

数字图像篡改盲检测综述

张怡暄^{1,2}, 赵险峰^{1,2}, 曹 纭^{1,2}

¹ 中国科学院信息工程研究所 信息安全国家重点实验室 北京 中国 100093

² 中国科学院大学 网络空间安全学院 北京 中国 100093

摘要 随着近些年成本低廉的高性能电子成像设备的不断普及和操作简单的数字图像编辑软件的广泛应用,人们制作一幅篡改图像已经变得越来越容易。这些技术使得人们很难察觉和辨识那些使用专业技术处理过的篡改图像的伪造痕迹,因而对包括新闻传播、司法取证、信息安全等诸多领域带来了严重的威胁,数字信息的安全性和可靠性也因此越来越受到国际社会的广泛关注。综上所述,开展针对数字图像篡改检测方法的研究有着极其重要的意义。本综述围绕数字图像篡改盲检测方法开展工作。首先,本文根据数字图像篡改检测方法所依赖的线索对篡改检测方法进行层次化分类,将图像篡改检测方法分为两个方面:基于成像内容及成像系统印记一致性的检测方法和基于篡改及 JPEG 重压缩痕迹的检测方法。然后,按照内容的来源和篡改操作所处的阶段,将以上两方面篡改检测方法进一步分为四个分组:基于成像内容一致性的检测方法、基于成像系统印记一致性的检测方法、基于篡改及其后处理痕迹的检测方法和基于 JPEG 重压缩痕迹的检测方法;又根据目前文献涉及话题的分布情况,再将四个分组细分为十二个分类:基于光照一致性的检测方法、基于特征提取与分类的检测方法、基于成像色差印记一致性的检测方法、基于自然模糊印记一致性的检测方法、基于成像系统噪声印记一致性的检测方法、基于彩色滤波阵列插值印记一致性的检测方法、基于几何变换及插值痕迹的检测方法、基于人为模糊痕迹的检测方法、基于中值滤波痕迹的检测方法、基于特征匹配的检测方法、基于对齐 JPEG 重压缩假设的检测方法和基于非对齐 JPEG 重压缩假设的检测方法。接着,本文梳理出每种分类的主干的思想脉络并对该类中重要的算法加以详尽分析和论述。除此以外,本文还对各类方法中典型的算法的性能做了比较,并归纳总结了在各种方法中常见的性能衡量标准和公开数据集,便于后续研究使用。最后,本文对各方法存在的问题加以总结,并对未来发展的趋势做出预测。希望此综述能够对数字取证有关的研究者提供参考文献的参考、研究方法上的启发和研究思路上的借鉴。

关键词 图像篡改; 盲检测; 成像内容; 成像系统; 篡改痕迹

中图分类号 TN915.08 DOI号 10.19363/J.cnki.cn10-1380/tn.2022.05.05

A Survey on Blind Detection of Tampered Digital Images

ZHANG Yixuan^{1,2}, ZHAO Xianfeng^{1,2}, CAO Yun^{1,2}

¹ State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

² School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100093, China

Abstract In recent years, with the popularity of cheap and advanced electronic imaging equipment and user-friendly image editing software, it is becoming easier for people to make tampered images. The traces of tampered images processed with professional techniques are hardly noticeable, which poses serious threats to many areas including news broadcasting, judicial forensics and information security, hence the security and reliability of digital information have been receiving more and more attention from all over the world. In summary, it is extremely important to carry out research on digital image tampering detection. This survey focuses on blind detection of tampered digital images. Firstly, according to the clues on which the digital image tampering detection method relies, the tampering detection method is hierarchically classified into two aspects: the one is consistency of imaging content and imprint of imaging system based method, the other is tampering trace and JPEG recompression trace based method. Then, according to the sources of the content and the stages of the tampering operation, the two aspects of tampering detection are further divided into four groups: consistency of image content based method, consistency of imprint of imaging system based method, tampering and its post processing based method and JPEG recompression based method. Based on the distribution of the topics in the current literature, the four groups can be further subdivided into 12 categories: consistency of illumination based method, feature extraction and classification based method, consistency of color difference imprint based method, consistency of natural blurring imprint based method, consistency of imaging noise imprint based method, consistency of color filter array interpolation imprint based method, geometric transformation and interpolation trace based method, artificial blurring trace based method, median filter trace based method, feature matching based method, aligned JPEG recompression hypothesis based method and non-aligned JPEG recompression hypothesis based method. What's more, we summarize the main ideas

通讯作者: 赵险峰, 博士, 研究员, zhaoxianfeng@iie.ac.cn。

本课题得到国家重点研发计划课题(No. 2019QY2202, No. 2020AAA0140000)资助

收稿日期: 2019-10-27; 修改日期: 2019-10-27; 定稿日期: 2022-03-25

in each category, with describing and analyzing important algorithms in each category. In addition, we compare the performance of typical algorithms in each category and summarize evaluation metrics and public datasets in all categories, which will be helpful for subsequent research. Finally, we sum up the shortcomings and predict the trend in each category. We hope that this survey can act as a reference and inspiration and can provide ideas for future researchers.

Key words image tampering; blind detection; imaging contents; imaging equipment; tampering traces

1 前言

随着互联网发展和智能手机的普及, 数字图像的制作日益便捷, 数字图像这种信息载体也在人们的生活中扮演着越来越重要的角色。图像可以对信息进行记录与传播, 它所记录事物的丰富程度远远超过其他信息载体。正因为如此, 人们为了达到特定的目的, 对图像中的内容进行修改以使他人相信这些被篡改后的信息。图像篡改技术是指利用一些手段修改或移除图像场景中真实

存在的内容, 或者给场景中增加一些实际上不存在的事物, 最后借助图像处理/编辑技术来掩盖有关篡改的痕迹。

1.1 图像篡改检测方法研究意义

事实上, 很多年以前图像篡改技术就用在了政治宣传等活动中。早在 20 世纪 30 年代, 前苏联领导人斯大林就使用图像篡改技术, 将他的政敌尼古拉·耶卓夫从一张照片中移除, 如图 1 所示。从图像中可以看出, 当时图像篡改技术已经非常高明, 篡改痕迹几乎无法被察觉。

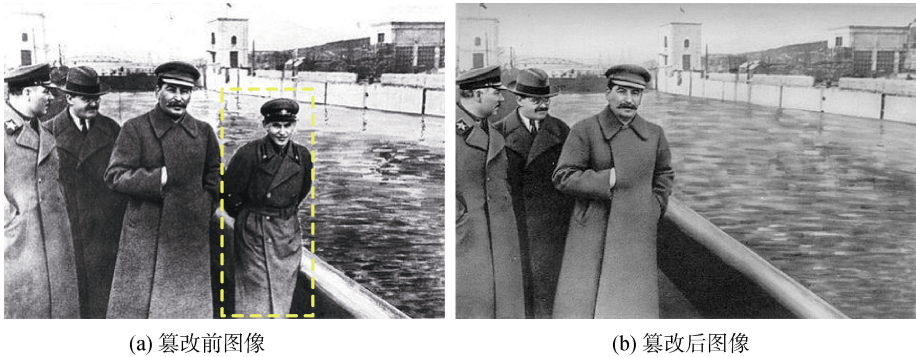


图 1 前苏联时期的一张篡改图像
Figure 1 A tampered image of Soviet era
(注: 黄色虚线内为篡改目标)

精心制作的篡改图像往往难以用肉眼分辨, 而且篡改图像的影响已经涉及人们日常生活的方方面面。图 2(a)展示的是影星简·方达与总统候选人约翰·克里在反越战集会时的一张合影, 事后证明该图是被伪造过的。该图曾在 2004 年美国大选期间广泛流传, 对克里的总统竞选之路带来了严重的影响。图 2(b)显示的是 2008 年伊朗对外发布的一张导弹试射图像, 后来有人发现其中一枚导弹与旁边的导弹非常相似, 最后证实该导弹确实是被篡改上去的, 此事件使得伊朗的声誉遭到国际社会的广泛质疑。图 2(c)显示的是黎巴嫩的一名摄影师哈吉拍摄的该国首都贝鲁特被空袭的情景, 该图发布后有人指责其中有被篡改过的痕迹。事后证实该摄影师为了使空袭的场景更佳震撼而对黑烟进行了渲染。

以图 1 当时的技术手段, 完成这样一幅精细而

又复杂的篡改图像一定是费力费时的。到了今天, 随着各种高性能、多用途且操作简单、价格低廉的成像设备, 图像处理设备以及图像编辑软件的普及, 篡改一张图像早已不像以前一样费时费力且需要大量的专业知识, 一个新手只需要经过一些简单的培训就可以在很短的时间内学会如何篡改一张图像。因此, 图像篡改检测技术持续人们的受到关注。

数字图像信息安全的重要性体现在诸多方面: 在司法取证领域, 需要真实的图像作为证据去给罪犯定罪或证明一个人的清白; 在科学研究领域, 需要真实的图像来说明一项科研成果的意义与价值; 在新闻报道中, 人们需要从真实的图像和视频中了解近期发生的大事; 在国防与国家安全等相关领域, 人们更是依赖真实的多媒体情报去进行决策以维护国家的安全。一旦有伪造的图像混入以上各领域, 将会给人们造成不可估量的损失。

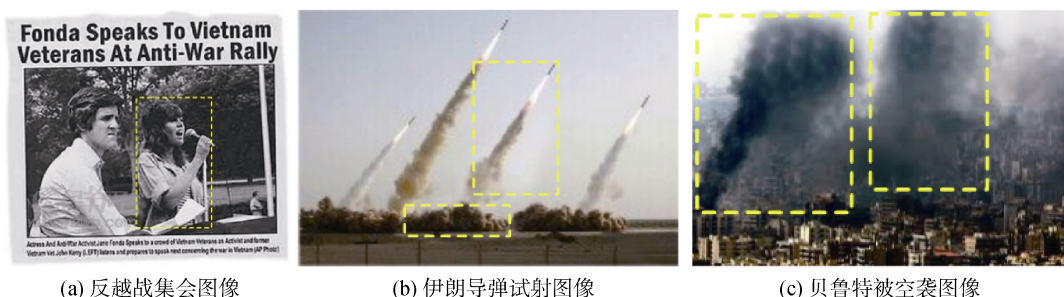


图 2 三张著名的篡改图像

Figure 2 Three well-known tampered images

(注: 黄色虚线内为篡改目标)

综上所述, 目前的图像篡改技术已经达到无法用肉眼进行分辨的水平, 且篡改图像对文化、政治、新闻传播等领域造成了极大的影响。随着图像编辑软件和篡改技术的不断进步, 人们迫切地需要更可靠、更全面的篡改检测技术的帮助去更好地识别伪造的图像。在未来, 数字图像篡改检测技术将会起到越来越重要的作用。近些年, 数字图像篡改检测话题受到很大关注, 研究文献日益增多, 有必要对其进行深入的分析和总结。

1.2 图像篡改检测方法的分类分析

图像篡改检测方式主要分为两种方式: 主动方式和被动方式。主动的篡改检测方式是指预先在图像中嵌入一些能够验证身份的信息, 例如水印等。在篡改检测阶段, 只需要验证预先嵌入的信息是否完整, 就可以判断该图像是否经过了篡改。被动的篡改检测(也称为盲检测)方式指的是在不知道图像任何先验知识的前提下, 利用图像的各种统计特性来进行篡改检测。

主动的篡改检测方式准确率高、速度快, 然而预先嵌入信息的前提条件在很多现实环境中难以达

成, 因此其应用范围非常有限。被动的篡改检测方式不需要任何先验信息, 只需要一张图像就能进行检测, 非常适合目前互联网盲检测的环境。基于以上原因, 本综述只讨论被动的篡改检测方式。在后面的论述中, 如未特殊说明, 本文所讲的“篡改检测技术和方法”, 含义都是指“图像篡改盲检测技术和方法”。

1.2.1 图像检测方法分类的背景

图像信息涉及图像生产和图像加工的不同环节, 也就是从图像的拍摄, 到图像的处理和存储, 再到最终在互联网上传播的各个环节, 图像篡改与这些产生与加工的过程是密不可分的。考虑到互联网是篡改图像的最普遍的生产和传播途径, 所以互联网图像的生命周期(如图 3 所示)是一条基本线索。

图 3 中绿色区域表示一幅真实图像从拍摄再到网上传播的全过程。首先, 自然景物产生或发射的光信息进入相机镜头(图 3 红色虚线内的部分)。随后, 这些信息将依次通过透镜系统、CFA 和 CCD 等环节(图 3 紫色虚线内的部分)而最终生成数字图像。成像系统的各个环节会在拍摄的图像中留下记录, 这些

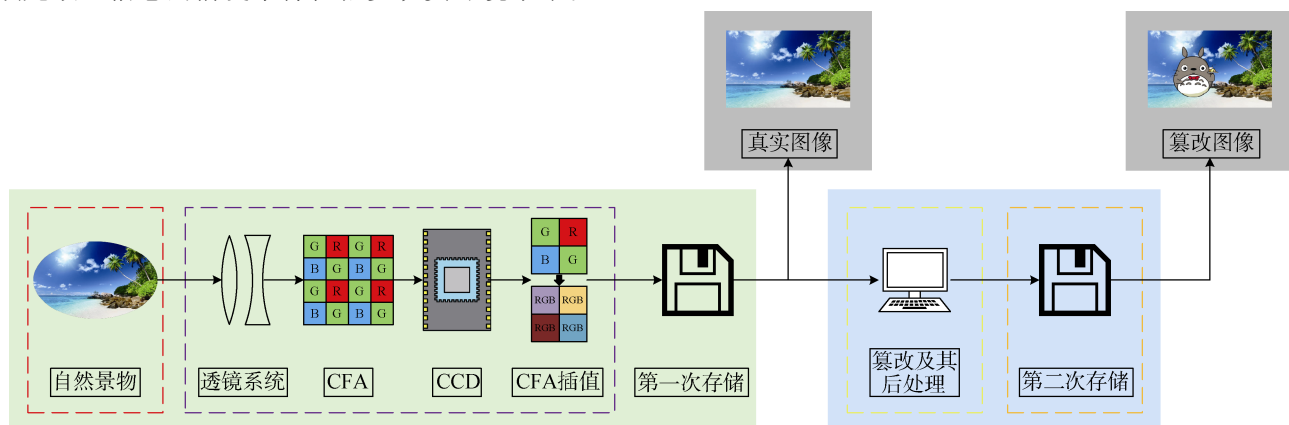


图 3 数字图像生命周期图

Figure 3 Life cycle of digital images

记录将会分布于获得的数字图像的每一个角落并可以成为判断其真伪“印记”。当一副自然图像拍摄完成后,如果有人对该图像进行篡改,就会对拍摄的自然景物和这些成像系统带来的印记造成不同程度的破坏。于是我们可以利用拍摄内容和这些印记的完整性来进行篡改检测。本文称这种方法为**基于成像内容以及成像系统印记一致性的检测方法**。

另一方面,尽管篡改者对自己的篡改的痕迹百般掩饰,所篡改的图像能达到肉眼难以分辨的程度,但还是会有微观的篡改痕迹被遗留在图像中,可以通过检测这些痕迹来发现篡改的事实。篡改者对图像经过篡改的过程如图 3 蓝色区域所示,图像被篡改的痕迹可能遗留在篡改及其后处理阶段(图 3 黄色虚线内的部分)或第二次存储阶段(图 3 橙色虚线内的部分)。本文称这种方法为**基于篡改及 JPEG 重压缩痕迹的检测方法**。

1.2.2 图像检测方法的分类层次

基于前节分析,图像篡改盲检测方法分为两个不同的方面:基于成像内容及成像系统印记一致性的检测方法和基于篡改及 JPEG 重压缩痕迹的检测方法,这是检测方法的分类的第一层次。

本文把第一个方面基于成像内容以及系统印记一致性的检测方法,按照一致性内容的来源,再分为两组方法:基于成像内容一致性的检测方法和基于成像系统印记一致性的检测方法,也就是下面介绍的第一组和第二组;而把第二个方面基于篡改及 JPEG 重压缩痕迹的检测方法,按照篡改操作所处的阶段,也再分为两组方法:基于篡改及其后处理痕迹的检测方法和基于 JPEG 重压缩痕迹的检测方法,也就是下面介绍的第三组和第四组。这是检测方法的分类的第二层次。

按照这样的思路,也根据目前文献涉及的话题分布情况,本文在以上四组的框架下,进一步具体分为 12 个分类。

第一组,基于成像内容一致性的检测方法,包括 2 类:基于光照一致性的检测方法和基于特征提取与分类的检测方法。相机所记录的场景都是来自于真实世界,真实世界的所有事与物都是满足一定规律的,而篡改会对这种规律造成不同程度的破坏,因此我们可以利用这些规律的存在性和完整性来进行篡改检测。

第二组,基于成像系统印记一致性的检测方法,包括 4 类:基于成像色差印记一致性的检测方法、基于自然模糊印记一致性的检测方法、基于成像系统噪声印记一致性的检测方法和基于彩色滤波阵列

插值印记一致性的检测方法。在成像的过程中,成像系统的各个环节都会在图像中留下特殊的印记,这些印记在篡改的过程中有可能被破坏掉,因此可以利用这些特殊印记的来进行篡改检测。

第三组,基于篡改及其后处理痕迹的检测方法,包括 4 类:基于几何变换及插值痕迹的检测方法、基于人为模糊痕迹的检测方法、基于中值滤波痕迹的检测方法和基于特征匹配的检测方法。在篡改的过程中,为了使篡改物体更加符合目标场景,篡改者往往会对篡改物体使用缩放、旋转、模糊等操作以及为了掩饰篡改痕迹而进行的其他操作。因此我们可以利用这些在篡改过程中及其后处理阶段留下的痕迹来进行检测。

第四组,基于 JPEG 重压缩痕迹的检测方法,包括 2 类:基于对齐 JPEG 重压缩假设的检测方法和基于非对齐 JPEG 重压缩假设的检测方法。JPEG 压缩会大大节省图像所占用的空间,大部分在互联网上传输的图像都是经过 JPEG 压缩的。而在篡改的过程中,可能会涉及多次 JPEG 压缩,重压缩操作在图像中留下的痕迹可以作为图像被篡改的证据。

这是检测方法分类的第三层次。

如图 4 所示,本文对图像篡改检测方法进行三个层次的划分:先将所有算法划分为两个方面,进一步划分为 4 个分组,更进一步划分为 12 个分类。每个层次具体的细节,还有未来发展的趋势,将在后续章节中叙述。

1.2.3 其他综述文献的概况与关系

图像篡改检测研究话题目前已经发表的综述主要有 8 篇^[2-9],它们对课题研究不同时期的进展起到了促进作用。综述[2-7]均发表于 2008—2014 年,缺少对最近几年新技术和新方法的总结,尤其是有关深度学习的方法。

综述[8]发表于 2017 年,该文对自然环境下被动定位方法进行了总结,除此以外,该文还对很多定位方法进行了复现并对各方法的性能做了详细的对比。但该文是一篇只针对篡改定位方法的综述,一般的篡改检测方法没有出现在该综述中。

综述[9]发表于 2018 年,该文是图像篡改检测领域中比较新的文献,涵盖了领域中大部分类型的方法,对所列举文献也有比较细致的描述。但该综述缺少对一些重要方法总结,例如有关模糊的检测算法。除此以外,该综述没有总结研究方法所使用的数据集和评价标准,也没有对各研究方法中不同的算法进行对比。

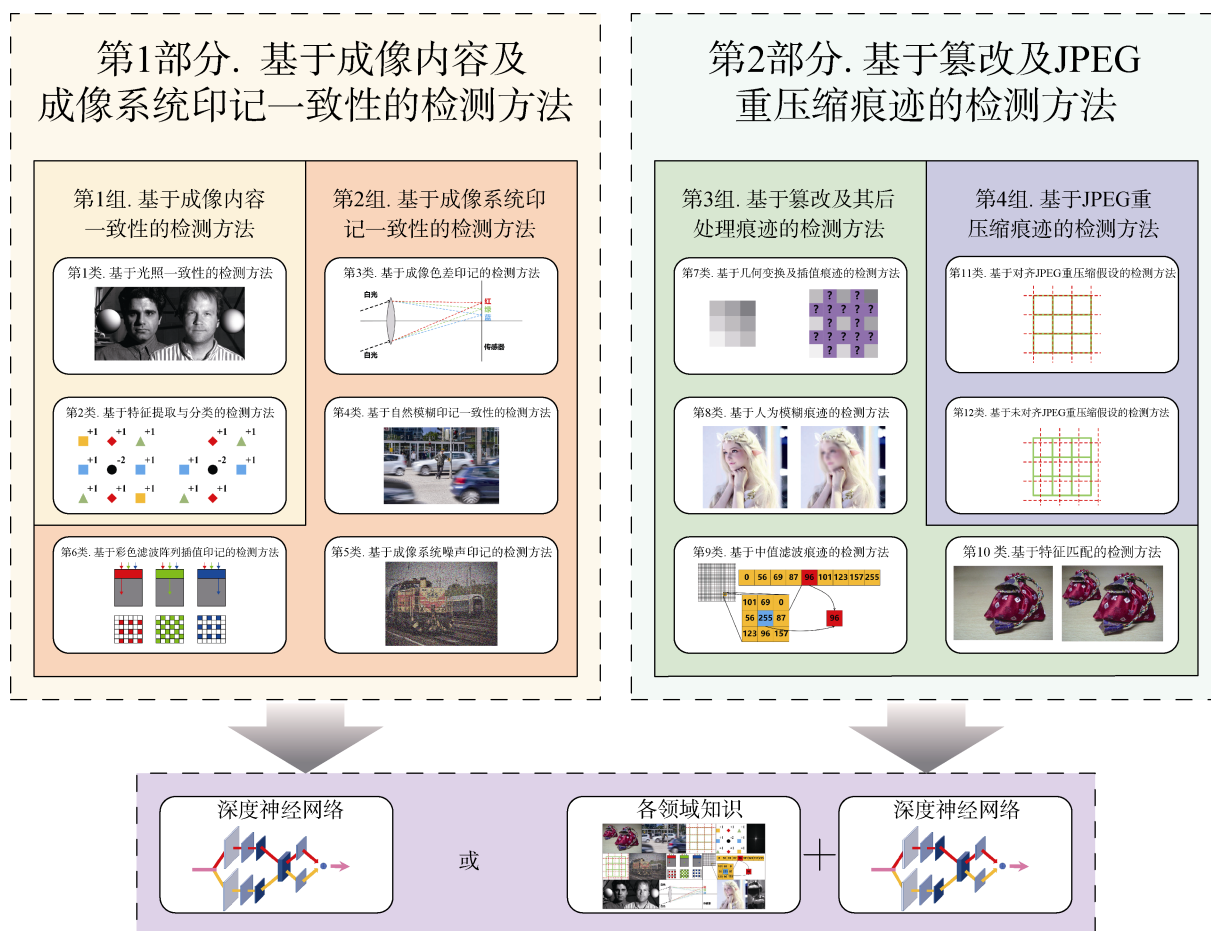


图 4 篡改检测方法的三个层次以及未来发展的趋势

Figure 4 Three levels of tampering detection and the trend of future development

(注: 左上角的图引用自文献[1])

