面向真实性鉴别的数字图像盲取证技术综述

吴 琼 [1] 李国辉 [1] 涂 丹 [1] 孙韶杰 [1]

摘 要 数字图像盲取证技术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, 正逐步成为多媒体安全领域新的研究热点, 且有着广泛的应用前景. 首先简要描述了图像盲取证技术要解决的问题和任务. 根据图像鉴别使用的取证特征, 将用于真实性鉴别的图像盲取证技术划分为三类: 基于图像伪造过程遗留痕迹的盲取证技术、基于成像设备一致性的盲取证技术和基于自然图像统计特性的盲取证技术, 然后分别阐述了这三类取证技术的基本特征和典型方法, 对不同算法进行了性能比较和总结. 最后综合近年来国内外学者在面向真实性鉴别的图像盲取证技术方面的主要研究成果, 探讨了图像盲取证技术存在的问题及未来研究方向.

关键词 真实性鉴别, 图像盲取证, 篡改检测, 多媒体安全
中图分类号 TP309

A Survey of Blind Digital Image Forensics Technology for Authenticity Detection

WU Qiong[1] LI Guo-Hui[1] TU Dan[1] SUN Shao-Jie[1]

Abstract Blind digital image forensics, a technology for detecting image authenticity and source without relying on any pre-extraction or pre-embedded information, is emerging as a new hotspot with broad prospect in the multimedia security field. First, the main problems and tasks of image forensics are briefly reviewed. Then, according to the forensics characteristics used by image authentication, the techniques of blind image forensics for authenticity detection are divided into three categories: techniques based on the traces left by the process of image forgery, techniques based on the consistency of imaging equipment, and techniques based on the statistical characteristics of natural images. The basic characteristics, typical methods, as well as performance comparison and analysis of various algorithms are summarized in detail for each category. So the latest related works are analyzed, and discussion on unresolved problems and future directions is presented.

Key words Authenticity detection, blind image forensics, tamper detection, multimedia security

数字图像广泛应用于日常生活和工作当中, 与此同时图像编辑和处理工具发展迅速, 一般用户和专业用户都很容易利用这些图像编辑工具, 修改图像内容而使得人眼难以辨别修改的痕迹. 我们能否相信发布的新闻照片? 能否把证人提交的照片作为证据? 能否把传送来的军事信息作为可靠情报? 这些问题促进了数字图像鉴别技术研究的发展. 归纳起来, 有三种技术手段可供其使用: 数字签名[1−3]、数字水印[4−7] 和新兴的数字图像盲取证技术. 前两种方法的一个共同特点是要求内容提供方必须对图像进行预处理, 如提取摘要或插入水印. 然而, 许多实际情况要求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, 这就是图像盲取证技术 (Blind digital image forensics)[8], 它是一种新的图像鉴别分析思路.

目前图像盲取证技术是一项前沿的研究领域, 国内外的研究均处于探索阶段, 其挑战性高, 创新空间大, 因此吸引了众多院校、研究机构及公司投入到该领域中来. 其中, 著名的 Dartmouth 学院[1]、Binghamton 大学[2]、Columbia 大学[3]和美国 Polytechnic 大学[4] 等成立了专门的数字媒体取证研究小组. 据称 Adobe 公司已与 Dartmouth 学院的图片真伪鉴别专家 Farid 合作, 将添加 Photoshop 外挂防伪工具来分辨照片的真伪. 国内在该领域的研究相比国外起步较晚, 中山大学、同济大学已经开始从事有关方面的研究. 图像盲取证技术可用于网络图像的真实性过滤、电子政务和电子商务系统中的文书和证书图像的鉴别、法律证据图像的取证、军事图像信息的鉴别等方面, 其应用前景非常广阔.

本文首先描述了数字图像盲取证技术要解决的问题、任务和基本框架, 并对现有方法进行了分类, 然后详细阐述了三类图像盲取证技术的基本特征和典型方法, 并进行了比较和总结. 最后讨论了面向真实性鉴别的图像盲取证技术目前存在的问题, 展望了其未来研究方向.

收稿日期 2007-09-10 收修改稿日期 2008-03-10
Received September 10, 2007; in revised form March 10, 2008
1. 国防科技大学信息系统与管理学院 长沙 410073
1. College of Information System and Management, National University of Defense Technology, Changsha 410073
DOI: 10.3724/SP.J.1004.2008.01458

[1] http://www.cs.dartmouth.edu/farid/research/tampering.html
[2] http://www.ws.binghamton.edu/fridrich/publications.html
[3] http://www.ee.columbia.edu/ln/dvmm
[4] http://isis.poly.edu/index.php?page=1&project=1089

1　数字图像盲取证的概念框架

Khanna 等指出, 数字图像盲取证技术要解决的问题包括以下几个方面[9−10]:

1) 能否确认一幅图像是由成像设备获取的真实照片, 还是电脑制作的图片, 或是由不同图像经过处理操作后的伪造图像?

2) 能否确认拍摄照片的成像设备类型及制造商? 能否分辨拍摄照片的成像设备 D、拍摄时间 T 以及拍摄地点 L?

3) 能否确认伪造图像的篡改区域和篡改程度?

4) 能否确认图像是否被修改以用于嵌入秘密信息?

上述问题仅仅是执法和分析机构调查可疑事件时经常要面对的一部分问题. 就确定图像真实性和来源的盲取证技术而言, 目前尚缺乏系统的方法论和理论体系, 关于图像盲取证的概括性描述, 可参考文献 [10−12]. 由图像盲取证技术要解决的问题可以看出, 其主要需完成两项任务:

1) 真实性鉴别 (也称防伪检测): 判断图像在最初获取之后是否遭受了某种形式的修改或处理.

2) 像源鉴别: 判断生成图像的数据获取设备. 这一类技术将图像与一系列具有共同特性的像源联系起来, 以便于将图像匹配给某一类型的来源设备, 关于像源鉴别技术可见文献 [13−17].

数字图像盲取证的基本框架如图 1 所示, 其同时适用于真实性鉴别和像源鉴别. 数字图像盲取证的关键技术包括以下五个方面: 图像特征提取与一致性模型建立、图像盲取证算法、图像分类与篡改区域定位、图像数据库以及相关的图像知识 (包括图像伪造知识、成像过程、统计特性等).

