

# An Image Tampering Detection Algorithm Based on the Posterior Probability and Color Filter Array Artifacts

Jinqiao Wei, Ying Wang

Electronic Information College, Qingdao University, Qingdao Shandong  
Email: [jinqiao199299@163.com](mailto:jinqiao199299@163.com), [wangmingzhou01@126.com](mailto:wangmingzhou01@126.com)

Received: Aug. 29<sup>th</sup>, 2017; accepted: Sep. 9<sup>th</sup>, 2017; published: Sep. 13<sup>th</sup>, 2017

---

## Abstract

Focused on the artifacts between the three primaries of an image introduced by the interpolation algorithm during its acquisition process, an image tampering detection algorithm based on the posterior probability and the color filter array artifacts is proposed. Firstly, the green channel component of the image is extracted, and the two-dimensional predictive filter is used to construct the predictive error function. Then the histograms' character of original and tampering images is analyzed, and then the Gaussian mixture statistical model is established. EM algorithm is applied to estimate the model parameters. Then the posterior probability of each sub-block as an original block is calculated, and the feature likelihood is defined and it is applied to every sub-block, so that the tampering-area map can be obtained to complete the detection. The simulation results show that the algorithm has strong robustness and can locate the image's tampered region more accurately.

## Keywords

Posterior Probability, Color Filter Array Artifacts, Gaussian Mixture Model, Likelihood Ratio, Image Tempering Detection

---

# 基于后验概率和滤色阵列特性的图像篡改检测算法

魏金巧, 王 英

青岛大学电子信息学院, 山东 青岛  
Email: [jinqiao199299@163.com](mailto:jinqiao199299@163.com), [wangmingzhou01@126.com](mailto:wangmingzhou01@126.com)

收稿日期: 2017年8月29日; 录用日期: 2017年9月9日; 发布日期: 2017年9月13日

## 摘要

针对图像采集过程中插值算法对图像三原色之间所引入的插值特性, 论文提出一种基于后验概率和滤色阵列特性的图像篡改检测算法。首先提取待测图像绿色通道分量, 引入二维预测滤波构建预测误差函数; 然后分块提取特征, 分析原始与篡改图像所提特征的直方图特性, 从而建立特征的高斯混合统计模型, 并借助EM算法估计模型参数; 计算子块特征作为原始块的后验概率, 定义似然率并应用到每个子块中, 从而可得到篡改区域映射图, 完成本次检测。仿真结果表明, 该算法具有较强的鲁棒性, 能够对图像篡改区域进行较准确地定位。

## 关键词

后验概率, 滤色阵列特性, 高斯混合模型, 似然率, 图像篡改检测

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

图像取证是一项多学科综合的科学, 其旨在获取与数字图像相关的重要“历史”信息, 主要包括其编码进程和可能的编辑修改过程[1]。随着图像编辑软件的日益强大, 人们能够很轻易地修饰和篡改一幅图像。因此, 鉴别图像的真伪及篡改区域的精细定位变得日趋重要。

T. Bianch 等[2]对 JPEG 图像压缩过程进行建模, 使用后验概率法估计量化步长, 最后通过篡改区域和未篡改区域量化表的非一致性来定位篡改区域。Swaminathan 等[3]提出了一种通过估计 CFA 模式和插值内核进行识别相机型号的方法, 紧接着又在文献[4]中将去马赛克模型中估计出的参数之间的非一致性作为篡改检测的依据。H. Farid 等[5]提出用期望最大化算法来估计去马赛克算法内插核参数, 实现了借助每个像素与其相邻像素相关性的概率图来检测篡改区域, 但这种方法需要预先知道图像篡改区域的尺寸。S. Vinoth 等[6]提出将反向传播神经网络作为非线性模型来描述 CFA 插值特性并进行分类的方法。Peng Shuang 等[7]提出一种利用自然图像颜色通道之间的相关性来检测图像篡改操作的方法, 但对经模糊篡改的检测存在一定的误差。