基于以上观察, 本文的目标是提供一篇更新更全面的综述, 对图像篡改盲检测的新旧方法进行更细致的总结和分析, 并结合目前相关领域的发展趋势, 对未来的研究提出一些建议。

1.3 综述的特点

本文的特点可以概括为以下三个方面:

(1) 与现有综述对篡改检测方法的分类不同, 本文结合不同篡改手段特点和互联网图像的生命周期, 发现图像信息从生产到加工的不同环节与图像篡改检测的方式具有很大的联系。通过深入分析, 本文将图像篡改检测算法分成了两个方面: 基于成像内容及成像系统印记一致性的检测方法和基于篡改及 JPEG 重压缩痕迹的检测方法。然后再将图像篡改检测方法进一步划分为 4 个分组, 最后又将 4 个分组划分为 12 个分类, 以系统性地对各类技术方法进行展示。

(2) 本文根据所划分的 12 个具体的类别, 分别对每类方法中的主要算法进行了详尽的描述和分析, 梳理了不同算法的差异、共同点以及传承关系。另

一方面, 由于不同的方法在适用的环境和使用的数据集上有明显差异, 无法进行分类间的对比, 本文对每类中典型算法的性能进行了对比, 主要包括准确率以及运行效率。另外, 本文还对篡改盲检测算法所使用的性能指标和公开数据集进行了总结和归纳。最后, 本文对每类方法中现存算法的优缺点进行了分析, 并对未来发展方向和趋势提供了必要的建议。

(3) 考虑到目前深度学习技术在图像处理领域的广泛应用。本文着重分析了与深度学习结合较为紧密的 3 种方法, 包括基于特征提取与分类的检测方法、基于中值滤波痕迹的检测方法和基于特征匹配的检测方法, 由此发现以上 3 种方法中基于深度学习的算法检测准确率和检测效率明显高于其他传统方法, 本文指出与深度学习技术相结合是未来图像篡改检测方法的发展趋势。对于目前还没有与深度学习有效结合的方法, 应当选取合适的切入点将深度学习技术应用到该方法中; 而对于已经用到深度学习技术的方法, 也可以通过改善网络结构和数据

集以进一步提升各算法的性能。

1.4 综述的结构

综述后续各节的内容如下。

本文第二节的内容,是关于第一组“基于成像内容一致性的检测方法”,各种话题技术的描述、分析和小结;第三节的内容,是关于第二组“基于成像系统印记一致性的检测方法”各种话题技术;第四节的内容,是关于第三组“基于篡改及其后处理痕迹的检测方法”各种话题技术;第五节的内容,是关于第四组“基于 JPEG 重压缩痕迹的检测方法”各种话题技术;第六节的内容,是关于篡改检测算法评价标准和数据集的分析,以及各方法中典型算法的对比;第七节的内容,是关于篡改检测技术的总结与展望;第八节,是整篇综述的总结。

2 基于成像内容一致性的检测方法

现实世界中出现的事物会被记录在拍摄的数字图像中,一张真实的图像所记录的内容必定满足一定的自然规律。一旦图像里的内容违背了这些规律,就可以认定该图像是被篡改过的。比如如果一张图像中出现了奥巴马和林肯坐在一起聊天,则该图像一定是被篡改过的。光照分析和图像的边角特征是图像内容的基本规律。光照一致性的检测方法利用的是光源在物体上所投射阴影的规律性,而特征提取与分类的检测方法利用的是真实图像中存在的自然的边和角的规律性。

本节把基于成像内容一致性的篡改检测方法分为两个分类:基于光照一致性的检测方法和基于特征提取与分类的检测方法。

2.1 基于光照一致性的检测方法

在使用相机拍摄照片的过程中,场景中的光照信息也被记录在了照片中。同一光源在物体上留下的明暗轮廓和在背景上投射的阴影均满足一定规律,因此可以利用这些信息来估计光源的位置和大小。一幅图像中不同物体估计出的光源信息的不一致可以当作该图像被篡改的证据。

文献[10]首次将光源不一致性用于篡改检测。该文利用光照在物体上形成的明暗分界轮廓来估计光源的方位和性质。文献[10]中所用方法只适合单光源的情况,而文献[11]所用方法可应用于多光源场景。文献[11]使用球面调和函数(Spherical harmonics, SH)来对多光源的场景进行建模。多光源场景中的物体拥有非常复杂的阴影和光照梯度(Lighting gradients),为了实施计算,可以对前提条件进行简化从而可用一个九维的模型或一个更为精简的五维

模型来对该场景进行建模。文献[12]利用光线经过物体投射在背景上的阴影来对光源进行估计。在该文中,光源被假设成一个单一点光源,作者使用线性规划方法来解决该问题。如果所建立的方程没有解,则说明图像是被篡改过的。并且在这种情况下,文献[12]中的方法还可以用来找出图像中光照不一致的区域。

图像是二维的,但图像所记录的自然场景是三维的。图像中的很多物体由于缺乏三维的参数而无法直接进行三维建模。不同于之前提出的基于二维模型计算光源的方法,文献[13]提出一种基于用户引导的三维光源建模方法。实验者经过一定练习后,就可以从一张图像中估计三维物体的法线,进而对场景中的光源进行三维建模。由于三维建模的方法消除了二维建模方法中存在的歧义性(Ambiguity),文献[13]中提出的三维建模的方法效果明显好于二维建模的方法。

文献[14-15]中所使用的方法也可以对多光源进行建模。文献[14]使用通用的三维模型来对人的头部区域进行建模,该方法中所用的模型可以自动与图像中人物的头部姿态进行对准,以此获得三维模型的法向量信息。实验者基于所获得三维模型信息就可以对光源进行有效估计。文献[15]首先证明了基于二维建模方式的篡改检测方法容易受到反取证方法的干扰。其次,该文献利用在三维场景中广泛应用的阴影恢复形状(Shape from shading, SFS)方法来对场景进行建模,该方法能直接从图像中获取三维模型因而更具有普遍性。

文献[16]借助光源在人的眼部形成的高光(Highlight)来估计光源信息,在场景中发现的由不同人眼的高光区域所建立起的光源的不一致可以当作图像被篡改的证据。但该方法的一大问题是由于光源在人眼睛的高光区域较小,篡改者可以用一些反取证手段来消除篡改的痕迹。

需要指出的是,文献[11,14]所用建模方法的前提条件是场景中的物体有恒定的反射比和严格凸的表面,但现实环境中的很多物体并不符合该假设,例如常见的人面部区域就并非一个严格凸的模型,在这种情况下使用文献[11,14]中的方法会带来很大误差。文献[1]将局部纹理以及几何特征融入到位置相关反射模型(Position dependent reflection model)里去,放宽了文献[11,14]中建模所需的前提条件,使得文献[1]所用方法与实际情况更为契合。文献[17]对文献[1]中提出方法的原理做了更详尽的阐释,并做了大量扩展实验,来验证该方法的有效性。实验

表明, 文献[1,17]中所用方法的性能要明显优于文献[14]和文献[15]所用的方法。

现阶段利用光照一致性进行篡改检测的方法已经可以对较为复杂的光场和环境进行建模, 然而还难以达到对各种实际场景进行完全模拟的程度。在以后的研究中, 还需要结合计算机视觉和计算机图像处理等相关知识, 更准确地对光源和场景进行建模。另外, 目前很多基于光照一致性的检测算法还有部分环节依旧需要人为操作, 未来的研究可以着力于提高这些算法的自动化程度。

2.2 基于特征提取与分类的检测方法

在对图像进行篡改的过程中, 会产生一些不连续、不自然的角、线和边缘。众所周知, 自然图像都是满足一定分布的, 而对图像的篡改操作也会破坏掉自然图像的原始分布。基于特征和分类器的方法能够利用设计的特征去检测一副图像是否符合自然分布, 从而判断该图像是否经过了篡改。

按照所使用特征的特点, 基于特征提取与分类的方法大致可以将其分为三类: 基于一般特征的方法、基于马尔科夫特征的方法和基于 SRM 特征的方法。其中一般特征指的是马尔科夫特征和 SRM 特征之外的其它特征。

2.2.1 基于一般特征的检测方法

在文献[18]中, 双相干性(Bicoherence)特征用来对拼接图像进行检测。文献[18]指出: 虽然双相干性特征在语音取证领域检测效果优异, 但如果直接将该特征应用于图像篡改检测, 效果并不理想。于是文献[18]使用了两种手段来优化该特征, 检测效果明显改善。此方法属于早期探索性的方法, 检测准确率有待提升。文献[19]使用希尔伯特-黄锲(Hilbert-Huang transform, HHT)变换来提取图像特征。HHT 是一种在信号处理领域中有广泛应用的特征, 不同于 DCT 和 FFT 有固定的基础函数, HHT 能够根据信号的内容来自动生成基础函数, 因此非常适合去处理非平稳和非线性信号。实验表明使用 HHT 特征^[19]来篡改图像进行检测, 准确率能够比使用双相干性特征^[18]提高 8%。

基于观察: (1)在篡改过程中引入的不连续的点、线和不一致的光照会使图像的相位一致性(Phase congruency)明显升高, (2)小波特性方程(Moments of wavelet characteristic function)在处理隐写分析(Steganalysis)问题上有良好的表现, 文献[20]利用 24 维的相位一致性特征和 96 维的瞬时特征(Moments feature)共计 120 维特征来进行篡改检测。实验表明该联合特征的检测准确率明显高于双相干

性特征^[18]。文献[21]将待检测图像进行 DCT 变换后, 使用能对图像纹理特点进行良好描述韦伯局部描述子(Weber local descriptor, WLD)特征来区分真实的和篡改的图像。实验表明, WLD 的特征维度和检测准确率都要优于 HHT^[19]和文献[20]中使用的特征。

由于(1)在图像的亮度通道有大量的语义信息, 因而篡改信息在亮度通道中被这些语义信息所掩盖; (2)而图像色度通道只有很少的语义信息。在色度通道中不连续、突兀的篡改边界会比物体的自然边界更加明显。文献[22]在色度通道中提取特征来进行篡改检测: 先对图像进行 YCbCr 分解, 然后在色度通道中(Cb 通道或 Cr 通道)提取截断的边缘图(Edge image)的灰度共生矩阵(Gray level cooccurrence matrix, GLCM)作为特征。最后, 使用提升特征选择(Boosting feature selection)方法来对特征进行筛选, 以减小计算复杂度。

游程矩阵(Run-length matrix)能够很好地反映出图像的纹理信息, 可以用来对图像进行篡改检测。文献[23]首先将图像进行去相关化处理以减小图像受平滑区域的影响程度, 接着从图像色度通道的游程矩阵中提取 4 个方向的游程运动数字(Run-length run-number, RLRN)向量来进行检测。文献[23]所做的大量实验表明, 在图像的色度通道提取的 RLRN^[23]效果要明显好于亮度通道。和文献[22-23]一样, 文献[24]首先将图像转化到色度通道以增强篡改边缘, 然后对其进行方向金字塔变换(Steerable pyramid transform, SPT), 以产生多方向和多尺度的子带。接着, 从产生的子带中提取局部二值模式(Local binary pattern, LBP)直方图, 再将这些直方图串接起来作为最终特征进行分类。SPT-LBP^[24]的组合可以充分利用图像多方向和多尺度的信息。文献[25]将能反映图像分布变化的矩特征和局部平均值分解(Local mean decomposition, LMD)特征结合起来, 并利用自适应增强(Adaboost)分类器来进行篡改检测。在哥伦比亚拼接数据集(DVMM)上的实验表明, LMD+矩特征^[25]的检测准确率明显高于文献[19]、文献[20]和文献[26]所使用的特征。

2.2.2 基于马尔科夫特征的检测方法

马尔科夫特征反映了每个像素和其临近像素之间的关系^[27]。拼接篡改图像中不自然的边界和模糊、插值等后处理手段, 会破坏自然图像邻接像素的分布特性。

马尔科夫特征提取的步骤为: 首先在待检测图像的横、纵方向以及主、副对角线方向上求取残差(如图 5 所示, 从上到下依次为横、纵方向以及主、副对

角线上的残差示意图), 然后对残差图进行截断以减小特征维度。最后计算邻接像素残差截断值的转移概率。该转移概率就是马尔科夫特征。

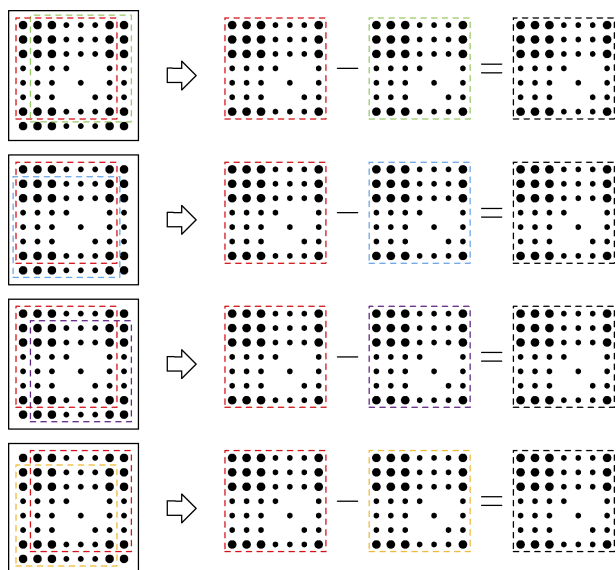


图5 马尔科夫特征提取方式^[27]

Figure 5 Extraction of Markov feature^[27]

多尺度块离散余弦变换(Multi-size block discrete cosine transform, MBDCT)是一种广泛应用于无线通信领域的方法, 文献[28]中使用 MBDCT 来进行篡改检测, 首先对图像进行 MBDCT 以得到该图像的一系列多尺度的表达, 接着从原图像和得到的 MBDCT 二维阵列中提取两种统计特征——马尔科夫转移概率(Markov transition probabilities)特征和特征方程的统计瞬时 (Statistical moments of characteristic functions)特征来进行篡改检测。MBDCT 可以利用图像相邻像素在不同尺度下的信息, 在哥伦比亚拼接数据集(DVMM)上达到了90%以上的检测准确率。文献[26]对文献[28]中的以下方面进行了改善: (1)不再使用的瞬时特征, 而只保留更有效率的马尔科夫特征; (2)使用马尔科夫特征同时考虑了块内和块间关系; (3)同时使用DCT和DWT中的马尔科夫特征来更好地表达图像不同尺度、方向和位置信息。最后, 为了减小计算复杂度, 文献[26]使用支持向量机递归式特征消除(Support vector machine recursive feature elimination, SVM-RFE)方法进行筛选以减小计算复杂度。实验表明, 在特征维度、检测准确率和计算时间上 DCT+DWT^[26]均明显优于 HHT^[19]。

不同于传统马尔科夫模型只对模型在一个方向的相关性进行统计, 文献[29]中使用的二维非因果马尔科夫模型(2-D noncausal Markov model)可以综合更多方向的信息因而可以更好地对二维图像信号进

行建模。文献[29]将提出的二维非因果模型用于 DCT 域和离散梅耶尔小波变换域(Discrete Meyer wavelet transform domain)来提取多域特征进行篡改检测, 实验表明该多域特征准确率高于文献[26]和文献[28]中所用特征。

在文献[30]中, 作者使用四元数 DCT(Quaternion DCT, QDCT)来提取特征。不同于一般的方法要先将图像转化到灰度通道, 文献[30]直接对彩色图像进行 QDCT 变换后在 QDCT 域中提取扩展马尔科夫特征(Expanded Markov feature)来进行篡改检测。由于 QDCT^[30]可以利用彩色图像所有通道的信息, 其检测准确率要明显高于文献[26]和文献[28]中所用特征。

文献[31]利用含有块间和块内关系的马尔科夫特征和轮廓波变换(Courlet transform)特征来进行篡改检测。轮廓波变换演变自小波变换, 可以多方向和多尺度地对信号进行分解。文献[32]提出了一种基于块 DWT(Block DCT, BDWT)来计算马尔科夫特征的方法。该方法避免了对整张图进行 DWT 的方式中不同区域的不同纹理和语义信息对 DWT 变换的影响, 且该方法相比于在整张图上计算 DWT 方式的计算代价更小。文献[32]先把图像分解成不重叠的图像块, 然后再对这些图像块进行三维 DWT 并提取马尔科夫特征来区分篡改的和真实的图像, 实验表明这种基于 DWT 特征的检测准确率要高于 DCT+DWT^[26]和 DCT+CT^[31]。

2.2.3 基于空域富模型(SRM)特征的检测方法

篡改图像与真实图像的差异主要表现在一些像素级的微观的差异上, 而不是语义级的、宏观的差异上。比如, 不连续的边缘或是由于平滑等后处理留下的肉眼不可见的篡改痕迹。

杰西卡·弗里德里希(Jessica Fridrich)提出了一个名为空域富模型(Spatial-domain rich model, SRM)^[33]的特征集来对图像进行隐写分析, 该特征集能够压制图像的内容信息并挖掘出相邻像素间微小的差异, 所以能够很好地反映在图像篡改中引入的细微噪声, 该特征集的两种典型特征如图 6 所示。基于篡改分析与隐写分析的相似性, 该特征集也在篡改检测的任务中也取得了良好的效果。

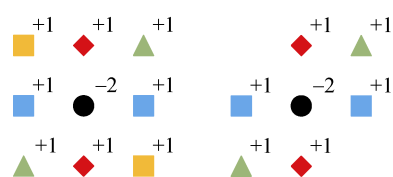


图6 两种典型的 SRM 特征^[33]

Figure 6 Two typical SRM features^[33]

文献[34]借助 SRM 特征来对篡改图像进行检测和定位, 该文中提出了有监督和无监督两种检测模式。在有监督的模式中, 需要用户预先指定一个未经篡改的区域去学习模型参数; 在无监督的模式中, 不需要用户事先指定区域, 只要利用期望最大化算法(expectation-maximization, EM), 篡改区域分割和模型参数学习可以同时进行。文献[35]首先从图像中提取出 SRM 残差图, 然后分别提取三邻域和五邻域的局部二值模式(Local binary pattern, LBP)特征和 LBP 的共生矩阵, 再对特征进行合并以降低计算复杂度。相比于 SRM^[33], 文献[35]中所用方法检测准确率高且特征维度更低。

文献[36]使用深度学习方法来进行篡改检测, 该文中设计的神经网络第一层使用的是从 SRM 中选出的 30 个卷积核, 这 30 个卷积核的参数固定, 不参与训练。该方法的效果远远超过之前的非深度学习方法, 在 CASIA V1.0 和 CASIA V2.0 数据集上甚至能达到 98% 以上的准确率。但文献[36]中方法只能对图像进行整体检测, 而不能实现对篡改区域的定位。文献[37-38]提出了对图像中篡改区域进行定位的方法, 该方法首先采取滑窗的方法在图像中提取重叠图像块, 再将提取出的图像块放入一个二分类神经网络中进行训练与分类, 再将图像块的分类结果融合成一张篡改概率图。文献[37]和文献[38]的具体实现细节的差异如下: 文献[37]中所用方法需要提取 5 种尺度的图像块并构造出五种尺度的篡改概率图。然后将该五张概率图使用简单线性迭代聚类(Simple linear iterative clustering, SLIC)分割法和条件随机场(Conditional random field, CRF)融合为一张最终的篡改概率图。而文献[38]中使用的二分类网络同时利用了图像块的空域信息和小波域信息, 使得对图像块的表达更为完善。

虽然以上描述的各种特征(一般特征、马尔科夫特征和 SRM 特征)目前已经能在单一的数据集上达到很高的检测准确率, 然而这些特征存在跨库准确率低, 篡改定位精度不高的问题。未来还需要设计更鲁棒的特征和性能更强的神经网络来对篡改图像进行检测和定位。

3 基于成像系统印记的检测方法

在数字图像成像的过程中, 图像信号会依次经过透镜、CCD、CFA 插值等环节, 这些环节都会在数字图像上产生特殊印记, 这些印记可能是硬件系统印在图像上的图案或是附着在图像上的噪声。虽然这些印记的能量非常微弱, 无法用肉眼察觉到,

但它们却可以被一些特殊的方法所捕捉到。对于一张真实的数字图像而言, 这些印记都是符合一定规律的。而在篡改的过程中, 篡改者很难做到让所有印记仍然保持原来的分布。

本节把基于成像系统印记的篡改检测方法分为 4 个分类: 基于成像色差印记一致性的检测方法, 基于自然模糊印记一致性的检测方法, 基于成像系统噪声印记一致性的检测方法和基于彩色滤波阵列插值印记一致性的检测方法。

3.1 基于成像色差印记一致性的检测方法

成像色差(Chromatic aberration)指的是成像过程中平行白光中不同颜色分量经过透镜后聚合到的焦点不一致的现象。这种不一致性既存在于和透镜垂直的平面上, 称为纵向色差(Longitudinal chromatic aberration), 如图 7(a)所示; 也存在于和透镜平行的平面上, 称为水平色差(Lateral chromatic aberration, LCA), 如图 7(b)所示。

