

数字图像复制粘贴篡改取证

邢文博^{1,2} 杜志淳¹

(华东政法大学刑事司法学院 上海 200042)¹ (南京森林警察学院刑事科学技术学院 南京 210046)²

摘要 数字图像在被复制粘贴篡改时会对被复制部分进行缩放、旋转,然后粘贴在图像的不同部位,从而形成篡改图像。用 sift 算法检测出图像中的 sift 关键点,通过关键点的向量找出相互匹配的关键点对。用随机采样一致性算法在匹配点对中随机采取 3 个匹配点对计算其仿射变换矩阵,并对匹配点对进行循环分类,分类完成后按照每一类的仿射变换模型对篡改图像进行仿射变换以及仿射变换的逆变换,然后求得篡改图像与其仿射变换图像的局部相关图、篡改图像与其逆仿射变换图像的局部相关图。将每一类匹配关键点对按照仿射变换关系分为两组,从每组的关键点位置设置二值图像并用结构元素进行膨胀,将二值图像膨胀位置在相关图像的相关值大于阈值的膨胀部分保留,小于阈值的膨胀部分舍去,迭代膨胀到二值图像不再膨胀为止。然后获取二值图像的边界,在原图像中标出复制粘贴篡改部分。实验表明该方法可以有效定位篡改图像中的复制粘贴区域。

关键词 数字图像取证, sift 关键点, RANSAC 算法, 局部相关图, 复制粘贴篡改

中图分类号 TP391 文献标识码 A

Digital Image Forensics for Copy and Paste Tampering

XING Wen-bo^{1,2} DU Zhi-chun¹

(School of Criminal Justice, East China University of Political Science and Law, Shanghai 200042, China)¹

(School of Criminal Science and Technology, Nanjing Forest Police College, Nanjing 210046, China)²

Abstract When a digital image is tampered with copy and paste, the copied part might be zoomed, rotated and then pasted in different parts of the image. The sift key points are detected in the image by SIFT algorithm, and the matched key points are found out by matching vector of the key points. The matched key points are classified by the affine transformation matrix which is calculated by Random Sample Consensus, namely randomly extracting three matched point pairs in matched point pairs. The affine transformation of the image and its inverse affine transformation are carried out. Then the local correlation map of the tampered image and its affine transformation image is obtained, and the local correlation map of the tampered image and its inverse affine transformation image too. Each class of matched key points is divided into two groups according to affine transformation relations. A binary image is set from the key point of each group and dilated with structural elements. The dilated position of the binary image is continued to have when its value in the local correlation map is greater than the threshold and is get rid of when its value in the local correlation map is less than the threshold. The binary image is dilated iteratively until it's no longer expands and its boundary is marked on the original image. Experiments show that this method can effectively locate the copy paste tampering areas in digital images.

Keywords Digital image forensics, Sift key points, RANSAC algorithm, Local correlation map, Copy-pasted forgery

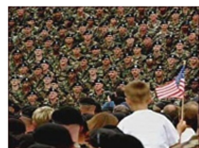
1 概述

图像是对客观实在的记录,“眼见为实”是人们最朴素的价值观。现在人们看到的大多是数字图像,但数字图像更容易被图像处理软件篡改,篡改痕迹也很难被人们发现,被篡改的数字图像改变了图像所记录的客观真实。将被篡改的图像作为新闻报道的图片、摄影的参赛作品、法庭的视听资料证据等,将会造成严重的影响。图 1(b)是 2004 年美国总统小布什在竞选时被人进行多次复制粘贴篡改的一张宣传照片,并在网络上流传。因此,当数字图像作为证据使用时,图像是否是对客观实在的真实记录即数字图像的真实性,就

成了人们最为关注的焦点。



(a) 原图像



(b) 多次篡改图像

图 1 小布什竞选时被篡改的宣传照片

图像的复制粘贴篡改操作是一种比较常见的图像篡改方法。它通过复制同一幅图像中的一部分内容,并将其直接粘贴到图像中的另一部分,以便掩盖图像中的内容,或者增加图

像中的内容。以掩盖图像内容为目的的篡改多是复制图像中的背景部分,一般复制图像中相似背景部分(如蓝天、沙滩、草地、墙面等)进行上下左右平移粘贴。以增加图像内容为目的的篡改多是复制图像中的前景物体,甚至对复制部分进行旋转、缩放、模糊、加噪,然后再选择图像中的目标部位进行粘贴,以达到自己期望的图像效果。

