

融合多特征的视频帧间篡改检测算法

肖辉^{1,2}, 翁彬^{1,2}, 黄添强^{1,2}, 普菡^{1,2}, 黄则辉³

(1. 福建师范大学数学与信息学院, 福建 福州 350007;

2. 福建省大数据挖掘与应用工程技术研究中心, 福建 福州 350007;

3. 世新大学, 台湾 台北 350108)

摘要: 传统的视频帧间被动取证往往依赖单一特征, 而这些特征各自适用于某类视频, 对其他视频的检测精度较低。针对这种情况, 提出一种融合多特征的视频帧间篡改检测算法。该算法首先计算视频的空间信息值并对视频进行分组, 接着计算视频帧间连续性 VQA 特征, 然后结合 SVM-RFE 特征递归消除算法对不同特征排序, 最后利用顺序前向选择算法和 Adaboost 二元分类器对排序好的特征进行筛选与融合。实验结果表明, 该算法提高了篡改检测精度。

关键词: 视频篡改检测; 融合算法; 特征选择; Adaboost 二元分类; 视频分组

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020007

Video inter-frame tampering detection algorithm fusing multiple features

XIAO Hui^{1,2}, WENG Bin^{1,2}, HUANG Tianqiang^{1,2}, PU Han^{1,2}, HUANG Zehui³

1. School of Mathematics and Information, Fujian Normal University, Fuzhou 350007, China

2. Fujian Research Center for Big Data Mining and Applied Engineering, Fuzhou 350007, China

3. Shih Hsin University, Taipei 350108, China

Abstract: Traditional passive forensics of video inter-frame tampering often relies on single feature. Each of these features is usually suitable for certain types of videos, while has low detection accuracy for other videos. To combine the advantages of these features, a video inter-frame tampering detection algorithm that could fuse multi-features was proposed. The algorithm firstly classified the input video into one group based on its space information and time information values. Then it calculated the VQA features that represented the video inter-frame continuity. These features were sorted by the SVM-RFE feature recursive elimination algorithm. Finally, the sorted features were fil-

收稿日期: 2019-07-10; 修回日期: 2019-10-03

通信作者: 黄添强, fjhtq@fjnu.edu.cn

基金项目: 国家重点研发计划专项基金资助项目 (No.2018YFC1505805); 应用数学福建省高校重点研究资助项目 (No.SX201803)

Foundation Items: The National Key Program for Developing Basic Science (No.2018YFC1505805), Applied Mathematics Fujian Provincial Key Laboratory Project (No.SX201803)

论文引用格式: 肖辉, 翁彬, 黄添强, 等. 融合多特征的视频帧间篡改检测算法[J]. 网络与信息安全学报, 2020, 6(1): 84-93.
XIAO H, WENG B, HUANG T Q, et al. Video inter-frame tampering detection algorithm fusing multiple features [J]. Chinese Journal of Network and Information Security, 2020, 6(1): 84-93.

tered and fused by the sequential forward selection algorithm and Adaboost binary classifier. Experimental results show that the proposed algorithm could achieve higher tampering detection accuracy.

Key words: video tamper detection, fusion algorithm, feature selection, Adaboost binary classification, video grouping

1 引言

数字视频为信息的主要载体之一,到2021年其在互联网上的视频流量将占消费者使用流量的85%^[1]。与此同时,大量的视频剪辑软件使视频篡改成为一项很轻松的工作,而这些篡改后的视频往往人眼很难分辨出来。然而,视频的真实性和完整性在新闻媒体、科学发现、法庭取证等领域都十分重要^[2]。因此,视频篡改检测已经成为近年来一个热门的研究方向^[3-5]。

视频的帧间篡改包含帧复制粘贴、帧插入和帧删除等操作,这些篡改方式易于实现,且对社会安全危害较大。因此,帧间篡改检测算法吸引了大量的研究者,近年来不断有新方法被提出。然而,这些方法往往通过提取视频的某个特征来识别篡改。而现实世界的视频是丰富多彩的,不同的光照环境、模糊程度和画面运动的快慢等都有可能对某种特征失效。所以,已有的检测算法一般对某类视频比较有效,而对其他视频的效果不好。将这些特征融合起来以提高篡改检测的效果是一个新的研究方向。

