

逆重心密度的半监督学习在视频篡改检测的应用^①

吴铁浩¹, 黄添强^{1,2}, 袁秀娟¹, 陈智文¹, 苏伟峰³

¹(福建师范大学 数学与计算机科学学院, 福州 350007)

²(福建师范大学 网络安全与密码技术福建省高校重点实验室, 福州 350007)

³(北京师范大学 香港浸会大学联合国际学院 理工科技学部, 珠海 519085)

摘要: 针对视频遭受非同源片段合成的篡改, 根据其特点以及数据特征, 提出一种改进的半监督学习算法, 并将这种算法应用在视频篡改取证研究. 首先, 以每一个视频帧作为样本, 提取其 R、G、B 分量, 对这三个分量分别提取噪声, 然后以这三个噪声模板的均值、方差等统计量构成样本属性. 最后, 用本文提出半监督学习算法对样本集聚类. 实验结果表明本文提出的算法能够有效检测视频是否由非同源片段合成.

关键词: 半监督聚类; 视频篡改; 模式噪声; 小波变换

Video Tamper Detection Based on Inverse Gravity Density Semi-Supervised Learning

WU Tie-Hao¹, HUANG Tian-Qiang^{1,2}, YUAN Xiu-Juan¹, CHEN Zhi-Wen¹, SU Wei-Feng³

¹(School of Mathematics and Computer Science, Fujian Normal University, Fuzhou 350007, China)

²(Key Laboratory of Network Security and Cryptography, Fujian Normal University, Fuzhou 350007, China)

³(Department of Computer Science and Technology, BUN-HKBU United International College, Zhuhai 519085, China)

Abstract: Aim at the condition that a video may be combined from non-same source segments, and according to its data characteristic, we propose an improved semi-supervised learning algorithm, and apply it to the research of detecting video's authenticity. First, use per-frame of a video as a sample, extracting their R, G, B components, and getting the noise of these three components. Then, getting these noise's mean and variance, and use it as attributes of the sample. Last, use this paper's algorithm cluster the samples. Experimental results show that this method can effectively identify whether a video is combined from non-same source segments.

Key words: semi-supervised clustering; video tamper; pattern noise; wavelet transform

在数字多媒体愈发普及的时代, 各种成像设备在各个领域的使用变得顺其自然. 随着先进的视频编辑软件的出现和不断升级, 修改一张图片或一段视频变得越来越容易. 这种技术一旦被非法者利用, 将可能对社会产生不可预期的影响. 因此, 数字视频篡改检测技术的研究成为当今急需快速发展的重要技术.

目前, 国内外学者对数字视频篡改取证的研究主要有两类: 一类是针对视频编码(比如 MPEG 编码)过程的特点. 因为被篡改的数字视频经常有经过重压缩的过程, 而经过重压缩的视频, 其帧序列的内在规律会遭到破坏, 所以可以通过分析特定格式视频帧序列的内在规律是否遭到破坏来检测. WANG^[1]等人提出

的针对 MPEG 格式编码的视频序列, 利用篡改前后的预测残差周期性变化的特点进行分析. 另外, 黄^[2]等人在 WANG^[1]的基础上, 并针对其存在的问题, 提出基于双向运动矢量的数字视频篡改取证算法, 大大提高了检测精度. 另一类是从现有的数字图像篡改取证算法^[3,4]的基础出发. 根据视频拍摄过程中引入的模式噪声, 或者视频帧序列的连续性及其他统计特性的特点进行分析. Hsu^[5]等人提出了基于模式噪声相关性的视频篡改检测算法, 首先提取视频每一帧的模式噪声, 然后将仅剩模式噪声的视频帧分块, 最后通过计算时间上相邻的块级相关性, 并构建高斯混合模型, 来鉴定数字视频是否被篡改. 王^[6]等人同样利用视频模式

① 基金项目: 国家自然科学基金(61070062,61073017); 福建省高校产学研合作科技重大项目(2012H6006); 福建省高校服务海西建设重点项目(2008HX200941-4-5); 福建省高等学校新世纪优秀人才支持计划(JA11038)

收稿时间:2013-01-08;收到修改稿时间:2013-01-29

噪声, 通过比较待鉴别帧的噪声与模式噪声之间的相关性, 利用实验经验阈值判断, 定位篡改区域. 另外, 黄^[7]提出一种利用模式噪声聚类分析的视频篡改检测研究, 在检测精度上取得不错的效果.