2 基于关键点匹配图像复制粘贴检测方法

图 2 是利用旋转缩放对图像进行复制粘贴篡改的模型, O 是图像的原始图像块, D_1 是对原始图像块 O 旋转放大后的复制图像块, D_2 是对原始图像块 O 旋转缩小后的复制图像块。

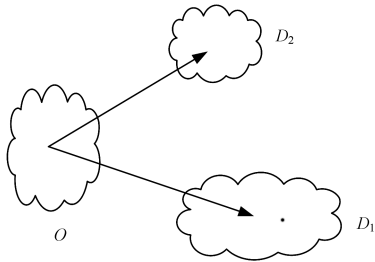


图 2 旋转缩放对图像进行复制粘贴篡改的模型

图像的复制粘贴篡改检测算法包括基于块匹配的检测算法和基于图像中关键点的检测算法^[1]。本文是基于关键点检测图像中的复制粘贴篡改。数字图像中的关键点有尺度不变特征变换算法 sift (Scale Invariant Feature Transform) 关键点^[2]、加速鲁棒性特征算法 surf (Speeded Up Robust Features) 关键点^[3], 以及 KAZE 算法关键点^[4]。他们共同的特征是检测出图像中的关键点具有旋转和缩放不变性。通过关键点向量的相似性找到匹配的关键点, 图像中复制粘贴篡改部分具有密集的匹配关键点。Huang^[5] 最先将图像的 sift 关键点用于图像的复制粘贴篡改检测。Ardizzone^[6] 对提取到关键点进行聚类, 并采用类匹配方法检测多次复制粘贴篡改区域。池静^[7] 等提出以 sift 特征点的聚类中心为种子点进行区域生长, 来检测图像中的复制粘贴篡改。Amerini^[8] 利用 J-Linkage 聚类检测图像中多次复制粘贴篡改部位。用聚类方法检测图像的复制粘贴篡改, 在确定特征点类别个数时存在用适当的距离阈值进行类别确定或根据散点图直观地确定类别个数。当复制粘贴部分相互重叠时, 聚类方法将不能正确区分复制粘贴部分的特征点。李云^[9] 等通过降低 sift 关键点的对比度阈值, 几乎不剔除对比度的极值点, 在平滑区域提取 sift 关键点。冢家昌^[10] 通过在图像的对空间提取 sift 关键点, 以及基于颜色不变量空间提取 sift 关键点, 来检测图像复制粘贴篡改。

本文对图像中检出的 sift 关键点进行匹配, 形成匹配点对集, 设定图像中复制粘贴部分的匹配点对阈值, 利用 RANSAC 算法从匹配点对集中随机选取 3 对匹配点对生成仿射变换模型进行分类, 用每一类的仿射变换模型对篡改图像进行仿射变换以及逆仿射变换, 再求得篡改图像与仿射变换图像以及逆仿射变换图像之间的相关图, 用篡改图像中复制部分与粘贴部分的匹配 sift 关键点生成两个二值图像, 并用结构元素对二值图像进行膨胀, 膨胀部分对应相关图的值大于设定阈值的保留, 小于阈值的舍弃, 迭代膨胀到二值图像不再膨胀为止, 获取篡改图像的复制粘贴区域, 并在篡改图像中定位。本文第 3 节具体说明算法的实现过程; 第 4 节介绍

算法编程并进行图像测试; 最后分得出结论。

3 篡改图像中 sift 匹配关键点对的 RANSAC 分类

尺度不变特征变换 (Scale Invariant Feature Transform, SIFT) 是 Lowe^[2] 于 1999 年提出并于 2004 年完善的基于尺度空间对图像平移、缩放、旋转等仿射变换保持不变的图像局部关键点描述向量。它通过建立图像的尺度空间, 并将尺度空间的极值点作为关键点, 利用关键点处的梯度值和方向为每个特征点指定方向参数, 并形成关键点的特征向量。