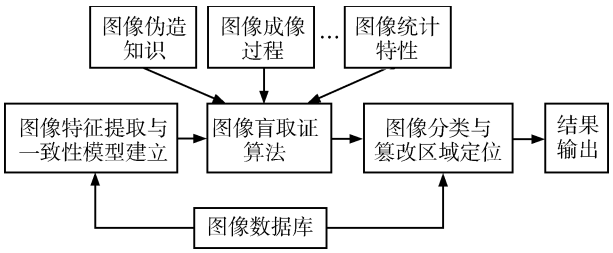

图 1　数字图像盲取证的基本框架

Fig. 1　Basic framework of blind digital image forensics

随着低成本和高分辨率数码相机的问世, 以及先进图像编辑和处理软件的出现, 数字图像可以很容易地被操纵和篡改, 因此真实性鉴别是图像盲取证中的一个重要任务, 也是本文综述的主题.

图像篡改过程通常涉及一系列加工步骤, 试图产生一致的视觉感受. 因此, 一幅伪造的图像或图像的一部分通常都会经历一些常用的图像处理操作, 例如重采样、颜色和亮度变化的补偿、细节损失 (如过滤、压缩、噪声) 等, 篡改过程可能留下一些人工痕迹, 或产生意想不到的相似性、特征偏差等异常. 另外, 自然图像通常是通过数据采集设备获取的, 如数码相机、便携式摄像机或扫描仪等, 这些设备产生的图像具有特殊一致性, 而不同设备具有不同的特性, 如不同镜头的光学畸变不同、不同相机使用的色彩滤镜矩阵 (Color filter array, CFA) 和相应的插值算法不同[18] 等, 这些一致性的规律可用于图像伪造检测和篡改定位. 最后, 图像可能在不同方面、不同程度上被篡改, 虽然许多篡改操作后的图像在视觉上无失真, 但是它们会影响图像的内在统计特性, 破坏自然图像的一致性.

因此根据图像盲取证所使用的取证特征不同, 本文将面向真实性鉴别的图像盲取证方法分为以下三类 (如图 2 所示):

1) 基于图像伪造过程的遗留痕迹进行盲取证;

2) 基于成像设备的一致性进行盲取证;

3) 利用自然图像的统计特性进行盲取证.

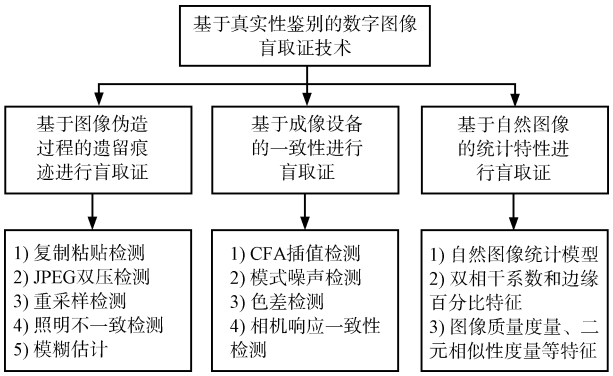

图 2　图像盲取证方法分类

Fig. 2　Classification of blind digital image approaches

下面将按图 2 所示的分类方法对面向真实性鉴别的图像盲取证技术展开详细论述.

2　基于图像伪造过程遗留痕迹的盲取证技术

该类技术试图选择能够描述伪造痕迹的取证特征来检测图像是否被篡改. 这些特征的形式, 可能是图像的特征偏离、图像模糊度或者是图像区域之间意想不到的相似性等. 目前的研究方法包括复制 – 粘贴检测、JPEG 双重压缩检测、重采样检测、照明条件不一致性检测、模糊估计等.

2.1　复制 – 粘贴检测方法

为了掩盖图像中某些重要目标或编造出不存在的场景, 常用的一种伪造手段是复制图像中的背景

区域来覆盖图像中的对象区域. 复制 – 粘贴类型的图像伪造, 虽然可能对篡改图像进行了一定的边缘处理, 但是图像中的复制区域和相对应的粘贴区域基本上相似. 基于这个特点, 可以通过寻找图像中存在的相似区域来检测图像的伪造痕迹. 穷举搜索法是一个显然的解决方法, 其优点是算法简单, 易于实现, 缺点是运算量大. 为了减少运算量, Fridrich 把对图像像素的点操作转化为块操作, 提出了一种对图像块的 DCT 量化系数进行字典排序的算法[19], 来检测图像复制伪造区域. Popescu 提出了一种类似的检测图像重复区域的算法[20]. 该算法对图像块使用主成分分析 (Principal component analysis, PCA), 将得到的降维特征作为该图像块的特征描述, 以减少特征空间的维数. 但是他们的算法中都是直接对原始图像分块, 然后进行降维特征提取, 而一次移动一个像素的滑窗操作使得图像块的数目非常大, 尤其是图像较大时. 为了进一步减少运算量, Li 从缩小图像尺寸和降低特征空间维数两方面考虑, 提出了一种基于小波和奇异值分解的图像复制区域检测算法[21]. 该算法利用小波变换提取的图像低频分量作为分析对象, 并使用奇异值分解提取小波低频图像特征进行降维, 然后对图像奇异值特征矩阵按行进行字典排序. 因为相似图像块的奇异值矢量相近, 排序后两个相近的矢量会在排序矩阵中相邻, 遍历排序后的矩阵, 并且配合图像块的偏移频率信息, 可检测出复制伪造区域.

文献 [19−21] 提出的算法都用到了字典排序, 排序矩阵的规模是影响计算复杂度的主要因素, 它的行数表示图像块的数目, 列数表示图像块特征维数. 表 1 列出了三种算法的比较结果, 并举例进行了说明. 以一幅 512×512 的灰度图像为例, 设定分块大小为 8×8, 采用一级 haar 小波变换, 从表 1 的第 3、4 行可以看出, Li 算法的图像块数目大约是其他两种算法的 1/4. Li 算法中图像块的特征维数是 8 维, 而 Fridrich 和 Popescu 算法中图像块的特征维数分别是 64 维和 32 维. 但是在采用相同的 8×8 分块大小情况下, Li 算法的定位精度较低, 为 16×16, 而 Fridrich 和 Popescu 算法的定位精度为 8×8, 见表 1 第 5 行.