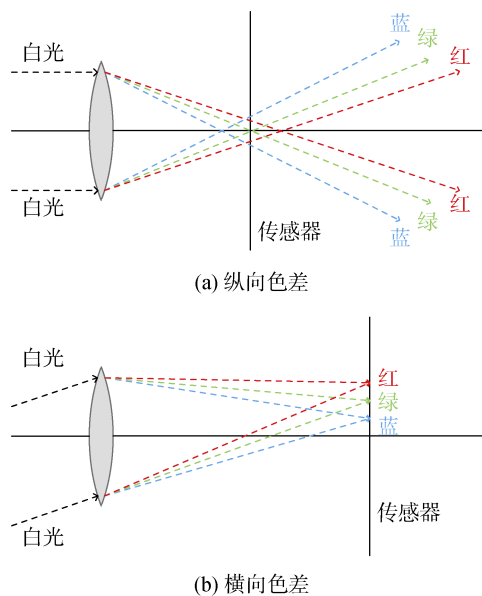


图 7 纵向与横向成像色差

Figure 7 Longitudinal and lateral chromatic aberration

成像色差记录了成像设备和成像过程的信息, 图像篡改往往会破坏掉这些信息。目前大部分基于色差进行篡改检测的方法都是关于 LCA 的。

文献[39]首次利用 LCA 方法来进行篡改检测, 该方法首先从三通道中选取两种不同颜色通道(例如红色通道与蓝色通道), 然后计算自然场景中同一个点的这两种颜色分量在成像面上的位置差, 该位置差值用箭头符号来表示, 其中箭头的方向代表色差的符号, 箭头的大小代表色差的大小。最终可以得

到整幅图像的色差分布。图 8 展示的是 LCA 位移向量场, 其中图 8(a)表明一幅真实图像的位移向量场, 我们可以看到真实图像的色差分布是符合一定规律的, 即所有箭头以辐射状围绕光学中心展开; 而图像的篡改过程会破坏这种规律, 如图 8(b)所示。然而, 文献[39]所用方法只根据色差的方向角来进行判断, 因而存在以下 3 个弊端: (1)在光学中心, 由于色差的幅度太小, 有可能无法检测到色差方向角; (2)如果将 A 区域的物体拼接到 B 区域上, 假如光学中心 O 与 AB 在同一条连线上, 则 A 区域和 B 区域色差的方向角完全相同, 会造成误判; (3)文献[39]使用的暴力迭代搜索(Brute-force iterative search)方法计算效率低下。

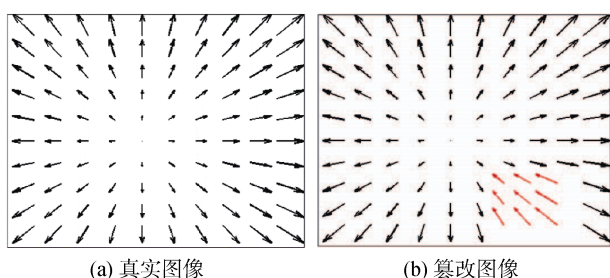


图 8 横向色差位移向量场^[40]
Figure 8 LCA displacement vector field^[40]

文献[41]指出: (1)不同类型手机中的成像设备构造会有很大不同, 因此造成的色差分布也会不同; (2)大多数手机相机的材质都比较廉价, 因此他们的色差都比较严重, 该文使用让经过校正的 R、G、B 通道之间的互信息(Mutual information)最大化的方式来对 LCA 的参数进行估计, 接着利用得到的 LCA 参数来对图像的成像设备进行判断。文献[39]使用的暴力迭代搜索的方式计算代价较高, 而文献[41]所用方法将搜索范围限制在距离光学中心的指定范围内, 这样可以显著地降低计算复杂度。但使用文献[41]所用方法得到的光学中心常常不准确, 会导致很高的误检率。文献[42]为减少计算代价, 对不同颜色通道的所有图像块之间的位移(Displacement)进行搜索以寻找一个相似度最高的位移来对 LCA 位移进行局部估计, 再使用全局模型来拟合这些局部估计。

文献[40]中使用一种基于梯度的模型来检测 LCA, 该模型能够刻画出局部和全局估计出的 LCA 的不一致性。篡改取证问题在该模型下转化成一个假设检验(Hypothesis test)问题, 从而可以得到更好的判断。另外, 文献[40]所用方法可以有效区分局部和全局位移的差异, 从而消除文献[39]中基于方向角方法的弊端。文献[43]在文献[40,42]的基础之上提

出改进方法: 在未被篡改的区域中将 LCA 的不连续性模拟成 0 均值、独立同分布的高斯随机噪声, 而在篡改的区域中将 LCA 的不连续性模拟成有偏置(Bias)的独立同分布高斯随机噪声。文献[43]中的策略可以进一步提高篡改检测的准确率。和文献[40]一样, 文献[43]也在梯度模型的作用下将篡改取证问题转化成了一个假设检验问题, 从而只需实验者简单设置一个阈值就可以进行决策。实验表明, 文献[43]所用方法的效果比文献[40]所用方法更为有效且远远超过文献[39,44]中使用的方法。除此以外, 文献[43]将在动态图像专家组(Moving picture experts group, MPEG)中广泛使用的一种运动向量的估计方法——菱形搜索(Diamond search), 用来对篡改区域的 LCA 进行估计, 使得检测效率比文献[42]所用方法高出两个数量级。

在文献[45]中, 作者提出了一种反取证的方法来隐藏篡改操作遗留在色差中的痕迹。虽然该反取证方法无法被文献[39]所用方法有效检测到, 但其篡改痕迹会被文献[46]所用方法检测到。在观察到: (1)在真实图像中, 任意两通道(蓝绿通道、红蓝通道、或者红绿通道)的频谱尖峰(Spectral peak magnitude, SPM)值的比率近乎一致, 而在篡改的图像中, 这些值却有很大偏差; (2)在真实图像中, 三通道 JPEG 谱峰值(Spectral peaks)相位角(Phase angle)之间的差异近乎为 0, 而在篡改图像中, 这些差异不为 0。文献[46]提出的利用通道间频谱尖峰比率和三通道 JPEG 谱峰值和相位角之间的差异的检测方法可以对文献[45]中提出的反取证方法进行有效检测。

以上利用成像色差进行篡改检测的方法均是利用 LCA 来检测的, 文献[47]利用 RGB 三通道的锐度信息来估计像素的纵向色差, 并通过纵向色差的不一致性来进行篡改检测。

紫色散射色差(Purple fringing aberration, PFA)指的是在成像过程中物体边缘处产生的紫色或蓝色的光晕现象。文献[44]提出一种利用 PFA 来进行篡改检测的方法, 该文中提到, PFA 形成的原因是: (1)电子外溢到光电二极管, 使得光子冲击传感器造成高对比度的边缘中出现模糊; (2)红外光到达传感器造成的边界模糊; (3)传感器单元边缘处的光冲击折射到微透镜(Micro lens)上并影响到邻接的单元。PFA 的蓝紫色光晕具有方向性。相对于图像中心而言, PFA 的方向是指向图像中心的, 它会出现暗物体的近端(或者亮物体的远端), 而篡改会破坏该规律性。虽然文献[44]中所用方法并不依赖特定的模型, 但当待检测的图像对比度较低时, 该方法的效果并不理

想。

虽然利用 LCA 进行篡改检测的方法有较高的准确率, 但单一的检测方法很容易被反取证方法所针对。目前的检测算法大多是利用 LCA 的, 未来应多开展利用 PFA 和纵向色差的篡改检测方法, 使得利用色差一致性进行篡改检测的方法能具有更强的鲁棒性和多样性。

3.2 基于自然模糊印记一致性的检测方法

自然模糊指的是成像过程中成像系统带来的模糊, 包括运动模糊(Motion blur)和离焦模糊(Defocus blur/Out-of-focus blur)两种。运动模糊是由于相机在拍摄照片的过程中产生抖动或被拍摄的物体运动过快而导致的, 离焦模糊是由于被拍摄的物体不在相机焦点上而引起的。在篡改的过程中, 篡改者很难让篡改物体与真实物体的自然模糊保持完全一致。

在基于自然模糊一致性的检测方法中, 首先计算出图像中不同物体的模糊种类以及方向、幅度等信息, 再结合场景信息来综合决策。图 9(a)(b)分别展示了运动模糊和离焦模糊的模糊核, 其中越偏白的区域代表像素值越大。可以看出, 运动模糊的模糊核和离焦模糊有明显的不同。

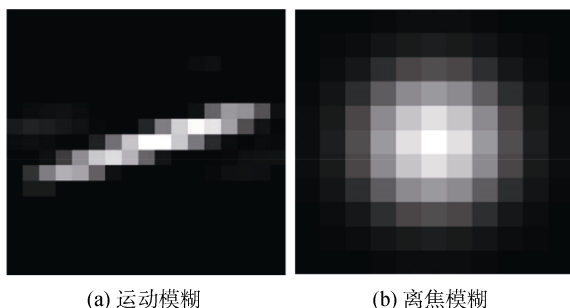


图 9 典型的自然模糊核^[48]
Figure 9 Typical natural blur kernel^[48]

文献[49]认为在成像过程中两个相对于相机处于同一深度的物体应该有相同的离焦模糊。如果一张图像违背了这一准则, 则该图像很可能是被篡改过的, 该文不同区域离焦模糊的估计和一致性判断均使用埃尔德-祖克尔(Elder-Zucker)方法。但文献[49]所用方法只能检测离焦模糊程度较低的图像, 文献[50]提出了一种可以检测高离焦模糊图像的方法。该方法需要先将图像分割成图像块, 再利用局部核重模糊(Reblur)的方法来估计图像中各区域的模糊程度, 最后结合各区域的模糊不一致性和深度信息来对篡改进行检测和定位。

以上所列举的算法都是基于离焦模糊的, 利用运动模糊的不一致性同样能够对篡改区域进行检测。

文献[51]认为运动模糊的方向可以通过图像中的物体估计出来, 而运动模糊的幅度却很难被估计, 因此篡改者很难让篡改物体与其他未篡改物体保持完全一致的运动模糊幅度。文献[51]利用谱抠图(Spectral matting)方法来对模糊进行估计, 而文献[52]则使用梯度方法来代替文献[51]中使用的谱抠图方法去估计物体的运动模糊, 使得检测拥有更快的速度。除此之外, 文献[52]还减少了文献[51]所用方法中所需的人为操作并对分割的技术做了优化。文献[53]使用具有方向性的高通滤波器来对运动模糊进行分析, 该文中给出了基于频域的和基于时域的分析方法。虽然该文只使用了两种高通滤波器来进行实验, 但其他的高通滤波器也可以很方便地融入到该系统之中。

如果能够同时识别离焦模糊和运动模糊, 并结合两种模糊的检测效果去对篡改区域进行定位和检测, 会有更好的检测效果。文献[54]提出一种有效的对离焦模糊和运动模糊的估计方法: 先根据像素奇异值(Singular value)找出模糊区域, 然后依据阿尔法(alpha)通道的约束(该约束不需要计算模糊核也不需要去模糊处理)来确定模糊类型。文献[48,55]指出, 如果图像中存在两个相对场景静止的物体, 如果只在一个物体中检测出了运动模糊, 则该图像是被篡改过的。

文献[55]首先从图像中提取出图像块来对局部模糊核进行估计, 接着使用聚合方法(Clustering method)将同一类型的图像块聚集到一起。对每个聚集区域的模糊进行分类后结合场景信息进行篡改检测。但是, 文献[55]所用方法只能处理对称离焦核(Symmetric out-of-focus)和一致运动模糊核(Uniform motion blur), 而在实际情况中存在更复杂的模糊核。文献[48]设计了一组新的特征并能够对更多类型的模糊核进行检测, 除此以外, 该文献使用了更先进的分割方法使得对篡改区域边缘的分割更为精确。

模糊检测需要先对模糊核进行估计, 而目前很多模糊核估计方法难以对实际问题需要的模糊核进行有效估计, 导致对篡改的检测出现严重偏差。在将来, 研究者应设计出更具有普遍性的模型来对模糊核进行估计。

3.3 基于成像系统噪声印记一致性的检测方法

成像系统带来的噪声会在图像信号中留下印记, 在一副未经篡改的图像中, 该印记通常是一致的。如果发现一张图像里估计出的某些区域系统噪声信号与其他地方不一致, 则表明该图像很有可能是被

篡改过的。

文献[56]利用图像带通域(Band-pass domain)峰度(Kurtosis)的规律和峰度与噪声特性的关系来进行篡改检测,一副图像中噪声的不一致性被当成篡改的证据。但该方法对于存在大量相似纹理的图像和经过强 JPEG 压缩的图像检测效果不佳。

值得注意的是,相机响应非一致性(Photo-response non-uniformity, PRNU)噪声是成像设备引入一种特殊的噪声,该噪声源于相机的传感装置的硅晶片在加工的过程中出现的一些小的瑕疵。这些瑕疵会在生成的图像中留下一些特殊的图案,这些图案就是 PRNU 噪声。该噪声与拍摄的内容和场景无关,只取决于成像所用的相机,可以当作相机的身份标识。

文献[57]介绍了两种利用 PRNU 进行篡改检测的方法:一种指定一个区域并对该区域进行篡改检测,另一种会自动检测出图像中的篡改区域。在品质因数为 70 的 JPEG 压缩图像中,文献[57]中所用方法依然能发挥良好效果。文献[58]对文献[57]中的方法做出了改进:首先使用最大似然准则(Maximum likelihood principle)估计出 PRNU,然后利用内曼-皮尔森准则(Neyman-Pearson criterion)将 PRNU 的检测问题转化为一个假设检验问题,最后使用最优检测统计量(Optimal detection statistics)来进行检测或识别相机的来源。实验表明该方法只需要很少的样本就能可以得到不错的效果,另外该方法也能有效地抵御缩放和有损压缩等操作。文献[59]中使用 BM3D 算法被用来代替米卡克(Mihcak)去噪滤波算法,使得检测准确率明显提高。文献[60]对文献[58]中的方法做出以下改进:(1)采用全局决策;(2)使用更加灵活的贝叶斯法则来决策;(3)使用马尔科夫随机场(Markov random field, MRF)建模;(4)使用非局部去噪算法,尤其是 BM3D 算法来提升数据质量;(5)使用凸优化(Convex-optimization),保证在有效时间收敛到最佳。基于以上改变,文献[60]所用方法比文献[58]所用方法的检测准确率有很大提升。但是文献[60]所用方法对图像中存在的小尺寸篡改的检测性能不佳,需要设计功能更强大的预测器以提升性能。

为了解决文献[58]提出的方法对于图像中存在的小尺寸篡改的检测效率不佳的问题,文献[61]采取了一种基于区域信噪比(Signal-to-noise ratio, SNR)的硬分割的方法来对图像进行预分割。文献[62]对文献[61]中的方法做出了改进,该文采取一种更为灵活的软分割的方式,该方式可以依据内容信息来对图

像进行分割,并在分析窗(analysis window)中采用自适应的权重。虽然文献[61]和文献[62]中所用方法相比之前提及的方法有很大改进,但这两种方法依然存在以下缺点:(1)过度依赖分割和先导图像(Pilot image);(2)无法对有遮挡的篡改(Occlusive forgeries)进行有效检测。为了改善文献[61]和文献[62]方法的弊端,文献[63]对滑窗从篡改边界的一边到另一边产生的决策变化进行建模,并依此调整决策阈值。

由于图像在拍摄的过程中都要经历 CFA 插值的过程,而 CFA 插值会对图像的 PRNU 信号产生很大的干扰。在文献[64]中,作者给出一种耦合-去耦合 PRNU(Couple-decoupled PRNU, CD-PRNU)信号提取方法。该方法首先将图像分解为 4 个子通道,并在每一个通道中分别提取 PRNU 噪声。接着,将 4 个通道提取的 PRNU 耦合起来的 CD-PRNU 作为特征进行篡改检测。该方法可以有效抵御 CFA 插值对 PRNU 信号的干扰。

在文献[65]中,作者结合成像系统的 PRNU 和多尺度融合方法来实现对篡改区域更精细的定位。作者利用滑窗(Sliding window)的方法对图像进行分析进而得到图像的篡改概率图,融合不同尺度的篡改概率图可以得到一个更全面的多尺度融合概率图。大量实验表明该多尺度融合计算方法的定位准确率明显高于单尺度的方法。

在利用噪声对图像进行篡改检测的所有方法中,研究最为广泛的当属 PRNU 噪声。由于利用 PRNU 对系统进行篡改检测的方法需要知道待检测图像的成像设备或由该设备拍摄的其他图像,大大限制了此方法的实际应用范围。另外,目前该方法对图像中存在的小范围篡改的检测效果也有待提升。

3.4 基于彩色滤波阵列插值印记一致性的检测方法

典型的光电传感器只能够收集到光线的强度信息,而不能确定是哪种波长的光。因此,相机在成像过程中 1 个像素只能记录到 1 种颜色的光信息。于是人们在光电传感器前边加上一组颜色滤镜,3 种颜色的滤镜交替排列,使得 3 种颜色的光信息穿插着地记录在图像中。该滤镜系统称为彩色滤波阵列(Color filter arrays, CFA)。不同的相机拥有不同的 CFA 插值模式,图 10(a)(b)展示的是两种典型的 CFA 阵列排列方式。

由于图像中的每个像素只允许 RGB 3 通道中 1 种颜色通道的光进入,初始成像的图像会有严重的马赛克效应。为了去除这种马赛克效应,要使用插值方法来计算每个像素中其他两个颜色通道的光信

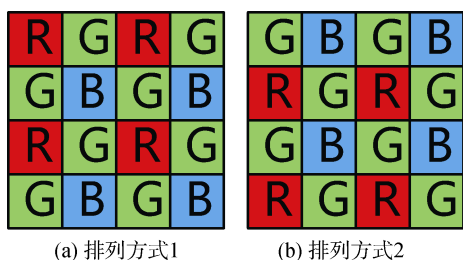


图 10 两种常见的 CFA
Figure 10 Two common CFAs

息, 这种插值在图像中留下的痕迹成为去马赛克痕迹(Demosaic artifact), 该痕迹可以作为图像篡改检测的依据: 成像设备的 CFA 插值信息记录在每张图像中, 一张未被篡改图像中的 CFA 插值在整张图像上应该是一致的, 而篡改操作会破坏这种一致性。

文献[66]详细介绍了 8 种 CFA 插值模式及其原理, 并指出使用简单的线性模型就可以很好地对这八种 CFA 模式进行模拟。文献[67]提出旋转、缩放、CFA 插值等重采样会给邻接像素带来强相关性并使用 EM 算法来估计插值的参数。文献[66]和文献[67]均属于此类方法早期的工作, 它们为后续研究工作的开展打下了基础。在文献[68]中, 作者提出了两种篡改检测的算法: CFA 模式数目估计(CFA pattern number estimation)法和 CFA 噪声分析(CFA based noise analysis)法。这两种方法优势是所用的特征不需要使用复杂的机器学习方法进行处理, 只需要设置一个简单的阈值就可以进行判断。观察到插值像素的噪声方差小于未被插值的像素, 文献[69]提出一种利用噪声方差(Noise variance)去估计 CFA 插值的方法, 实验结果表明该方法的检测效果好于文献[66]中提出的方法。但文献[69]和文献[68]中的方法, 在图像经过缩放和强 JPEG 压缩等后处理操作后, 检测效果欠佳。

文献[70]指出被插值的图像的二阶导数信号中存在着周期性, 并提出一种利用该周期性进行篡改检测的算法, 该算法对整数因子(integer factor)和非整数因子的插值方法都能有效检测。另外, 该算法不仅能够判断一张图像是否被插值过, 还可以计算插值的系数。文献[71]提出一种利用文献[70]中观察到的周期性来判断一幅图像是否是电脑渲染图像(Photorealistic computer generated images, PRCG)的方法: 先将图像放入高通滤波器以压制图像的低频信息, 以此来增强去马赛克效应带来的周期性信号, 接着使用傅里叶分析去检测对角线方差的周期性, 进而就可以判断该图像是否被 CFA 插值过。文献[72]结合了文献[67]和文献[70]中使用的特征来对图像的

成像设备进行检测。该文献指出, 文献[67]提出的有关二阶导数周期性的方法更适合于对图像中的平滑区域进行检测, 而文献[70]提出的有关 EM 算法的方法更适合于对图像中不平滑的区域进行检测。文献[72]先根据平滑程度将图像分成不同区域, 再分别利用不同的方法进行检测。

文献[73]中提出一种非侵入式的(Non-intrusive)的成像设备识别方法。该方法只需要利用成像设备的几张输出图像, 就可以通过线性逼近和局部纹理分析来计算该成像设备的 CFA 插值系数, 进而识别出成像设备。文献[74]使用偏微分相关模型(Partial derivative correlation models)和反向分类算法来进行篡改检测。该文中使用的偏微分相关模型充分利用了去马赛克效应在通道内和通道间的关系, 因而能够在多种后处理操作下有效识别不同的 CFA 插值, 实验表明该方法的性能远远超过文献[73]的方法。

文献[75]设计了一种通过检测局部 CFA 插值痕迹来进行篡改检测与定位的方法。利用该方法可以计算出一张精确到 8×8 像素的篡改概率图, 以此实现高精度的篡改定位。实验证明该方法的检测效果要明显优于文献[68]所用的方法, 但在图像经过低品质 JPEG 压缩的情况下文献[75]所用方法的定位精度不太理想。在文献[76]中, 作者使用基于马尔科夫转移概率矩阵(Markov transition probability matrix, MTPM)的高阶分析, 来进行篡改的检测与定位, 因为 MTPM 可以充分挖掘篡改带来的统计偏差(Statistical aberration)。因为文献[76]中所用方法充分利用了高阶统计信息, 使得该方法比文献[68]、文献[71]和文献[75]所用方法的错误率低且消耗时间少。文献[77]首先对图像使用高斯混合模型进行建模, 然后对得到的噪声图像使用 7×7 的高斯窗进行滤波, 最后使用在滤波后图像中提取的特征进行篡改检测与定位, 实验表明该方法的定位效果比文献[75]中所用方法更为精准。