面对错综复杂的图像篡改方式, 很多方法的鲁棒性很差甚至会失效。本文从图像生成的过程出发, 通过分析数码相机的一般成像体系(如图 1)发现, 图像经过滤色阵列(Color Filter Array, CFA)插值后三颜色分量具有很强的相关性, 而且这种相关性会在图像的原始区域表现得非常明显, 但在被篡改过的图像区域却会消失[8]。所以, 可以此作为判断图像篡改与否的依据。

基于文献[5] [6] [7]以及对成像体系的分析, 发现 CFA 插值过程引入的噪声会使图像篡改检测的精度降低, 而且一般的算法对模糊篡改后的平滑区域的检测总会有误差。所以, 本文提出一种基于后验概率和滤色阵列特性的图像篡改检测方法, 对篡改区域实行较高精度的定位。首先提取待测图像绿色通道分量, 引入二维预测滤波构建预测误差函数; 然后分块提取特征, 分析原始与篡改图像所提特征的直方图

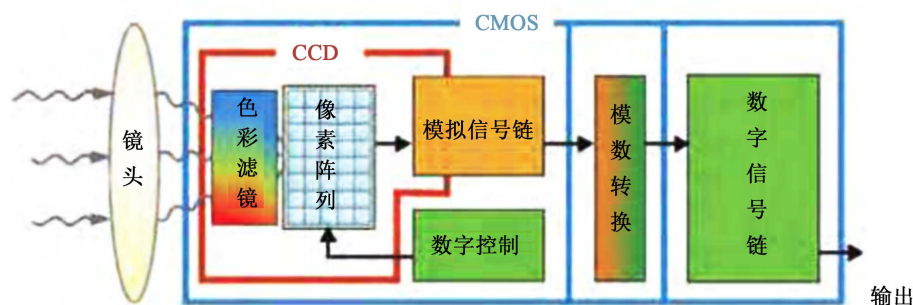


Figure 1. Digital camera general imaging system

图 1. 数码相机一般成像体系

特性, 从而建立特征的高斯混合统计模型(Gaussian Mixture Model, GMM), 并借助 EM (Expectation Maximization)算法估计模型参数; 计算块特征作为原始块的后验概率, 定义似然率并应用到每个子块中, 从而可得到篡改区域映射图, 映射图上每个像素的灰度值分别对应于待测图像上该子块的似然率值, 该值越小, 表明该区域被篡改的可能性越高, 而被篡改过的区域将以灰度值极低的黑色区域显现出来, 完成本次检测。

## 2. 滤色阵列

数码相机的图像传感器通过 CFA 来感应不同光的强度。现存的阵列, 包括 Bayer CFA, Mosaic CFA, Stripe CFA, 其中 Bayer CFA 作为最经典的阵列, 应用最为广泛, 它交替使用一组红色和绿色滤镜及一组绿色和蓝色滤镜, 其中绿色像素是红、蓝像素数量的两倍, 如图 2(a)所示。这是由于人眼对于绿色更为敏感, 且绿色在可见光谱中占据着最重要和最宽的位置。所以本文的研究以绿色通道分量为焦点。

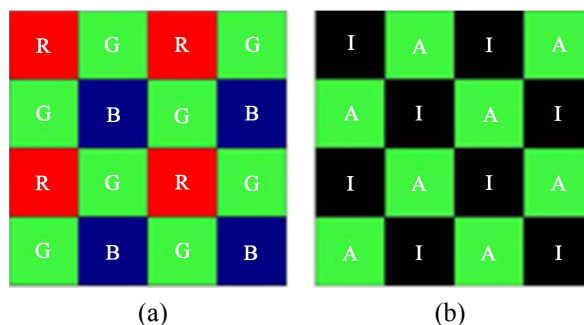
Bayer CFA 传感器的原始输出图像的每个像素只有红、绿、蓝中的一种颜色分量, 俗称“马赛克图像”, 因此, 必须采用插值算法将单色的马赛克图像转换成逼真的全彩色图像, 这个过程通常被称为“去马赛克(Demosaicing)”。插值通常在缺失值的邻域应用一个加权矩阵(核)来实现, 常见的 CFA 插值算法分两类: 一类是以双线性 and 双三次插值[9]为代表的非自适应插值算法, 另一类则是以梯度插值[10]为代表的自适应插值算法。但无论采用哪种插值算法, 数码相机拍摄出的原始自然图像往往都会存在一定的滤色阵列特性, 也称“去马赛克效应”, 而被修改过的图像区域则不具备这些特性。