为解决上述问题,本文提出了一种融合多特征的视频帧间篡改检测算法。针对待检测的视频,该算法首先计算多个特征,然后对这些特征排序并筛选,最后得到的融合检测结果往往要优于基于单个特征的方法。此外,事先将视频分组,再进行特征融合来提高融合算法的检测精度。

2 相关工作

2.1 帧间篡改检测

视频帧间篡改的检测已有不少研究成果。Chao等^[6]提出了一种基于光流特征的帧间检测方法。使用小窗口移动计算第一帧和最后一帧以及相邻帧之间的光流,出现帧删除或插入时,光流的高度不一致,用二分查找方案来检测插入篡改,并应用双自适应阈值来检测删除篡改。然而,光流特征对篡改帧数少的视频的检测精度不高。Wu

等^[7]提出了基于速度场特征的检测方法。速度场特征是由粒子图像测速(PIV)的关键点计算相邻帧并估计它们的位移,根据速度场序列计算相应的相对因子序列,最后用广义极值学习偏差(ESD)来识别篡改类型并定位,速度场特征随压缩比例的增大,检测精度会大大降低。Liu等^[8]将RGB颜色空间转化为2D对立色度空间,使用Zernike矩来计算二维对立色度空间的Zernike色度变换矩(ZOCM)执行粗略检测,利用相邻帧ZOCM的差异性来提取异常点。粗检测过程计算速度快,但会存在误检测现象,最后使用Tamura粗糙度特征来进行精细计算得到更准确的帧复制粘贴位置,篡改区域移动速度过快时,Tamura粗糙度特征的检测精度不高。Wang等^[9]分析了MPEG-1,2压缩视频,当视频序列遭受伪造(帧删除/插入),一些帧从一个GOP组删除或插入另一个GOP组时,会出现较大的运动估计误差。它依赖于P帧残差特征产生的周期性伪像,用傅里叶变换检查P帧残余误差造成的峰值从而判断出伪造视频。基于此基础,Aghamaleki等^[10]提取了具有空间约束的算法来检测P帧残差特征量化误差的丰富区域,减少运动对P帧残差的影响,继而用小波变换丰富频域中的量化误差轨迹。利用P帧残差特征的缺点在于对删除整个GOP组的伪造痕迹无法被量化出来。Li等^[11]提出了一种基于结构相似度(MSSIM)特征的新算法,由于重复帧之间的相似度值高于正常帧之间的相似度值,通过测量短子序列之间的时间相似性度量策略来检测帧复制粘贴篡改。文献[12]同样使用MSSIM特征,基于相邻帧之间MSSIM商具有连续性,使用两次切比雪夫不等式及阈值法对提出的MSSIM商特征进行异常点检测,从而实现对视频帧插入和删除的篡改检测和定位。MSSIM利用了视频帧之间的连续性,用作视频篡改检测时有较好的效果,本文将其融入提出的算法中。上述的帧间篡改检测技术通常只使用某个特定的特征来寻找异常点。然而,现实世界的视频往往具有不

同的特点,如不同的光照、颜色、亮度和运动的剧烈程度等。因此,基于单个特征的篡改检测方法,一般只对某类视频有较好的效果,而对另一类视频检测效果较差。本文的思路是建立一个算法来融合多个不同特征。

2.2 视频质量评估

视频质量评估(VQA, video quality assessment)模型针对视频每帧的序列进行计算,随时间记录汇总每帧的测量质量,以评估整个视频的质量。VQA特征已有不少,表1列出了目前已有的质量评估指标,包括MSE、PSNR、SSIM^[13]、MSSIM^[14]、VSNR^[15]、VIF^[16]、VIFP^[17]、UQI、IFC^[18]、NQM^[19]、WSNR^[20]、SNR、VSI^[21]、FSIM^[22]、IWSSIM^[23]、GSM^[24]、MAD^[25]、SRSIM^[26]、RFSIM^[27]。以上的VQA方法都是基于计算视频帧间相似度信息来进行视频质量测量。其中,MSSIM特征已经被应用于视频帧间篡改检测并取得较高的检测率^[11-12]。本文不仅测试将其他VQA特征应用于视频帧间篡改检测的可行性,而且将它们融合到一个统一的算法中。