2.2 JPEG 双重压缩检测方法

JPEG 图像双重压缩是指一幅 JPEG 图像被解压, 然后用另一个量化表重新压缩存储. 在使用图像编辑软件创作伪造图像的过程中, 一幅 JPEG 图像在伪造处理结束后有可能再一次被压缩, 并且使用的压缩质量因子不同于原始图像的压缩质量因子. 这种 JPEG 双重压缩会使图像 DCT 变换系数的直方图产生周期性模式. Popescu 利用这种周期性模式来检测和质疑图像的真实性[22]. Fridrich 在双重压缩检测的基础上, 进一步研究了从 JPEG 双重压缩图像中估计出原始量化表的方法[23]. 戴蒙提出了使用抖动模式来分析与检测图像 JPEG 双重压缩, 并给出了满足 DCT 量化系数直方图抖动模式的充分必要条件[24]. 更进一步, He 通过审查图像 DCT 变换系数的双重量化效应, 不仅能检测出恶意修改的 JPEG 图像, 还实现了图像篡改区域的定位功能[25]. 但是, He 方法只对部分内容被修改的篡改图像有效, 而对全局的合成图像不适用.

需要指出的是, 双重压缩并不能绝对说明图像是伪造的, 因为它是很多情况下为了节省存储空间而进行的合法操作. 因此 JPEG 双重压缩只能作为图像伪造检测时的一种间接证据.

2.3 图像重采样检测方法

图像编辑过程中, 常常涉及到重采样操作, 如旋转、缩放、平移等, 由于图像重采样往往伴随插值操作, 使像素之间的相关性发生变化, 通过检测重采样引起的原始信息变化规律, 可以判别图像是否被修改. Popescu 采用期望最大化 (Expectation maximization, EM) 算法来检测图像是否经历过重采样操作[26]. 如果是重采样图像, 其实质是原始图像信号和周期信号的叠加, 则 EM 算法输出的概率图的傅里叶频谱图中会出现规律性的亮点; 而非重采样的原始图像, 其傅里叶频谱图中不会出现亮点. 同济大学的朱秀明对 Popescu 的模型[26] 进行了改进, 将原始图像分布代替均匀分布并增加了先验概率的迭代[27]. 同时, 通过增加小补偿量的方法来避免 EM 算法可能遇到的奇异点, 并且该方法将检测图像重采样的技术扩展到检测图像是否被 JPEG 压缩过.

表 1 复制 – 粘贴检测算法比较

Table 1 Comparison of copy-paste detection approaches

以 512×512 的灰度图像为例	Fridrich 算法	Popescu 算法	Li 算法
图像特征描述	DCT & 量化	PCA & 量化	DWT & SVD
8×8 块的数目 (排序矩阵行数)	$(512-8+1)^2 = 255\,025$	$(512-8+1)^2 = 255\,025$	$(256-8+1)^2 = 62\,001$
特征维数 (排序矩阵列数)	64	32	8
定位精度	8×8	8×8	16×16

2.4 光照条件不一致性检测方法

创作一幅合成图像, 例如通过剪接造成两个人合影的假象, 往往很难完全吻合由定向光照物体 (如太阳) 带来的照明效果. 因此光照不一致性成为揭示图像篡改痕迹的一个有力依据. Johnson 提出对图像提取闭合边界, 沿着闭合边界将图像分成若干局部块, 估计局部块的二维光源方向, 然后根据光源方向是否一致来检测图像伪造情况[28], 其中图像强度 I 可用式 (1) 表示

$$I(x,y) = R(\boldsymbol{N}(x,y)\cdot\boldsymbol{L}) + A \tag{1}$$

其中, R 代表恒量反射值, $\boldsymbol{L}$ 代表点光源方向的三维矢量, $\boldsymbol{N}$ 代表点 (x,y) 的曲面法线的三维矢量, A 代表恒量环境光. 图 3 为该算法的检测结果, 其中图 3 (a) 为真实图像, 图中人物的反射光源方向分别为 98 度和 95 度 (从左至右); 图 3 (b) 为篡改图像, 图中人物的反射光源方向分别为 123 度和 86 度 (从左至右), 显然不满足定向照明条件的一致性. 该算法的缺点是需要人工提取图像边界. 另外, 如果图像伪造区域和原始区域的表面不满足朗伯反射的假设, 或者图像拍摄于阴天, 难以确定定向的光源, 该算法将失效. 随后 Johnson 又提出使用图像中人物眼睛的镜面反射来估计光源的方向或照明条件[29], 而图像中照明条件的不一致性可以用来揭示图像伪造痕迹, 尤其是检测多人合影的合成图像伪造类型. 文献 [30] 提出在不需要评估光源的情况下检测图像的照明一致性, 其基本理论是基于球面频率不变量, 并且假设已知对象的几何形状.

(a) 真实图像　　(b) 篡改图像

(a) Authentic image　　(b) Tampered image

图 3　Johnson 照明条件不一致性检测算法的实验结果[28]

Fig. 3　The experimental results of light inconsistency detection approach proposed by Johnson[28]

2.5 模糊估计方法

图像篡改, 如人脸替换、肤质平滑、全景摄影中都不可避免地要用到模糊操作, 以获得无缝的伪造图像. 模糊操作是图像篡改中常用的一种处理操作, 图像处理中主要包括两种形式的模糊操作: 散焦模糊和高斯模糊, 它们用来减少伪造图像不连续的程度或删除无用的缺陷, 最终产生一个似是而非的图像. 因此如果图像中探测到附加的模糊过程, 那么该图像极有可能被修改过. 基于这个想法, Hsiao 提出利用图像频域知识进行模糊估计, 揭示模糊区域作为可能被篡改的图像区域[31]. Sutcu 提出利用图像小波变换系数的规律性来估计图像边缘的清晰度/模糊度[32]. 该检测方法是基于这样的假设: 如果图像经历了某种类型的伪造, 那么伪造区域的平均清晰度/模糊度将会与非篡改区域不同, 通过度量这种差异可以检测和定位图像篡改区域.