所提取绿色通道的像素构成如图 2(b)所示, 获得像素和插值像素分别对应于晶格 A 和晶格 I。如果图像是原始的自然图像, 则晶格 A 和 I 的绿色像素值具有一定的规律, 即晶格 A 处绿色像素的预测误差方差要高于晶格 I 处; 相反, 如果图像被篡改过, 这种规律被破坏, 则两个晶格的预测误差方差会变成相似的值。因此, 可通过评估两类不同晶格处绿色像素值的预测误差方差的不平衡性, 来检测 CFA 插值特性的存在与否, 从而判定图像是否受到篡改。

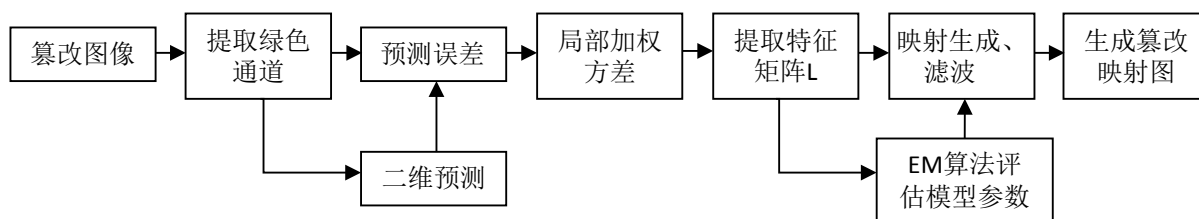
## 3. 算法介绍

本文所提算法的检测流程如图 3 所示。检测算法主要包含: 1) 提取待测图像绿色通道分量; 2) 计算预测误差及其加权方差; 3) 建立篡改特征统计模型; 4) 估计特征模型参数; 5) 引入后验概率模型, 定义似然率, 输出篡改映射图。

这里引入权重因子和二维预测器来构建预测误差及其加权方差的统计模型, 从模型中的不重叠像素块中提取出特征。如果这个特征明显大于零值, 则认为它是一个原始像素块; 反之, 如果特征很接近于零值, 则认为该子块为一个篡改块。



**Figure 2.** Bayer CFA and distribution of acquired and interpolated pixels of green channel  
**图 2.** Bayer CFA 及绿色通道获得和插值像素分布



**Figure 3.** Detecting process of the proposed algorithm  
**图 3.** 所提算法的检测流程

### 3.1. 特征提取

待测图像的绿色通道分量为  $s(x, y)$ ,  $(x, y) \in Z^2$ , 引入二维预测滤波, 得其预测误差为:

$$e(x, y) = s(x, y) - \sum_{u, v=0} k_{u, v} s(x+u, y+v) \quad (1)$$

其中  $k_{u, v}$  为二维预测滤波,  $\forall (u, v)$  为像素  $s(u, v)$  周围像素的横、纵坐标。在理想情况下, 可令  $k_{u, v} = h_{u, v}$ , 其中  $h_{u, v}$  为插值算法的插值内核。由于相机的插值算法往往是未知的, 所以令  $k_{u, v} \neq h_{u, v}$ 。

