干细胞的图片都是基于这两个干细胞的基础上篡改伪造得到的。而且其在2004年2月发表在《科学》杂志上关于首例人类克隆胚胎细胞的文章也是伪造而来的。美国《科学》杂志不仅撤销了黄禹锡2004年和2005年所发表的两篇论文,黄也因此次事件辞去了首尔大学教授的职务。但是,其造成的恶劣的影响却远远没有结束,而且这也并不是唯一的学术

造假的案例。据美国《细胞生物学杂志》主编迈克·罗斯纳证实在他们已经录用的手稿中至少有百分之二十的文章图片是经过重新编辑修改过的。

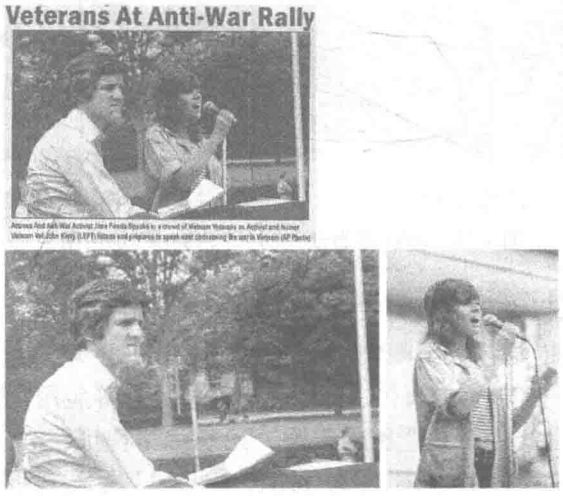
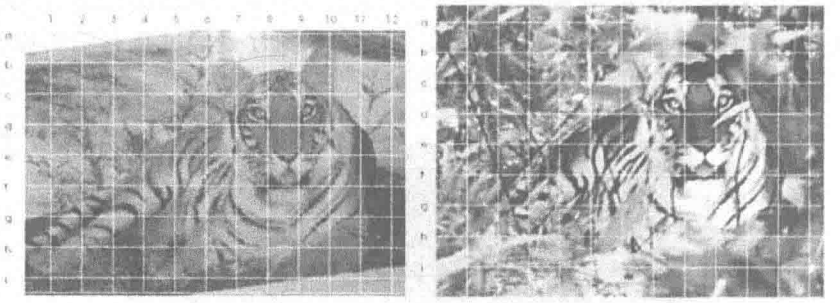


图 1-2 2004 年美国总统大选期间网络流传的伪造图片

2007 年 10 月 12 日,我国陕西省林业厅公布了一张野生华南虎照片,当时引起了各方人士的注意及质疑。后经多方专家的证实,这是一张利用数码相机与胶片照相机拍摄之后又经过 PS(Adobe Photoshop)处理过的图片。该伪造图像的纹理部分来自于一张中国传统的老虎年画。直到 2008 年 6 月 29 日,当事人周正龙以及涉案的相关人员被相关部门处理后,此事才得以平息。至此,闹得沸沸扬扬的"周老虎"案终于尘埃落定,但是,其造成的负面影响却远远没有停止。甚至连美国的《科学》杂志在此期间也刊登了相关文章来报道此事。图 1-3 是 2007 年轰动一时的"华南虎"伪造图片。



(a)传统年画　　(b)“华南虎”伪造图片

图 1-3　2007 年轰动一时的“华南虎”伪造图片

所以,从以上几个案例来看,那些篡改和伪造的图像,不仅能够在一定程度上歪曲、混淆甚至颠倒事物的客观本质和事实真相,而且这些篡改伪造图像一旦被官方媒体、科学研究、法律证物等应用,将对政治和社会的和谐安定产生极其严重的影响。比如,“华南虎”事件不仅会降低行政部门的社会公信力,还可能会助长一些不正之风,帮助某些人达到一些不可告人的秘密。而美伊战争期间广为流传的伊拉克士兵的篡改图片会进一步激化西方世界与伊斯兰世界矛盾,从而为世界和平蒙上了更深的阴影。就在美国 9·11 恐怖袭击爆炸事件发生之后不久,美国著名的《USA Today》就报道在 1998 年的东非美国大使馆炸弹袭击事件中。本·拉登等恐怖分子就曾经将聊天网站图片作为恐怖袭击目标的地点和具体地图传输信息的工具。由此可见,那种传统的“眼见为实”观念已经脱离了时代的轨道。