在基于视频模式噪声的篡改检测的众多算法中, 利用数据挖掘^[8]的聚类算法^[8-14]比采用相关性分析进行数据处理的检测精度更高, 更重要的是, 可以避免相关性阈值对检测结果的影响. 另外, 聚类是属于无监督的自动学习过程, 聚类效果会受数据集本身的影响. 为了进一步提高视频篡改检测精度, 本文提出一种基于半监督学习的视频篡改检测技术. 半监督学习是利用少量的标注样本和大量的未标注样本进行训练和分类的过程. 考虑到视频的篡改通常不可能发生在视频的起始和结尾处. 所以, 可以将一段待测视频的起始和结尾处的若干帧作为后期半监督学习的带标注样本.

本文主要内容包括: 第一部分对视频特征的相关研究; 第二部分介绍本文提出的 IGSSSL 算法; 第三部分是算法流程及可行性分析; 第四部分给出相关的实验结论及分析.

1 相关研究

1.1 视频的模式噪声

每一部摄像机都有属于自己的唯一标识, 它就是模式噪声^[6,7,15]. 噪声产生的原因是: 由于数码摄像机内部电路以及 CCD 存在非理想性, 导致生成帧图像过程中夹杂着多种误差. 这些误差最终形成噪声. 噪声主要由两部分构成: 随机噪声和模式噪声. 任意两部摄像机拍摄同一场景, 所得的视频的模式噪声都会有所差异. 而同一部相机拍摄同一场景, 所得的模式噪声十分接近. 由于噪声的随机部分符合高斯分布, 所以通过对同一视频的多帧图像的噪声取平均可以有效抵消噪声的随机噪声, 从而得该摄像机的模式噪声.

单帧图像的噪声表示为:

$$N_t = I_t - f_t \tag{1}$$

视频的模式噪声表示为:

$$n = \frac{\sum_{t=0}^{T-1} (I_t - f_t)}{T} \tag{2}$$

其中, N_t 表示视频 t 时刻的帧图像噪声, I_t 为视频 t 时刻的帧图像, f_t 为视频 t 时刻的帧图像经过低通滤波

后的去噪图像. T 为视频总帧数. 单帧图像去噪过程如图 1 所示:

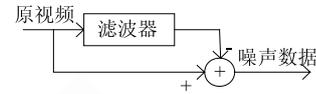


图 1 单帧图像去噪过程

1.2 视频的三色分离特征

彩色图像中的任一种颜色是以 R(红)、G(绿)、B(蓝)为基色, 通过 R、G、B 混合得到, 即任何一个像素的颜色值都可以用 R、G、B 三个值来描述. 也就是说, 任一种颜色都是由 RGB 三种单色共同作用的结果. 如果将图片的每一个像素的颜色的三基色成分提取出来, 重新构成三幅基色图像, 再对三副基色图像分别提取图像特征. 由于每一个特征都是三色分离的, 显然这些特征比未三色分离的图像特征更能代表原始图像的特征. 一幅图像的噪声的均值和方差可以用来表示该图像的特征. 所以, 经过三色分离的图像就有六个特征, 分别为噪声的红色分量的均值和方差; 噪声的绿色分量的均值和方差; 噪声的蓝色分量的均值和方差.

本文关于视频的三色分离特征的提取步骤如下:

- 1) 提取原视频的帧序列, 并对帧序列进行去噪处理得到每一帧的噪声.
- 2) 将步骤 1 得到的每一帧噪声通过三色分离器得到三基色噪声图像序列.
- 3) 最后, 分别求取三基色噪声图像序列的每一帧图像的均值和方差特征.