因为余量局部平稳, 并且预测误差  $e(x, y)$  的方差在插值或获得像素的邻域像素位置可以被逐个局部估计, 所以预测误差  $e(x, y)$  具有局部平稳性, 假设其在  $(2K+1) \times (2K+1)$  窗口内有效, 则其局部加权方差可定义为:

$$\sigma_e^2 = \frac{1}{c} \left[ \left( \sum_{i, j=-K}^K \alpha_{ij} e^2(x+i, y+j) \right) - (\mu_e)^2 \right] \quad (2)$$

其中  $\mu_e = \sum_{i, j=-K}^K \alpha_{ij} e(x+i, y+j)$  为预测误差的局部方差均值;  $c = 1 - \sum_{i, j=-K}^K \alpha_{ij}^2$  为一尺度因子, 可使参数评估更准确, 例如对于每一个像素都有  $E[\sigma_e^2(x, y)] = \text{var}[e(x, y)]$ ;  $\alpha_{ij}$  是合适的权重因子:

$$\alpha_{ij} = \alpha'_{ij} / \sum_{i, j} \alpha'_{ij} \quad (3)$$

$$\alpha'_{ij} = \begin{cases} W(i, j), & \text{if } e(x+i, y+j) \in e(x, y) \\ 0, & \text{else} \end{cases} \quad (4)$$

式中,  $W(i, j)$  为一  $(2K+1) \times (2K+1)$  的高斯窗, 其标准偏差为  $K/2$ 。

若待测图像大小为  $N \times N$ , 将其分割成  $B \times B$  大小的非重叠子块,  $B$  与 Bayer CFA 的周期相关, 最小块尺寸为  $2 \times 2$ 。任意位置  $(k, l)$  处的子块  $B_{k, l}$  ( $k, l = 0, \dots, (N/B) - 1$ ) 都是由获得像素和插值像素的非相交集  $B_{A_{k, l}}$ 、 $B_{I_{k, l}}$  构成, 由此提取特征  $L$  并定义为:

$$L(k,l) = \log \left[ \frac{GM_A(k,l)}{GM_I(k,l)} \right] \quad (5)$$

$$GM_A(k,l) = \left[ \prod_{i,j \in B_{k,l}} \sigma_e^2(i,j) \right]^{\frac{1}{|B_{k,l}|}} \quad (6)$$

$$GM_I(k,l) = \left[ \prod_{i,j \in B_{k,l}} \sigma_e^2(i,j) \right]^{\frac{1}{|B_{k,l}|}} \quad (7)$$

其中,  $GM_A(k,l)$  和  $GM_I(k,l)$  分别为获得像素和插值像素预测误差局部加权方差的几何均值。式(5)中的  $L(k,l)$  是非连续的, 可用于评估预测误差局部方差的平衡性。若待测块是未被篡改过的原始块, 则其获得像素的局部方差要远大于插值像素的局部方差, 即特征  $L(k,l)$  的期望值为一个正量; 反之, 当图像块被篡改过, 则获得像素与插值像素局部方差之间的差异消失, 特征  $L(k,l)$  的期望值为 0。

### 3.2. 建立特征模型

为了准确描述提取出的特征, 需建立一个简单的统计模型, 对每个子块  $B_{k,l}$ , 应用贝叶斯的方法可评估其 CFA 插值特性出现的概率。假设存在 CFA 插值特性时为  $M_1$ , 不存在 CFA 插值特性时为  $M_2$ 。已知篡改图像和未篡改图像中提取出的特征  $L$  的统计直方如图 4 所示。

经分析, 特征  $L(k,l)$  在这两种假设下均服从高斯分布。若分块尺寸  $B$  固定, 可以用下面这两个条件概率密度函数来描述所提特征:

$$P_r\{L(k,l)|M_1\} = N(\mu_1, \sigma_1^2) \quad (8)$$

$$P_r\{L(k,l)|M_2\} = N(0, \sigma_2^2) \quad (9)$$

$\forall k,l=0, \dots, (N/B)-1$ ,  $P_r\{L(k,l)|M_1\}$  和  $P_r\{L(k,l)|M_2\}$  分别表示  $L(k,l)$  在假设  $M_1$ 、 $M_2$  下的先验概率。 $\mu_1 > 0$  代表该区域存在 CFA 插值特性,  $\mu_1 = 0$  则代表该区域 CFA 插值特性已被擦除。如果一幅图像包含了一些 CFA 特征被破坏的篡改区域, 则以上两种假设均存在, 此时特征  $L(k,l)$  服从高斯混合分布(如图 4(b)示)。为了估计出该高斯混合模型的参数, 引入 EM 方法来估计参数  $\mu_1, \sigma_1, \sigma_2$  的值。