3.1 图像中的 sift 特征点

人类在识别一个物体时, 对一定距离范围内不同旋转角度的物体都能正确识别, 这就是尺度旋转不变性。对图像进行各种高斯模糊可以模拟实现这种尺度不变性, 所以首先构造图像的尺度空间。为了得到同一图像在各分辨率下的不变特征点, 需要对图像进行降采样处理, 得到一系列像素缩小的图像, 即图像金字塔。每一层图像由上一层分辨率长宽的一半即 $1/4$ 的像素组成。然后对图像金字塔中的每层图像用不同的高斯方差参数做高斯模糊, 使得每层金字塔有多张高斯模糊图, 形成图像的多尺度空间表示。金字塔每层的多张图像合成为一组, 每组有多张图像。对图像金字塔每组的每张高斯图像的相邻层进行差分, 得到高斯差分图像。其次, 在尺度空间中检测极值点。在寻找高斯差分图像的极值点时, 每一个像素都需要和与它在同一尺度以及相邻尺度的所有相邻点进行比较, 当其为极大值或极小值时即为关键点。由于这些关键点是在离散空间得到的, 而离散空间是对连续空间进行采样的结果, 因此在离散空间得到的关键点不一定是真正意义上的关键点。通过泰勒展开式多次迭代求得关键点的精确位置, 进而求得关键点的对比度值, 通过设置对比度阈值, 删除低对比度关键点。通过 Hessian 矩阵将主曲率比值大于一定阈值的边缘点删除, 得到优化后的关键点。接着, 要使关键点的特征向量具有旋转不变性, 就必须确定关键点的方向。关键点的方向是利用其邻域像素的梯度分布特性来确定, 再对图像的梯度直方图进行插值拟合处理, 以求得关键点局部结构的稳定方向。在构造关键点特征向量时, 将坐标轴旋转为关键点的方向, 从而实现特征向量的旋转不变性。最后, 生成关键点的描述向量。将图像的坐标轴旋转为关键点的方向, 在以关键点为中心的 16×16 的像素区域内以 4×4 像素为一个种子点, 共形成 16 个种子点, 求得每一个像素的梯度幅值与梯度方向, 每个种子点的梯度区域由 0 度到 360 度划分为 8 个方向区间, 每个区间为 45 度, 即每个种子点有 8 个方向的梯度强度信息。每个关键点产生 $4 \times 4 \times 8 = 128$ 个数据, 形成 128 维的 sift 特征向量。关键点的特征向量用于关键点的匹配。

3.2 关键点的特征向量匹配

关键点的匹配就是通过比较关键点特征向量之间的欧氏距离来实现, 用最近邻距离与次近邻距离的比值小于确定的阈值来确定关键点的相互匹配。在 Matlab 中, 为了提高计算效率, 利用关键点特征向量的点乘来代替特征向量之间的距离计算。在特征向量夹角比较小时, 特征向量夹角之间的比值近似于特征向量距离之间的比值。通过设置特征向量夹角比值的阈值来确定最近邻向量的匹配。设 α, β 是经过归一化处理的特征向量的特征向量, θ 是两个关键点特征向量之间的夹角, 则:

$$\alpha \cdot \beta = \cos \theta$$

$$\theta = \arccos(\alpha \cdot \beta)$$

在同一幅图像的复制粘贴检测中,假设检测到 n 个关键点,从 n 个关键点中取出一个关键点,将它与 n 个关键点的特征向量进行点乘,通过反余弦变换求得它们之间的夹角,然后进行降序排列,形成 $d_1, d_2, d_3, \dots, d_n$ 。排在第一个位置的夹角为 0,因为是关键点的检测向量与自己的夹角;排在第二位的是候选匹配向量夹角。在 sift 特征比对中,候选向量的夹角与比对向量夹角的比值小于一定阈值时才能确定候选向量为相似向量,这两个向量表示的关键点才是匹配关键点。sift 关键点匹配算法中设置的匹配向量夹角与比对向量夹角的比值小于 0.6,即 $d_2/d_3 < 0.6$,才能确定其关键点相互匹配。为了检测篡改图像中的多次复制粘贴篡改部分,需将 $d_1, d_2, d_3, \dots, d_n$ 序列中的所有匹配向量检测出来,用 d_i 表示 d_1, d_2, \dots, d_n 中的任意一项, i 取值为 2 到 $n-1$ 。从 d_2 开始,所有满足 $d_i/d_{i+1} < 0.6$ 的生成 d_i 的两个关键点为匹配关键点。然后将剩余关键点进行同样的运算,得到匹配点对集。