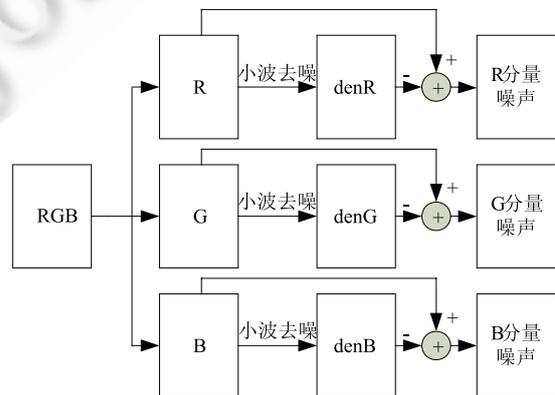


图 2 RGB 图像噪声三色分离过程

2 IGSSSL 算法

在机器学习领域中, 传统的学习方法有两种: 监

督学习和无监督学习. 半监督学习^[8,16]是介于两者之间的学习方法. 其主要思想是利用少量带标签样本来指导大量无标签样本的训练和分类.

本文将视频帧经过三色分离后提取的噪声的均值与方差作为检测视频是否被篡改的特征数据. 理论上, 对于一段完整的原始视频, 这些特征数据将满足一定的分布而紧密地聚集在一起. 而一旦视频遭到异源篡改, 这些异源的插入帧的特征数据会与原始视频的特征数据由于分布不同而相互分离. 结果是会出现一些离群点, 或者是会出现几个无交集簇. DBSCAN 算法是一种基于密度的无监督学习算法. 正是由于 DBSCAN 是基于密度的, 它能够发现任意形状的簇, 更可贵的是, 它能够发现噪声. DBSCAN 的这种优点, 恰恰适合于视频的异源篡改检测. 图 3 为 DBSCAN 聚类过程实例.

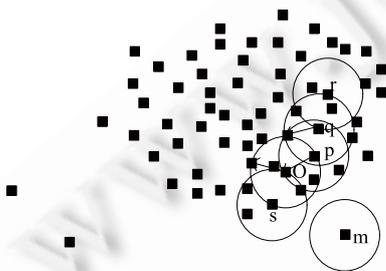


图 3 DBSCAN 聚类过程

在参数设置合理的情况下, DBSCAN 完全能够对待测视频进行异源篡改检测. 然而, 美中不足的是, 传统的 DBSCAN 算法中, 邻域半径(ϵ)和最小邻域数目(minpts)这两个参数对聚类的结果影响很大, 特别是 ϵ 的影响更是差之毫厘就可能产生不同的结果.

由上述分析的视频特征数据的特点, 以及 DBSCAN 算法的优缺点, 提出一种改进的基于 DBSCAN 的半监督学习算法 IGSSSL(Inverse Gravity Density Semi-Supervised Learning), 利用已知少量类标签样本, 首先估计一个最优邻域半径 ϵ_{ps} , 然后对全数据集进行 DBSCAN 聚类. 在理论上, 这一算法将比直接使用 DBSCAN 算法聚类的精度要高.

IGSSSL 算法的相关定义:

① ϵ 邻域: 给定对象半径 ϵ 内的邻域称为该对象的 ϵ 邻域.

② 最优邻域半径 ϵ_{b} : 能够使得聚类效果最佳的邻域半径.

③ 核心对象: 如果对象的 ϵ 邻域至少包含最小数目 minpts 的对象, 则称该对象为核心对象. 图 3 中, minpts 等于 5, 所以 r, q, p, o 均为核心对象, s 不是核心对象.

④ 直接密度可达: 给定一个对象集合 D , 如果对象 p 在对象 q 的 ϵ 邻域内, 而 q 是一个核心对象, 就称对象 p 是从对象 q 出发直接密度可达的. 如图 3 中, q 是 r 直接密度可达的.

⑤ 间接密度可达: 存在一个对象链 $p_1, p_2, \dots, p_n, p_1=q, p_n=q$, 对于 i , p_{i+1} 是从 p_i 关于 ϵ 和 minpts 直接密度可达的, 则称对象 p 是从对象 q 关于 ϵ 和 minpts 密度可达的. 如图 3 中, s 是 r 间接密度可达的.

⑥ 密度相连: 存在对象 $o \in D$, 使对象 p 和 q 都是从 o 关于 ϵ 和 minpts 密度可达的, 称对象 p 到 q 是关于 ϵ 和 minpts 密度相连的.

⑦ 簇: 基于密度可达性的最大的密度相连对象的集合.

⑧ 噪声(离群点): 不包含在任何基于密度的簇中的对象称为噪声(离群点).