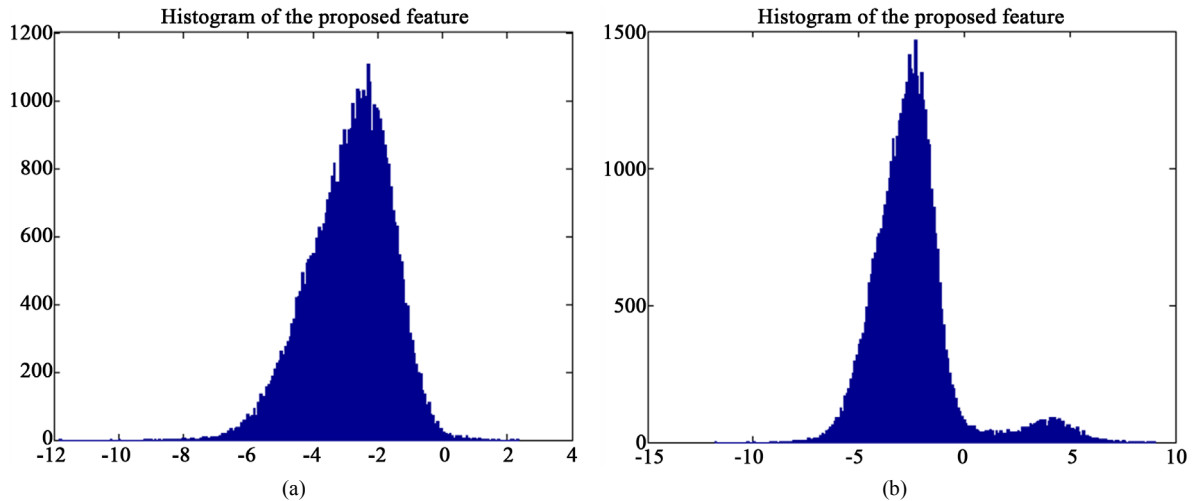


Figure 4. Histograms of proposed features. (a) feature histogram of original image; (b) feature histogram of tampered image  
图 4. 所提特征统计直方图。(a)原始图像特征统计直方图; (b)篡改图像特征统计直方图



### 3.3. 生成篡改映射图

本算法最终的目标是针对每一个子块, 都能基于 CFA 插值特性的存在与否来获得它是原始块的概率, 并以不同子块灰度值的高低显示, 从而来定位出篡改区域。根据(8)式和(9)式, 且已知先验概率  $P_r\{M_1\} = P_r\{M_2\} = 1/2$ , 则可得该子块作为原始块的后验概率为:

$$P_r\{M_1|L(k,l)\} = \frac{P_r\{L(k,l)|M_1\}}{P_r\{L(k,l)|M_1\} + P_r\{L(k,l)|M_2\}} \quad (10)$$