3.3 匹配点对的 RANSAC 算法分类

在同一幅复制粘贴篡改的图像中,除了复制粘贴区域有相互匹配的匹配关键点外,还会有部分非复制粘贴区域的具有相互匹配的关键点。复制粘贴区域的匹配关键点之间存在仿射变换关系,而非复制粘贴之间的匹配关键点将不存在仿射变换关系,用 RANSAC(RANdom SAmple Consensus)算法对存在仿射变换关系的匹配关键点进行分类,可以检测出图像中的多次复制粘贴区域,并滤除非复制粘贴区域的匹配关键点。

RANSAC^[11],即随机采样一致性统计算法。它是通过对一组包含异常数据的样本数据集进行随机采样计算出数据的数学模型,再通过计算符合数据模型的数据,得到有效样本数据的算法。

由于匹配关键点之间是相互匹配关系,设 a, b 是两个相互匹配的关键点, a 与 b 匹配,同理 b 也与 a 匹配, a, b 是一对匹配点。在所有的关键点都完成匹配后,会形成 a 与 b 匹配以及 b 与 a 匹配的两个匹配点对,按照匹配点的坐标将匹配点对调整成从左到右、从上到下的顺序,删除匹配对中重复匹配的匹配点对。仿射变换的数据模型需要 3 个匹配点对来确定。用 RANSAC 算法随机采取 3 个匹配点对,因为匹配点对的两个匹配点分别取自篡改图像中的复制部分与粘贴部分,如果篡改图像中复制部分的 3 个点与粘贴部分 3 个点不在同一条直线上,则复制部分的 3 个点与粘贴部分 3 个点各自构成三角形,则根据这 3 个匹配点对求出一个符合仿射变换的变换模型。对所有的匹配点对从左面或上面的关键点进行仿射变换,得到其仿射变换后点的位置,再与其匹配的关键点的位置进行比较。通过设定的误差阈值判定符合仿射变换的匹配点对,当符合该仿射变换模型的匹配点对数超过设定的阈值时,将所有符合该仿射变换模型的匹配点对从匹配点对集中提取出来作为一类;然后对剩余的匹配点对再用同样的算法提出符合另一个仿射变换的所有匹配点对,并作为第二类,直到剩余的匹配点对少于设定的阈值,或随机采样次数达到设定的值,退出 RANSAC 算法对匹配点对的分类。

3.4 篡改图像中复制粘贴区域的定位

篡改图像中复制粘贴区域之间存在仿射变换关系以及逆仿射变换关系,根据 3.3 节求得的仿射变换关系对篡改图像

进行仿射变换以及逆仿射变换,得到仿射变换图像以及逆仿射变换图像。篡改图像中的粘贴部分与仿射变换图像的对应区域将有较高的相关性,篡改图像的复制区域将与逆仿射变换图像的对应区域有较高的相关性。将篡改图像、仿射变换图像、逆仿射变换图像转换为灰度图像,求得篡改图像与仿射变换图像的局部相关图,再求得篡改图像与逆仿射变换图像的局部相关图。图像局部的相关性由下面的相关性公式给出。

$$\gamma = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}}$$

其中, γ 表示 $m \times n$ 图像块 A 与 $m \times n$ 图像块 B 的相关性, \bar{A} 表示图像块 A 的像素均值, \bar{B} 表示图像块 B 的均值。

按照相关性公式将篡改图像分别与仿射变换图像以及逆仿射变换图像以 3×3 的图像块求得相关图。将每一类匹配点对按照仿射变换关系分成对应的两个关键点组,并对每一个组关键点的坐标进行取整,生成关键点二值图像。对关键点二值图像用 3×3 的结构元素进行膨胀,将二值图像膨胀位置在相关图中的相关值大于阈值的膨胀区域保留,小于阈值的膨胀区域舍去(本文设置阈值为 0.7),然后对二值图像重复进行上述过程,直到二值图像不再膨胀为止。然后获取二值图像的边界,在原图像中标出复制粘贴篡改区域。