不同于传统的针对单通道 CFA 插值进行检测的算法, 文献[78-79]提出了针对通道间(Inter-channel)CFA 插值进行检测的方法。单通道 CFA 插值算法需要对 3 个通道分别进行插值, 而通道间的 CFA 插值算法利用 RGB 3 个通道的频域相关性进行插值, 这种插值方式会使得 RGB 3 个通道频谱之间有较强的相关性, 而每个通道中相邻像素之间的相关性较弱。文献[78]指出, CFA 插值图像在不同颜色通道高频区域的频谱差与未插值图像完全不同。文献[79]利用 CFA 在图像重插值前后的色度失真以及在频谱中的改变来进行篡改检测。虽然文献[79]中

所用方法在未经 JPEG 压缩的图像中的检测准确率略低于文献[78]中使用的方法,但在经过 JPEG 压缩的图像中的准确率要明显高于文献[78]中使用的方法。

虽然目前基于 CFA 的篡改检测方法已经能够实现篡改区域的精确定位,但当图像中存在分布一致的区域(Uniform region)或者尖锐区域(Sharp region)的时候,还是会有很高的虚警率,新的研究应当围绕这个缺陷来进行。

4 基于篡改及其后处理痕迹的检测方法

在篡改数字图像和掩饰篡改痕迹的过程中,往往会在篡改图像中留下痕迹。这些痕迹虽然常常无法被肉眼所识别,但却可以被一些专门设计的方法检测出来。

本节把篡改检测方法分为 4 个分类:基于几何变换及插值痕迹的检测方法,基于人为模糊痕迹的检测方法,基于中值滤波痕迹的检测方法和基于特征匹配的检测方法。

4.1 基于几何变换及插值痕迹的检测方法

在将一张图像里的物体(或对象)拼接到另一张图像的过程中,为了让物体在新场景中看起来更加自然,篡改者往往会对这些物体施加一些几何变换,例如旋转,缩放,歪斜(Skew)等。图像在几何变换的过程中经常会涉及插值操作,而这些插值操作插值留下的痕迹虽然肉眼不可见,但却可以用一些微观的方法将其检测出来。

基于文献[67]中的观察,文献[80]描述了 4 种利用插值痕迹来进行篡改检测的方法:基于频域的检测方法分别利用 DCT 高通滤波和小波变换来检测插值痕迹;基于空域的检测方法分别利用二阶差分(Second difference)和二阶差分的零交点(Zero crossing)来检测插值痕迹。文献[70]和文献[80]中的方法在对经过歪斜的和旋转等几何变换图像检测效率不佳,而文献[81]中的方法却能够适用于各种几何操作。文献[81]着力于分析插值操作给图像信号及其导数的协方差带来的周期性特质,进而提出一种能适应于旋转、缩放和歪斜等各种几何变换的篡改检测方法,该方法还能够用来估计这些几何变换的参数。文献[82]在文献[70]的基础上展现了插值信号及其导数中存在的周期性,并把该理论拓展到二维空间。除此以外,文献[82]中的方法还能自动检测图像中几何变换的痕迹。相比于文献[67]中的方法,文献[82]中的方法更易于使用,并且不需要对参数进行初始化。

不同于大多数使用全局核的线性滤波器去检测残差信号中的插值痕迹的方法,文献[83]使用一些精心设计的行预测器和列预测器来检测插值信号的周期性,实验表明该方法对缩放操作引起的插值有很好检测效果。但是,对于除缩放外其他几何操作,该方法却很难适用。

奇异值分解(Singular value decomposition, SVD)是一种在信号处理领域中常用的分析工具,它能反映出图像像素之间一些重要的关系。文献[84]指出,插值操作会改变图像原有的线性相关性,而 SVD 恰好可以使这些相关性凸显出来,于是提出一种基于 SVD 的检测方法。该方法计算过程简单且检测效果良好,但在物体经过旋转或者图像被严重的噪声影响后,检测效果欠佳。文献[85]指出当可以使用样本过少或当被分析的图像包含结构性或周期性的样式时,文献[83]和文献[84]中方法的检测效果会受到严重影响。于是,在文献[85]中,作者提出一种基于 SVD 来对上采样进行检测的方法。该方法不需要 SVM 分类器且能适用于小尺寸的图像。该文没有给出检测下采样的具体方法,只提供了一些解决思路。

文献[86]和文献[87]分别提出了一种专门针对于上采样操作的篡改检测方法。在文献[67]的方法中,因为使用了固定尺寸的检测窗,只能对一些特定分数(Fraction)系数的上采样插值进行检测,且不能估计出上采样所用的系数。在文献[86]中,作者使用插值矩阵重建(Resampling matrix construction, RMC)来计算特定分数插值系数的插值矩阵,之后使用归零掩模求导(Zeroing Mask Derivation)来对插值创建归零掩模,进而对插值进行检测和对插值系数进行估计。该方法有效地解决了文献[67]中存在的问题。文献[87]使用了一种基于随机矩阵理论(Random matrix theory, RMT)和子空间分解(Subspace decomposition)的上采样检测和采样系数估计的方法,该方法能有效对上采样引起的插值操作进行检测。大量实验表明,文献[87]中方法对插值的检测效果明显超过文献[85]和文献[86]中使用的方法。

因为下采样操作不会给图像留下周期性的痕迹,目前的检测方法无法对下采样的图像进行有效检测。

4.2 基于人为模糊痕迹的检测方法

当一张图像被拼接到另一张图像中时,在拼接处经常会呈现出明显的篡改边缘。篡改者往往会对这些边缘进行模糊处理以掩饰这些篡改痕迹。一张图像中存在的人为模糊的痕迹可以作为该图像被篡改的有力证据。

文献[88]结合图像的先验知识和距离信息, 利用一种基于 DCT 变换的方法来检测一副图像中的人为模糊痕迹。保边平滑滤波(Edge preserving smoothing filtering)可以使人为模糊的边缘变得更尖锐, 而数学形态学(Mathematical morphology)方法可以在消除自然边缘的同时保留这些被保边平滑滤波锐化后的模糊边缘。文献[89]利用以上两种方法来对图像中的人为模糊进行检测, 虽然文中的检测效果图说明了该方法的有效性, 但文中没有给出有关人为模糊存在的定量的度量标准, 该问题在文献[90]中得到了解决。文献[90]提出一种基于边缘模糊估计和一致性检验的人为模糊检测方法, 该方法利用模糊曲线本身和线性拟合(Linear fitting)后的模糊曲线之间的差值来进行检测。

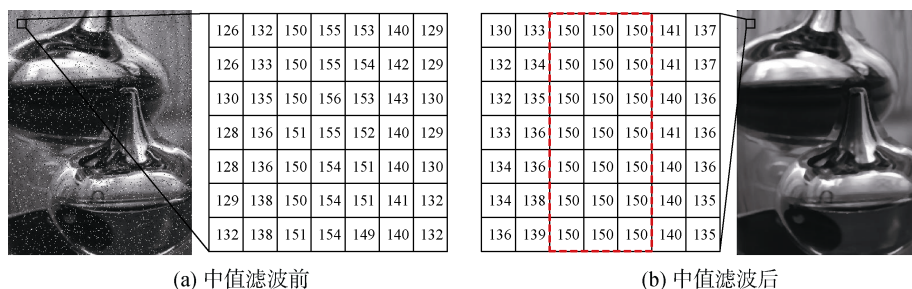
在文献[91]中, 邻接像素的相似度被用作权重去调节局部熵, 进而利用加权局部熵(Weighted local entropy, WLE)来检测人为模糊的边缘。文献[92]首先使用两维的非下采样轮廓波系数(Non-subsampled contourlet coefficients)特征和四维的相位一致性

(Phase congruency)特征共计六维特征来进行模糊检测, 进而利用局部特征来判断该模糊属于人为模糊还是自然的离焦模糊。实验表明该方法能有效检测到人为模糊痕迹。

基于人为模糊的检测方法需要有效区分人为模糊和各种自然模糊, 当一片区域受到多种模糊的影响时, 现存方法还难以有效地对不同类型的模糊进行有效识别和区分。

4.3 基于中值滤波痕迹的检测方法

中值滤波是一种常见的图像后处理方式, 它可以在保留图像原有的边缘的同时对图像进行去噪处理, 如图 11 所示。图 11(a)中的图像经过 5×5 的中值滤波后, 其结果如图 11(b)所示, 我们可以看到此时噪声对图像的影响已经大大降低, 而图像的细节大致保存完好。由于中值滤波是一种非线性的操作, 它可以抹去线性操作留下的痕迹, 让针对这些线性操作的取证手段无法发挥作用。所以说, 有关中值滤波的检测方法也是图像取证领域非常重要的一部分。



(a) 中值滤波前

(b) 中值滤波后

图 11 中值滤波的去噪效果以及条纹效应

Figure 11 Denosing effect and streaking artifact of median filtering

文献[93]介绍了在中值滤波图像中存在的条纹效应(Streaking artifact): 由于中值滤波图像中所有的像素值均和滤波前图像的某一像素值对应, 且由于相邻像素的滤波窗彼此有大部分的重叠区域, 在中值滤波图像中会出现一些像素值相同或相近的图像块, 这种现象就是条纹效应, 如图 11(b)所示。我们可以看到, 图 11(a)中的图像经过中值滤波后, 在图 11(b)中出现了大片像素值相同的区域。

文献[94]指出, 基于像素一阶残差的特征可以对未经 JPEG 压缩的图像进行有效检测, 但却无法有效检测经过 JPEG 压缩的图像。而一种在隐写分析领域中强有力的特征——减性像素邻接矩阵(Subtractive pixel adjacency matrix, SPAM)特征却可以有效检测出经过 JPEG 压缩的图像中存在的的中值滤波。

文献[95]指出, 在计算中值滤波的过程中相邻像

素的滤波模板存在许多相同的值, 这会给滤波后的相邻像素带来较强的相关性。为了捕捉这种相关性, 文献[95]提出了 44 维的中值滤波特征(Median filtering feature, MFF)。MFF 在检测高品质因数 JPEG 图像的性能上和 SPAM^[94]相差不大, 但在低品质因数的 JPEG 图像上 MFF 的检测效果要明显好于 SPAM。另外 MFF 的维度也远低于 SPAM, 能使计算复杂度大大降低。

文献[96]使用二阶局部三元模式(Local ternary pattern, LTP)来检测中值滤波在图像中留下的痕迹, 该特征能够有效反映出中值滤波所带来的纹理变化。除此以外文献[96]使用了核主成分分析(Kernel principal component analysis, KPCA)来降低特征维度以减小计算复杂度。实验表明, 文献[96]中的方法在检测准确率和计算效率上都要好于 SPAM^[94]和

MFF^[95]。

文献[97]中使用在纹理丰富的区域上统计的一阶残差图中 0 值出现的概率, 来进行中值滤波检测。该方法在图像经过一定程度后处理操作的情况下也能够对中值滤波进行有效检测; 但在图像被严重的均值滤波或高斯滤波处理后, 该方法的检测效果欠佳。文献[98]使用基于边缘的预测矩阵(Edge based prediction matrix, EBPM)来进行中值滤波的检测。中值滤波对噪声的压制特性、对边缘的良好保存特性以及给邻域像素带来的相关性, 都可以在 EBPM 中良好体现。文献[99]详尽分析了中值滤波图像在不同作用域下的统计特征, 并使用局部相关特征(Local correlation features, LCF)和全局概率特征(Global probability features, GPF)以及两者结合而来的 GLF 特征, 来检测中值滤波。文献[100]做了更多扩展实验来验证 GLF 特征的有效性。实验表明, GLF 在对 JPEG 压缩图像的检测准确率高于 SPAM^[94]和 MFF^[95]。

文献[101]提出了中值滤波残差的概念(Median filter residual, MFR), MFR 是指原图像和滤波处理后图像之间的差值。该文献使用自动回归(Auto regressive, AR)模型来对 MFR 进行拟合, 以获取 MFR 的统计特性, 进而得到的 AR 模型系数。MFR 可以有效检测到图像中的中值滤波痕迹, 甚至在 JPEG 品质因数低至 30 的图像中还能有良好表现。文献[102]设计了一套神经网络来检测图像中存在的中值滤波, 该神经网络中的第一层用来提取 MFR, 之后的卷积层和池化层来对深度特征进行提取。

为了解决在小尺寸图像上检测效果不佳的问题, 文献[103]使用最近邻插值(Nearest neighbor interpolation)来放大和中值滤波图像及其原图像之间的差值。另外, 该文设计了一种网络 MFNet, 该网络包含一种全新的卷积层——Mlp 卷积层。相比于普通的卷积层, Mlp 卷积层可以增强网络的非线性能力, 这和中值滤波的非线性特质非常吻合。实验表明, MFNet 的检测准确率要明显高于文献[102]中使用的网络, 另外 MFNet 还能对尺寸小至 16*16 的图像进行有效检测。

文献[104]利用局部二值模式(Local binary pattern, LBP)和像素差分矩阵(Pixel difference matrix, PDM)来进行检测。LBP 可以对微特征的统计特性进行量化, PDM 则可以有效反映中值滤波对图像像素产生的作用。文献[105]利用多方向的中值滤波残差差分(Median filtering residual difference, MFRD)和

AR 模型, 来检测图像中存在中值滤波痕迹, 该方法对小图和 JPEG 压缩后的图像也能有效检测。

相比于其他方法, 基于中值滤波痕迹的检测方法与深度学习结合较为紧密, 已能实现端到端的检测。结合深度神经网络, 目前基于中值滤波痕迹的检测方法, 能对尺寸较小的、经过有损压缩和包含噪声的图像进行有效检测。未来可以对网络结构继续优化, 以进一步提升检测准确率。

4.4 基于特征匹配的检测方法

基于特征匹配的检测方法主要针对复制粘贴这种特殊的篡改方式。复制粘贴篡改是将一张图的某物体复制并移动到该图的另一个地方。基于特征匹配的方法首先在图像中各区域提取出局部特征, 然后逐一匹配与该局部特征相似的其他局部特征, 以发现篡改图像中存在的相似区域。在图像被复制粘贴篡改的过程中, 被篡改的物体可能会被旋转、缩放、镜像、模糊等操作所处理, 这将会大大增加检测的难度。

目前主流的复制粘贴篡改检测方法分为两个子类: 基于图像块匹配的检测方法和基于关键点匹配的检测方法。基于图像块匹配的检测方法对于图像块提取特征进行匹配, 而基于关键点匹配的检测方法对于图像中存在的关键点提取特征进行匹配, 这是两种方法的最大不同。基于块匹配的检测方法和基于关键点的检测方法流程图如图 12 所示, 其流程大致如下: (1)首先是预处理操作, 很多方法需要在灰度图上进行检测, 对于这些方法我们首先要将彩色图转化为灰度图; (2)对于基于图像块匹配的检测方法, 要将图像分割成重叠的图像块, 而对于基于关键点匹配的检测方法, 要在图像中提取出关键点; (3)在提取出的图像块或者关键点中提取特征; (4)对提取的特征按照相似度进行排序, 找出与每个特征最相似的其它特征; (5)最后, 根据得到特征的相似性关系, 来进行篡改检测与定位。

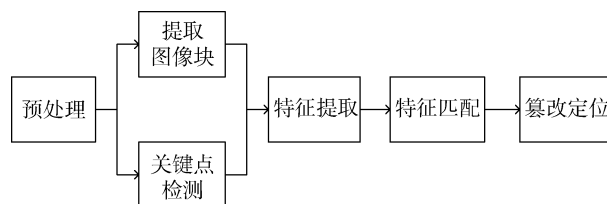


图 12 基于块匹配的方法、基于关键点匹配的复制粘贴篡改检测方法流程图^[107]

Figure 12 Flow chart of block-based and keypoint-based copy-move tampering detection methods^[107]

最近, 文献[106]提出一种基于深度神经网络的

检测方法, 该方法可以自动对输入图像进行检测, 不需要手工设计特征, 且该方法在检测准确率和检测效率比传统的两种方法更有优势。

4.4.1 基于块匹配的检测算法

文献[108]中在图像块中提取 DCT 特征来进行匹配, 并利用字典排序来提高特征匹配的效率。基于文献[108]中的方法, 文献[109]使用主成分分析(Principal component analysis, PCA)来降低图像块所提特征的维度。文献[110]提出了一种改进的 DCT 特征, 该特征比文献[108]中所用特征的复杂度更低, 另外该特征在图像经过加性高斯白噪声 AWGN(Additive White Gaussian Noise)、高斯模糊和 JPEG 压缩的处理后也有不错的检测效果。文献[111]使用二维傅里叶变换(Two dimensional Fourier Transform, 2D-FT)来提取图像块的特征, 该方法解决了文献[110]中不能对多次复制的粘贴篡改(Multiple copy move forgery)进行有效检测的问题。

文献[112]使用计数布隆过滤器(Counting bloom filters)来代替文献[108-109]中使用的字典排序, 以提高计算效率。除此之外, 文献[112]使用对数极坐标系(Log-polar coordinate)下的傅里叶-梅林变换(Fourier-Mellin transform, FMT)来对图像块提取特征, 所得 FMT 特征对加性噪声、有损压缩和模糊等操作都很鲁棒, 并对平移和尺度变化均具有不变性。观察到 FMT 特征无法对较大角度旋转变换后的物体进行有效检测, 文献[113]在 FMT 的基础上使用向量腐蚀滤波器(Vector erosion filter)来对距离向量进行聚合, 该方法能够对经过较大角度旋转变换的篡改物体进行有效检测。除此之外, 文献[113]还使用一种基于连接的处理方式(Link processing), 来取代计数布隆过滤器中所用的哈希值计数(Hash value counting)方法, 以提高计算效率。

文献[114]从图像块中提取出模糊瞬间不变性(Blur moment invariants)特征来进行检测。相比于文献[108]和文献[109]中的方法, 模糊瞬间不变性特征能对被模糊处理过或有较强对比变化(Contrast change)的图像有较为理想检测效果。另外, 文献[114]还使用主成分变换(Principal component transformation, PCT)来降低特征的维度。

FMT 特征^[112]无法对经过较大角度旋转变换的物体进行有效检测。在文献[115]中, 作者使用泽尔尼克瞬时(Zernike moments)特征来进行检测, 该特征能对一定程度的 JPEG 压缩、模糊和高斯加性白噪声具有鲁棒性, 且在篡改物体经受较大角度旋转的情况下依然会有良好的检测效果, 但该特征在对于经

过仿射变换(Affine transform)的篡改物体检测效果欠佳。块匹配(Patch match)算法是一种近似最近邻搜索算法, 可以用来降低特征匹配环节的计算复杂度。文献[116]利用块匹配算法进行检测, 大大提升了检测速度。

文献[117]结合快速傅里叶变换(Fast Fourier transform, FFT)、奇异值分解(Singular value decomposition, SVD)和主成分分析(Principal component analysis, PCA)来提取特征。相比于文献[110-111]所用方法需要对多个参数进行调节, 文献[117]中的方法不用对任何参数进行调节就可以达到很高的检测准确率。实验表明, 该方法对于经过噪声和模糊处理后的图像也有很好的检测效果。

4.4.2 基于关键点匹配的检测算法

尺度不变特征变换(Scale invariant feature transform, SIFT)特征是一种在图像处理领域有着广泛应用的特征, 该特征有以下优点: (1)对缩放、旋转等几何变换具有不变性, 且对噪声、光照变化也很鲁棒; (2)表达力丰富, 提出的特征具有很强的分辨力。

文献[118]使用 SIFT 特征来对复制粘贴篡改进行检测。相比于块匹配的方法, 该方法对旋转、缩放等几何变换和 JPEG 压缩、加性噪声都非常鲁棒。但使用文献[118]中的方法只能得到匹配的特征点, 而使用文献[119]中的方法不但可以估计出被篡改所经过的变换, 还能利用相关图恢复出篡改区域完整的边界线。但文献[119]的方法不能对多次复制粘贴篡改(Multiple copy move forgery)进行有效检测, 而文献[120]中的方法可以处理多个复制-粘贴的篡改, 并能够有效地估计出篡改物体所经过几何变换的参数。文献[121]利用 SIFT 特征在图像中提取关键点, 并使用中心对称局部二值模式(Center symmetric-local binary pattern, CS-LBP)来描述提取的关键点, 再使用最优节点优先(Best-bin-first, BBF)策略和 k-d 树(K-dimensional tree)来对提取的特征进行匹配。实验表明, 该方法准确率与文献[118]中的方法相当且计算耗时远低于方法[118]中的方法。

加速稳健特征(Speeded up robust features, SURF)是另一种常见的关键点描述算子, SURF 是对 SIFT 特征的改进, 相比 SIFT 计算复杂度更低。文献[122]和文献[123]分别提出了一种利用 SURF 特征进行检测的方法。文献[122]中的方法利用 k-d 树对特征进行排序来提升特征匹配的效率, 但该方法对小区域的篡改检测效果不佳。文献[123]中的方法无法自动定位出篡改的区域。文献[124]利用 SURF 特征和聚合

式层次聚类(Hierarchical agglomerative clustering, HAC)将匹配到的关键点聚合成区域。实验表明 HAC 能大大减小区域聚合所需的时间, 然而所创建区域的准确率还有待提高。

ORB(Oriented FAST and rotated BRIEF)可以快速地特征点进行描述和提取, 它的计算速度远远超过 SIFT 和 SURF。在文献[125]中, 作者使用经过缩放的 ORB(Scaled ORB)特征来描述特征点, 该特征对缩放、旋转等几何变换和 JPEG 压缩、高斯模糊、高斯白噪声等后处理都很鲁棒。实验表明, 文献[125]所用的方法的检测性能和计算速度都优于文献[120]和文献[123]中的方法。

4.4.3 基于深度神经网络检测算法

在文献[106]中, 作者设计了一个名为 BusterNet 的二分支卷积神经网络来对复制粘贴篡改进行检测, 两个分支分别为相似检测分支和篡改检测分支。其中相似检测分支利用文献[126]中的匹配模块来对图中相似的区域进行定位; 而篡改检测分支可以看成是一个特殊的分割网络, 来对复制粘贴篡改中的被篡改区域进行定位; 最后的融合模块用来融合两个分支的检测结果。不同于其他非深度学习的方法, 该方法可以端到端地对复制粘贴篡改进行检测和定位。另外, 该方法首次能够区分复制粘贴篡改的源区域和篡改后的区域。大量实验表明该方法的检测准确率和效率相比于非深度学习的方法有很大提高, 且该方法在 JPEG 压缩、加性噪声、模糊等环境下的鲁棒性也优于传统方法。