也可改写成:

$$P_r\{M_1|L(k,l)\} = \frac{1}{1 + \xi(L(k,l))} \quad (11)$$

$\xi(L(k,l))$  为  $L(k,l)$  的似然率, 定义式为:

$$\xi(L(k,l)) = \frac{P_r\{L(k,l)|M_2\}}{P_r\{L(k,l)|M_1\}} \quad (12)$$

(11)式和(12)式具有相同的统计意义。将(12)式作用于图像的每个子块中, 则可得到一映射图, 图中的每个像素值分别对应于一个  $B \times B$  块的似然率, 且似然率越低, 则该子块为篡改块的概率越高。然而, 似然率图往往包含通道噪声, 为了消除这些噪声, 需在更大的块上计算其累计特征值, 块大小为  $C \times C$ , 且满足  $C = n \times B$ 。假设  $M_1$  和  $M_2$  之间条件独立, 则可定义累计似然率为:

$$\xi_{\text{accumu}}(L(k,l)) = \frac{\prod_{k,l} P_r\{L(k,l)|M_2\}}{\prod_{k,l} P_r\{L(k,l)|M_1\}} \quad (13)$$

在实际篡改图像中, 篡改区域通常是连通的。为了进一步提高定位性能, 对生成的似然图先取对数后应用一个中值滤波器, 既能保证数值的稳定性, 还能将这些关联区域都强调出来。

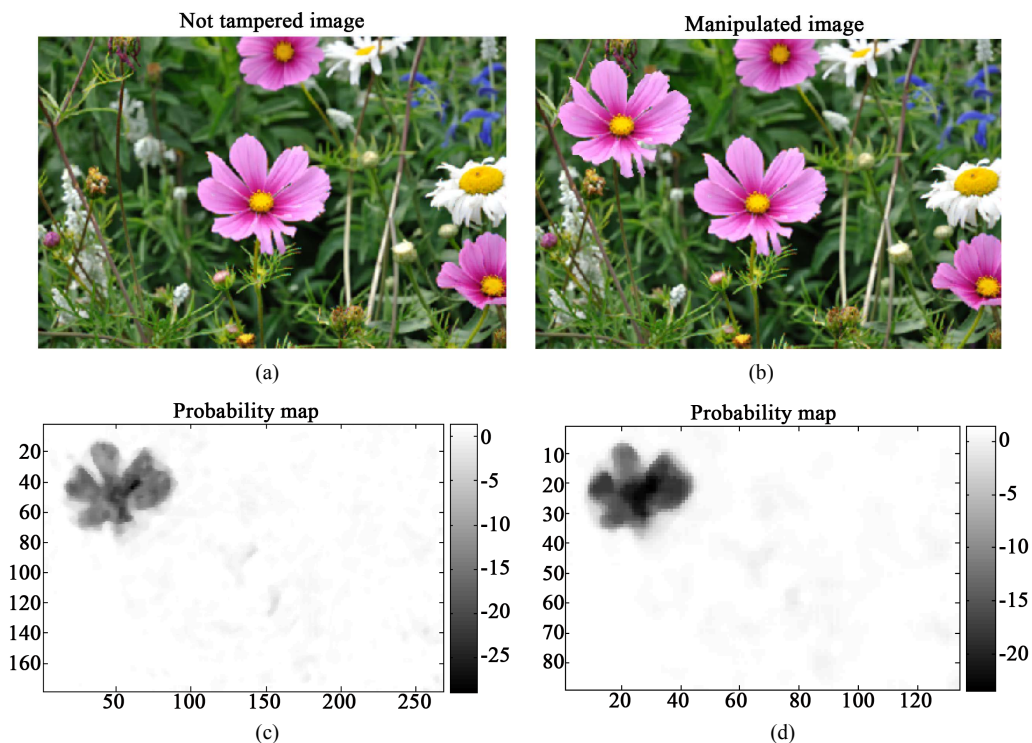
## 4. 仿真结果及分析

为了检测本文算法对图像篡改区域的定位性能, 在篡改图像库中进行实验。原始样本均由 Nikon D90 获得, 经 PS 软件对样本进行复制-粘贴篡改, 同时对图像做后处理操作, 如平滑, 滤波, 尺度、旋转等可更好的移除篡改区域的 CFA 特征。关键参数设置:  $\alpha = 0.5$ ,  $\sigma_2^2 = \sigma_1^2/10$ ,  $C = B = 2$ 。与文献[11]所提出方法进行了检测效果的对比, 图 5 和图 6 为两种方法仿真结果的对比图。