4 实验结果与分析

4.1 Matlab 编程检测

sift 特征点的检测是用 siftDemoV4^[12] Matlab 程序进行的,求得 sift 特征点的位置以及特征点的向量表示,并对特征向量进行归一化。对于每一个特征向量,求其与其他特征向量的余弦夹角,并进行升序排序,从最小夹角开始,将其与紧挨着的夹角进行比较,并将比值小于一定阈值的两个向量设定为匹配向量。舍弃匹配向量距离小于 10 个像素的匹配点。匹配完成后,形成匹配点对矩阵。因为被匹配的 sift 特征点与其匹配的特征点会再次形成匹配点对,所以将匹配点对调整成从左到右的匹配点对形式,删除重复的匹配点对。再将匹配点对的位置进行标准化处理,便于用 RANSAC 算法对匹配点对进行分类。

匹配点对数量比较集中的区域就是图像复制粘贴区域的 sift 匹配关键点,这些点对之间存在仿射变换关系。用随机采样一致性算法在匹配的点对中随机选取 3 个匹配点对,求得仿射变换关系,对匹配点对中左面的关键点进行仿射变换,求其仿射变换点的位置,并与其对应的匹配关键点的位置进行比较。在设定的误差范围内(因为对匹配点的位置进行了归一化,本文将其设置为 0.05),符合仿射变换关系的匹配点对达到设定阈值的匹配点数时(由于图 23 中的匹配点对少,所以对图 23 检测时设置为 5,对其他图像进行检测时设置为 10),认定这种仿射变换关系成立,并将这些匹配点对划归一类单独存储。在匹配点对中删除符合这种仿射变换关系的匹配点对,并对剩余的匹配点对再次进行随机采样,寻找仿射变换关系式,直到剩余的匹配点对小于设定的阈值(本文设置阈值为 3)或随机采样次数达到设定的值(本文设置为 1000)时,结束对匹配点对寻找仿射变换的关系。图 3 是算法的流程图。

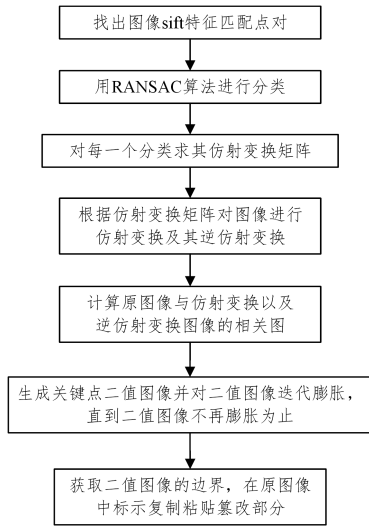


图 3 复制粘贴检测算法流程图

下面以网络上流行的小布什竞选时的复制粘贴篡改照片(图像大小为 370×278)^[13] 来为例说明该算法的检测过程。图 4 是 siftDemoV4 算法检测到的 sift 关键点。图 5 是 sift 关键点的匹配结果。图 6 是 RANSAC 算法检测到的一类仿射变换匹配点。图 7 是 RANSAC 算法检测到的另一类仿射变换匹配点。图 8 是 RANSAC 算法检测到的两处粘贴部分的匹配关键点。图 9 是剩余的不符合仿射变换关系的匹配关键点。图 10 是篡改图像的一类仿射变换图像与逆仿射变换图像。图 11 是篡改图像的灰度图像分别与一类仿射变换与逆仿射变换灰度图像的相关图。图 12 是篡改图像中的一类复制区域和粘贴区域的匹配关键点。图 13 是篡改图像中的一类复制区域和粘贴区域。图 14 是篡改图像中另一类仿射变换与逆仿射变换图像。图 15 是篡改图像分别与另一类仿射变换与逆仿射变换图像的相关图。图 16 是篡改图像中另一类复制区域和粘贴区域的匹配关键点。图 17 是篡改图像中的另一类复制区域和粘贴区域。图 18 是篡改图中粘贴部分仿射变换与逆仿射变换图像。图 19 是篡改图分别与粘贴部分仿射变换与逆仿射变换图像的相关图。图 20 是篡改图像中两个粘贴区域的关键点。图 21 是篡改图像中的粘贴区域。图 22 是篡改图像中的复制区域及粘贴区域。图 23 是复制粘贴篡改图像的检测结果。