目前网络上广泛传播的数字产品图片主要包括法律、医学、新闻、商业、军事、重要文件等等,这些信息如果被恶意地非法篡改、伪造或者攻击,那么将会给传播者和接收者带来巨大甚至无法估量的损失。毫无疑问,数字图像取证技术的应用覆盖了社会的各个领域。

由于这些篡改过的图像一般在视觉上是根本无法觉察到的,如果这些虚假伪造的图片被恶意地传播或者利用,那么这将会严重扰乱社会的稳定性,对我们的社会和个人产生的负面影响将是无法估量的。所以如何将数字图像作为有效的司法证据显得越来越重要。尽管现有的技术

可以对一些篡改过的图片进行相关的鉴定,但是仍然还有许多关键问题没有得到相应的解决,对数字图像的完整性和真实性的鉴别在技术上尚处于初期阶段。这就彰显出了数字图像取证技术的必要性与迫切性。由此可见,数字图像的取证技术无论对科技还是人类的社会生活的各个领域,都有不可忽视的作用,具有不可估量的研究意义。所以研究数字图像取证技术的发展对整个社会个人乃至全世界的安全以及世界和平稳定有着重大意义。

二、数字图像取证技术概述

数字图像取证技术(Digital Image Forensics)是指对数字图像的篡改、伪造和隐秘进行分析、鉴别和认证,是通过对图像的统计特性进行分析研究来判断数字图像内容的真实性、完整性和原始性。目前已有的图像真实性取证技术有主动取证技术和被动取证技术两大类。

三、数字图像主动取证技术

截止到目前,现有的发展以及应用比较成熟的数字图像主动取证技术包括:鲁棒性数字水印(Robust Watermarking)防伪取证技术、脆弱数字水印(Fragile Watermarking)防篡改取证技术、数字签名(Signature Based Image Authentication,SBIA)以及数字指纹(Digital Fingerprint)为代表的主动取证技术等等。现有的大部分主动取证技术所采用的基本思想都是先通过对图像进行预处理,在目标图像中嵌入或者添加具有标志性信息,并在此基础上对数字图像内容以及来源进行真实性和完整性鉴别。比如数字签名、数字水印(Digital Watermarking)都属于数字图像主动取证技术的范畴。

(一)数字水印主动取证技术

数字水印主动取证技术是迄今为止发展以及应用都较为成熟的一类数字图像主动取证技术。数字图像水印主动取证技术主要应用于版权保护、数据认证和信息传输等等。数字图像水印主动认证技术的现有的经典算法有 Schyndel 主动取证算法、扩展频谱主动取证算法、Patchwork 主动取证算法、NEC 算法以及基于 JPEG(Joint Photographic Experts Group)和 MPEG(Moving Pictures Experts Group)压缩标准的数字水印算法。目前大部分的数字图像水印取证算法都是先将图像中重要