从图 5 可看出, 两种方法都能定位出篡改区域, 但文献[11]方法篡改区域的边界污染较严重, 即检测的精度不够高, 本文算法对篡改区域的边界轮廓定位得很精确, 且本文算法和文献[11]方法输出映射图的最低似然率分别为-28.8988 和-23.4673, 也充分说明了本文算法检测性能的优越性; 图 6 中的篡改图像后期经过了模糊处理, 虽然两种算法的检测均存在较明显的误判区域, 即对天空处平滑区域的检测失效, 俩算法在篡改区域输出的最低似然率分别为-7.8577 (本文)和-6.5299 (文献[11]), 但本文算法较文献[11]算法而言, 依旧能够更清晰地将真正的篡改区域定位出来。综上可得, 本文所提算法较文献[11]算法的检测性能更好, 且精确度达到更高。

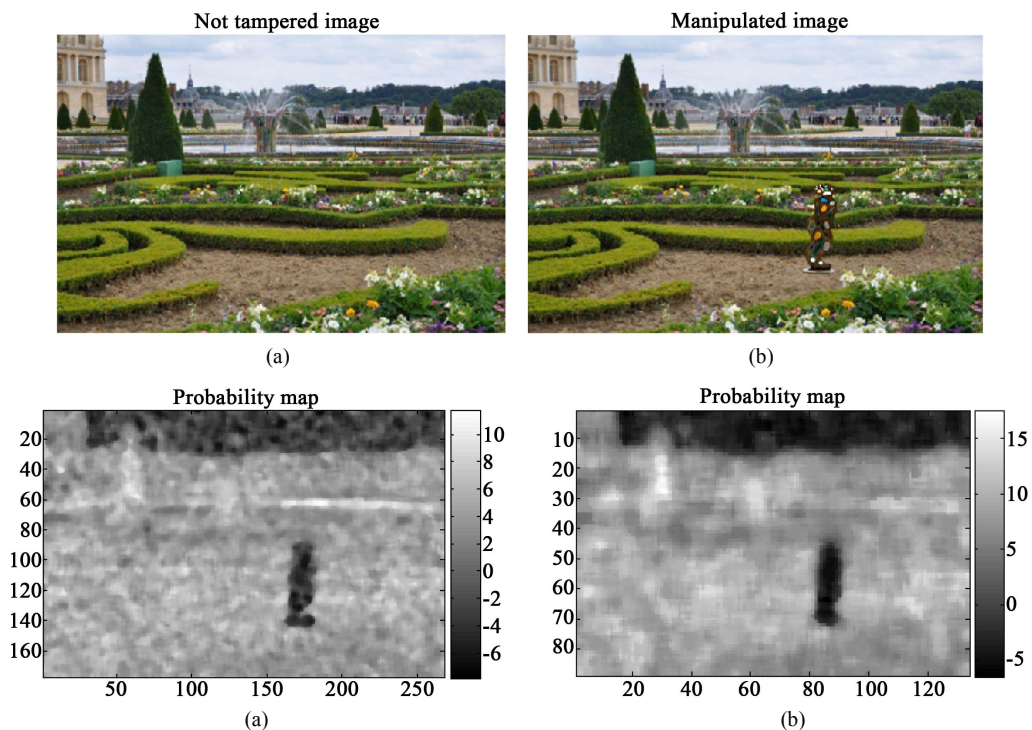
## 5. 结语

针对当前数字图像被动取证算法易受噪声影响致使检测精度不高情况, 本文提出了基于后验概率和滤色阵列特性的图像篡改检测算法。对每个子块中提取的特征引入其后验概率, 然后建立高斯混合模型,



**Figure 5.** Localization of tampered image area. (a) original image; (b) tampered image; (c) detecting result of proposed algorithm; (d) detecting result of algorithm in [11]