图 4 被篡改图中检测到的 sift 特征点

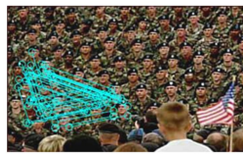


图 5 相互匹配的 sift 特征点

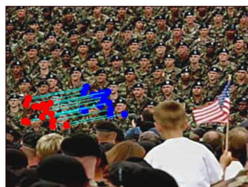


图 6 RANSAC 算法检测到的一类仿射变换匹配点

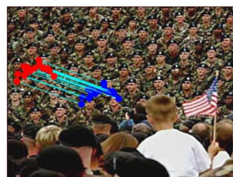


图 7 RANSAC 算法检测到另一类仿射变换匹配点

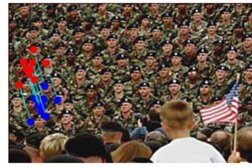


图 8 RANSAC 算法检测到的粘贴部分的仿射变换匹配点

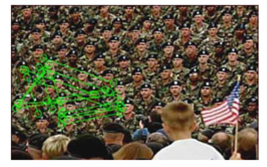
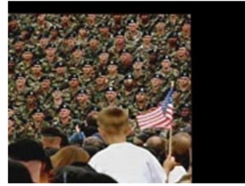


图 9 剩余的不符合仿射变换的匹配点

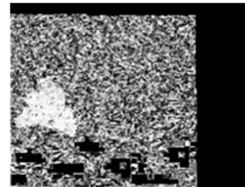


(a) 仿射变换图像

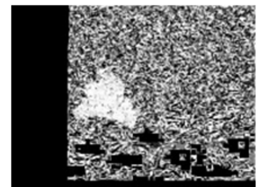


(b) 逆仿射变换图像

图 10 篡改图像的一类仿射变换图像与逆仿射变换图像

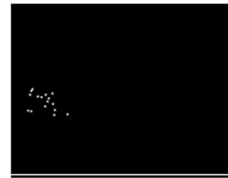


(a) 篡改图像与仿射变换图像相关图



(b) 篡改图像与逆仿射变换图像相关图

图 11 篡改图像分别与一类仿射变换与逆仿射变换图像相关图



(a) 篡改图像中的粘贴区域关键点



(b) 篡改图中的复制区域关键点

图 12 篡改图像中一类复制区域和粘贴区域的匹配关键点

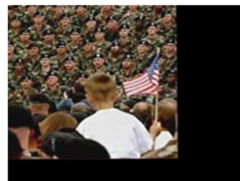


(a) 篡改图像中的粘贴区域

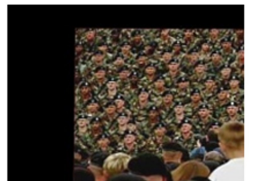


(b) 篡改图中的复制区域

图 13 篡改图像中的一类复制区域和粘贴区域

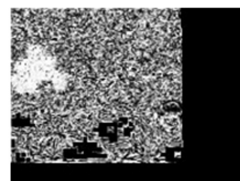


(a) 仿射变换图像



(b) 逆仿射变换图像

图 14 篡改图像中另一类仿射变换与逆仿射变换图像



(a) 篡改图像与仿射变换图像相关图



(b) 篡改图像与逆仿射变换图像相关图

图 15 篡改图像分别与另一类仿射变换与逆仿射变换图像相关图



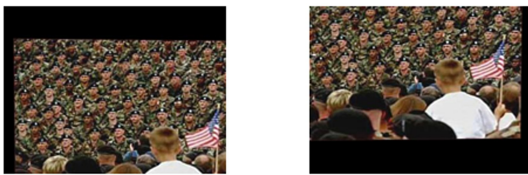
(a)篡改图像中的粘贴区域关键点 (b)篡改图像中的复制区域关键点

图 16 篡改图像中另一类复制区域和粘贴区域的匹配关键点



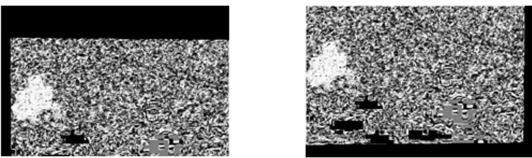
(a)篡改图像中的粘贴区域 (b)篡改图中的复制区域

图 17 篡改图像中的另一类复制区域和粘贴区域