IGSSSL 算法分为两个步骤进行: 一是由少量带标签的数据经过逆重心密度算法得到最优邻域半径 ϵ_{b} , 如图 4 所示. 二是将第一步得到的最优邻域半径作为 DBSCAN 的参数, 并设置最小邻域数 minpts , 然后对全数据集聚类.

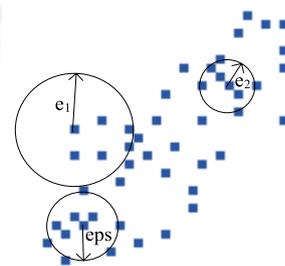


图 4 少量带标签数据点估计最佳邻域半径 ϵ_{ps}

算法具体步骤如下:

第一步: 确定邻域半径 ϵ_{ps}

输入: 已知类标签的数据集 LabeledSet、 minpts

输出: 最佳 ϵ_{ps}

算法: 1. 计算带标签数据点中, 每个点的 minpts 邻域半径

ϵ_i

2. 将 $\{\epsilon_i\}$ 升序排列

3. 对于集合 $E=\{e_1, e_2, \dots, e_n\}$, 求 $eb=\arg\{(e_1+e_2+\dots+e_b)-(e_{b+1}+\dots+e_n)\}$

4. eb 即最佳 eps

第二步: DBSCAN 聚类

输入: 已知和未知类标签的混合数据集 Set、minpts、eps

输出: 聚类结果

算法: 1.Repeat

2. 从混合数据集 Set 中抽出一个未处理的点;

3. IF 抽出的点是核心点

THEN 找出所有从该点密度可达的对象, 形成一个

簇;

ELSE 抽出的点是边缘点(非核心对象),

跳出本次循环, 寻找下一个点;

4. UNTIL 所有的点都被处理.

3 算法流程及可行性分析

众所周知, 视频是由视频帧序列所组成的, 视频帧的快速切换使得画面产生动态效果. 然而, 要让视频看起来自然流畅, 必须保证视频的视频帧率要大于 25fps.

对于普通的视频, 其帧率通常在 25fps-30fps. 可以认为, 一段视频的有效部分在中间. 换句话说, 一段视频的开始前几秒和结束前几秒一般不会被篡改. 于是就可以利用这一点, 将视频的开始和结束前 1-2 秒的视频帧提取出来来指导半监督聚类, 从而达到篡改检测的目的.

3.1 算法流程

本文利用三色分离的模式噪声作为数据集, 并结合提出的基于逆重心密度的半监督学习的视频篡改检测算法的具体过程如下:

1) 首先将待检测视频经过分离去噪器, 得到三色分离并且无噪的视频帧序列.

2) 将 1)过程的到的无噪的三色分离视频帧序列经过减法器得到三色分离的视频帧噪声序列.

3) 接着对于每一个分量的视频序列分别求取每一帧的均值和方差, 得到一个 6 维的数据集.

4) 然后根据本文的假设, 提取该数据集的前 2 秒和后 1 秒的数据(大约 75 帧), 作为带标签的数据, 并利用 IGDSSL 算法处理这些数据得到最佳邻域半径 eb .

5) 最后将 eb 作为参数, 并设置最小邻域个数 $minpts$, 通常为 6, 再利用 DBSCAN 算法对全数据集聚类得到检测结果.

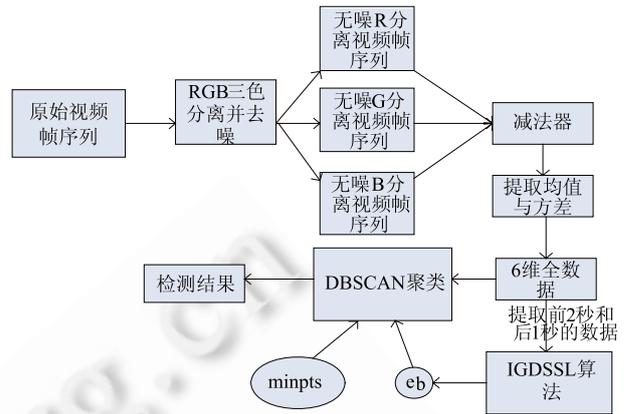


图 5 本文算法流程图

3.2 可行性分析

与现有的基于模式噪声的视频篡改检测的算法相比, 由于本文对视频进行 RGB 三色分离, 所以在数据特征提取的时间效率上, 本文的提出的算法比传统的算法要低. 但是, 从数据特征点的有效性上, 由于 RGB 三色分离, 使得数据特征彼此间互不干扰, 因此, 数据特征的鲁棒性比传统算法要好.