基于块匹配的方法存在计算复杂度高, 对于几何变换不鲁棒的缺点。而基于关键点匹配的方法对于存在过多平滑区域或者相似区域的图像检测效果不佳。最新出现的基于深度学习的方法可以自动实现端到端的网络, 且实验表明该方法的检测准确率和计算效率都优于非深度学习的方法。未来可进一步优化深度神经网络结构以提升检测的速度和准确率。

5 基于 JPEG 重压缩痕迹的检测方法

JPEG 压缩能使图片在占用很少空间的前提下拥有较高的质量, 因此是目前最流行的图像传输和存储格式。图像在篡改过程中可能会涉及到多次 JPEG 压缩, 而 JPEG 重压缩会在图像中留下痕迹, 可以根据这些痕迹去进行篡改检测。

利用 JPEG 重压缩痕迹进行篡改检测的方法分为两大类: 基于对齐 JPEG 重压缩假设的检测 (Aligned double JPEG, A-DJPG)方法和基于非对齐

JPEG 重压缩假设的检测(Nonaligned double JPEG, NA-DJPG)方法。A-DJPG 代表两次 JPEG 压缩的网格完全对齐, 如图 13(a)所示, NA-DJPG 代表两次 JPEG 压缩的网格有偏差, 如图 13(b)所示。

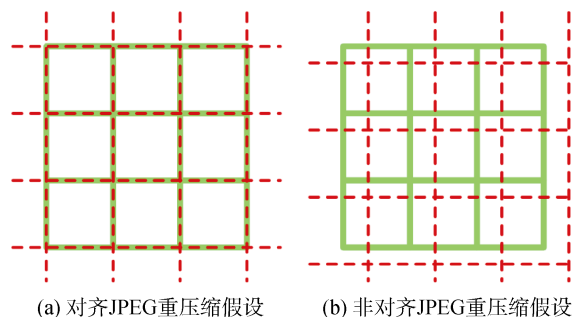


图 13 两种 JPEG 重压缩假设
Figure 13 Two JPEG recompression hypotheses

我们借助图 14 来描述与 JPEG 重压缩痕迹检测有关的图像篡改方法: 假设我们将图 14 中 A 图像中的斑马拼接接到 B 图像上, 成为篡改图像 C。在篡改图像 C 中, 无论是真实区域还是篡改的斑马区域, 都会经历两次重压缩。A-DJPG 假设图 C 中的斑马区域在篡改的过程中, 会经历旋转、缩放变换或模糊等操作, 以至于第一次 JPEG 压缩的痕迹已经很难被检测到。这时我们认为图像 C 中真实的区域会呈现出两次 JPEG 压缩的痕迹, 而篡改的斑马区域只有一次 JPEG 压缩的痕迹。NA-DJPG 假设斑马区域在从图像 A 移到图像 C 的后, 计算 DCT 系数的 8*8 网格没有对齐, 此时 C 中篡改的斑马区域会经过两次网格未对齐的 JPEG 压缩, 而 C 中真实的区域会经历两次网格彼此对齐的 JPEG 压缩。

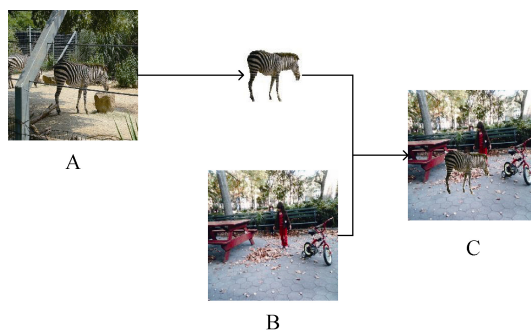


图 14 图像拼接篡改的流程图
Figure 14 Flow chart of image splicing

总之, 在 JPEG 图像的篡改的过程中, 真实区域和篡改区域由于经历的压缩过程有差异而导致 DCT 系数呈现不同的分布, 我们可以利用这一特性来进行篡改检测。

本节把篡改检测方法分为两个分类: 基于对齐 JPEG 重压缩假设的检测方法和基于非对齐 JPEG 重压缩假设的检测方法。

5.1 基于对齐 JPEG 重压缩假设的检测方法

文献[127-128]属于 A-JPEG 前瞻性的工作。在这两篇文献中, 作者介绍了图像经过双重 JPEG 压缩后产生的 DQ(Double quantization)效应出现的原因和表现。只经历一次 JPEG 压缩的图像将会服从拉普拉斯分布(或广义柯西分布), 该分布在 0 处为峰值, 在其左右两侧单调下降。而如果图像经过两次 JPEG 压缩, 且两次压缩的品质因数 Q1 与 Q2 处于某种比例关系时, 其 DCT 系数将会周期性地呈现谷值和峰值, 这种现象称为 DQ 现象。

文献[129]利用文献[127-128]中的原理, 结合 3 通道的总共 192 个 DCT 分布直方图和每个 8*8 图像块的 DCT 信息, 使用贝叶斯推断(Bayesian inference)来计算图像块的篡改后验概率图(Block posterior probability map, BPPM)。最后, 将从整张图像的篡改概率图中提取的特征放入 SVM, 来判断该图像是否被篡改过。在文献[129]中, 图像块的篡改概率只考虑相对于真实图像的量化 DCT 分布。文献[130]认为该概率应同时考虑相对于真实图像和篡改图像的量化 DCT 分布, 因此对篡改概率的计算方式做出了调整。实验结果表明, 文献[130]所用方法的检测效果要明显优于文献[129]。文献[131]是在文献[130]的基础上, 对篡改概率图使用粒子群算法(Particle swarm optimization, PSO)进行优化, 从而使得图像中篡改区域和真实区域能够自动划分出来, 进一步提高了检测准确率。

本福特法则(Benford's law)描述的是在自然数据中以 0~9 为首的数出现的概率, 它可以用来检验统计数字是否是伪造的。文献[132]指出, 如果一张图像只经过了一次 JPEG 压缩, 则其 DCT 系数的分布符合本福特法则; 而如果该图像经过了多次压缩, 则该法则会被破坏。文献[133]中指出, 不是所有 JPEG 压缩的交流(Alternate current, AC)系数都符合本福特法则, 于是使用了一种基于单个交流系数值的模式基首位特征(Mode based first digit features, MBFDF)来代替文献[132]中使用的全局首位特征(Global first digit features, GFDF), 来进行 JPEG 重压缩检测。实验表明, 文献[133]中所用方法的检测准确率要明显高于文献[132]中的方法。此外, MBFDF 还可以对双重 JPEG 压缩中的第一次压缩系数进行估计。文献[134]指出, JPEG 重压缩将会破坏信号直方图的连续性并留下周期性的痕迹, 文献设计了 3

种特征来对 JPEG 重压缩进行检测, 最后使用非线性的 SVD 分类器和线性的 Fisher 线性判别(Fisher linear discriminant, FLD)分类器, 对特征进行分类。实验表明, 文献[134]中使用的特征要优于文献[132]中的特征。

以上所述方法只在两次 JPEG 压缩质量因子不相同的情况下有效。在 JPEG 重压缩中, 如果两次压缩的质量因子相同, 需要用其他方法才能进行有效检测。文献[135]和文献[136]分别提出了专门针对同质量因子的 JPEG 重压缩检测的方法。文献[135]认为: (1)在 JPEG 编码的过程中, 会在三个环节给图像带来误差, 即量化环节、截断环节和取整环节; 而一张图像经过相同质量因子的一次和两次压缩后的 DCT 系数有所不同, 这正是由这三个环节引入的误差所导致的; (2)在同质量因子的情况下, $N+1$ 次 JPEG 压缩和 N 次压缩 DCT 系数差距要大于 N 次和 $N-1$ 次的差距。基于以上两点观点, 文献[135]设计了一种基于随机抖动策略(Random perturbation strategy)的方法来区分同质量因子的情况下经过 2 次和 1 次 JPEG 重压缩的图像, 该方法还可以对 3 次甚至 4 次 JPEG 重压缩的图像进行检测。文献[136]首先构造出 JPEG 图像误差块(Error blocks), 然后从经过取整(Rounding)和截断(Truncating)的误差块中, 提取出 13 维的基于误差的统计特征(Error based statistical features, EBSF)而来进行检测。文献[136]所用方法体现了取整和截断误差块的幅度信息, 在 3 个数据集上的实验表明, 该方法的检测效果比文献[135]中的方法效果更好。

5.2 基于非对齐 JPEG 重压缩假设的检测方法

文献[137]使用块效应特征矩阵(Blocking artifact characteristics matrix, BACM)来对重压缩图像进行检测。该文献指出, 在经过一次 JPEG 压缩的图像中, BACM 是对称的, 而 NA-DJPG 会破坏掉这种对称性。

文献[138]指出, BACM 容易受到图像语义信息的干扰, 因此不能当作区分 NA-DJPG 的可靠的度量标准, 该文提出使用块效应周期性特征(Periodic blocking artifacts)来检测 NA-DJPG 的思路。块效应周期性特征提取自相邻 DCT 块量化误差的差值, 比 BACM 更少受到语义信息的干扰。文献[139]详尽分析了 JPEG 图像在空域和变换域的周期性特质, 并使用空域和变换域的联合特征来对 JPEG 重压缩进行检测。值得说明的是, 该方法对 A-DJPG 和 NA-DJPG 均可以进行检测。相比于块效应周期性特征^[138], 文献[139]提出的特征更不容易受到图像语义

信息的影响,也能深入挖掘 JPEG 图像的块效应特质。

文献[140]在文献[137]的基础上,利用一种新的噪声卷积模型来对 (Noisy convolutive mixing mode)NA-DJPG 进行建模。文献[140]指出, NA-DJPG 会破坏块 DCT 之间的独立性,因此设计了一种有关独立成分分析(Independent component analysis, ICA)的方法进行检测。实验表明,文献[140]所用方法的检测准确率要明显高于 BACM^[137]。文献[141]指出,如果按照第一次 JPEG 压缩的网格对 NA-DJPG 图像的块 DCT 系数进行计算,该 DCT 系数会呈现出整数周期性(Integer periodicity)。基于以上观察,该文设计了一种检测方法:从 DCT 系数中提取出特征后,只需要设置一个简单的阈值检测器,就可以对图像进行 NA-DJPG 检测,而不用像文献[137-140]那样需要使用分类器对复杂的特征进行分类。

文献[142]提出一种新的统计模型,使用该模型可以对图像中的 A-DJPG 和 NA-DJPG 进行建模,并可以计算出每个 8*8 DCT 块被重压缩的概率,从而实现篡改区域的定位。文献[143]详尽分析了 NA-DJPG 对 DCT 系数的影响,并提出一种能模拟篡改图像中 NA-DJPG 的模型,使用该模型可以有效估计 NA-DJPG 带来的量化噪声,进而实现对 NA-DJPG 的检测与定位。实验表明,文献[143]所用方法的检测准确率要高于文献[141]和文献[142]中方法的检测准确率。

基于 JPEG 重压缩痕迹的检测方法(对齐假设与未对齐假设)的一个缺陷,是对互联网中存在的其它格式图像无法检测。此外,这类方法只有当第一次压缩的品质因数 Q1 和第二次压缩的品质因数 Q2 满足一定比例的时候才较为有效。

6 篡改检测方法评价标准与数据集分析

在第 2~5 节关于篡改检测方法技术细节描述的基础上,为了全面展示有关检测方法的性能,本节描述与篡改检测方法的有关的评价标准和公开的数据集情况,也通过技术指标分析对典型方法的性能进行对比。

6.1 篡改检测方法评价标准

检测方法评价标准需要使用样本的归类量,包括真正类、假负类、真负类和假正类。对于一个二分类问题,依据真实标签将样本分为正负两类。如果一个正样本被分为正类,则称为真正类(True positive, *TP*);如果一个正样本被分为负类,则称为

假负类(False negative, *FN*);如果一个负样本被分为负类,则称为真负类(True negative, *TN*);如果一个负样本被分为正类,则称为假正类(False positive, *FP*)。

表 1 描述了篡改检测方法涉及的 15 个评价标准,包括正阳性率(True positive rate, *TPR*)、假阳性率(False positive rate, *FPR*)、正确率(Accuracy, *ACC*)、漏检率(Missed Detection, *MD*)、虚警率(False Alarm, *FA*)、接受操作特性曲线(Receiver Operating Characteristic Curve, *ROC*)、曲线下面积(Area Under Curve, *AUC*)、准确率(Precision, *P*)、召回率(Recall, *R*)和 *F1* 值(*F1*, *F1Score*)、*PR* 曲线(*PR Curve*)、平均精度(Average Precision, *AP*)、平均精度均值(Mean Average Precision, *mAP*)、交并比(Intersection over union, *IoU*)和马修斯相关系数(Matthews correlation coefficient, *MCC*)的标准及其使用说明。

6.2 篡改检测方法公开数据集分析

本节通过表 2 对如下 23 个篡改检测方法公开数据集进行分析,介绍了每个数据集的名称、数据集适用方法、数据集内部细节和获取方式。

6.3 篡改检测技术指标分析

如下 10 个小节描述了 12 个分类方法中 10 个分类的技术指标分析。

6.3.1 “基于光照一致性检测方法”的技术指标分析

表 3 展示的是在 MultiPIE 数据集中阴影突出度(shading prominence) α 为各值时 3 种算法的 AUC,在该表中我们可以看出, L3DMM^[17]的性能要明显优于 Full 3D Environment^[14]和 SFS^[15]。

6.3.2 “基于特征提取与分类检测方法”的技术指标分析

表 4 展示的是在 CASIA V2.0 数据集中 3 种算法的特征维度和检测准确率、特征提取时间。在表中我们可以看出虽然算法 Markov in QDCT^[30]的特征维度较高,但是相比其他两种算法,该算法的检测准确率高且特征提取时间短。

表 5 展示的是在 DVMM 数据集中 4 种算法的特征维度和检测准确率,在表中我们可以看出算法 Markov in DCT & CT^[31]的维度远远低于 2-D Noncausal Markov^[29],且在四种算法中检测准确率较高

表 6 展示的是在 DVMM 数据集中三种算法的检测准确率,在表中我们可以看出 SRM+10layer CNN^[36]的检测准确率最高。

表 1 篡改检测方法评价标准

Table 1 Evaluation metrics of tampering detection methods

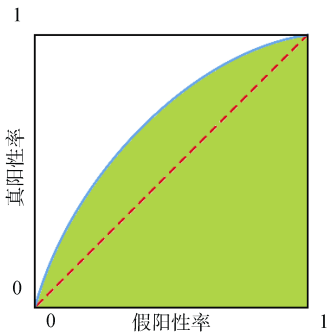
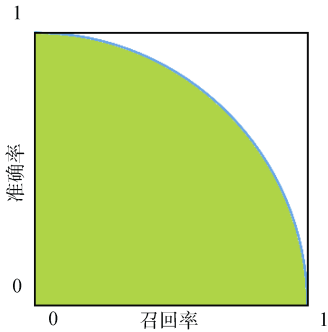
评价标准:	计算方式	使用说明
正阳性率(True positive rate, TPR), 假阳性率 FPR (False positive rate, FPR)	$TPR = \frac{TP}{TP + FN} \quad (1)$ $FPR = \frac{FP}{FP + TN} \quad (2)$	<p>TPR 代表所有正样本中被正确分类的比例; FPR 代表所有的负样本中被错误分类的比例。 TPR 越高代表分类器性能越强, FPR 越低代表分类器性能越强。 TPR 和 FPR 的值在 0 与 1 之间。</p>
正确率(Accuracy, ACC)	$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$	<p>ACC 代表所有样本分类的正确率。 ACC 越高代表分类器的性能越强。 ACC 的值在 0 与 1 之间。该度量标准不太适用于两类样本不均衡的情况。</p>
漏检率(Missed Detection, MD) 虚警率(False Alarm, FA)	$MD = \frac{FN}{TP + FN} \quad (4)$ $FA = \frac{FP}{TN + FP} \quad (5)$	<p>MD 代表正样本被误判的比例, 而 FA 代表负样本被误判的比例。 MD、FA 越低代表分类器性能越强。 MD 和 FA 的值均在 0 与 1 之间。</p>
接受操作特性曲线(Receiver Operating Characteristic Curve, ROC), 曲线下面积(Area Under Curve, AUC)	 <p>图 15 ROC 曲线 Figure 15 ROC curve</p>	<p>ROC 和 AUC 反映了一个二分类器的分类性能, 可以用来解决 ACC、TPR、FPR 等指标的单个局限, 从而更全面地对分类器的性能进行度量。如图 15 所示, 蓝色的曲线就是一条 ROC, 而绿色区域的面积就是该 ROC 对应的 AUC, 红色虚线代表随机分类器的 ROC。 AUC 面积越大, 代表一个分类器的分类性能越好。 AUC 的值在 0 和 1 之间。</p>
准确率(Precision, P)、召回率(Recall, R)和 $F1$ 值($F1$, Score)	$P = \frac{TP}{TP + FP} \quad (6)$ $R = \frac{TP}{TP + FN} \quad (7)$ $F1 = \frac{2 * P * R}{P + R} = \frac{2 * TP}{2 * TP + FN + FP} \quad (8)$	<p>$F1$ 值常用来度量在各类别样本不均衡时一个分类器的性能。 $F1$ 值越高, 代表一个分类器的分类性能越好。 $F1$ 取值在 0 到 1 之间。</p>
PR 曲线(PR Curve), 平均精度(Average Precision, AP), 平均精度均值(Mean Average Precision, mAP)	 <p>图 16 PR 曲线 Figure 16 PR curve</p>	<p>如图 16 所示, 蓝色曲线为一个分类器的 PR 曲线, 该曲线的横坐标为召回率, 纵坐标为准确率。而 AP 指的是绿色区域的面积, mAP 指的是所有类别 AP 的平均值。当各类别的样本比较均衡时, AP 与 AUC 差别不大; 当各类别样本不均衡时, AP 比 AUC 更能体现出一个分类器的性能。另外, AP 度量很好地解决了 $F1$ 值度量的单个局限问题。 AP、mAP 取值均在 0 到 1 之间。</p>
交并比(Intersection over union, IoU)	$IoU = \frac{TP}{FP + TP + FN} \quad (9)$	<p>IoU 值常用来度量在样本不均衡时一个分类器的性能。 IoU 值越高, 代表一个分类器的分类性能越好。 IoU 常用来度量图像篡改定位的精度。 IoU 取值在 0~1 之间。</p>
马修斯相关系数(Matthews correlation coefficient, MCC)	$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (10)$	<p>MCC 用来描述预测分布和实际分布之间的相关性, 常用来度量在各类别样本不均衡时一个分类器的性能。 MCC 值越高, 代表一个分类器的分类性能越好。 MCC 取值在 -1~1 之间。</p>

表 2 篡改检测方法公开数据集

Table 2 Public datasets of tampering detection methods

数据集名称	数据集适用方法	数据集介绍	数据集获取方式
DSO-1 ^[144]	2.1 基于光照一致性的检测方法	200 张分辨率为 2014*1536 的室内和室外场景图像。其中真实图像和篡改图像各 100 张。	https://recodbr.wordpress.com/cod-e-n-data/#dso1_dsi1
DSI-1 ^[144]	2.1 基于光照一致性的检测方法	真实图像和篡改图像各 25 张。	https://recodbr.wordpress.com/cod-e-n-data/#dso1_dsi1
Yale B ^[145]	2.1 基于光照一致性的检测方法	45 种光照条件, 9 种头部姿态, 10 个人, 总共 4050 张人脸图像。	http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
Multi-PIE ^[146]	2.1 基于光照一致性的检测方法	包括 337 个人物, 从 15 种视角以及 19 种光照条件下进行拍摄。	http://multipie.org
DVMM ^[147]	2.2 基于特征提取与分类的检测方法/4.4 基于特征匹配的检测方法	933 张真实的和 912 张拼接或复制粘贴篡改的图像, 图像尺寸为 128*128	http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm
CASIA V1.0 ^[148]	2.2 基于特征提取与分类的检测方法/4.4 基于特征匹配的检测方法	800 张真实的和 921 张拼接的或者复制粘贴的篡改图像, 所有图像均为 JPEG 压缩图像, 尺寸为 374*256。	https://www.kaggle.com/sophatvathana/casia-dataset
CASIA V2.0 ^[148]	2.2 基于特征提取与分类的检测方法/4.4 基于特征匹配的检测方法	7491 张真实的和 5123 张拼接或复制粘贴篡改图像, 大部分图像为 JPEG 压缩图像, 少量为 BMP 和 TIFF 格式。	https://www.kaggle.com/sophatvathana/casia-dataset
COVER ^[149]	2.2 基于特征提取与分类的检测方法	共 200 张图像, 100 张真实的和 100 张经过复制粘贴篡改的图像。数据集中提供了篡改图像的遮罩(mask)。	https://github.com/wenbihan/coverage
NIST16 ^[150]	2.2 基于特征提取与分类的检测方法	含有三种篡改操作: 拼接、复制粘贴和移除的高清数据集, 数据集中提供了篡改区域的遮罩(mask)。	https://www.nist.gov/itl/iad/mig/nim-ble-challenge-2017-evaluation/
Columbia Un-compressed ^[151]	2.2 基于特征提取与分类的检测方法/4.2 基于人为模糊痕迹的检测方法	363 张尺寸在 757*568 到 1152*768 之间的拼接数据集, 照片来自于四个相机, 数据集的篡改区域遮罩(mask)已经给出。	http://www.ee.columbia.edu/ln/dvmm/downloads/authspluncmp/
Carvalho ^[144]	2.2 基于特征提取与分类的检测方法	200 张 png 格式的图像, 其中真实图像和篡改图像分别有 100 张	http://www.ic.unicamp.br/~tjose/files/database-tifs-small-resolution.zip
Dresden Image ^[152]	3.1 基于成像色差印记一致性的检测方法/4.3 基于中值滤波痕迹的检测方法	由 73 个相机拍摄得到的 14000 多张图像, 包括户内和户外场景。	http://forensics.inf.tu-dresden.de/dresden_image_database/
BOWS2 ^[153]	4.3 基于中值滤波痕迹的检测方法	10000 张尺寸为 512*512 的灰度图像。所有图像均是通过对自然图像的剪裁和缩放而得来的。	http://bows2.ec-lille.fr/
BOSS RAW ^[154]	4.3 基于中值滤波痕迹的检测方法	9000 张未经压缩的图像, 尺寸从 3008*2000 到 5212*3468	http://boss.gipsa-lab.grenoble-inp.fr/BOSSRank/index.php?mode=VIEW&tmpl=materials
BOSSbase1.01 ^[155]	4.3 基于中值滤波痕迹的检测方法	10000 张尺寸为 512*512 的图像	http://agents.fel.cvut.cz/stegodata/
UCID ^[156]	4.3 基于中值滤波痕迹的检测方法/5.1 基于对齐 JPEG 重压缩假设的检测方法/5.2 基于未对齐 JPEG 重压缩假设的检测方法	1338 张未经压缩的图像, 尺寸为 384*512 或者 512*384	见文献[156]
NRCS ^[157]	4.3 基于中值滤波痕迹的检测方法/5.1 基于对齐 JPEG 重压缩假设的检测方法/5.2 基于未对齐 JPEG 重压缩假设的检测方法	图像尺寸为 1500*2100, 图像格式为 JPEG 或 TIFF	http://photogallery.nrcs.usda.gov/res/sites/photogallery/
CoMoFod ^[158]	4.4 基于特征匹配的检测方法	该数据集包含 260 个图像集合, 每个集合有原图, 篡改图像和两个篡改遮罩(mask)。其中, 篡改区域被平移、旋转、缩放、组合、扭曲、加性噪声、JPEG 压缩、模糊等操作处理过。	http://www.vcl.fer.hr/comofod/download.html
Image Manipulation ^[107]	4.4 基于特征匹配的检测方法	48 张复制粘贴篡改的图像, 篡改区域经过 JPEG 压缩、旋转、缩放等操作。尺寸从 420*300 到 3888*2592。	https://www5.cs.fau.de/research/data/image-manipulation