2.3 多特征融合

Liu等^[28]通过融合图像质量评估特征(IQA, image quality assessment)来实现客观图像质量评估,根据失真类型将图像分为3到5组,使用机器学习方法训练预测模型,在6个代表性数据库中进行了测试,结果表明提出的融合方法比现有的IQA方法性能好。Lin等^[29]融合VQA特征进行客观视频质量评估,将视频按压缩率和调整大小分组,在每组内融合几个VQA特征预测感知质量,通过机器学习减少了内容多样性并提高融合性能,在MCL-V数据集中进行测试,结果表明融合VQA特征比单个VQA特征的评估效果更好。上述特征融合方法均是针对质量评估的。在篡改检测方面,Huang等^[30]提出了融合音频的多通道方法。该方法利用篡改视频文件的视频通道时会影响音频通道的原理,取得了较好的检测效果。然而,许多视频常常没有音频通道(如监控视频),这就限制了该方法的应用范围。Shanableh等^[31]从视频比特流中提取多特征来对MPEG视频进行检测,得到了较好的效果。然而,他们只

用到了MPEG视频中P帧的一些简单特征,而本文的算法和所选用的特征则不受限于具体的视频编码格式。而且,他们的方法仅对帧删除的篡改进行了测试,而本文会对复制、插入和删除等篡改都进行测试。此外,本文还提出利用视频分组来改进特征融合的效果。

表1 质量评估指标
Table 1 Quality evaluation index

指标索引	简称	中文全称	英文全称
m1	MSE	均方误差	mean-square error
m2	PSNR	峰值信噪比	peak signal-to-noise ratio
m3	SSIM	结构相似性	structural similarity index
m4	MSSIM	多尺度结构相似性	multiscale-SSIM
m5	VSNR	视觉信噪比	visual signal-to-noise-ratio
m6	VIF	视觉信息保真度	visual information fidelity
m7	VIFP	基于像素的视觉信息保真度	pixel-based VIF
m8	UQI	全局质量指标	universal quality index
m9	IFC	信息保真度测量	information fidelity criterion
m10	NQM	噪声质量检测	noise quality measure
m11	WSNR	加权信噪比	weighted signal-to-noise ratio
m12	SNR	信噪比	signal-to-noise ratio
m13	VSI	视觉显著索引	a visual saliency induced index
m14	FSIM	特征相似性	feature-similarity
m15	IWS-SIM	信息内容加权的结构相似性	information content weighted SSIM
m16	GSM	梯度相似性度量	gradient similarity based metric
m17	MAD	最显著失真	most apparent distortion
m18	SRSIM	频谱残差相似度	spectral residual based similarity
m19	RFSIM	基于Riesz变换的特征相似度	Riesz transforms based feature similarity

3 帧间篡改检测融合算法

融合算法的目的是融合多个特征,获得比利用单一特征检测精确度更高的视频篡改取证算法。但视频特征数量多,将所有可能的组合遍历一遍显然是不可取的。本文使用特征递归消除算法(SVM-RFE)对特征进行排序,然后用文献[32]

中提到的顺序前向选择 (SFS, sequential forward method selection) 算法和 Adaboost 二元分类器对排序好的特征逐个融合, 每次留下能提高当前得分的特征, 这样只要遍历一遍特征集就能得到一组满意的结果, 实验表明该方法可以提高融合算法的检测精度。融合算法流程如图 1 所示。