(a)仿射变换图像 (b)逆仿射变换图像

图 18 篡改图像中粘贴部分的仿射变换与逆仿射变换图像



(a)篡改图像与仿射变换图像相关图 (b)篡改图像与逆仿射变换图像相关图

图 19 篡改图像分别与粘贴部分的仿射变换与逆仿射变换图像相关图



(a)粘贴部分 1 的匹配关键点 (b)粘贴部分 2 的匹配关键点

图 20 篡改图像中两个粘贴区域的关键点



(a)篡改图像中的粘贴区域 1 (b)篡改图像中的粘贴区域 2

图 21 篡改图像中的粘贴区域



图 22 篡改图像中的复制粘贴区域

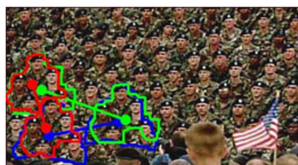


图 23 复制粘贴篡改图像的检测结果

图 24 是本文算法对网上流行的广场鸽复制粘贴篡改图

像(图像大小为 500×349)^[14] 的检测结果。



图 24 广场鸽图像的检测结果

图 25 是对 CoMoFoD 图像库^[15] 中第 140 幅篡改图像的检测结果。

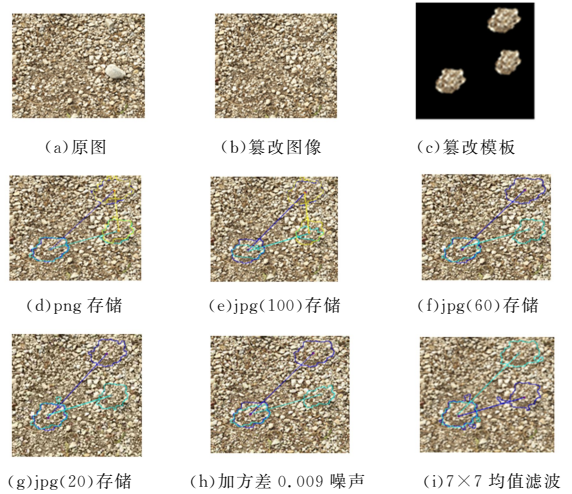


图 25 CoMoFoD 图像库中第 140 幅篡改图像的检测

结束语 本文算法是通过随机采样一致性算法对篡改图像中相互匹配的 sift 关键点对进行分类,每一类是一次复制粘贴操作。由于复制粘贴区域之间存在着仿射变换关系,对篡改图像进行仿射变换以及逆仿射变换,通过求篡改图像与仿射变换图像以及逆仿射变换图像的灰度相关图,准确定位图像的粘贴、复制区域。对篡改图像中没有检测到足够的 sift 匹配关键点的复制粘贴篡改图像,本文算法将不能进行有效的检测。导致图像中 sift 匹配关键点减少的原因,是图像像素成倍缩小,或者图像中复制粘贴区域模糊或过度平滑。在对匹配关键点用随机采样一致性算法进行分类的过程中,图像的有损压缩或者图像的过度缩放,致使部分匹配点的位置出现偏移,这些点之间的仿射变换关系不符合图像中复制粘贴部分之间的仿射变换关系,有效剔除图像中的这些匹配点,提高复制粘贴篡改部分的定位精度,是我们下一步的工作方向。

参考文献

[1] WARBHE A D, DHARASKAR R V, THAKARE V M. A Survey on Keypoint Based Copy-paste Forgery Detection Techniques[J]. Procedia Computer Science, 2016, 78: 61-67.

[2] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[3] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded Up Robust Features[J]. Computer Vision & Image Understanding, 2006, 110(3): 404-417.