的信息，也就是那些可以代表所有版权的标志性信息、图形图像的内容、音频数据、随机序列等等一系列有效信息隐藏于图像中。

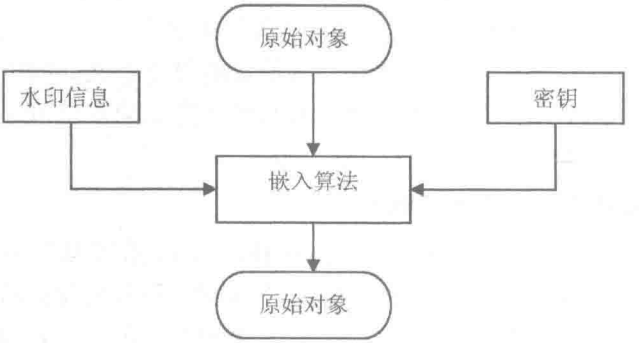


图 1-4　数字水印嵌入模型

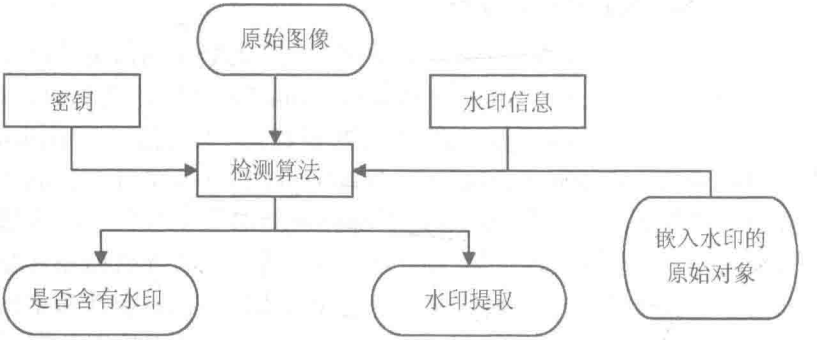


图 1-5　数字水印提取模型

由于数字水印技术针对的认证目的以及目标的差异，所以其对篡改图像的敏感度要求也不一致。根据对图像修改的不同程度可以把现有的数字水印取证分为以下几大类：

(1)脆弱数字水印：这种主动取证算法能够检测出对图像任意的篡改。由于这种取证技术注重数字图像的数据整体特性，所以这类取证技术不允许对图像进行任何修饰篡改等编辑操作。如果对图像进行了任意的修饰编辑等操作，那么从图像中重新恢复出来的嵌入的数字水印也随之发生相应的变化。人们根据恢复出来的脆弱性水印的破坏状态和

程度来判断数据是否被篡改过以及篡改的程度和位置。所以可以将其作为攻击过程和属性判断的依据。

(2)半脆弱性数字弱水印:这种主动取证技术是以保护图像内容信息的传递为主要目的,所以对常见的一些非恶意的数字图像有损压缩、格式转换、噪声去除等操作也会产生很高的虚警概率。虽然这类数字水印主动取证技术具有一定的稳健性,但是其对脆弱水印对像素变化的敏感性要差一些。

(3)图像内容的认证技术:由于在某些应用场合中许多用户仅仅对数字图像的部分内容或者其视觉效果感兴趣。因此如果对图像的编辑操作没有影响到图像的主要内容或视觉效果的前提下,都将被认为是可接受的。即使不能允许针对数字图像视觉效果进行任何改动,这一类系统的鲁棒性仍然会比前面两类要好一些。

(4)鲁棒性数字水印技术:这种取证技术难以被去除,而且其抗干扰能力十分强。这种主动取证技术要求能够经受各种常见的数字图像编辑软件处理和各种典型的水印攻击。这种取证技术主要应用在数字作品著作权信息的标识等领域中。

虽然目前数字水印取证技术的准确率比较高,但是由于数字水印相机的消费市场前景并不光明。而且大部分的自然图像由于它们本身没有嵌入或者预先嵌入水印等其他的信息而根本无法用上述方法进行相关取证。所以数字水印取证技术的应用局限性很大。

(二)数字签名主动取证技术

数字签名技术是数字图像主动取证发展较为成熟的技术之一,其又可以称为电子签名技术。这种主动取证技术基本思路是利用数据电文中的电子形式包含或者附加信息来用于对鉴别目标进行鉴别。截止到目前,所有的数字签名技术都是基于公共密钥体制基础值上的建立起来的。传统的密码学签名方法必须对数据传输过程中的各个环节都要进行认证。一方面这种认证方式增加了数据处理的计算量,另一方面也使认证的过程变得更加复杂。除此之外,传统惯用的哈希函数方法已经对篡改操作定位失效。如果认证失败,传统的密码学数据认证技术无法定位出篡改的位置。所有的检测目标数据都将会被当作没有实际有效的