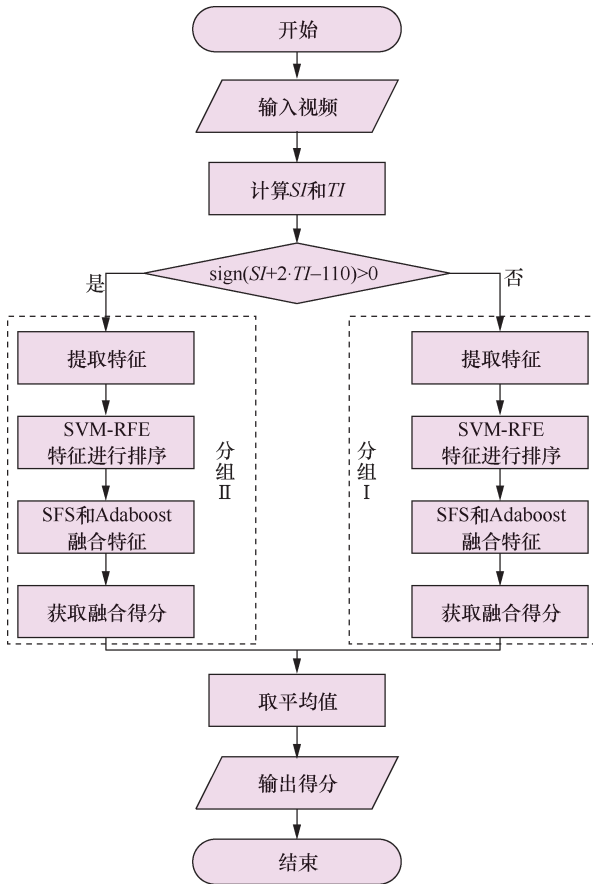


图1 融合算法流程
Figure 1 Fusion algorithm flow

3.1 视频分组

将类似内容的视频分类到同一组当中, 可以在每一组中构建更加准确的检测模型。在每一组中, 希望有足够多的样本和视频来学习和训练。根据实验的情况分析发现, 利用空间信息 (SI) 和时间信息 (TI) 将视频分成两组, 可以取得比不分组更好的检测效果。

国际电信联盟组织 (ITU-T) 给出了空间信息和时间信息的定义^[33]。SI 用于度量图片中的空间细节。空间更复杂的场景的 SI 值一般会更高。

TI 用于衡量视频序列时间变换。高运动序列中的图片的 TI 值通常会更高。SI 基于 Sobel 滤波器, 首先用 $[Sobel(F_n)]$ 滤波器对时间 n 处的视频帧 F_n (亮度平面) 进行滤波, 然后计算经过 Sobel 滤波帧像素上的标准偏差。对视频序列上每一帧重复该操作, 并产生场景空间信息的时间序列。选择时间序列中最大值来表示场景的空间信息。以上过程可以用式(1)表示。

$$SI = \max_{\text{time}} \{std_{\text{space}}[Sobel(F_n)]\} \quad (1)$$

其中, F_n 表示视频帧序列的第 n 帧, $[Sobel(F_n)]$ 表示的是对视频帧进行 Sobel 滤波, std_{space} 表示取标准偏差, $\max_{\text{time}}$ 表示取最大值。

TI 基于视频帧之间的运动差异特征 $M_n(i, j)$, 是连续两帧之间相对位置上的 (亮度平面的) 像素值之差。在视频帧序列第 n 帧处, $M_n(i, j)$ 定义为

$$M_n(i, j) = F_n(i, j) - F_{n-1} \quad (2)$$

其中, $F_n(i, j)$ 是视频帧序列第 n 帧 i 行 j 列处的像素值。

计算 TI, 先对视频帧序列提取的运动差异特性 $M_n(i, j)$, 再计算标准偏差。TI 为沿着时间轴计算 $M_n(i, j)$ 的空间标准偏差的最大值, 可以用式(3)表示。

$$TI = \max_{\text{time}} \{std_{\text{space}}[M_n(i, j)]\} \quad (3)$$

利用以下分割线将视频分为两组, 如式(4)所示。

$$Gx = \text{sign}(aSI_x + bTI_x + c) \quad (4)$$

其中, x 为视频样本, SI_x 和 TI_x 分别为对应视频的 SI 和 TI 值。图 2 展示了 SULFA^[27] 视频库中的视频分组结果。其中, 为了使数据样本均匀分布到每一组中, 取 $a=1$, $b=2$, $c=-110$ 。当 G_x 值为 -1 时, 将视频分为组 I; 当 G_x 值为 1 时, 将视频分为组 II。

3.2 特征提取

在分组之后, 分别对每组里的视频提取特征。因为要检测视频的帧间篡改, 所以计算视频的帧间 VQA 特征。首先将视频分解成帧序列, 然后逐帧提取两帧间的 19 个 VQA 特征 (见表 1)。有以下帧序列:

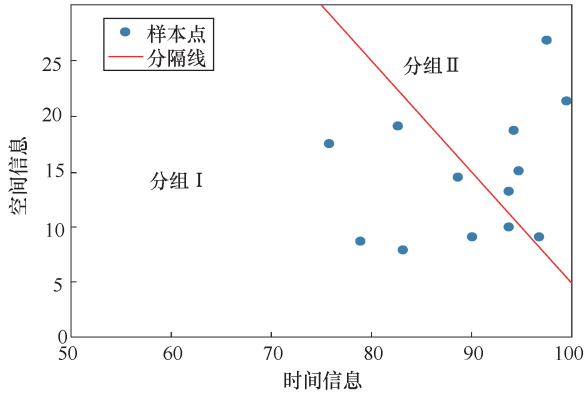


图2 视频分组
Figure 2 Video grouping

$$F = \{F_1, F_2, F_3, \dots, F_i, \dots, F_n\} \quad (5)$$

对于 n 个连续的视频帧，可以提取 $n-1$ 次帧间相关特性，每次都提取 19 个 VQA 特征，如式(6)所示。

$$x_i = f_{\text{VQA}}(F_i, F_{i+1}), i=1, 2, \dots, n-1 \quad (6)$$

其中， f_{VQA} 代表所有的 VQA 特征提取方法， x_i 表示相邻帧 F_i 和 F_{i+1} 所提取的 19 个 VQA 特征。 x_i 作为模型的输入样本。在未被篡改的原始视频帧序列中，相邻帧间具有很强的相关性；而在被篡改的视频帧附近，相关性会减弱，此时某些 VQA 特征会有明显变化。在这 19 个 VQA 特征中，SSIM、MSSIM、VIF、VIFP、UQI、VSI、FSIM、IWSSIM、GSM、SRSIM、RFSIM 的特征值介于 0 到 1 之间，特征值越接近 1 表示相邻帧相关性越高。MSE 值越小说明两帧之间的相关性越高。而 PSNR、VSNR、IFC、NQM、WSNR、SNR、MAD 特征值越大则说明两帧之间相关性越高。在实验过程中，所有数据的数值被归一化为 0 到 1 之间。

3.3 特征排序与融合

在提取特征后，首先对这些特征排序，然后根据排序的结果进行融合。排序的目的是尽量让样本中对融合性能有正面贡献的特征排在前面。本文使用特征递归消除算法对特征进行排序^[34]，该方法使用特征值的大小作为排名标准。SVM-RFE 算法过程如算法 1 所示。算法输入样本集 $X = (x_1, x_2, \dots, x_n)$ ； $y = \{1, 0\}$ ，其中 n 为样本数，每个样本 x_i 表示的是两帧之间提取的 VQA 特征，包含了表 1 中展示的所有特征； y 为样本标签，

正常点处提取的样本用 1 表示，异常点处提取的样本用 0 表示。初始化特征序列 s （包含表 1 中的所有特征）和特征排序序列 r （初始值为空）。首先选取训练样本 X_s ，包含 s 中的所有特征。然后使用选取的样本 X_s 训练 SVM 分类器。接着根据训练 SVM 分类器得到最优化权重向量 w 。然后使用最优化权重向量 w 计算候选特征的排序标准 c ，排序标准 c 越小说明对应的特征在当前训练样 X_s 中是最不重要的。之后选出排序标准 c 中得分最小的排序特征 f 。然后将特征 f 放入特征排序列表 r 的表头，即本轮选取的特征 f 都是已选特征的排序列表 r 中的最佳特征。最后从特征序列 s 中删除本轮选中的最不重要的特征 f 。重复以上步骤，直到特征序列 s 为空。

算法 1 SVM-RFE 特征排序算法

输入 $X = (x_1, x_2, \dots, x_n)$ ； $y = \{1, 0\}$

输出 r

初始化 特征序列 $s = [1, 2, \dots, n]$ ；特征排序列表 $r = []$

- 1) 获取候选特征集合的训练样本包含 s 中的所有特征： $X_s = X(:, s)$
- 2) 训练分类器： $a = \text{SVM-train}(X_s, y)$
- 3) 计算候选特征的权重向量： $w = \sum_k \alpha_k y_k x_k$
- 4) 计算候选特征排序标准： $c_i = (w_i^2)$
- 5) 找到排序得分最小的特征： $f = \arg \min(c)$
- 6) 更新特征排序列表，将当前循环得到的特征 $s(f)$ 加入表 r 的表头： $r = [s(f), r]$
- 7) 从特征序列中去除排序得分最小的特征： $s = s - s(f)$
- 8) 重复步骤 1)~7) 直到候选特征序列 s 为空： $s = []$