另外, 本文所针对的被篡改的视频是由多段非同源视频片段合成的, 且主片段的长度远大于其他片段的长度. 因此, 可以预知, 理想的聚类结果应该是有一个大类, 加几个小类或一个大类加一些噪声. 由于本文提出的 IGDSSL 算法是对 DBSCAN 算法的改进, 所以具有 DBSCAN 这种基于密度的算法的一切优点: 可以发现任意形状的簇. 并且, 利用了少量已知标签的样本得到了最优邻域半径, 因此大大提高了检测精度.

4 实验及对比结果分析

为了评估 IGDSSL 算法的检测精度, 分别对实验数据集和真实视频进行实验. 实验环境如下: 2.0GHZ CPU, 操作系统为 Windows XP, 开发环境为 matlab + java, java 编译运行环境为 jdk6.0.

实验数据集共有样本 4986 个, 其中噪声样本 30 个, 每个样本有 5 维, 均由 weka 随机生成. 实验首先选取前 1-50 号样本作为已知类标签样本, 然后利用这些样本和 IGDSSL 算法评估出数据集的最佳邻域半径. 最后利用得到的最佳邻域半径和 DBSCAN 聚类并检测出噪声样本. 通过实验, 得到实验数据集样本的最佳邻域半径为 0.362837. 为了评估 IGDSSL 算法的可行性和准确性, 实验过程中同时针对不同的邻域半径

检测结果(邻域个数为 6). 实验结果如图 6 至图 9 所示:

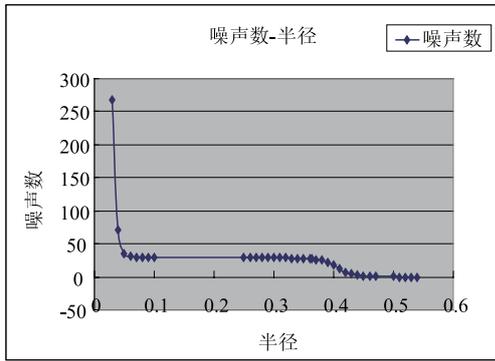


图 6 噪声数随半径变化图

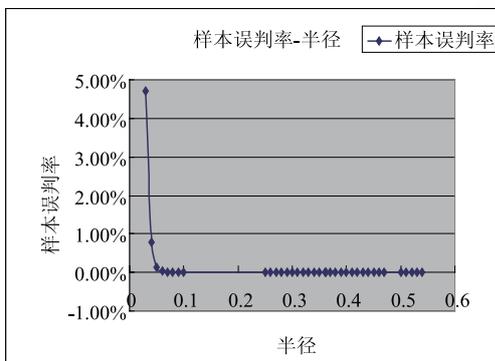


图 7 准确率随半径变化图

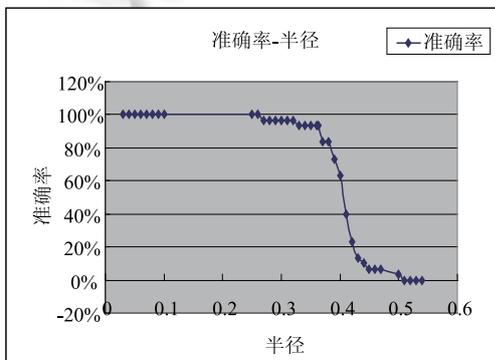


图 8 样本误判率随半径变化图

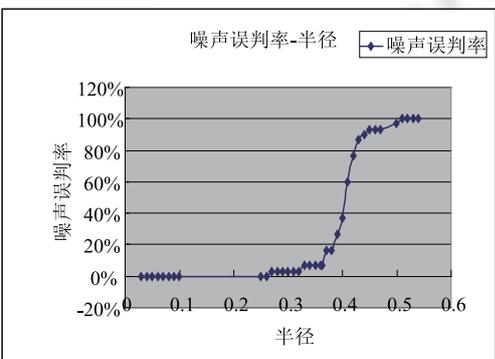


图 9 噪声误判率随半径变化图

从图 7 可知, 半径在 0.37 处骤降, 到 0.42 时, 准确率已降至 23.3%. 根据本文提出的 IGSSSL 算法, 得到最佳邻域半径 0.362837, 此时准确率为 93.30%. 另外, 从图 8、图 9 可知, 半径在 0.362837 处, 样本误判率为 0%, 噪声误判率为 6.7%, 均控制在较低水平.

从实验数据的实验结果: IGSSSL 算法可以根据已知的少量类标签样本估算出数据集的最佳邻域半径. 从而以较高的准确率检测出噪声样本.

图 11 至图 13, 是当邻域半径为最佳邻域半径, 而邻域个数变化的结果. 从结果图可知就, 当邻域个数大于 4 时, 准确率和误判率均保持稳定.

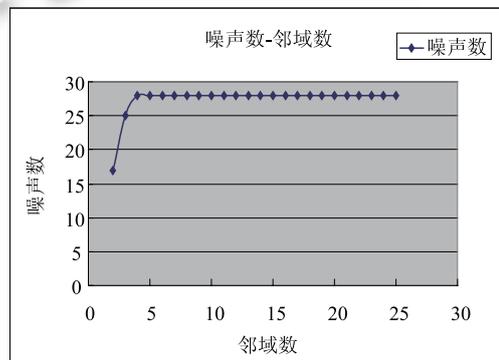


图 10 噪声数随邻域个数变化图

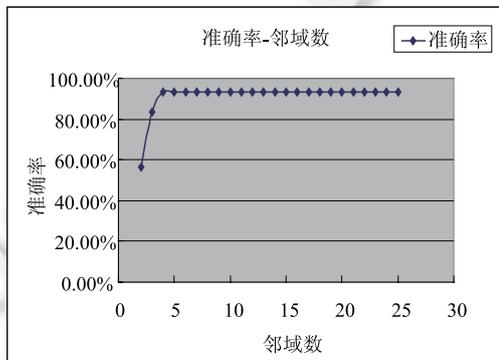


图 11 准确率随邻域个数变化图

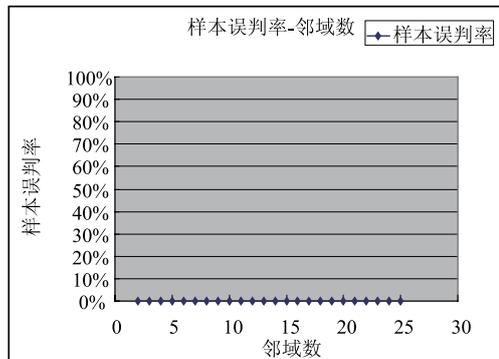


图 12 样本误判率随邻域个数变化图

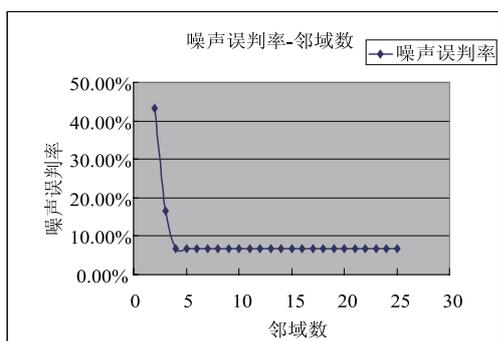


图 13 噪声误判率随邻域个数变化图

从以上两个实验可以得出以下结论: 邻域半径对于 DBSCAN 算法的影响很大, 而当邻域半径固定时, 邻域个数对于 DBSCAN 算法几乎不影响。由此说明, IGDSSL 算法能够在少量带标签样本的指导下, 以较高的准确率检测出噪声样本。

接下来, 用真实视频数据 1 进行实验, 图 14 列出部分视频帧。实验首先提取视频每一帧的三色分离噪声数据, 分别求取均值和方差后, 得到 6 维样本共 283 个, 其中已知视频中有 10 帧被篡改, 也就是说噪声样本有 10 个。然后利用提出的 IGDSSL 算法对实验数据进行检测, 得到最佳邻域半径为 0.33853, 检测结果如表 1 所示。