垃圾信息处理。所以,在数字图像签名技术中已经不能再直接应用。因此国内外许多研究机构和学者针对数字图像特征又重新对数字图像签名取证技术进行了相关研究。现有数字图像签名方法主要区别在于数字图像特征的提取。这些特征包括灰度直方图特性、边缘特征、特征点、分块灰度矩特征、DCT 系数和 DWT 系数等等。到目前为止,不同研究机构以及学者从不同的角度入手针对数字签名技术取得了大量研究成果。

(三)数字图像主动取证技术的局限性

虽然数字图像主动取证技术可以用来识别文件、音乐或者是图片等的最初来源,也可以应用到版权保护等等。但是在使用数字水印进行主动取证时,需要验证方提取出预先在原始图像中嵌入的水印或者添加的信息,然后再经第三方认证之后才能对图像的真实性和完整性进行鉴别。而且就目前情况来看目前流传于各个领域的数码照片中不含预先嵌入的数字水印或者数字摘要等信息。同样的数字签名技术也需要预先嵌入验证信息以及第三方的认证。而且在实际中不可能实现对每张图片都嵌入信息,并且嵌入信息之后图像的质量会不同程度地下降。由此虽然数字图像主动取证技术检测的准确率很高,但是在实际应用中局限性很大,所以这种传统的数字图像取证技术大大地限制了图像取证技术应用的范围。

四、数字图像被动盲取证技术

数字图像被动盲取证技术(Blind Digital Image Forensics),是近些年来发展起来的一个全新的前沿领域。国内外对数字图像被动盲取证技术的研究都尚处于初级阶段。但是在国内外的众多研究机构和学者的极大关注下,数字图像被动取证技术的发展十分迅速。迄今为止,许多重要的国际会议和组织已经设置了相应的杂志或者分会对数字图像被动盲取证技术进行深入研究,一些可行有效的数字图像被动取证算法也相继被提出。相对于数字图像主动取证技术而言数字图像被动取证技术的局限性更小,实用性更强。数字图像被动盲取证技术的基本思想是在不依赖任何先验信息的条件下来对图像内容的真伪性和图像来源进行鉴别。所以数字图像被动取证技术具有更高的应用价值。但是由于

目前尚处于研究初级阶段，所以数字图像被动取证技术面临着很多挑战性的问题，但是其创新空间大，应用前景也非常地广泛，因此本课题的研究不仅具有较高的理论价值而且其实际的应用价值也非常大。

现有的数字图像被动盲取证技术大体可以分为三大类：基于图像伪造过程的遗留痕迹进行被动取证，例如：图像复制-粘贴篡改取证、JPEG双压取证、图像重采样取证、图像照明不一致取证、图像模糊估计取证；基于成像设备的一致性进行被动取证，例如：CFA插值取证、数码相机模式噪声取证、色差取证、相机相应一致性取证；基于自然图像的统计特性进行被动取证，例如：自然图像统计模型、双相干系数和边缘百分比特征、图像质量度量、二元相似性度量等特征。因为分析图像时不具备任何先验信息，所以难以选择用来取证图像伪造的特征。所以数字图像被动取证技术面临着巨大的挑战。尽管如此，国内外的许多研究团队和学者从不同的角度对数字图像被动盲取证技术进行了相关研究，并取得了一定的成果。