对输出的特征排序列表 r ，使用顺序前向选择算法对特征进行融合。思路是依次添加待选特征，直到最终选出得分最高的一组特征集。算法过程展示在算法 2 中，其中，特征集 $M = \{m_j | j = 1, \dots, N\}$ ，1 到 N 的排序是按照 SVM-RFE 算法输出的特征排序列表 r 排序的。判断特征 m_j 是否加入子集 M_s 的目标函数 J 为 Adaboost 二元分类器。Adaboost 分类器的工作原理是先从初始训练样本中训练出一个基学习器，再根据基学习器的表现

对样本分布进行调整, 使在先前做错的样本在后续的学习器中受到更多的关注, 然后基于调整分布的样本来训练下一个基学习器; 重复进行这个步骤, 直到弱学习器学习的次数达到预设的阈值 T 时, 最后将这 T 个学习器通过结合策略进行结合, 得到最终的强学习器。Adaboost 的优点在于作为分类器时, 分类精度高, 在本文的实验结果中也证明了这一点。因此, 在本文的工作中用 Adaboost 分类器作为目标函数 J 判断当前特征是否加入特征子集 M_s 。当 $J(M_n + m_j) > J(M_n)$ 时, 特征 m_j 加入子集 M_s , 直到遍历完所有的特征, 选出一组得分最高的特征作为最终的融合特征。

算法 2 顺序前向选择

输入 SVM-RFE 排序好的特征集合 $M = \{m_j | j=1, \dots, N\}$

输出 M_s

1) 设置特征子集 $M_s = \{\emptyset\}$

2) $m^* = m_j$

3) 判断 $J(M_s + m^*) > J(M_s)$

4) 如果大于则更新 $M_s = M_s + m^*$

5) $j = j + 1$; 返回 2)

使用算法 1 对特征进行排序, 再使用算法 2 选择特征并逐个融合特征, 最终每个分组得到一组特征集。在所有样本中使用融合特征集获取融合样本, 将样本分为训练样本和测试样本。使用融合测试样本训练 Adaboost 二元分类器, 获得一个融合算法的分类模型。将测试样本输入融合分类模型得到每组融合模型得分。为了方便分组融合模型与不分组融合模型的比较, 使用平均法取两个分组的平均值作为融合算法的得分, 最后得到整个融合算法的视频取证得分。得分越高说明融合算法检测帧间篡改的精度越高。

4 实验结果与分析

4.1 实验设置

本文使用了华南理工的视频篡改检测数据库 (VFDD, video forgery detection database) Version1.0^[35]和 SULFA (surrey university library for forensic analysis)^[36]视频库中的 220 个视频对融合算法的有效性进行验证。实验中使用 Matlab

R2016b 提取视频特征, 使用 Scikit-learn^[37]库进行特征选择以及分类。实验环境为 Intel Core i5-4590 CPU 3.30 GHz、8 GB 内存、Windows 10 系统。

4.2 实验样本来源

首先构造一个视频帧间篡改的数据集。该数据集包含 44 686 条样本, 其中正负样本各占一半。视频内容包括日常生活场景, 如行人走路、汽车在公路行使、做体育运动 (踢球、打羽毛球、打乒乓球、跑步)、自然风景等。图 3 展示了帧间篡改操作。其中, 图 3(a)为原始帧序列 (以第 1 到第 12 帧为例); 图 3(b)显示了帧插入篡改操作, 其中 x 、 y 、 z 为插入帧; 图 3(c)显示了帧删除篡改操作, 其中第 7、8、9 帧 (用白色表示) 已从视频帧序列中删除; 图 3(d)显示了帧复制粘贴篡改操作, 第 10 到第 12 帧被复制粘贴到了第 3 帧和第 4 帧之间。上述 3 种帧间篡改操作会导致视频帧序列发生变化, 在构造样本时, 篡改点处对应的是负样本。如图 3(b)中, 编号为 3 的帧和编号为 x 的帧之间是异常点, 编号为 z 的帧与编号为 4 的帧之间也是异常点; 图 3(c)中, 编号为 6 的帧与编号为 10 的帧之间是异常点。图 3(d)中, 编号为 3 的帧与其相邻的编号为 10 的帧之间是异常点, 编号为 12 与相邻的编号为 4 的帧之间是异常点。在异常点处的样本是负样本。帧间篡改的视频异常点比正常点要少得多。然而, 训练分类模型一般要求正负样本数相当才能训练出好的分类模型。因此, 在生成数据集时, 每生成一个正常点, 就随机构造一个异常点。构造方法如下, 假设原始视频帧序列是从 F_1 到 F_n , 如果正常点 F_1 和 F_2 之间提取特征作为正样本, 则同时随机选取一帧 F_i , F_1 与 F_i 构成负样本。这样, 对于一个原始视频可以生成等量的正样本和负样本。考虑到视频篡改的帧数一般超过 5 帧, 本文选取的随机帧 F_i 与当前帧的间隔在 5~35 帧之间。