续表

数据集名称	数据集适用方法	数据集介绍	数据集获取方式
MICC-F220 ^[120]	4.4 基于特征匹配的检测方法	110 张真实的和 110 张篡改的图像, 图像尺寸从 722*480 到 800*600。	http://www.micc.unifi.it/download/s/MICC-F220.zip
MICC-F600 ^[159]	4.4 基于特征匹配的检测方法	448 张真实的图像和 152 张篡改的, 图像尺寸从 800*533 到 3888*2592 不等。	http://www.micc.unifi.it/download/s/MICC-F600.zip
MICC-F2000 ^[120]	4.4 基于特征匹配的检测方法	1300 张真实图像和 700 张篡改图像, 尺寸为 2048*1536, 篡改区域经过了平移、旋转、缩放等操作。	http://www.micc.unifi.it/download/s/MICC-F2000.zip
SYSU/ OurLab ^[135]	5.1 基于对齐 JPEG 重压缩假设的检测方法	1128 张图像, 尺寸为 512*512	见文献[135]

表 3 基于光照一致性检测方法性能比较^[17]Table 3 Performance comparison of illumination consistency based detection method^[17]

检测算法	AUC		
	$\alpha=0.4$	$\alpha=0.6$	$\alpha=0.8$
Full 3D Environment ^[14]	0.870	0.893	0.910
SFS ^[15]	0.834	0.877	0.895
L3DMM ^[17]	0.921	0.956	0.966

表 4 基于特征提取与分类检测方法性能比较^{1[30]}Table 4 Performance comparison 1 of feature extraction and classification based detection method^[30]

检测算法	特征维度	检测准确率(%)	特征提取时间(s)
MBDCT ^[28]	266	84.86	4.479
Markov in DCT & DWT ^[26]	100	89.76	4.376
Markov in QDCT ^[30]	972	92.38	3.61

表 5 基于特征提取与分类检测方法性能比较^{2[31]}Table 5 Performance comparison 2 of feature extraction and classification based detection method^[31]

检测算法	特征维度	检测准确率(%)
MBDCT ^[28]	266	90.16
Markov in DCT & DWT ^[26]	100	93.53
2-D Noncausal Markov ^[29]	14240	93.36
Markov in DCT & CT ^[31]	200	94.10

表 6 基于特征提取与分类检测方法性能比较^{3[36]}Table 6 Performance comparison 3 of feature extraction and classification based detection method^[36]

检测算法	检测准确率(%)
2-D Noncausal Markov ^[29]	93.36
Markov in DCT & DWT ^[26]	93.55
SRM+10layer CNN ^[36]	96.38

6.3.3 “基于成像色差印记一致性检测方法”的技术指标分析

表7展示的是在文献[43]自建数据集中重采样因

表 7 基于成像色差印记一致性检测方法性能比较^[43]Table 7 Performance comparison of color aberration imprint consistency based method^[43]

检测算法	检测率		
	$P_{Fa}=0.01$	$P_{Fa}=0.05$	$P_{Fa}=0.10$
Angular Error ^[39]	0.10	0.38	0.57
Noise Model ^[40]	0.44	0.77	0.87
Diamond Search ^[43]	0.76	0.87	0.91

子 $u=5$ 的情况下虚警率 P_{Fa} 为各值时各算法的检测率, 在该表中我们可以看出在相同条件下, 算法 Diamond Search^[43]的检测率明显高于其它两种算法。

6.3.4 “基于自然模糊印记一致性检测方法”的技术指标分析

表8展示的是在文献[48]自建数据集中对于各尺寸图像块各算法的离焦模糊/运动模糊的检测准确率, 通过该表我们可以看出算法 MAP^[48]在各种情况下的检测准确率明显高于其它两种方法。

6.3.5 “基于彩色滤波阵列插值印记一致性检测方法”的技术指标分析

表9展示的是在文献[76]自建数据集中进行篡改检测各算法的最小决策错误率(Minimum decision error)和平均处理时间, 通过该表我们可以看出算法 Singh^[76]的最小决策错误率最低且运算的平均处理时间最少。

6.3.6 “基于几何变换与插值痕迹检测方法”的技术指标分析

表10展示的是在 Dresden Image 数据集里由 Nikon 相机拍摄的 1317 张图像中, 当图像块尺寸为 32*32, 插值核为各种不同类型时, 错误接受率(False Acceptance Rate, FAR)小于 1%情况下各算法的真阳性率(True Positive Rate, TPR), 在该表中我们可以看出算法 RMT^[87]在 4 种不同插值核的条件下真阳性率均高于其他 3 种算法。

表 8 基于自然模糊印记一致性检测方法性能比较^[48]Table 8 Performance comparison of natural blurring imprint consistency based method^[48]

检测算法	检测准确率(%)		
	图像块尺寸 64*64 无重叠	图像块尺寸 64*64 有重叠	图像块尺寸 128*128 有重叠
Directional HF ^[53]	81.7	81.8	79.0
SVD ^[54]	81.2	84.3	83.0
MAP ^[48]	92.8	94.3	91.3

表 9 基于彩色滤波阵列插值印记一致性检测方法性能比较^[76]Table 9 Performance comparison of color filter array interpolation imprint consistency based method^[76]

检测算法	最小决策错误率	平均处理时间(s)
DM ^[68]	0.3532	207.81
GC_L ^[71]	0.3103	205.03
GC_B ^[71]	0.3158	203.63
Ferrara ^[75]	0.2618	193
Singh ^[76]	0.1305	167.36

6.3.7 “基于中值滤波痕迹检测方法”的技术指标分析

表 11 展示的是, 在由 BOSSbase 1.01、UCID、BOSS RAW、Dresden Image、NRCS 构成的混合数

据集中, 32*32 图像块的不同 JPEG 品质因数和滤波核大小的情况下 4 种算法的检测准确率, 在该表中我们可以看到算法 MFR+CNN^[102]在各种条件下检测的准确率均优于其他 3 种算法。在表 11 中 QF 代表 JPEG 图像的品质因数, KS 代表滤波核大小。

6.3.8 “基于特征匹配检测方法”的技术指标分析

表 12 展示的是在 MICC-F220 数据集中各算法的检测假阳性率(False positive rate, *FPR*)、真阳性率(True positive rate, *TPR*)和每张图像检测时间, 在该表中我们可以看到算法 SURF+HAC^[124]算法相比于其他 3 种算法假阳性率低且检测耗时少, 但真阳性率过低; 而 SIFT^[120]算法假阳性率和检测效率与 SURF+HAC^[124]算法相差不大, 且 SIFT^[120]算法拥有很高的真阳性率。

表 10 基于几何变换与插值痕迹检测方法性能比较^[87]Table 10 Performance comparison of geometric transformation and interpolation trace based method^[87]

检测算法	真阳性率			
	Linear 插值核	Catmull-Rom 插值核	B-spline 插值核	Lanczos 插值核
Linear Predictor ^[83]	0.5882	0.6727	0.0226	0.4325
SVD ^[85]	0.8897	0.8296	0.9749	0.8067
Zeroing Mask ^[86]	0.9383	0.8032	0.9942	0.7873
RMT ^[87]	0.9946	0.9911	0.9966	0.9781

表 11 基于中值滤波痕迹检测方法性能比较^[102]Table 11 Performance comparison of median filter trace based method^[102]

检测算法	检测准确率(%)			
	QF=70, KS=3*3	QF=90, KS=3*3	QF=70, KS=5*5	QF=90, KS=5*5
MFF ^[95]	73.99	80.32	82.49	85.91
AR ^[101]	75.63	83.52	80.80	86.26
GLF ^[99]	78.15	85.43	87.28	91.57
MFR+CNN ^[102]	79.42	87.71	88.65	93.21

表 12 基于特征匹配检测方法性能比较 1^[124]Table 12 Performance comparison 1 of feature matching based method^[124]

检测算法	FPR(%)	TPR(%)	每张图像检测时间(s)
DCT ^[108]	84	89	294.69
PCA ^[109]	86	87	70.97
SIFT ^[120]	8	100	4.94
SURF+HAC ^[124]	3.64	73.64	2.85

表 13 基于特征匹配检测方法性能比较 2^[106]Table 13 Performance comparison 2 of feature matching based method^[106]

检测算法	检测准确率(%)	每张图像检测时间(s)
Zernike Moments ^[115]	39.08	5.11
Dense-Field ^[116]	46.82	1.78
BusterNet ^[106]	75.98	0.62

表 13 展示的是在 CASIA V2.0 数据集中 3 种算法的检测准确率和每张图像的检测时间, 在该表中我们可以看出算法 BusterNet^[106] 相比其它两种算法检测准确率高且耗时较少。

6.3.9 “基于对齐 JPEG 重压缩假设检测方法”的技术指标分析

表 14 展示的是在 UCID、NRCS、SYSU 3 个数

据集中, 当 JPEG 品质因数为 70、80 时两种同质量因子对齐 JPEG 重压缩假设检测方法的检测准确率。在该表中我们可以看出算法 EBSF^[136] 在 3 个数据集及两种品质因数 JPEG 图像中的检测准确率均要高于算法 Perturbation Strategy^[135]。在表 14 中 QF 代表 JPEG 图像的品质因数。

6.3.10 “基于未对齐 JPEG 重压缩假设检测方法”的技术指标分析

表 15 展示的是在 UCID、NRCS 构成的混合数据集中, 在第一次压缩品质因数 QF1 与第二次压缩品质因数 QF2 为各值时, 图像块大小为 128*128 时各算法的检测准确率, 在该表中我们能够看到算法 Adaptive DCT^[143] 在 QF1 和 QF2 为各值时检测准确率均高于其他 4 种算法。在表 14 中 QF1 和 QF2 分别代表第一次和第二次 JPEG 压缩的品质因数。

表 14 基于对齐 JPEG 重压缩假设检测方法性能比较^[136]Table 14 Performance comparison of aligned JPEG recompression hypothesis based method^[136]

检测算法	检测准确率(%)					
	数据集 UCID		数据集 NRCS		数据集 SYSU	
	QF=70	QF=80	QF=70	QF=80	QF=70	QF=80
Perturbation Strategy ^[135]	73.65	85.80	69.21	78.84	66.40	76.37
EBSF ^[136]	80.81	95.73	76.27	95.23	86.64	94.30

表 15 基于未对齐 JPEG 重压缩假设检测方法性能比较^[143]Table 15 Performance comparison of non-aligned JPEG recompression hypothesis based method^[143]

检测算法	检测准确率(%)			
	QF1=70		QF1=80	
	QF2=70	QF2=80	QF2=70	QF2=80
BACM ^[137]	52.64	53.76	50.11	50.99
IPM ^[141]	53.55	64.37	50.30	52.79
Block-Grained Analysis ^[142]	53.97	64.07	50.44	52.42
Periodicity Analysis ^[139]	51.57	54.28	50.05	51.00
Adaptive DCT ^[143]	72.26	95.48	59.82	71.59

7 图像篡改检测方法总结与展望

本文按照数字图像篡改检测技术所依赖的线索将各种方法分为两个方面, 进一步划分为 4 个分组, 更进一步划分为 12 个分类。基于 12 个分类图像篡改检测技术发展内容的分析, 本节对篡改检测方法进行总结及展望。

7.1 图像篡改检测方法总结

基于本文前面对 12 个分类中现有的算法的细致描述和分析, 下面将对各个分类方法涉及的技术进行总结。由于对齐假设与未对齐假设的基于 JPEG

重压缩痕迹的两个分类的相似性, 本文将这两种检测方法一起进行总结。

基于光照一致性的检测方法的总结: ①很多对场景建模的方法都假设场景中存在的平行光源且只有一个光源, 未来的研究者需要着力开发适用于各种复杂光场的建模方法。②虽然目前有很多利用光源一致性来进行篡改检测的算法, 但所有方法都对建模中的光场有很严格的前提条件。例如, 有的方法要求场景中的物体必须是严格凸(Convex)的且有恒定的反射率(Constant reflectance)^[11,14], 有的方法要求场景中的物体必须是理想散射(Lambertian

reflection)^[17]的。一旦这些标准达不到,使用这些方法就会造成很大误差。③现有的基于光照一致性的检测方法还不能做到完全自动化,很多方法的有着非常复杂的中间环节且需要人工进行干预。以后的研究工作需要着力于简化基于光照一致性的检测方法的中间环节。

基于特征提取与分类的检测方法的总结: ①现有特征已经能够在单一的数据集上达到很高的检测准确率,但在跨数据集的情况下检测准确率还不太理想。未来需设计一些更为鲁棒的特征,使得能对于各种数据集都有较高的检测准确率。②无论是传统的机器学习的方法,还是深度学习的方法,都需要大量的样本来训练模型。一旦所能使用的样本数量过少,这些方法学习到的模型就很容易产生过拟合。目前所存在的数据集在质量和数量上还有所欠缺,希望在将来开发出一些质量较高的数据集。③虽然目前很多方法能达到很高的检测准确率,但大多方法的篡改定位精度还有待提升。

基于成像色差印记一致性的检测方法的总结: 相比早先提出的方法,近些年提出的利用成像色差进行篡改检测的方法在检测准确率和计算效率上都有不错的提升: 例如在对局部 LCA 偏移量进行计算的过程中,在相似度计算环节中使用菱形搜索算法^[43]让计算效率大大提高。目前基于色差的检测方法依然存在以下问题: ①现有算法中,大多数都是针对横向色差来进行检测的,而利用纵向色差的算法非常之少。文献[45]中所提出的反取证方法能避免被传统的横向色差方法所检测到,可见,仅仅利用一种色差模式来检测篡改的方式是不可靠的。研究者在未来的研究中应当开发一些利用纵向色差进行检测的算法,以增强该类算法的多样性。②如果在整张图像中篡改的区域太小,会导致篡改区域的统计量对图像整体色差分布影响过小而使得检测出现高漏检率;如果篡改的区域过大,会对全局估计的模型产生较大的影响,也会影响检测的准确率。所以,只有当篡改区域占整张图像的比例在合适的范围内时,现有的基于成像色差的检测方法才能有较高的准确率。

基于自然模糊印记一致性的检测方法的总结: ①检测模糊的一个重要环节是对模糊核进行估计。目前很多算法中所假设的模糊核形式过于简单,不能对真实存在的一些模糊核进行精确模拟,以至影响检测的准确率。在将来应开发出更具普适性模型来对模糊核进行估计,以提高检测的准确率。②如果检测出同一深度的物体具有不同的离焦模糊,则

可以认为该图像是被篡改过的。然而,如果两个物体处在不同深度,则无法通过模糊与深度信息来判断图像是否被篡改过。希望在未来的研究中找到真实图像中深度信息与模糊程度的对应关系,使得能够利用不在同一深度物体的模糊程度进行检测。

基于成像系统噪声印记一致性的篡改检测方法的总结: ①因为基于 PRNU 噪声的篡改检测方法需要知道待检测图像对应的成像设备或用该设备拍摄的其它图像,所以该方法在现实中的应用很局限;②如果图像的空间分辨率(Spatial resolution)不够高,则基于 PRNU 印记一致性的篡改检测方法很难对图像中存在的小区域的篡改进行有效检测。分割方法可以提高图像的空间分辨率,未来的研究可以借助更强大的分割方法来对图像中存在的小区域篡改进行有效检测。

基于彩色滤波阵列插值印记一致性的篡改检测方法的总结: 虽然目前基于 CFA 印记一致性的方法已经能够实现对篡改区域的精确定位,但当图像中存在大量分布一致的区域(Uniform region)或者尖锐区域(Sharp region)的时候,该检测方式可能会出现比较高的虚警率。

基于几何变换与插值痕迹的检测方法的总结: ①由于下采样操作不会给邻接像素带来周期性的相关关系,所以目前的基于插值痕迹的检测方法不能对图像中的下采样操作进行检测;②很多基于频域的插值检测方法很大程度上受到可用样本数量的制约,当样本过于少时,这类方法的性能会明显下降。

基于人为模糊痕迹的检测方法的总结: 在利用人为模糊痕迹进行篡改检测的方法中,如何对模糊的种类和强度进行判断显得非常重要。目前的研究方法可以在一定程度上区分人为模糊、离焦模糊和运动模糊。但如果一个区域同时受到好几种模糊的影响,现存的检测方法还不能对该区域存在的所有模糊类型进行有效判断。

基于中值滤波痕迹的检测方法的总结: 该方法是目前与深度学习结合较好的几种方法之一,已经能够实现对端对端的检测。基于深度学习的中值滤波痕迹检测算法能够在小尺寸图像经过有损压缩和噪声处理的情况下对图像进行有效的检测。未来可以继续优化网络结构使检测准确率进一步提高。

基于特征匹配的检测方法的总结: ①为了得到精确的匹配结果,基于块匹配的篡改检测方法会密集地在图像中提取重叠图像块,这会使该类方法拥有非常高的计算复杂度,部分算法采用 PCA 和 SVD

来降低特征复杂度。基于块匹配检测的另一个特点是对于旋转、缩放等几何变换不鲁棒。②基于关键点匹配的检测方法计算速度快, 且提取的特征能够在图像经过旋转、缩放等几何变换和 JPEG 压缩、加性噪声等后处理操作的情况下有不错的检测准确率。但当图像中存在大量平滑区域或高度相似的区域时, 使用该类方法会带来很高的错误率。③基于神经网络的检测方法是近年来出现的新方法, 相比于传统的两种方法, 该方法不需要手工设计特征, 可以对图像进行端到端的检测。另外, 该方法也不需要像两种传统方法一样需要对很多阈值进行调节才能得到令人满意的检测结果。大量实验表明, 基于深度学习的方法在准确率和计算效率上都要优于两种传统方法。

基于 JPEG 重压缩痕迹的检测方法(对齐假设与未对齐假设)的总结: ①虽然互联网上存在的大多数图像都是 JPEG 格式的, 但也有部分是以其他格式储存的, 这时这类方法便无法进行检测; ②绝大多数方法都是通过统计图像各块 DCT 的系数来进行篡改检测的, 如果篡改区域太过于小的话, 则篡改对图像整体 DCT 分布的影响有限, 此时篡改痕迹很难被检测到; ③很多检测算法只有当知道第一次压缩系数时才会有效, 然而在现实环境中, 第一次压缩系数很难知道; ④目前很多算法只有当第一次 JPEG 压缩的品质因数 Q1 和第二次 JPEG 压缩的品质因数 Q2 满足特定关系的时候才能有效地进行检测。希望将来的研究可以逐步放宽此条件的约束。

另外, 基本上各类检测方法均容易收到有损压缩、模糊、插值等后处理操作的影响, 希望在未来能设计出更多更鲁棒的能适应各种复杂环境的新方法。

近几年, 随着数据集的扩充和计算能力的不断增强, 有理由相信在未来将有更多功能更强的新方法出现。

7.2 图像篡改检测方法发展展望

随着这几年科学技术日新月异地发展和硬件设备计算性能地显著提升, 数字图像取证技术也有了长足的进步, 并出现了许多新的派系和分支。尽管不断地有新技术新方法的提出, 但仍有很多问题没有得到有效解决。

鉴于数字图像取证技术的复杂性以及多样性, 我们仍然需要依靠相关各学科(例如信号处理、计算机图形学、计算机视觉等)相关领域知识, 不断结合实际问题, 发展新方法。另一方面, 如图 15 所示, 领域知识与深度神经网络进行结合, 也是篡改检测

新技术未来重要的发展方向。

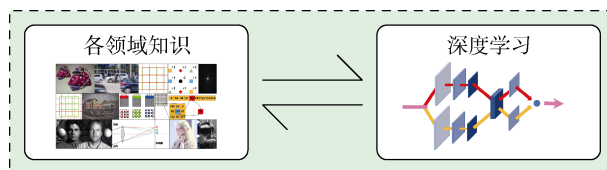


图 15 图像篡改检测未来发展趋势

Figure 15 The future development trend of image tampering detection

如下从两个侧面对看好的篡改检测方法进行展望, 其中侧面一为基于各领域知识的检测方法, 而侧面二为领域知识与深度神经网络结合的检测方法。

7.2.1 基于各领域知识的检测方法展望

本文已经对各方法的主干脉络和国内外发展现状做出了详细的总结和分析, 在此基础上, 本文对篡改检测方法未来发展的方向提出一些自己的看法, 希望能给今后的研究者提供一些方法上的启发和思路上的借鉴。