3 基于成像设备一致性的盲取证技术

图像在采集过程中通常会引入一些强区分性的特征, 这些特征对整幅图像是一致的, 其一致性的规律可用于图像伪造检测和篡改定位. 目前的研究方法包括 CFA 插值检测、传感器模式噪声检测、色差检测、相机响应常态性和一致性检测等.

3.1 CFA 插值检测方法

大多数中低档数码相机拍摄的彩色照片是通过单个传感器结合色彩滤镜矩阵 CFA 来获取的. 最常用的 CFA 是 Bayer 矩阵, 由 Bayer 矩阵获取的 CFA 图像仅采集了彩色照片中三分之一的样本点, 如图 4 所示, 而其余三分之二的样本点需要用插值方法进行填充. 插值法使得彩色图像的样本点之间存在着特定的关联关系, 而图像篡改过程则很有可能破坏或改变这些关联关系.

$r_{1,1}$	$g_{1,2}$	$r_{1,3}$	$g_{1,4}$	$r_{1,5}$	$g_{1,6}$	$\cdots$
$g_{2,1}$	$b_{2,2}$	$g_{1,3}$	$b_{1,4}$	$g_{1,5}$	$b_{1,6}$	$\cdots$
$r_{3,1}$	$g_{3,2}$	$r_{3,3}$	$g_{3,4}$	$r_{3,5}$	$g_{3,6}$	$\cdots$
$g_{4,1}$	$b_{4,2}$	$g_{4,3}$	$b_{4,4}$	$g_{4,5}$	$b_{4,6}$	$\cdots$
$r_{5,1}$	$g_{5,2}$	$r_{5,3}$	$g_{5,4}$	$r_{5,5}$	$g_{5,6}$	$\cdots$
$g_{6,1}$	$b_{6,2}$	$g_{6,3}$	$b_{6,4}$	$g_{6,5}$	$b_{6,6}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

图 4　Bayer 矩阵获取的 CFA 图像[33]

Fig. 4　A CFA image obtained from a Bayer matrix[33]

Popescu 用一个简化的线性模型来表示 CFA 插值的周期关联性, 然后使用 EM 算法来量化和评估一幅图像中的关联关系的存在性[33]. Long 认为用线性模型来表示 CFA 插值规律过于简单, 提出将 CFA 插值带来的图像像素间的空间关联用一个均方差模型表示, 并基于该模型从每个颜色通道获得一个系数矩阵, 然后将系数矩阵用 PCA 降维后送入前向反馈 BP 神经网络中, 分类真伪图像[34]. Swaminathan 认为大多数篡改操作都可以近似为

线性和非线性的组合, 提出一种去卷积的算法来检测图像在最初获取之后的后续操作, 从而验证图像的真实性[35], 并在此基础上设计了一个基于相似性度量和阈值的分类器来区分篡改图像和真实图像.

3.2 模式噪声检测方法

相机拍摄照片过程中会对图像附加上该相机特有的模式噪声, 而缺乏模式噪声的区域可以被认为是伪造区域. Lukas 提出计算图像区域的残留噪声和参考模式噪声的关联, 然后对所有区域的关联关系进行概率统计分析, 确定阈值, 概率小于阈值的区域被认为不存在模式噪声, 则断定该区域可能被修改过; 否则认为该区域是真实的[36]. 实验表明提出的算法对质量因子为 70 % 以上的 JPEG 压缩图像能够可靠地辨识出篡改区域. 但是, 该方法要求伪造检测时, 必须具备拍摄照片的相机或者一组由该相机拍摄的照片.

3.3 色差检测方法

事实上, 所有的光学成像系统都会对图像引入多种失真. 色差就是一类很常见的光学成像失真, 它是由于光学系统不能很好地聚焦不同波长的光线而导致的. 当图像被篡改时, 色差会被破坏, 从而导致整幅图像中色差的不一致. 因此色差的不一致性可以作为图像被篡改的一种证据.

Johnson 提出利用横向色差来检测图像篡改的方法[37]. 横向色差对不同波长的光远近成正比, 它可建模为一个颜色通道对于其他颜色通道的颜色扩充或收缩, 通过求颜色通道的互信息的最大值来估计模型参数. 实验结果表明该方法只在篡改区域较小的时候, 检测效果比较理想, 如果篡改区域较大或是全局修改, 则检测效果不理想. 该方法中只利用了横向色差模型, 进一步的研究可加入纵向色差模型或其他的光学失真模型, 从而提高篡改检测的灵敏性和准确性.

3.4 相机响应常态性和一致性检测方法

图像辐照度 r 与最终输出图像的强度 R 是相关的, 两者的关系可以用相机响应函数 f 来表示

$$R = f(r) \tag{2}$$

其中 f 可以是 Gamma 变换或是线性指数模型等形式. 相机响应函数 (Camera response function, CRF) 是最常用的相机特性之一, 它将电荷耦合器件 (Charge-coupled device, CCD) 传感器感应到的辐照度转换成亮度值, 最后以底片或数字形式记录下来. 由于不同的相机有不同的响应功能, 因此 CRF 可以作为一种自然特征来确定相机类型. CRF 可以从单幅图像中计算获得, 其不一致性可作为图像篡改的一种证据.

Hsu 提出了一个基于几何不变量和相机一致性特征来区分真实与拼接图像的半自动分类方法[38], 该方法利用各区域的像素计算几何不变量, 然后使用几何不变量来估计每个区域的 CRF, 使用交叉拟和技术来检测所有的 CRF 是否相互一致. 实验结果表明, 该算法检测真实图像获得了 87 % 的准确率, 检测拼接图像获得了 90 % 的准确率. 其缺点在于运算耗时, 计算特征需要 11 分钟, 而使用支持向量机 (Support vector machine, SVM) 训练至获得最佳参数则需要 5 个小时. Lin 提出了基于相机响应函数的常态性和一致性的篡改图像检测方法[39], 如果图像的形成不匹配相机响应函数的常态性, 则恢复的逆响应函数就会出现异常或不一致.