4.3 分类器的比较与选择

一组样本的众多分类器的得分如表 2 所示。从表中可以看出 AdaBoost 分类器的效果较佳, 优于其他分类效果。选择 AdaBoost 作为分类器时, 分类精度高, 作为简单的二元分类器时, 构造简单, 且不容易发生过拟合。

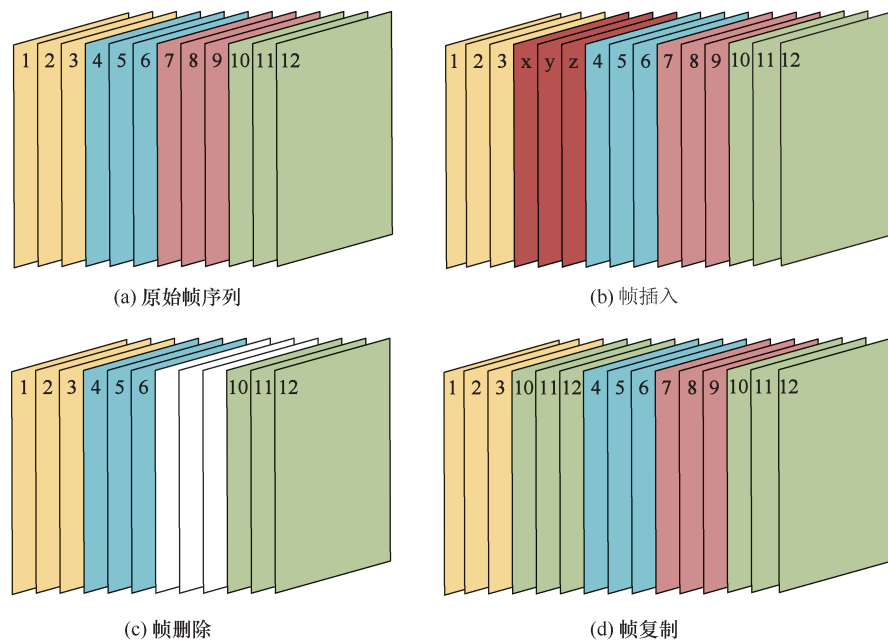


图 3 帧间篡改操作

Figure 3 Interframe tamper operation

表 2 分类器得分
Table 2 Classifier score

神经网络	决策树	K 最近邻	逻辑回归	朴素贝叶斯	随机森林	支持向量机	AdaBoost
0.862 465	0.724 582	0.868 036	0.885 097	0.862 465	0.823 12	0.868 036	0.903 552

4.4 实验结果与分析

大量的实验结果表明融合所有特征往往不能达到最好的效果。因此,一般不需要融合所有特征,只需要找到一个较佳的融合数量即可。在表 3 中第 1 到 19 栏展示了逐一添加特征的结果,第 20 栏是本文融合算法添加特征的结果,并绘制了结果曲线,如图 4 所示,以便更好地展示融合效果。表 3 中索引号 1 到 19 栏展示的是先对特征进行排序,按照排序结果依次添加特征后的得分,其过程中没有排除任何特征。最后一栏(索引号 20)展示的是使用本文中提到的算法 2(只添加能提高当前检测精度的特征)融合特征的结果。表 3 中第一个所选的特征是给定样本下得分最高的特征,在增加一个特征后,其性能反而下降了。图 4 可以清晰地看出当特征数从 2 增加到 4 时,性能有大幅提升。当融合特征数为 4 时有最佳性能。特征数大于 4 之后,性能有所下降也有所上升,一般保持在一个比较平稳的水平。最后一栏只融合了部分特征也获得了较高的得分。从图 4 中可以看出融合方法确实有利于提高检测精度且