(1) 如何区分对图像的恶意篡改和无恶意篡改(例如图像润饰)将是未来研究的重点之一。前者是为了歪曲事实, 而后者只是为了改变图像的视觉效果。两者造成的危害完全不同, 所以要对两种类型的篡改加以区分。但目前存在的很多算法都不能有效区分以上两种性质截然不同的篡改方式, 希望在将来能够研究出有效区分以上两种篡改方式的新算法。

(2) 随着近些年不断有新的检测算法出现, 研究者迫切需要高质量的公共数据集去评测各算法的性能。但是在一些方法中(例如基于彩色滤波阵列插值印记一致性的检测方法)目前没有公开的公共数据集, 在另一些算法中(例如特征提取与分类的检测方法), 目前存在的公共数据集数量较少或者质量较差。希望研究者在未来的研究中能多制备一些高质量的公共数据集以供科研所用。

(3) 现存的很多算法仅仅能判断一张图片篡改与否, 然而在很多实际应用场景中, 仅仅得出一张图片是否被篡改过的结论并不能令人信服, 人们更关心的是篡改图片的什么部位经过了什么样的篡改, 这对暴露篡改者的篡改意图有很大帮助。所以新提出的算法不仅应该能区分真实的和篡改的图片, 还应该能够定位图像篡改的具体位置。

(4) 现存的利用深度学习来解决篡改检测问题所设计的网络大多来自于计算机视觉任务, 这些网

络大多是为视觉任务服务的,可能会更倾向于提取出图像的语义信息而不是隐藏在语义信息下的篡改痕迹。如果要开发利用深度学习解决篡改检测问题的新方法,应该结合篡改检测的特点对神经网络进行适当的修改,以达到更高的检测准确率。

(5) 在近几年里,随着各种电子成像设备尤其是手机端成像设备的大力发展,数字图像的分辨率越来越高,这也对处理图像的算法和硬件设备提出了更高的要求。图像篡改检测新算法应当着重对计算效率进行优化,以满足高清图像检测的需要。

(6) 目前利用 deepfake、face2face 等方法制作的篡改视频对很多行业造成很大威胁,人们急需有效的检测算法来对这些篡改视频进行检测。由于视频的高压缩率,成像系统的印记和篡改痕迹会随着压缩变得无影无踪。在本文提及的 12 种针对图像的篡改检测方法中,只有基于光照一致性的方法受压缩的影响较小,未来可以考虑将该类方法应用于 deepfake、face2face 等篡改视频的检测任务中。

7.2.2 领域知识与深度网络结合的检测方法展望

随着深度学习方法的不断完善和深度学习工具的成熟化,再加上各种新数据集的支撑,很多深度学习方法的检测准确率和检测效率都达到一个新的高度。目前与深度学习结合得较为紧密方法有基于特征提取与分类的检测方法、基于中值滤波痕迹的检测方法和基于特征匹配的检测方法。

表 16 展示的是基于特征提取与分类方法在 DVMM 数据集中的检测准确率。表 17 展示的是基于中值滤波痕迹的检测方法在 BOSSbase 1.01、UCID、BOSS RAW、Dresden Image、NRCS 构成的混合数据集中,32*32 图像块的不同 JPEG 品质因数和滤波核大小的情况下的检测准确率。 QF 代表 JPEG 图像的压缩品质因数, KZ 代表卷积核大小。表 18 展示的是基于特征匹配检测方法在 CASIA V2.0 数据集中的检测准确率和每张图像的检测时间。表 19 展示的是检测网络在 NIST16、Columbia、

COVER、CASIA(CASIA V2.0 上训练,CASIA V1.0 上检测)上篡改定位 $F1$ 值。表 20 展示的是基于特征匹配的方法在 CASIA V2.0 数据集中的篡改定位 $F1$ 值。

表 16 基于深度学习算法与非深度学习算法性能比较 1^[36]

Table 16 Performance comparison 1 of deep learning based algorithms and other algorithms^[36]

检测算法	检测准确率(%)
2-D Noncausal Markov ^[29]	93.36
Markov in DCT & DWT ^[26]	93.55
SRM+10layer CNN^[36]	96.38

(注:加粗字体显示的方法为深度学习方法)

可以看出,深度学习的方法的检测准确率明显高于非深度学习的方法。在表 18 中我们可以看出深度学习的方法的计算效率相对于非深度学习方法也有明显的优势。

在表 19 和表 20 中我们可以看出未结合深度学习的方法定位精度远低于基于深度学习方法。另一方面,我们还可以看出基于深度学习方法的定位效果也有很大的提升空间。

近期提出的与深度学习有关的篡改检测方法有:CNNs^[36-38,160,164]、堆栈自编码器(stacked auto-encoder, SAE)^[165-166]和长短期记忆(long short-term memory, LSTM)网络^[167-168]。

相信在未来会提出更加优化的网络结构以进一步提升篡改定位的准确率。

8 总结

随着近些年科学技术的快速发展,制作一张效果逼真的篡改图片已经越来越容易。篡改图像会对包括学术研究、司法取证、新闻传播等诸多领域带来严重的影响,人们迫切需要成熟的检测技术去识别篡改图像,因此针对图像篡改检测开展的研究将会非常有意义。鉴于目前关于图像篡改检测的综述

表 17 基于深度学习算法与非深度学习算法性能比较 2^[102]

Table 17 Performance comparison 2 of deep learning based algorithms and other algorithms^[102]

检测算法	检测准确率(%)			
	$QF=70$ $KZ=3*3$	$QF=90$ $KZ=3*3$	$QF=70$ $KZ=5*5$	$QF=90$ $KZ=5*5$
MFF ^[95]	73.99	80.32	82.49	85.91
AR ^[101]	75.63	83.52	80.80	86.26
GLF ^[99]	78.15	85.43	87.28	91.57
MFR+CNN^[102]	79.42	87.71	88.65	93.21

(注:加粗字体显示的方法为深度学习方法)

表 18 基于深度学习算法与非深度学习算法性能比较 3^[106]

Table 18 Performance comparison 3 of deep learning based algorithms and other algorithms^[106]

检测算法	检测准确率(%)	每张图像检测时间(s)
Zernike Moments ^[115]	39.08	5.11
Dense-Field ^[116]	46.82	1.78
BusterNet^[106]	75.98	0.62

(注: 加粗字体显示的方法为深度学习方法)

文献需要扩展和更新, 例如缺乏对新文献的总结或缺少一些必要的分析, 因此本文对最新的相关工作

进行了归纳总结, 希望能对数字取证有关的研究者提供研究文献的参考、研究方法上的启发和研究思路上的借鉴。

本文依据图像篡改检测方法所依赖的线索将这些方法依次分为 2 个方面和 4 个分组, 再根据方法所涉及到的细节, 分为 12 个具体的类别。本文还对每类方法中现有的算法做了详细的分析的论述, 并对各方法中典型的算法的性能做了比较。另外, 本文还总结了在各篡改检测方法中所使用的性能指标和公开数据集。

表 19 基于深度学习算法与非深度学习算法性能比较 4^[160]

Table 19 Performance comparison 4 of deep learning based algorithms and other algorithms^[160]

检测算法	定位 FI 值			
	NIST16	Columbia	COVER	CASIA
ELA ^[161]	0.236	0.470	0.222	0.214
NOI ^[162]	0.285	0.574	0.269	0.263
CFA1 ^[75]	0.174	0.467	0.190	0.207
MFCN^[163]	0.571	0.612	—	0.541
RGB-N^[160]	0.722	0.697	0.437	0.408

(注: 加粗字体显示的方法为深度学习方法)

表 20 基于深度学习算法与非深度学习算法性能比较 5^[106]

Table 20 Performance comparison 5 of deep learning based algorithms and other algorithms^[106]

检测算法	定位 FI 值(%)
Zernike Moments ^[115]	16.40
Dense-Field ^[116]	25.43
BusterNet^[106]	45.56

(注: 加粗字体显示的方法为深度学习方法)

本文对提出的十二类方法中的各种算法目前存在的优缺点做了总结。目前方法存在的主要问题有: (1)应用范围有限; (2)易受到 JPEG 压缩、模糊等后处理影响; (3)数据集数据量不够充足或数据质量不高; (4)对小范围篡改检测定位效果不佳。希望研究者能在未来的研究中对目前存在的问题加以改善。

最后, 本文还对各方法未来发展的趋势做了展望。通过对目前与深度学习得较为紧密的 3 种方法中各算法的性能分析, 我们发现基于深度学习的方法无论在检测准确率还是在检测速度上都明显优于其它的方法。本文指出了未来各学科发展的方向是将各领域知识与深度学习有效地结合。对于目前已经与深度学习有效结合的方法, 未来可以进一步对具体算法和网络的结构加以改进以提升检测的准确率和效率。对于目前还没有与深度学习有效结合的

方法, 未来可以寻找合适的切入点使深度学习工具与该方法有机地结合。

致谢 在此向本文成文中给予指导的老师、提供帮助的同学和给本文提出建议的评审专家表示感谢。

参考文献

[1] Peng B, Wang W, Dong J, et al. Improved 3D Lighting Environment Estimation for Image Forgery Detection[J]. *2015 IEEE International Workshop on Information Forensics and Security*, 2015: 1-6.

[2] Wang W, Dong J and Tan T. A survey of passive image tampering detection[C]. *International Workshop on Digital Watermarking*, 2009: 308-322.

[3] Piva A. An Overview on Image Forensics[J]. *ISRN Signal Processing*, 2013, 2013: 496701.

[4] Qazi T, Hayat K, Khan S U, et al. Survey on Blind Image Forgery Detection[J]. *IET Image Processing*, 2013, 7(7): 660-670.

[5] Birajdar G K, Mankar V H. Digital Image Forgery Detection Using Passive Techniques: A Survey[J]. *Digital Investigation*, 2013, 10(3): 226-245.

[6] Liu L. Survey on Passive Digital Image Authenticity Check Techniques[J]. *Computer Engineering and Applications*, 2009, 45(26): 1-4.

(柳林. 被动式数字图像真实性检测技术综述[J]. *计算机工程与应用*, 2009, 45(26): 1-4.)

[7] Wu Q, Li G H, Tu D, et al. A Survey of Blind Digital Image Foren-

- sics Technology for Authenticity Detection[J]. *Acta Automatica Sinica*, 2008, 34(12): 1458-1466.
(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): 1458-1466.)
- [8] Zampoglou M, Papadopoulos S, Kompatsiaris Y. Large-Scale Evaluation of Splicing Localization Algorithms for Web Images[J]. *Multimedia Tools and Applications*, 2017, 76(4): 4801-4834.
- [9] Lin X, Li J H, Wang S L, et al. Recent Advances in Passive Digital Image Security Forensics: A Brief Review[J]. *Engineering*, 2018, 4(1): 29-39.
- [10] Johnson M K, Farid H. Exposing Digital Forgeries by Detecting Inconsistencies in Lighting[C]. *The 7th workshop on Multimedia and security - MM&Sec '05*, 2005: 1-10.
- [11] Johnson M K, Farid H. Exposing Digital Forgeries in Complex Lighting Environments[J]. *IEEE Transactions on Information Forensics and Security*, 2007, 2(3): 450-461.
- [12] Kee E, O'Brien J F, Farid H. Exposing Photo Manipulation with Inconsistent Shadows[J]. *ACM Transactions on Graphics*, 2013, 32(3): 1-12.
- [13] Carvalho T, Farid H, Kee E R. Exposing Photo Manipulation from User-Guided 3D Lighting Analysis[C]. *Proc SPIE 9409, Media Watermarking, Security, and Forensics 2015*, 2015, 9409: 940902.
- [14] Kee E, Farid H. Exposing Digital Forgeries from 3-D Lighting Environments[J]. *2010 IEEE International Workshop on Information Forensics and Security*, 2010: 1-6.
- [15] Fan W, Wang K, Cayre F, et al. 3D lighting-based image forgery detection using shape-from-shading[C]. *The 20th European Signal Processing Conference*, 2012: 1777-1781.
- [16] Johnson M K, Farid H. Exposing Digital Forgeries through Specular Highlights on the Eye[M]. *Information Hiding*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 311-325.
- [17] Peng B, Wang W, Dong J, et al. Optimized 3D Lighting Environment Estimation for Image Forgery Detection[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 12(2): 479-494.
- [18] Ng T T, Chang S F, Sun Q B. Blind Detection of Photomontage Using Higher Order Statistics[C]. *2004 IEEE International Symposium on Circuits and Systems*, 2004: V.
- [19] Fu D D, Shi Y Q, Su W. Detection of Image Splicing Based on Hilbert-Huang Transform and Moments of Characteristic Functions with Wavelet Decomposition[C]. *Digital Watermarking*, 2006: 177-187.
- [20] Chen W, Shi Y Q, Su W. Image Splicing Detection Using 2D Phase Congruency and Statistical Moments of Characteristic Function[C]. *Proc SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007, 6505: 281-288.
- [21] Liu X X, Li F, Xiong B. Image Splicing Detection Using Weber Local Descriptors[J]. *Computer Engineering and Applications*, 2013, 49(12): 140-143.
(刘晓霞, 李峰, 熊兵. 基于韦伯局部特征的图像拼接检测[J]. *计算机工程与应用*, 2013, 49(12): 140-143.)
- [22] Wei W, Dong J, Tan T N. Effective Image Splicing Detection Based on Image Chroma[C]. *2009 16th IEEE International Conference on Image Processing*, 2009: 1257-1260.
- [23] Zhao X, Li J, Li S, et al. Detecting digital image splicing in chroma spaces[C]. *International Workshop on Digital Watermarking 2010*: 12-22.
- [24] Muhammad G, Al-Hammadi M H, Hussain M, et al. Image Forgery Detection Using Steerable Pyramid Transform and Local Binary Pattern[J]. *Machine Vision and Applications*, 2014, 25(4): 985-995.
- [25] Huang J W, Zhou D K, Yang X, et al. Image Splicing Detection Based on Local Mean Decomposition and Moment Features[J]. *Electronic Measurement Technology*, 2017, 40(4): 146-151.
(黄经纬, 周大可, 杨欣, 等. 基于局部均值分解和矩特征的图像拼接检测[J]. *电子测量技术*, 2017, 40(4): 146-151.)
- [26] He Z W, Lu W, Sun W, et al. Digital Image Splicing Detection Based on Markov Features in DCT and DWT Domain[J]. *Pattern Recognition*, 2012, 45(12): 4292-4299.
- [27] Shi Y Q, Chen C H, Chen W. A Markov Process Based Approach to Effective Attacking JPEG Steganography[M]. *Information Hiding*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006: 249-264.
- [28] Shi Y Q, Chen C H, Chen W. A Natural Image Model Approach to Splicing Detection[C]. *The 9th workshop on Multimedia & security - MM&Sec '07*, 2007: 51-62.
- [29] Zhao X D, Wang S L, Li S H, et al. Passive Image-Splicing Detection by a 2-D Noncausal Markov Model[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 25(2): 185-199.
- [30] Li C, Ma Q, Xiao L M, et al. Image Splicing Detection Based on Markov Features in QDCT Domain[J]. *Neurocomputing*, 2017, 228: 29-36.
- [31] Zhang Q B, Lu W, Weng J. Joint Image Splicing Detection in DCT and Contourlet Transform Domain[J]. *Journal of Visual Communication and Image Representation*, 2016, 40: 449-458.
- [32] Zhang Q B, Lu W, Wang R X, et al. Digital Image Splicing Detection Based on Markov Features in Block DWT Domain[J]. *Multimedia Tools and Applications*, 2018, 77(23): 31239-31260.
- [33] Fridrich J, Kodovsky J. Rich Models for Steganalysis of Digital Images[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(3): 868-882.
- [34] Cozzolino D, Poggi G, Verdoliva L. Splicebuster: A New Blind Image Splicing Detector[J]. *2015 IEEE International Workshop on Information Forensics and Security*, 2015: 1-6.
- [35] Li Y, Zhong L, Li J. Detection of Image Splicing Forgery Based on LBP and Co-Occurrence Matrix[J]. *Journal of Wuhan University (Natural Science Edition)*, 2015, 61(6): 517-524.
(李燕, 钟磊, 李健. 基于 LBP 和共生矩阵的图像拼接篡改检测[J]. *武汉大学学报(理学版)*, 2015, 61(6): 517-524.)
- [36] Rao Y, Ni J Q. A Deep Learning Approach to Detection of Splicing and Copy-Move Forgeries in Images[J]. *2016 IEEE International Workshop on Information Forensics and Security*, 2016: 1-6.
- [37] Liu Y Q, Guan Q X, Zhao X F, et al. Image Forgery Localization Based on Multi-Scale Convolutional Neural Networks[C]. *The 6th ACM Workshop on Information Hiding and Multimedia Security*, 2018: 85-90.
- [38] Shi Z N, Shen X J, Kang H, et al. Image Manipulation Detection and Localization Based on the Dual-Domain Convolutional Neural Networks[J]. *IEEE Access*, 2018, 6: 76437-76453.
- [39] Johnson M K, Farid H. Exposing Digital Forgeries through Chro-

- matic Aberration[C]. *The 8th workshop on Multimedia and security*, 2006: 48-55.
- [40] Mayer O, Stamm M. Improved Forgery Detection with Lateral Chromatic Aberration[C]. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016: 2024-2028.
- [41] van L T, Emmanuel S, Kankanhalli M S. Identifying Source Cell Phone Using Chromatic Aberration[C]. *2007 IEEE International Conference on Multimedia and Expo*, 2007: 883-886.
- [42] Gloe T, Borowka K, Winkler A. Efficient Estimation and Large-Scale Evaluation of Lateral Chromatic Aberration for Digital Image Forensics[C]. *Proc SPIE 7541, Media Forensics and Security II*, 2010, 7541: 62-74.
- [43] Mayer O, Stamm M C. Accurate and Efficient Image Forgery Detection Using Lateral Chromatic Aberration[J]. *IEEE Transactions on Information Forensics and Security*, 2018, 13(7): 1762-1777.
- [44] Yerushalmy I, Hel-Or H. Digital Image Forgery Detection Based on Lens and Sensor Aberration[J]. *International Journal of Computer Vision*, 2011, 92(1): 71-91.
- [45] Mayer O, Stamm M C. Anti-Forensics of Chromatic Aberration[C]. *Media Watermarking, Security, and Forensics 2015*, 2015: 94090M.
- [46] Mayer O, Stamm M C. Countering Anti-Forensics of Lateral Chromatic Aberration[C]. *The 5th ACM Workshop on Information Hiding and Multimedia Security*, 2017: 15-20.
- [47] Chen Z Y, Fang Z. Detection of Digital Image Forgery Based on Chromatic Aberration[J]. *Journal of Applied Sciences*, 2015, 33(6): 604-614.
(陈竺益, 方针. 基于色像差特性的图像篡改检测[J]. *应用科学学报*, 2015, 33(6): 604-614.)
- [48] Bahrami K, Kot A C, Li L D, et al. Blurred Image Splicing Localization by Exposing Blur Type Inconsistency[J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(5): 999-1009.
- [49] Wang X, Xuan B, Peng S L. Digital Image Forgery Detection Based on the Consistency of Defocus Blur[C]. *2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2008: 192-195.
- [50] Bahrami K, Kot A C, Fan J Y. Splicing Detection in Out-of-Focus Blurred Images[J]. *2013 IEEE International Workshop on Information Forensics and Security*, 2013: 144-149.
- [51] Kakar P, Natarajan S, Ser W. Detecting Digital Image Forgeries through Inconsistent Motion Blur[C]. *2010 IEEE International Conference on Multimedia and Expo*, 2010: 486-491.
- [52] Kakar P, Sudha N, Ser W. Exposing Digital Image Forgeries by Detecting Discrepancies in Motion Blur[J]. *IEEE Transactions on Multimedia*, 2011, 13(3): 443-452.
- [53] Chen X G, Yang J, Wu Q, et al. Directional High-Pass Filter for Blurry Image Analysis[J]. *Signal Processing: Image Communication*, 2012, 27(7): 760-771.
- [54] Su B L, Lu S J, Tan C L. Blurred Image Region Detection and Classification[C]. *The 19th ACM international conference on Multimedia - MM'11*, 2011: 1397-1400.
- [55] Bahrami K, Kot A C. Image Tampering Detection by Exposing Blur Type Inconsistency[C]. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014: 2654-2658.
- [56] Lyu S W, Pan X Y, Zhang X. Exposing Region Splicing Forgeries with Blind Local Noise Estimation[J]. *International Journal of Computer Vision*, 2014, 110(2): 202-221.
- [57] Lukáš J, Fridrich J, Goljan M. Detecting Digital Image Forgeries Using Sensor Pattern Noise[C]. *SPIE Proceedings, Security, Steganography, and Watermarking of Multimedia Contents VIII*, 2006: 6072.
- [58] Chen M, Fridrich J, Goljan M, et al. Determining Image Origin and Integrity Using Sensor Noise[J]. *IEEE Transactions on Information Forensics and Security*, 2008, 3(1): 74-90.
- [59] Chierchia G, Parrilli S, Poggi G, et al. On the Influence of Denoising in PRNU Based Forgery Detection[C]. *The 2nd ACM workshop on Multimedia in forensics, security and intelligence*, 2010: 117-122.
- [60] Chierchia G, Poggi G, Sansone C, et al. A Bayesian-MRF Approach for PRNU-Based Image Forgery Detection[J]. *IEEE Transactions on Information Forensics and Security*, 2014, 9(4): 554-567.
- [61] Chierchia G, Parrilli S, Poggi G, et al. PRNU-Based Detection of Small-Size Image Forgeries[C]. *2011 17th International Conference on Digital Signal Processing*, 2011: 1-6.
- [62] Chierchia G, Cozzolino D, Poggi G, et al. Guided Filtering for PRNU-Based Localization of Small-Size Image Forgeries[C]. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014: 6231-6235.
- [63] Lin X F, Li C T. Refining PRNU-Based Detection of Image Forgeries[C]. *2016 Digital Media Industry & Academic Forum*, 2016: 222-226.
- [64] Li C T, Li Y. Color-Decoupled Photo Response Non-Uniformity for Digital Image Forensics[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(2): 260-271.
- [65] Korus P, Huang J W. Multi-Scale Analysis Strategies in PRNU-Based Tampering Localization[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 12(4): 809-824.
- [66] Popescu A C, Farid H. Exposing Digital Forgeries in Color Filter Array Interpolated Images[J]. *IEEE Transactions on Signal Processing*, 2005, 53(10): 3948-3959.
- [67] Popescu A C, Farid H. Exposing Digital Forgeries by Detecting Traces of Resampling[J]. *IEEE Transactions on Signal Processing*, 2005, 53(2): 758-767.
- [68] Dirik A E, Memon N. Image Tamper Detection Based on Demosaicing Artifacts[C]. *2009 16th IEEE International Conference on Image Processing*, 2009: 1497-1500.
- [69] Takamatsu J, Matsushita Y, Ogasawara T, et al. Estimating Demosaicing Algorithms Using Image Noise Variance[C]. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010: 279-286.
- [70] Gallagher A C. Detection of Linear and Cubic Interpolation in JPEG Compressed Images[C]. *The 2nd Canadian Conference on Computer and Robot Vision*, 2005: 65-72.
- [71] Gallagher A C, Chen T. Image Authentication by Detecting Traces of Demosaicing[C]. *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008: 1-8.
- [72] Bayram S, Sencar H T, Memon N. Classification of Digital Cam-