基于成像设备一致性的盲取证技术首先提取与成像设备相关的特征, 然后通过检测特征的全图一致性或特征的存在性来判断图像是否被篡改, 由于各算法测试相机类型差异、图像数据库不同等原因, 这里不作算法间的横向比较.

4 基于自然图像统计特性的盲取证技术

该类技术指定了一系列对图像篡改敏感的特征, 首先通过分析大量已知的原始图像和篡改图像来确定这些特征的区分阈值, 这些阈值被保存为参考值. 然后通过测量待检测图像的特征值与参考值的偏差情况, 来判断图像是否遭受篡改. 这些方法大多依靠分类器来作出决策. 目前的研究方法包括利用自然图像统计模型、双相干特征和边缘百分比特征、图像质量度量、二元相似性度量等特征进行图像取证.

4.1 自然图像统计模型方法

Lyu 提出使用基于正交镜象滤波器 (Quadrature mirror filter, QMF) 金字塔分解的自然图像统计模型方法来分类摄影图像 (Photographic images, PIM) 和计算机生成图像 (Photorealistic computer graphics, PRCG)[40]. 该方法提取彩色图像 QMF 分解后各个子带和方向上分解系数的四阶统计特征 (均值、方差、偏度、峰度). 同时考虑 QMF 分解后各个子带之间的关联关系, 进一步计算分解系数和其空间邻域像素位置、相邻尺度、相邻方向、其他颜色通道相同位置上的分解系数之间的四阶线性预测误差特征. 用上面提取的特征进行 PIM 和 PRCG 的分类. 在计算机生成图像的分类误报率控制在 1 % 的情况下, 使用 SVM 分类技术, 67 % 的自然图像被正确分类出来. 但是, 该统计模型容易被图像攻击破坏, 从而影响分类的效果.

随后 Lyu 对统计模型进行了改进, 在基于 QMF 金字塔分解的自然图像统计模型中增加了图像的局

部角谐波 (Local angular harmonic, LAH) 分解和提取基于旋转不变的相位统计特性[41]. 同时利用局部幅度统计和局部相位统计特性, 结合 SVM 分类技术, 在计算机生成图像的分类误报率控制在 1 % 的情况下, 将摄影图像的分类精度提高到了 74.3 %.

4.2 双相干系数和边缘百分比特征方法

Ng 指出在所有涉及图像非法编辑的操作中, 图像拼接被认为是最根本、最主要的操作[42]. 受之前双相干特征用于检测人类语音剪接的启发, Ng 首先使用双相干幅度和相位特征进行拼接图像检测. 双相干特征是一种三阶矩谱和有效检测二次相位耦合 (Quadratic phase coupling, QPC) 的技术, 其一维信号 $x(t)$ (其傅里叶变换形式为 $X(\omega)$) 的双相干系数的数学表示形式为

$$b(\omega_1,\omega_2)=\frac{\mathrm{E}[X(\omega_1)X(\omega_2)X^*(\omega_1+\omega_2)]}{\sqrt{\mathrm{E}[|X(\omega_1)X(\omega_2)|^2]\mathrm{E}[|X(\omega_1+\omega_2)|^2]}} \tag{3}$$

其中, X^* 表示 X 的伴随矩阵. 双相干的一个重要特性是 QPC 的敏感性. 人类语音信号具有低的 QPC, 而图像原本有非零级谱能量, 这意味着图像中原来就存在高的 QPC. 因此简单地将双相干特征直接应用于探测图像拼接效果并不理想, 检测精度仅为 62 % (随机判断的精度为 50 %). 为了进一步加强检测效率, Ng 又提出了两个基本的方法, 即刻画对双相干敏感的图像特征和估计拼接不变量, 并由此导出了三个新的特征: 双相干幅度和相位变化的预测残差特征, 以及边缘百分比特征[8]. 最后, 将改进的特征送入 SVM 分类器进行分类. 结果表明双相干特征结合边缘百分比特征可以显著提高图像拼接的检测准确率, 由 62 % 提升至 72 %. 但是, 该算法中使用的图像数据是 128×128 的块, 而不是一幅完整的、有意义的图像, 而且要求拼接图像没有进行后处理操作.

4.3 图像质量度量和二元相似性度量等特征方法

受文献 [43] 使用图像质量度量 (Image quality metrics, IQMs) 来预测图像中隐秘信号的存在性的启发, Avcibas 提出使用图像质量度量, 即广义矩特征来度量原图像和处理图像之间的失真, 探索图像多个方面的质量情况[44]. 其中原始图像用待检测图像的模糊版本来替代, 然后将度量特征送入分类器, 以区分原始图像和编辑过的修改图像. 实验表明对网络下载图像进行亮度调整、对比度增强和混合操作的分类准确率分别为 69.2 %、74.2 % 和 80 %, 平均分类准确率达到 74 %. 该算法的缺点是待检测的篡改图像的修改块不能小于 100×100 像素.

Bayram 提出了一种基于图像邻位平面的检测方法[45]. 该方法的基本思路是原始图像和篡改图像在位平面之间的相关性以及位平面内部的二元纹理特征方面是不相同的. 这一图像本质特征的变化可以通过监测位平面的量子空间矩得到, 这些特征称为二元相似性度量 (Binary similarity measures, BSMs). 同时使用顺序浮动前向搜索 (Sequential floating forward search, SFFS) 算法选择最佳特征, 使用线性回归分类器进行原始图像和修改图像的分类. 实验表明 BSM 特征结合线性分类能够可靠地检测出由 Photoshop 工具编辑的大部分伪造图像, 分类准确率为 75.5 %.

随后 Bayram[46] 将 BSMs、IQMs 和高阶小波统计特性 (Higher order wavelet statistics, HOWS) 融合起来, 形成综合取证特征, 融合特征集的分类效果好于单个特征的分类效果.