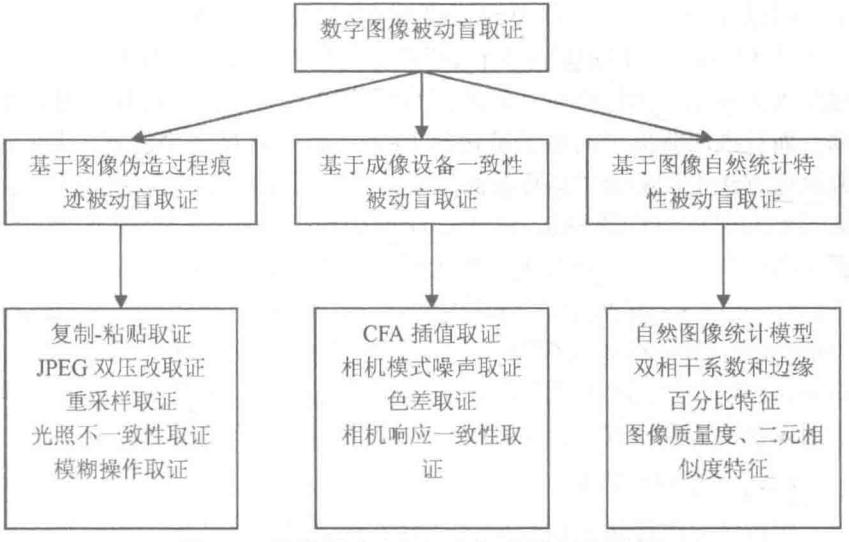


图 1-6 图像被动盲取证技术研究框架示意图

五、数字图像被动盲取证技术的国内外研究状况

(一)数字图像被动盲取证技术国外研究现状

近些年来,国内外的许多研究机构和学者对数字图像被动盲取证技术作了相关的研究并取得了一定的成果。国外对于数字图像被动盲取证技术的研究起步相对较早,其中主要以美国为主,德国、土耳其等其他国家也包括在内。美国 Dartmouth 大学以图像取证专家 Hany Farid 为核心的数字图像被动取证研究团体,针对各种数字图像的伪造手段的特点采用统计的方法提出了一系列的数字图像被动盲取证算法。并且得到了美国自然科学基金和美国国家安全部门的全力支持。该团队研究的数字图像被动盲取证算法包括利用统计特性实现对区域复制(region duplication)的取证、数字图像重采样(resampling)篡改操作取证、光源方向不一致性取证、彩色滤波插值图像(color filter interpolation)取证等等。这些取证算法不仅针对不同形式的图像篡改伪造取证取得了显著的成果,而且从不同的角度实现了对篡改数字图像的被动盲取证,为以后的数字图像被动盲取证工作奠定了良好的基础。美国 SUNY Binghamton 大学的 J.Fridrich 教授领导的研究团队从早期的水印主动取证研究逐渐地深入到利用数码相机传感器的固有模式噪声对图像的真伪性进行鉴别。而且该团队提出的基于量化的 DCT 系数的数字图像复制-粘贴篡改取证方法不仅降低了运算量而且其鲁棒性问题也得到了很好的解决。美国 Columbia 大学的 Shih-Fu Chang 团队致力于建立一个完整的在线数字图像认证系统。并且成功地实现了在线图像的来源认证。除此之外,美国 New Jersey 大学的 Yun Q.Shi 教授、美国 Purdue 大学的 Edward J.Delp 教授、Maryland 大学的 Ray Liu 以及美国 Polytechnic 大学的 Mehdi Kharrazi 教授领导的研究团队分别从该领域不同的角度不同方向对数字图像被动盲取证技术做了相关研究。

(二)数字图像被动盲取证技术国内研究现状

国内数字图像被动盲取证技术的研究相对于国外起步比较晚。计算机取证技术的概念直到 2001 年才被引入到国内。其中在中国广州中山大学召开的第六届国际数字水印学术会议(IWDW2007)是以数字图像主动取证为专题的。直到 2006 年举办的第六届全国信息隐藏暨多媒

体信息安全学术研讨会(CINW2006)才首次将数字取证(Digital Forensics)列入征文的主要内容。2007年召开的数字图像盲取证的专题报告,在国内掀起了研究数字取证的热潮。虽然对数字图像被动取证技术在国内研究起步相对比较晚,但是近几年来发展的速度却十分的迅速,并且也已经取得了丰硕的成果。其中大连理工大学的信息安全研究中心、国防科技大学、中山大学多媒体安全研究所、北京邮电大学信息安全部、上海师范大学以及同济大学计算机系等研究机构都对数字图像被动盲取证技术做了相关研究并取得了一定的成果。