**图 5.** 图像篡改区域定位。(a)原始图像；(b)篡改图像；(c)本文算法检测结果；(d)文献[11]算法检测结果



**Figure 6.** Localization of tampered image area. (a) original image; (b) tampered image; (c) detecting result of proposed algorithm; (d) detecting result of algorithm in [11]

**图 6.** 图像篡改区域定位。(a)原始图像；(b)篡改图像；(c)本文算法检测结果；(d)文献[11]算法检测结果

定义似然率, 输出篡改映射图并中值滤波去噪, 最后输出的映射值越低表明被篡改的概率越大。实验结果显示该算法不仅能准确检测和定位篡改区域, 而且具有较强的鲁棒性。但是对图像模糊篡改进行检测时, 在平滑区域存在误检, 这也是后续研究工作中的重点之一。

## 致 谢

本研究是在我的导师王教授的亲切关怀和悉心指导下完成的。王教授不仅在学业上给我以精心指导, 同时还在思想、生活上给我以无微不至的关怀, 在此谨向她致以诚挚的谢意和崇高的敬意。我还要感谢同在一个实验室的小伙伴们, 正是由于你们的帮助和支持, 我才能克服一个一个的困难和疑惑, 直至本文的顺利完成。最后我还要感谢培养我长大含辛茹苦的父母, 谢谢你们!

最后, 再次对关心、帮助我的老师和同学表示衷心地感谢!

## 参考文献 (References)

- [1] 周琳娜, 王东明. 数字图像取证技术[M]. 北京: 北京邮电大学出版社, 2008: 1-6.
- [2] Bianchi, T. and Piva, A. (2012) Image Forgery Localization via Block-Grained Analysis of JPEG Artifacts. *IEEE Transactions on Information Forensics and Security*, 7, 1003-1017. <https://doi.org/10.1109/TIFS.2012.2187516>
- [3] Swaminathan, A., Wu, M. and Liu, K.R. (2007) Nonintrusive Component Forensics of Visual Sensors Using Output Images. *IEEE Transactions on Information Forensics and Security*, 2, 91-106. <https://doi.org/10.1109/TIFS.2006.890307>
- [4] Swaminathan, A., Wu, M. and Liu, K.J.R. (2008) Digital Image Forensics via Intrinsic Fingerprints. *IEEE Transactions on Information Forensics and Security*, 3, 101-117. <https://doi.org/10.1109/TIFS.2007.916010>
- [5] Popescu, A.C. and Farid, H. (2005) Exposing Digital Forgeries in Color Filter Array Interpolated Images. *IEEE Transactions on Signal Processing*, 53, 3948-3959. <https://doi.org/10.1109/TSP.2005.855406>
- [6] Vinoth, S., et al. (2013) Neural Network Modeling of Color Array Filter for Digital Forgery Detection Using Kernel LDA. *Procedia Technology*, 10, 498-504.
- [7] 彭双, 彭圆圆, 肖长燕. 基于 CFA 插值的图像篡改检测算法[J]. 传感器与微系统, 2015, 32(6): 0141-0144.
- [8] 张雯, 李学明. 改进的基于颜色滤波阵列特性的篡改检测[J]. 计算机工程与应用, 2009, 45(6): 176-179.
- [9] Ng, T.T., Chang, S.F. and Sun, Q. (2004) Blind Detection of Photomontage Using Higher Order Statistics. *Proceeding of the 2004 International Symposium on Circuits and Systems*, Columbia, 23-26 May 2004, 688-691.
- [10] Ng, T.T. and Chang, S.F. (2004) A Model for Image Splicing. *Proceedings on 2004 International Conference on Image Processing*, Singapore, 24-27 October 2004, 1169-1172.
- [11] Ferrara, P., Bianchi, T., Rosa, D.A., et al. (2012) Image Forgery Localization via Fine-Grained Analysis of CFA Artifacts. *IEEE Transactions on Information Forensics and Security*, 7, 1566-1577. <https://doi.org/10.1109/TIFS.2012.2202227>

### 期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)