(c) 主视频帧



(d) 主视频帧

图 14 部分视频帧



(a) 异源视频帧



(b) 异源视频帧

表 1 真实视频 1 实验结果

邻域个数	聚类数	噪声数	准确率	误判率
3	2	0	96.47%	3.53%
4	2	0	96.47%	3.53%
5	2	0	96.47%	3.53%
6	2	0	96.47%	3.53%
7	2	0	96.47%	3.53%
8	2	0	96.47%	3.53%
9	2	0	96.47%	3.53%
10	2	0	96.47%	3.53%
11	1	10	100%	0%
12	1	10	100%	0%
13	1	10	100%	0%
14	1	10	100%	0%
15	1	10	100%	0%
16	1	10	100%	0%
17	1	10	100%	0%
18	1	10	100%	0%

由表 1 数据可知, 当邻域个数设定为小于 10 时, 聚类结果为 2 类, 其中未被篡改的视频帧归为一类, 被篡改的视频帧归为另一类. 当邻域个数设定为大于 10 时, 此时聚类结果为 1 类, 被篡改的视频帧被判定为噪声. 由此可得, IGDSSL 算法是一种适合检测视频数据是否被篡改的算法.

再用真是拍摄视频 2 进行实验, 图 15 列出部分视频帧. 和真实视频 1 实验的方法一样, 首先提取视频每一帧的三色分离噪声数据, 分别求取均值和方差后, 得到 6 维样本共 243 个, 其中已知视频中有 28 帧被篡改, 也就是说噪声样本有 28 个. 然后利用提出的 IGDSSL 算法对实验数据进行检测. 得到最佳邻域半径为 0.2195, 检测结果如表 2 所示.



(a) 主视频帧



(b) 异源视频帧

图 15 部分视频帧

表 2 真实视频 2 实验结果

邻域个数	源视频帧检	源视频帧误	准确率
	测数	检数	
2	202	13	93.95%
3	202	13	93.95%

4	202	13	93.95%
5	202	13	93.95%
6	200	15	93.02%
7	200	15	93.02%
8	188	27	87.44%

由表 1 和表 2 数据可以得出以下结论: 本文提出的算法能够以较高的精度检测出视频是否遭遇异源帧插入篡改.

本文在 DBSCAN 基础上提出 IGDSSL 算法, 并将其应用在视频异源篡改检测领域. 在实验部分, 图 6 到图 13 以及表 1 和表 2 充分说明了, 直接使用 DBSCAN 算法进行检测的话, 必须多次调整邻域半径, 甚至每一次调整的幅度小到 0.01. 而用本文提出的 IGDSSL 算法, 无需调整邻域半径即可较准确地检测出来.

5 结语

本文针对视频模式噪声数据的特点: 未被篡改的视频只能聚成一类, 被篡改的视频或聚成一类, 外加一些噪声点; 或聚成大小两类. 提出一种基于 DBSCAN 的半监督聚类算法, 由少量已知类标签的视频数据样本来估算出最佳的邻域半径, 然后用经典的 DBSCAN 算法对数据集进行聚类, 从而检测出视频的异常帧. 另外, 本文采用三色分离的视频模式噪声作为特征数据, 进一步净化了数据, 从而使得算法的检测精度提高了. 从实验数据集和真实数据集的实验结果中, 我们可以看出, 本文提出的算法与传统的 DBSCAN 算法相比, 本文算法无需盲目的设定邻域半径即可获得较高的检测精度.

参考文献

- 1 Wang W, Farid H. Exposing digital forgeries in video by detecting double MPEG compression. Proceeding MM&Sec '06 Proc. of the 8th workshop on Multimedia and security New York: ACM. 2006: 37-47.
- 2 黄添强, 陈智文. 基于双向运动矢量的数字视频篡改鉴定. 山东大学学报(工学版), 2011, 41(4): 13-19.
- 3 Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. Information Forensics and Security, IEEE Transactions on, 2006, 1(2): 205-214.
- 4 Li CT, Li Y. Digital camera identification using colour-decoupled photo response non-uniformity noise pattern. Proc.