六、数字图像被动盲取证技术面临的主要问题

尽管到目前为止,数字图像被动盲取证技术取得了不少的成果,也解决了不少的数字图像被动盲取证技术面临的难题,但是仍然还有许多亟待解决的难题和巨大的挑战。

第一,现有的基于自然图像统计特性的被动盲取证算法大部分对训练样本有较强的依赖性。例如J.Fridrich教授团队提出的利用数码相机模式噪声模式来对数字图像的来源进行取证的算法。由于现有许多数字图像被动盲取证算法的图像特征的提取基本完全依赖于图像的内容本身。所以截止到目前,大部分的基于数字图像统计性的被动盲取证算法实效性往往离不开预先训练好的样本。如果不考虑原始自然图像和篡改图像对比训练样本数据的支持,这些基于数字图像统计性的被动盲取证算法将失效。到目前为止,对训练样本的依赖是数字图像被动取证技术的研究仍然无法攻克的难题。

第二,现有的基于非统计特性的图像被动盲取证算法的适应性都相对较弱。大多数情况下,这些算法只能针对图像特定的某种篡改操作手段才能进行有效取证。而且这些算法大都以图像经历的篡改方式、种类预先做了特定的要求为前提。一般情况下,首先假设图像经历的篡改操作为单一的篡改手段,从而不需要考虑篡改过后的后续处理工作从而使得检测算法具有一定的局限性。而且现有的数字图像被动检测算法基本都会存在对某些多种篡改方式结合的取证失效的情况。考虑到目前的数字图像篡改技术往往具有多样性结合性,所以针对性太强的图像被动盲取证算法对于那些多种伪造技术结合的伪造篡改图像的取证效果

往往达不到理想的预期效果。

第三，虽然数字图像被动取证技术已经取得了一定的成果，各种检测算法也相继被提出。但是截止到目前，针对全面的数字图像篡改伪造的数字图像盲取证系统仍然没有建立起来。这也将大大限制了数字图像取证技术的实际应用。所以，如何建立健全的数字图像被动取证技术检测系统是数字图像被动取证技术发展的最终目标。

第二章　图像被动取证技术综述

一、引言

数字图像被动取证，也叫图像盲取证，是一种在鉴别数字图像时不需要借助签名或水印等信息的认证技术。从取证目的来看，数字图像盲取证技术要解决的问题主要有两个：图像篡改检测和图像来源辨识。下图所示为数字图像盲取证的一般框架。

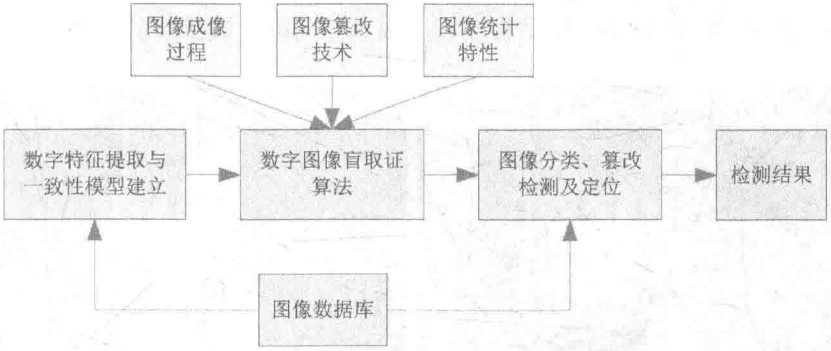


图 2-1　数字图像盲取证的一般框架

二、图像篡改检测

尽管大多数篡改图像不会留下视觉痕迹，但却不可避免地会破坏自然图像数据之间内在特征的统一性，引起图像某些特征的变化，图像篡改检测就是通过检测图像特征的变化或不一致性，来判断图像的真实性和完整性。图像篡改的手段很多，篡改检测的方法也有很多种，每种检测方法的切入点和使用范围也各有不同，要检测图像篡改，首先要对图像篡改技术有所了解。下面，我们将首先简单介绍图像篡改技术，然后再对图像篡改检测算法进行概述。