- era-Models Based on Demosaicing Artifacts[J]. *Digital Investigation*, 2008, 5(1/2): 49-59.
- [73] Swaminathan A, Wu M, Liu K J R. Nonintrusive Component Forensics of Visual Sensors Using Output Images[J]. *IEEE Transactions on Information Forensics and Security*, 2007, 2(1): 91-106.
- [74] Cao H, Kot A C. Accurate Detection of Demosaicing Regularity for Digital Image Forensics[J]. *IEEE Transactions on Information Forensics and Security*, 2009, 4(4): 899-910.
- [75] Ferrara P, Bianchi T, de Rosa A, et al. Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(5): 1566-1577.
- [76] Singh A, Singh G, Singh K. A Markov Based Image Forgery Detection Approach by Analyzing CFA Artifacts[J]. *Multimedia Tools and Applications*, 2018, 77(21): 28949-28968.
- [77] Peng S, Peng Y Y, Xiao C Y. Image Tampering Detection Algorithm Based on CFA Interpolation[J]. *Transducer and Microsystem Technologies*, 2015, 34(6): 141-144.
(彭双, 彭圆圆, 肖昌炎. 基于CFA插值的图像篡改检测算法[J]. *传感器与微系统*, 2015, 34(6): 141-144.)
- [78] Zhang X L, Fang Z, Zhang X P. Forgery Detection via Inter-Channel Correlation of CFA Images[J]. *Journal of Applied Sciences*, 2015, 33(1): 87-94.
(张晓琳, 方针, 张新鹏. 利用通道间相关性的CFA图像盲取证[J]. *应用科学学报*, 2015, 33(1): 87-94.)
- [79] Su W X, Fang Z. Identifying Image Authenticity Based on CFA Inconsistency of Interpolation Characteristics[J]. *Journal of Applied Sciences*, 2019, 37(1): 33-40.
(苏文煊, 方针. 基于CFA插值特性不一致的图像真伪鉴别[J]. *应用科学学报*, 2019, 37(1): 33-40.)
- [80] Prasad S, Ramakrishnan K R. On Resampling Detection and Its Application to Detect Image Tampering[C]. *2006 IEEE International Conference on Multimedia and Expo*, 2006: 1325-1328.
- [81] Mahdian B, Saic S. Blind Authentication Using Periodic Properties of Interpolation[J]. *IEEE Transactions on Information Forensics and Security*, 2008, 3(3): 529-538.
- [82] Mahdian B, Saic S. On Periodic Properties of Interpolation and Their Application to Image Authentication[C]. *Third International Symposium on Information Assurance and Security*, 2007: 439-446.
- [83] Kirchner M. Linear Row and Column Predictors for the Analysis of Resized Images[C]. *The 12th ACM workshop on Multimedia and security*, 2010: 13-18.
- [84] Wang R, Ping X J. Detection of Resampling Based on Singular Value Decomposition[C]. *2009 Fifth International Conference on Image and Graphics*, 2009: 879-884.
- [85] Vázquez-Padín D, Comesaña P, Pérez-González F. An SVD Approach to Forensic Image Resampling Detection[C]. *2015 23rd European Signal Processing Conference*, 2015: 2067-2071.
- [86] Kao Y T, Lin H J, Wang C W, et al. Effective Detection for Linear Up-Sampling by a Factor of Fraction[J]. *IEEE Transactions on Image Processing*, 2012, 21(8): 3443-3453.
- [87] Vázquez-Padín D, Pérez-González F, Comesaña-Alfaro P. A Random Matrix Approach to the Forensic Analysis of Upscaled Images[J]. *IEEE Transactions on Information Forensics and Security*, 2017, 12(9): 2115-2130.
- [88] Hsiao D Y, Pei S C. Detecting Digital Tampering by Blur Estimation[J]. *First International Workshop on Systematic Approaches to Digital Forensic Engineering*, 2005: 264-278.
- [89] Zhou L N, Wang D M, Guo Y B, et al. Blur Detection of Digital Forgery Using Mathematical Morphology[M]. *Agent and Multi-Agent Systems: Technologies and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 990-998.
- [90] Cao G, Zhao Y, Ni R. Edge-based blur metric for tamper detection[J]. *Journal of Information Hiding and Multimedia Signal Processing*, 2010, 1(1): 20-27.
- [91] Zhang C, Zhang H B. Detecting Digital Image Forgeries through Weighted Local Entropy[C]. *2007 IEEE International Symposium on Signal Processing and Information Technology*, 2007: 62-67.
- [92] Liu G J, Wang J W, Lian S G, et al. Detect Image Splicing with Artificial Blurred Boundary[J]. *Mathematical and Computer Modelling*, 2013, 57(11/12): 2647-2659.
- [93] Bovik A. Streaking in Median Filtered Images[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1987, 35(4): 493-503.
- [94] Kirchner M, Fridrich J. On Detection of Median Filtering in Digital Images[C]. *Proc SPIE 7541, Media Forensics and Security II*, 2010, 7541: 371-382.
- [95] Yuan H D. Blind Forensics of Median Filtering in Digital Images[J]. *IEEE Transactions on Information Forensics and Security*, 2011, 6(4): 1335-1345.
- [96] Zhang Y J, Li S H, Wang S L, et al. Revealing the Traces of Median Filtering Using High-Order Local Ternary Patterns[J]. *IEEE Signal Processing Letters*, 2014, 21(3): 275-279.
- [97] Cao G, Zhao Y, Ni R R, et al. Forensic Detection of Median Filtering in Digital Images[C]. *2010 IEEE International Conference on Multimedia and Expo*, 2010: 89-94.
- [98] Chen C L, Ni J Q. Median Filtering Detection Using Edge Based Prediction Matrix[M]. *Digital Forensics and Watermarking*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 361-375.
- [99] Chen C L, Ni J Q, Huang R B, et al. Blind Median Filtering Detection Using Statistics in Difference Domain[M]. *Information Hiding*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 1-15.
- [100] Chen C L, Ni J Q, Huang J W. Blind Detection of Median Filtering in Digital Images: A Difference Domain Based Approach[J]. *IEEE Transactions on Image Processing*, 2013, 22(12): 4699-4710.
- [101] Kang X G, Stamm M C, Peng A J, et al. Robust Median Filtering Forensics Using an Autoregressive Model[J]. *IEEE Transactions on Information Forensics and Security*, 2013, 8(9): 1456-1468.
- [102] Chen J S, Kang X G, Liu Y, et al. Median Filtering Forensics Based on Convolutional Neural Networks[J]. *IEEE Signal Processing Letters*, 2015, 22(11): 1849-1853.
- [103] Tang H S, Ni R R, Zhao Y, et al. Median Filtering Detection of Small-Size Image Based on CNN[J]. *Journal of Visual Communication and Image Representation*, 2018, 51: 162-168.
- [104] Niu Y K, Zhao Y, Ni R R. Robust Median Filtering Detection Based on Local Difference Descriptor[J]. *Signal Processing: Image Communication*, 2017, 53: 65-72.
- [105] Peng A J, Kang X G. Median Filtering Forensics Based on

- Multi-Directional Difference of Filtering Residuals[J]. *Chinese Journal of Computers*, 2016, 39(3): 503-515.
(彭安杰, 康显桂. 基于滤波残差多方向差分的中值滤波取证技术[J]. *计算机学报*, 2016, 39(3): 503-515.)
- [106] Wu Y, Abd-Almageed W, Natarajan P. BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization[C]. *Computer Vision – ECCV 2018*, 2018: 168-184.
- [107] Christlein V, Riess C, Jordan J, et al. An Evaluation of Popular Copy-Move Forgery Detection Approaches[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(6): 1841-1854.
- [108] Fridrich A, Soukal B, Lukáš A. Detection of copy-move forgery in digital images[C]. *Digital Forensic Research Workshop*, 2003.
- [109] Popescu A, Farid H. Exposing digital forgeries by detecting duplicated image regions, *Dept. Comput. Sci.*, Dartmouth College, Tech. Rep. TR20, 2004.
- [110] Huang Y P, Lu W, Sun W, et al. Improved DCT-Based Detection of Copy-Move Forgery in Images[J]. *Forensic Science International*, 2011, 206(1/2/3): 178-184.
- [111] Ketenci S, Ulutas G. Copy-Move Forgery Detection in Images via 2D-Fourier Transform[C]. *2013 36th International Conference on Telecommunications and Signal Processing*, 2013: 813-816.
- [112] Bayram S, Taha Sencar H, Memon N. An Efficient and Robust Method for Detecting Copy-Move Forgery[C]. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009: 1053-1056.
- [113] Li W H, Yu N H. Rotation Robust Detection of Copy-Move Forgery[C]. *2010 IEEE International Conference on Image Processing*, 2010: 2113-2116.
- [114] Mahdian B, Saic S. Detection of Copy-Move Forgery Using a Method Based on Blur Moment Invariants[J]. *Forensic Science International*, 2007, 171(2/3): 180-189.
- [115] Ryu S J, Lee M J, Lee H K. Detection of Copy-Rotate-Move Forgery Using Zernike Moments[M]. *Information Hiding*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010: 51-65.
- [116] Cozzolino D, Poggi G, Verdoliva L. Efficient Dense-Field Copy-Move Forgery Detection[J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(11): 2284-2297.
- [117] Huang D Y, Huang C N, Hu W C, et al. Robustness of Copy-Move Forgery Detection under High JPEG Compression Artifacts[J]. *Multimedia Tools and Applications*, 2017, 76(1): 1509-1530.
- [118] Huang H L, Guo W Q, Zhang Y. Detection of Copy-Move Forgery in Digital Images Using SIFT Algorithm[J]. *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, 2008, 2: 272-276.
- [119] Pan X Y, Lyu S W. Detecting Image Region Duplication Using SIFT Features[C]. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010: 1706-1709.
- [120] Amerini I, Ballan L, Caldelli R, et al. A SIFT-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery[J]. *IEEE Transactions on Information Forensics and Security*, 2011, 6(3): 1099-1110.
- [121] Liu D, Hu Y J, Liu B B. Copy-Paste Forgery Detection Using SIFT Key-Points and CS-LBP Descriptor[J]. *Journal of Hefei University of Technology (Natural Science)*, 2012, 35(3): 325-330.
(刘丹, 胡永健, 刘琲贝. 联合 SIFT 特征点和 CS-LBP 特征描述子的复制粘贴篡改检测[J]. *合肥工业大学学报(自然科学版)*, 2012, 35(3): 325-330.)
- [122] Shivakumar B, Baboo S. Detection of region duplication forgery in digital images using SURF[J]. *International Journal of Computer Science Issues*, 2011, 8(4): 199.
- [123] Xu B, Wang J W, Liu G J, et al. Image Copy-Move Forgery Detection Based on SURF[C]. *2010 International Conference on Multimedia Information Networking and Security*, 2010: 889-892.
- [124] Mishra P, Mishra N, Sharma S, et al. Region Duplication Forgery Detection Technique Based on SURF and HAC[J]. *The Scientific World Journal*, 2013, 2013: 267691.
- [125] Zhu Y, Shen X J, Chen H P. Copy-Move Forgery Detection Based on Scaled ORB[J]. *Multimedia Tools and Applications*, 2016, 75(6): 3221-3233.
- [126] Wu Y, Abd-Almageed W, Natarajan P. Deep Matching and Validation Network: An End-to-End Solution to Constrained Image Splicing Localization and Detection[C]. *The 25th ACM international conference on Multimedia*, 2017: 1480-1502.
- [127] Lukáš J, Fridrich J. Estimation of primary quantization matrix in double compressed JPEG images[C]. *Digital forensic research workshop*, 2003: 5-8.
- [128] Popescu A, Farid H. Statistical tools for digital forensics[C]. *International workshop on information hiding*, 2004: 128-147.
- [129] Lin Z C, He J F, Tang X O, et al. Fast, Automatic and Fine-Grained Tampered JPEG Image Detection via DCT Coefficient Analysis[J]. *Pattern Recognition*, 2009, 42(11): 2492-2501.
- [130] Bianchi T, de Rosa A, Piva A. Improved DCT Coefficient Analysis for Forgery Localization in JPEG Images[C]. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011: 2444-2447.
- [131] Duan X T, Peng T, Li F F, et al. Blind Separation of Tampered Images Based on JPEG Double Compression Properties[J]. *Journal of University of Jinan (Science and Technology)*, 2017, 31(2): 87-96.
(段新涛, 彭涛, 李飞飞, 等. 基于 JPEG 重压缩特性的篡改图像盲分离[J]. *济南大学学报(自然科学版)*, 2017, 31(2): 87-96.)
- [132] Fu D D, Shi Y Q, Su W. A Generalized Benford's Law for JPEG Coefficients and Its Applications in Image Forensics[C]. *Proc SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007, 6505: 574-584.
- [133] Li B, Shi Y Q, Huang J W. Detecting Doubly Compressed JPEG Images by Using Mode Based First Digit Features[J]. *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008: 730-735.
- [134] Feng X Y, Doërr G. JPEG Recompression Detection[C]. *SPIE Proceedings, Media Forensics and Security II*, 2010: 75410J.
- [135] Huang F J, Huang J W, Shi Y Q. Detecting Double JPEG Compression with the Same Quantization Matrix[J]. *IEEE Transactions on Information Forensics and Security*, 2010, 5(4): 848-856.
- [136] Yang J Q, Xie J, Zhu G P, et al. An Effective Method for Detecting Double JPEG Compression with the Same Quantization Matrix[J]. *IEEE Transactions on Information Forensics and Security*, 2014, 9(11): 1933-1942.
- [137] Luo W Q, Qu Z H, Huang J W, et al. A Novel Method for Detecting Cropped and Recompressed Image Block[C]. *2007 IEEE In-*

- ternational Conference on Acoustics, Speech and Signal Processing - ICASSP '07, 2007: II-217.
- [138] Chen Y L, Hsu C T. Image Tampering Detection by Blocking Periodicity Analysis in JPEG Compressed Images[J]. *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008: 803-808.
- [139] Chen Y L, Hsu C T. Detecting Recompression of JPEG Images via Periodicity Analysis of Compression Artifacts for Tampering Detection[J]. *IEEE Transactions on Information Forensics and Security*, 2011, 6(2): 396-406.
- [140] Qu Z H, Luo W Q, Huang J W. A Convolutional Mixing Model for Shifted Double JPEG Compression with Application to Passive Image Authentication[C]. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008: 1661-1664.
- [141] Bianchi T, Piva A. Detection of Nonaligned Double JPEG Compression Based on Integer Periodicity Maps[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(2): 842-848.
- [142] Bianchi T, Piva A. Image Forgery Localization via Block-Grained Analysis of JPEG Artifacts[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(3): 1003-1017.
- [143] Wang S L, Liew A W C, Li S H, et al. Detection of Shifted Double JPEG Compression by an Adaptive DCT Coefficient Model[J]. *EURASIP Journal on Advances in Signal Processing*, 2014, 2014: 101.
- [144] de Carvalho T J, Riess C, Angelopoulou E, et al. Exposing Digital Image Forgeries by Illumination Color Classification[J]. *IEEE Transactions on Information Forensics and Security*, 2013, 8(7): 1182-1194.
- [145] Georgiades A S, Belhumeur P N, Kriegman D J. From few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 643-660.
- [146] Gross R, Matthews I, Cohn J, et al. Multi-PIE[J]. *Image and Vision Computing*, 2010, 28(5): 807-813.
- [147] Ng T, Chang S, Sun Q. A data set of authentic and spliced image blocks, Columbia University, ADVENT Technical Report, 203-2004, 2004.
- [148] Dong J, Wang W, Tan T N. CASIA Image Tampering Detection Evaluation Database[C]. *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 2013: 422-426.
- [149] Wen B H, Zhu Y, Subramanian R, et al. COVERAGE—A Novel Database for Copy-Move Forgery Detection[C]. *2016 IEEE International Conference on Image Processing*, 2016: 161-165.
- [150] Nist nimble 2016 datasets, <https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation/>, 2016.
- [151] Hsu Y F, Chang S F. Detecting Image Splicing Using Geometry Invariants and Camera Characteristics Consistency[C]. *2006 IEEE International Conference on Multimedia and Expo*, 2006: 549-552.
- [152] Gloe T, Böhme R. The 'Dresden Image Database' for Benchmarking Digital Image Forensics[C]. *The 2010 ACM Symposium on Applied Computing - SAC10*, 2010: 1584-1590.
- [153] Bas P, Furon T. Break Our Watermarking System July 2007, <http://bows2.gipsa-lab.inpg.fr>, 2nd ed, 2007.
- [154] Available:<http://boss.gipsa-lab.grenoble-inp.fr/BOSSRank/index.php?mode=VIEW&tmpl=materials>.
- [155] Bas P, Filler T, Pevný T. "Break our Steganographic System": The Ins and Outs of Organizing BOSS[M]. *Information Hiding*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 59-70.
- [156] Schaefer G, Stich M. UCID: An Uncompressed Color Image Database[C]. *Proc SPIE 5307, Storage and Retrieval Methods and Applications for Multimedia 2004*, 2003, 5307: 472-480.
- [157] NRCS Photo Gallery, <http://photogallery.nrcs.usda.gov>, accessed Sep. 2012.
- [158] Tralic D, Zupancic I, Grgic S, et al. CoMoFoD—New Database for Copy-Move Forgery Detection[J]. *Proceedings ELMAR-2013*, 2013: 49-54.
- [159] Amerini I, Ballan L, Caldelli R, et al. Copy-Move Forgery Detection and Localization by Means of Robust Clustering with J-Linkage[J]. *Signal Processing: Image Communication*, 2013, 28(6): 659-669.
- [160] Zhou P, Han X T, Morariu V I, et al. Learning Rich Features for Image Manipulation Detection[C]. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 1053-1061.
- [161] Krawetz N, Solutions H[J]. A Picture's Worth, *Hacker Factor Solutions*, 2007, 6: 2.
- [162] Mahdian B, Saic S. Using Noise Inconsistencies for Blind Image Forensics[J]. *Image and Vision Computing*, 2009, 27(10): 1497-1503.
- [163] Salloum R, Ren Y Z, Jay Kuo C C. Image Splicing Localization Using a Multi-Task Fully Convolutional Network (MFCN)[J]. *Journal of Visual Communication and Image Representation*, 2018, 51: 201-209.
- [164] Bayar B, Stamm M C. A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer[C]. *The 4th ACM Workshop on Information Hiding and Multimedia Security*, 2016: 5-10.
- [165] Zhang Y, Goh J, Win L, et al. Image Region Forgery Detection: A Deep Learning Approach[C]. *SG-CRC*, 2016: 1-11.
- [166] Zhang Y, Thing V L L. A Semi-Feature Learning Approach for Tampered Region Localization across Multi-Format Images[J]. *Multimedia Tools and Applications*, 2018, 77(19): 25027-25052.
- [167] Bappy J H, Roy-Chowdhury A K, Bunk J, et al. Exploiting Spatial Structure for Localizing Manipulated Image Regions[C]. *2017 IEEE International Conference on Computer Vision*, 2017: 4980-4989.
- [168] Bappy J H, Simons C, Nataraj L, et al. Hybrid LSTM and Encoder-Decoder Architecture for Detection of Image Forgeries[J]. *IEEE Transactions on Image Processing*, 2019, 28(7): 3286-3300.



张怡暄 于 2013 年在华中科技大学光信息科学与技术专业获得学士学位, 现在中国科学院大学信息工程研究所信号与信息处理专业攻读硕士学位, 研究兴趣包括人工智能、多媒体取证技术。Email: zhangyixuan@iie.ac.cn



赵险峰 中国科学院信息工程研究所研究员, 中国科学院大学网络空间安全学院教授, 博士生导师。2003 年于上海交通大学获博士学位, 研究方向为信息隐藏、多媒体取证与内容安全分析等。任 IJDCF、IWDW 等期刊、会议的编委、主席或委员, 任中国电子学会通信与信息安全专委会、中国图象图形学会多媒体取证与安全专委会等学术组织的委员。曾承担国家自然科学基金、国家重点研发计划、中科院战略性先导专项、部委专项等任务 40 余项, 在 IEEE TIFS、ACM IH & MMSec 等本领域重要刊物和会议上发表论文 150 余篇, 获得与申请专利 29 项, 撰写或参与撰写著作 5 部, 主持研制的系统有重要应用, 获保密科学技术奖(部级)一等奖、中科院“朱李月华”优秀教师、ACM IH & MMSec 最佳论文奖等荣誉。Email: zhaoxianfeng@iie.ac.cn



曹纭 于 2012 年在中国科学院软件研究所获得博士学位。现任中国科学院信息工程研究所副研究员。研究领域为多媒体内容安全。研究兴趣包括: 隐写与隐写分析、数字内容取证等。Email: caoyun@iie.ac.cn