(下转第 102 页)

在信息隐藏技术的水印方法分类中,按可见性,可分为不可见和可见水印两类,其中不可见水印强调的是隐藏特性。在本文方法中,采用了网页标签常见的 id,并且和其它标签正常的 id 相混合,识别是采用不易引起视觉敏感的单双引号。其实,在方法中,没有把单双引号用于水印表示,只是识别标记。因此,像在属性值前后添加空格或取值等于符号前后添加空格,也可用于识别的标记,这也在一定程度上增加了水印方法的透明性。

3 结论

随着互联网的普及应用,作为互联网常见的信息传递媒体之一-网页,关于它的安全保护技术也变得越来越重要。对于网页水印技术,因网页本身的格式特性,相较其它水印技术,有其自身的弱点,尝试采用不同方法是网页水印技术将来积极发展的方向。在本文中,使用 Shamir 门限秘密共享,提高了嵌入水印的安全性和可靠性,并利用了网页文件标签的 id 值来嵌入水印,嵌入容量可较大,但可能存在因门限 n 值过大,而使得在较小的单个网页的适合点处水印完全嵌入较困难。未来设想是改进扫描算法,找出与网页存在关联链接的其它网页并用来水印嵌入。另外,也将进一

步考虑如何利用结合网页标签的 name 名称属性来表示水印的方法。

参考文献

- 1 丁伟.基于 Web 网页的文本水印技术的研究[学位论文].武汉:武汉理工大学,2012.
- 2 傅瑜,王保保.文本水印附加空格编码方法的实现及其性能.长安大学学报,2002,22(5):85-87.
- 3 Low SH, Maxemchuk NF, Lapone AM. Document Identification for Copyright Protection using Centroid Detection. IEEE Trans Communications,1998,46(3):372-383.
- 4 万唯一.基于数字水印的网页防篡改技术研究[学位论文].成都:西南交通大学,2010.
- 5 Zhao QJ, Lu HT. A PCA-based watermarking scheme for tamper-proof of web pages.Pattern recognition,2005,38(8):1321-1323.
- 6 雷红艳,邹汉斌.基于 Shamir 秘密共享的隐私保护分类算法.计算机工程与设计,2010,31(6):1271-1273.
- 7 Rivest R. The Md5 Message-Digest Algorithm.RFC. RFC Editor, 1992.
- 8 孙鹏.网页水印技术研究[学位论文].上海:上海交通大学,2010.
- 9 督聚类算法.计算机工程与应用,2010,46(8):123-126.
- 11 赵凤,焦李成,刘汉强,等.半监督谱聚类特征向量选择算法.模式识别与人工智能,2011,24(1):48-56.
- 12 王玲,薄列峰,焦李成.密度敏感的半监督谱聚类.软件学报,2007,18(10):2412-2422.
- 13 肖宇,于剑.基于近邻传播算法的半监督聚类.软件学报,2008,19(11):2803-2813.
- 14 宋凌,李枚毅,李孝源.一种新的半监督入侵检测算法.计算机应用,2008,28(7):1781-1783.
- 15 武伟,詹玲超.基于数码相机固有特性的篡改检测.微计算机信息,2007,23(28).
- 16 黄添强,秦小麟,叶飞跃.基于方形邻域的离群点查找新方法.控制与决策,2006,21(5):123-126.

(上接第 97 页)

of 2010 IEEE International Symposium on Circuits and Systems (ISCAS),2010.Paris: IEEE.

5 Hsu CC, Hung TY, Lin CW, et al. Video forgery detection using correlation of noise residue.2008:IEEE.

6 王俊文,刘光杰,张湛,等.基于模式噪声的数字视频篡改取证.东南大学学报(自然科学版),2008,38(A02):13-17.

7 黄添强,吴铁浩,袁秀娟,等.利用模式噪声聚类分析的视频非同源篡改检测.计算机科学与探索,2011,5(10):914-920.

8 韩家炜,坎伯.数据挖掘:概念与技术.北京:机械工业出版社,2001

9 周志华,王珏.机器学习及其应用.北京:清华大学出版社,2007.

10 赵倩,尚学群,王淼.基于 seeds 集和频繁项集挖掘的半监