4.4 分类算法归纳

表 2 对基于图像统计特性的几种分类方法, 从其特征、分类器、分类准确率等方面进行了归纳, 并总结出每个算法适合检测的图像类型. 总体来说, 使用综合特征要比单个特征的分类效果好, 但是特征维数大, 计算更为复杂. 另外, 每个盲取证算法侧重检测的图像篡改类型不尽相同. 从表 2 中不难发现, 面向真实性鉴别的图像盲取证算法的效率取决于如下两个主要方面: 1) 选择合适的盲鉴别特征; 2) 选择恰当的取证算法或分类算法.

表 2　基于自然图像统计特性的分类方法归纳

Table 2　Summary of the classification approaches based on statistical characteristics of natural images

文献	统计特征描述	分类器设计	分类准确率 (%)	适合检测的图像类型
文献 [40]	基于 QMF 分解的图像幅度特征	非线性 SVM 分类器	67 (误报率为 1%)	区分自然图像与计算机生成图像
文献 [41]	基于 QMF 和 LAH 分解的图像幅度和相位特征	非线性 SVM 分类器	74 (误报率为 1%)	区分自然图像与计算机生成图像
文献 [8]	双相干特征	SVM 分类器	62	拼接图像
文献 [8]	双相干、预测残差、边缘百分比特征	SVM 分类器	72	拼接图像
文献 [44]	图像质量度量 IQMs 特征	线性回归分类器	74	亮度和对比度增强、旋转、缩放操作
文献 [45]	二元相似性度量 BSMs 特征	线性回归分类器	75.5	图像放大、旋转、亮度增强、模糊

5 存在的问题与研究展望

5.1 存在的问题

前面介绍了近年来国内外学者在面向真实性鉴别的图像盲取证技术方面取得的主要研究成果, 从这些研究现状可以看出, 图像盲取证技术目前仍然存在如下主要问题:

1) 图像盲取证方法的针对性太强.

本文第 2～4 节描述的所有单个取证技术都不能对图像真伪给出全面的、确定的结论. 现有的一些方法, 如 Popescu 提出的统计工具箱中的一系列方法[22] 比较零散, 都是针对特定图像伪造类型、特定应用场合的, 不能普遍适用多应用的图像环境. 实际中, 一幅篡改图像通常综合使用了多种伪造手段, 因此单一攻击类型取证算法的检测效果可能受到影响. 如何将这些算法进行有效的融合, 或者寻求更为综合的、区分性更强的鉴别特征, 是有待解决的一个关键问题. 数字图像盲取证技术的研究具有很大的挑战性, 由于分析图像时不具备任何先验知识, 难以确定用来检测图像伪造的特征, 因此应首先基于一些确定的、简单的图像篡改进行分析, 然后将这些基本的图像伪造检测算法进行融合, 最终的决策需要将众多的图像盲取证技术 (包括继续研究新的方法) 结合起来, 从而对被检测图像给出一个综合的鉴别结果.

2) 图像盲取证方法的鲁棒性较差.

目前多数的图像盲取证方法对有损压缩、随机噪声叠加等操作的鲁棒性较差. 就图像盲取证方法检测的图像文件格式而言, 多数的算法还只能检测无压缩的原始图像, 或者对压缩图像真伪分类的误差较大. 而实际中广泛使用的图像是压缩格式的, 如常用的 JPEG 图像. 另外, 加入不可感知水印的图像, 可能会被一些脆弱的图像盲取证技术误判为伪造图像. 因此应该提高图像盲取证算法对压缩图像、水印图像等的鲁棒性.

3) 缺乏公用的图像测试数据库.

目前公用的图像测试数据库比较缺乏, 导致图像盲取证系统性能评价方法不统一, 难以对各种篡改图像取证技术的性能进行定量比较. 目前公开的仅有 Columbia 大学建立的图像拼接检测评估数据库、摄影图像和真实感计算机图形数据库[47]. 为了定量、客观地比较各种篡改图像描述和检测技术, 需要加强公用图像数据库的建设, 统一系统评测方法和规范.

5.2 研究展望

图像盲取证技术是一个多学科综合的研究问题, 它涉及了计算机视觉、信号处理、计算机图形学、机器学习、成像传感器、模式识别等领域的知识. 随着盲取证和其他相关技术的发展, 未来对图像盲取证技术的研究可以集中在以下几方面:

1) 完善图像盲取证的理论研究.

数字图像盲取证方面的研究刚刚起步, 其概念、理论和方法还不够全面和清晰. Ng 就图像盲取证的框架及关键技术进行了系统的分析, 提出了一个图像盲取证引擎的通用体系结构[11]. 今后图像盲取证技术的研究应该侧重于完善理论、提高盲取证算法的鲁棒性以及建立相关评价标准.

2) 建立自然图像一致性模型与有效的盲鉴别特征表示与描述机制.

选择合适的盲鉴别特征是影响图像取证算法效率的重要因素之一. 自然图像应当具有一致的能量谱、光照模型、色彩插值的周期性等[48], 真实图像一般是相对平滑的信号, 例如 CCD 相机有一个光学低通滤波器用来减少图像的折叠失真现象; 而伪造图像通常会出现一些不一致性, 例如光照不一致、高阶统计量出现锯齿、色彩周期性模型不存在等, 这些不一致性就可以作为鉴别图像真伪的依据. 因此针对大量的、不同类型和不同来源的真实图像尝试建立一致性的真实性模型与有效的盲鉴别特征表示与描述机制, 包括自然图像质量模型、自然图像篡改模型、光照一致性模型、色彩周期模型、对象边缘一致性模型和成像一致性模型等是非常必要的.

3) 研究基于对象的图像盲取证技术.

从以往相关研究来看, 图像盲取证算法主要是应用了某些底层信号特征来鉴别图像. 实际上, 图像与一般的数字信号不同, 它不仅仅是数字比特信号, 而且具有视觉内容和空间结构内容. 一幅图像认为被篡改更多关注的是图像对象内容的语义发生了改变, 影响了人们对图像内容的理解和认识. 而不改变图像内容的比特改变, 则不应该认为是篡改, 这就将加入水印的图像排除在篡改范围之外. 受此启发, 图像真实性鉴别可尝试从图像内容对象的添加、删除、修改三种普遍的篡改类型的角度出发, 通过分析图像质量一致性是否被破坏来检测和判断出图像是真实的还是可能被篡改的.

4) 实现自动化与多层次的图像真实性鉴别能力.

一个理想的图像盲取证系统应该是全自动的, 并且能够提供多层次的分析. 但是目前提出的图像盲取证技术大部分是半自动的或是粗糙层次的真伪图像检测或分类.

如果说自动化程度和检测精度之间是相互矛盾的, 那么设计图像盲取证系统应该分为两个层次: 第一个层次是自动化的, 支持实时、快速的图像真伪检测和过滤; 第二个层次是半自动化的, 支持对选定图像集进行详细分析和取证, 能够在已被恶意篡改的

图像中查出被攻击的区域.

6 结束语

面向真实性鉴别的图像盲取证研究的是如何在不依赖任何先验信息的情况下为鉴别图像真伪提供有力证据的技术, 其关键是找到充分、可靠、有说服力的证据来证明图像是否发生篡改. 本文首先对图像盲取证的概念框架、分类方法进行了简单介绍, 然后从三个方面阐述了图像盲取证技术, 并对当前图像盲取证研究方向上存在的问题和发展趋势作了比较详细的分析和探讨. 该研究领域还存在大量的问题和挑战, 深入的研究将可以获得更多的原创性的研究成果.

References

1 Friedman G L. The trustworthy digital camera: restoring credibility to the photographic image. *IEEE Transactions on Consumer Electronics*, 1993, **39**(4): 905−910

2 Lin C Y, Chang S F. A robust image authentication method distinguishing JPEG compression from malicious manipulation. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, **11**(2): 153−168

3 Lu C S, Liao H Y M. Structural digital signature for image authentication: an incidental distortion resistant scheme. *IEEE Transactions on Multimedia*, 2003, **5**(2): 161−173

4 Celik M U, Sharma G, Saber E, Tekalp A M. Hierarchical watermarking for secure image authentication with localization. *IEEE Transactions on Image Processing*, 2002, **11**(6): 585−595

5 Zhu B B, Swanson M D, Tewfik A H. When seeing isn′t believing. *IEEE Signal Processing Magazine*, 2004, **21**(2): 40−49

6 Wu Jin-Hai, Lin Fu-Zong. Image authentication based on digital watermarking. *Chinese Journal of Computers*, 2004, **27**(9): 1153−1161
(吴金海, 林福宗. 基于数字水印的图像认证技术. 计算机学报, 2004, **27**(9): 1153−1161)

7 Wu Q, Li G H, Tu D. An image authentication watermarking with self-localization and recovery. In: Proceedings of the 11th Joint International Computer Conference. Chongqing, China: World Scientific, 2005. 960−963

8 Ng T T, Chang S F, Sun Q B. Blind Detection of Digital Photomontage Using Higher Order Statistics, Advent Technical Report 201-2004-1, Columbia University, June 2004

9 Khanna N, Mikkilineni A K, Martone A F, Ali G N, Chiu G T C, Allebach J P. A survey of forensic characterization methods for physical devices. In: Proceedings of the 6th Annual Digital Forensics Research Workshop. Lafayette, USA: Elsevier, 2006. 17−28

10 Sencar H T, Memon N. Overview of state-of-the-art in digital image forensics. Part of *Indian Statistical Institute Platinum Jubilee Monograph Series* Titled *Statistical Science and Interdisciplinary Research*. USA: World Scientific Press, 2008

11 Ng T T, Chang S F, Lin C Y, Qibin Sun Q B. Passive-blind image forensics. In: Multimedia Security Technologies for Digital Rights. New York: Elsevier, 2006

12 Lanh T V, Chong K S, Emmanuel S, Kankanhalli M S. A survey on digital camera image forensic methods. In: Proceedings of 2007 IEEE International Conference on Multimedia and Expo. Beijing, China: IEEE, 2007. 16−19

13 Dirik A E, Sencar H T, Memon N. Source camera identification based on sensor dust characteristics. In: Proceedings of 2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics. Washington D. C., USA: IEEE, 2007. 1−6

14 Dehnie S, Sencar H T, Memon N. Digital image forensics for identifying computer generated and digital camera images. In: Proceedings of 2006 IEEE International Conference on Image Processing. Atlanta, USA: IEEE, 2006. 2313−2316

15 Mehdi K L, Sencar H T, Memon N. Blind source camera identification. In: Proceedings of 2004 IEEE International Conference on Image Processing. Singapore: IEEE, 2004. 709−712

16 Tsai M J, Wu G H. Using image features to identify camera sources. In: Proceedings of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. Toulouse, France: IEEE, 2006. 297−300

17 Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 2006, **1**(2): 205−214

18 Adams J, Parulski K, Spaulding K. Color processing in digital cameras. *IEEE Micro*, 1998, **18**(6): 20−30

19 Fridrich J, Soukal D, Lukas J. Detection of copy-move forgery in digital images. In: Proceedings of Digital Forensic Research Workshop. Cleveland, USA: Springer, 2003. 1−10

20 Popescu A C, Farid H. Exposing Digital Forgeries by Detecting Duplicated Image Regions, Technical Report TR2004-515, Department of Computer Science, Dartmouth College, 2004

21 Li G H, Wu Q, Tu D. Sun S J. A sorted neighborhood approach for detecting duplicated regions in image forgeries based on DWT and SVD. In: Proceedings of 2007 IEEE International Conference on Multimedia and Expo. Beijing, China: IEEE, 2007. 1750−1753

22 Popescu A C. Statistical Tools for Digital Forensics [Ph. D. dissertation], Department of Computer Science, Dartmouth College, 2005

23 Fridrich J, Lukas J. Estimation of primary quantization matrix in double compressed JPEG images. In: Proceedings of Digital Forensic Research Workshop. Cleveland, USA: IEEE, 2003

24 Dai Meng, Lin Jia-Jun, Mao Jia-Fa. The analysis and detection of double JPEG compression. *Journal of Image and Graphics*, 2006, **11**(11): 1619−1622
(戴蒙, 林家骏, 毛家发. JPEG 二次压缩的分析与检测. 中国图象图形学报, 2006, **11**(11): 1619−1622)

25 He J F, Lin Z C, Wang L F, Tang X O. Detecting doctored JPEG images via DCT coefficient analysis. In: Proceedings of 9th European Conference on Computer Vision. Graz, Austria: Springer, 2006. 423−435

26 Popescu A C, Farid H. Exposing digital forgeries by detecting traces of resampling. *IEEE Transactions on Signal Processing*, 2005, **53**(2): 758−767

27 Zhu Xiu-Ming, Xuan Guo-Rong, Yao Qiu-Ming, Tong Xue-Feng, Shi Yun-Qing. Resampling detection in information forensics. *Journal of Computer Applications*. 2006, **26**(11): 2596−2597
(朱秀明, 宣国荣, 姚秋明, 童学锋, 施云庆. 信息取证中图像重采样检测. 计算机应用, 2006, **26**(11): 2596−2597)

28 Johnson M K, Farid H. Exposing digital forgeries by detecting inconsistencies in lighting. In: Proceedings of the 7th Workshop on Multimedia and Security. New York, USA: ACM, 2005. 1−10

29 Johnson M K, Farid H. Exposing digital forgeries through specular highlights on the eye. In: Proceedings of the 9th International Workshop on Information Hiding. Saint Malo, France: Springer, 2007. 311−325

30 Mahajan D, Ramamoorthi R, Curless B. A theory of frequency domain invariants: spherical harmonic identities for BRDF/lighting transfer and image consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(2): 197−213

31 Hsiao D Y, Pei S C. Detecting digital tampering by blur estimation. In: Proceedings of the 1st International Workshop on Systematic Approaches to Digital Forensic Engineering. Taipei, China: IEEE, 2005. 264−278

32 Sutcu Y, Coskun B, Sencar H T, Memon N. Tamper detection based on regularity of wavelet transform coefficients. In: Proceedings of 2007 IEEE International Conference on Image Processing. San Antonio, USA: IEEE, 2007. 397−400

33 Popescu A C, Farid H. Exposing digital forgeries in color filter array interpolated images. *IEEE Transactions on Signal Processing*, 2005, **53**(10): 3948−3959

34 Long Y J, Huang Y Z. Image based source camera identification using demosaicking. In: Proceedings of the 8th Workshop on Multimedia Signal Processing. Victoria, USA: IEEE, 2006. 419−424

35 Swaminathan A, Wu M, Liu K J R. Image tampering identification using blind deconvolution. In: Proceedings of 2006 IEEE International Conference on Image Processing. Atlanta, USA: IEEE, 2006. 2309−2312

36 Lukas J, Fridrich J, Goljan M. Detecting digital image forgeries using sensor pattern noise. In: Proceedings of the SPIE. San Jose, USA: SPIE, 2006. 362−372

37 Johnson M K, Farid H. Exposing digital forgeries through chromatic aberration. In: Proceedings of the 8th Workshop on Multimedia and Security. Geneva, Switzerland: ACM, 2006. 48−55

38 Hsu Y F, Chang S F. Detecting image splicing using geometry invariants and camera. In: Proceedings of 2006 IEEE International Conference on Multimedia and Expo. Toronto, Canada: IEEE, 2006. 549−552

39 Lin Z C, Wang R R, Tang X O, Shum H Y. Detecting doctored images using camera response normality and consistency analysis. In: Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 1087−1092

40 Lyu S W, Farid H. How realistic is photorealistic? *IEEE Transactions on Signal Processing*, 2005, **53**(2): 845−850

41 Lyu S W. Natural Image Statistics for Digital Image Forensics [Ph. D. dissertation], Department of Computer Science, Darthmouth College, 2005

42 Ng T T, Chang S F. A model for image splicing. In: Proceedings of 2004 International Conference on Image Processing. Singapore, Singapore: IEEE, 2004. 1169−1172

43 Avcibas I, Memon N, Sankur B. Steganalysis using image quality metrics. *IEEE Transactions on Image Processing*, 2003, **12**(2): 221−229

44 Avcibas I, Memon N, Sankur B, Ramkumar M. A classifier design for detecting image manipulation. In: Proceedings of 2004 International Conference on Image Processing. Singapore, Singapore: IEEE, 2004. 2645−2648

45 Bayram S, Avcibas I, Sankur B, Memon N. Image manipulation detection with binary similarity measures. In: Proceedings of the 13th European Signal Processing Conference. Antalya, Turkey: Curran Associates, 2005. 752−755

46 Bayram S, Avcibas I, Sankur B, Memon N. Image manipulation detection. *Journal of Electronic Imaging*, 2006, **15**(4): 1117−1138

47 Ng T T, Chang S F. A Data Set of Authentic and Spliced Image Blocks, Advent Technical Report 203-2004-3, Columbia University, 2004

48 Srivastava A, Lee A B, Simoncelli E P, Zhu S C. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 2003, **18**(1): 17−33

吴 琼 国防科技大学信息系统与管理学院博士研究生. 主要研究方向为数字水印与图像取证. 本文通信作者.
E-mail: wuqiong_nudt@126.com
(**WU Qiong** Ph. D. candidate at the College of Information System and Management, National University of Defense Technology. Her research interest covers digital watermark and image forensics. Corresponding author of this paper.)

李国辉 国防科技大学信息系统与管理学院教授. 主要研究方向为多媒体安全.
E-mail: guohli@nudt.edu.cn
(**LI Guo-Hui** Professor at the College of Information System and Management, National University of Defense Technology. His main research interest is multimedia security.)

涂 丹 国防科技大学信息系统与管理学院副教授. 主要研究方向为图像处理.
E-mail: tudan1971@163.com
(**TU Dan** Associate professor at the College of Information System and Management, National University of Defense Technology. His main research interest is image processing.)

孙韶杰 国防科技大学信息系统与管理学院博士研究生. 主要研究方向为图像压缩与处理. E-mail: sshj_mil@126.com
(**SUN Shao-Jie** Ph. D. candidate at the College of Information System and Management, National University of Defense Technology. His research interest covers image compression and processing.)