使用本文的算法不需要融合所有特征也可以选择一组效果较佳的特征。高维数据性能有所下降而不是改善的原因有两点:特征增加意味着噪声和误差也随之增加;样本数据量不足以获得统计上合理和可靠的估计。如果选取一组样本,要找到最佳的特征组合,需要遍历所有可能,显然,这是不可取的。那么,使用本文提到的顺序前向选择算法,可以找到一组性能较佳的特征,而只需遍历一遍所有特征,且复杂度低。

对本文所提融合算法与单一 VQA 方法在华南理工视频篡改检测数据库和 SULFA 视频库的精确度进行比较。结果如表 4 所示。表中第一列展示了方法的名称。其中 MMF 为本文的融合算法,MMF 分有分组和无分组两种,第二列是 VFDD 库中每种方法的得分,第三列是 SULFA 库中每种方法的得分,得分越高表示检测精度越好。在 VFDD 视频库中表现最好的单一特征是 IFC(得分 0.918 623),而在 SULFA 视频库中表现最好的单一特征为 VSNR(得分 0.666 803);同时也可以看出 IFC 特征在 SULFA 视频库中的得分

表 3 融合步骤得分
Table 3 Fusion step score

索引	选入特征	融合框架得分
1	m9	0.958 257
2	m9、m12	0.902 498
3	m9、m12、m2	0.962 523
4	m9、m12、m2、m11	0.983 242
5	m9、m12、m2、m11、m1	0.971 359
6	m9、m12、m2、m11、m1、m5	0.971 054
7	m9、m12、m2、m11、m1、m5、m10	0.970 75
8	m9、m12、m2、m11、m1、m5、m10、m17	0.972 273
9	m9、m12、m2、m11、m1、m5、m10、m17、m19	0.969 226
10	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16	0.968 007
11	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18	0.969 835
12	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8	0.972 273
13	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13	0.969 835
14	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4	0.972 882
15	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4、m14	0.973 187
16	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4、m14、m3	0.974 711
17	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4、m14、m3、m6	0.966 789
18	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4、m14、m3、m6、m7	0.966 484
19	m9、m12、m2、m11、m1、m5、m10、m17、m19、m16、m18、m8、m13、m4、m14、m3、m6、m7、m15	0.970 445
20	m9、m16、m6、m7、m15	0.971 054

只有 0.481 829，相比于其他特征的表现有明显的不足，同样 VSNR 在 VFDD 视频库中的表现相对于其他特征也有明显的不足。从表 4 中可以看出融合算法在两个数据库中的性能表现都是最优的。VFDD 数据库中，MMF（无分组）融合的特征为 IFC、VIF、MAD、IWSSIM、GSM、VSI、SSIM。MMF（分组后）将视频分为两组，第一组融合特征为 IFC、NQM、PSNR、SNR、VSNR、WSNR；第二组的融合特征为 IFC、MSE、NQM、PSNR、VSNR、UQI、SSIM。分组与未分组的融合方法性能几乎无差别；在 SULFA 中，MMF（无分组）融合的特征为 IFC、NQM。MMF（分组后）将视频分为两组，第一组融合特征为 IFC、IWSSIM、SRSIM；第二组的融合特征为 WSNR、SNR、PSNR。分组后的融合方法比未分组的融合方法在性能上有显著的提升。这说明正确的分

组有利于融合算法性能的提升。在时间上效率上，融合算法比单一特征多花费 200 s。相比于检测效果的提升，额外的开销时间在可接受范围之内。

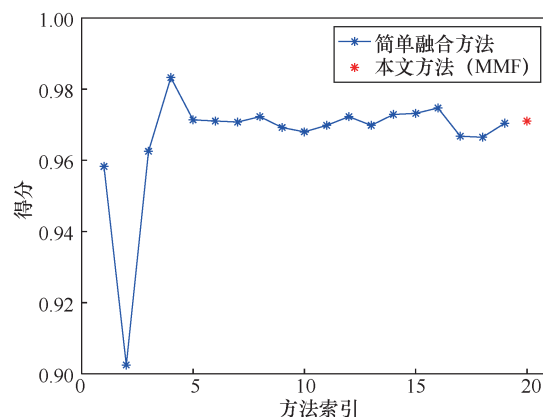


图 4 融合方法曲线
Figure 4 Fusion method graph

表4 融合模型性能比较
Table 4 Performance comparison of fusion models

方法名称	VFDD 库中方法的得分	SULFA 库中方法的得分
MSE	0.685 269	0.564 72
PSNR	0.685 215	0.566 762
SSIM	0.702 852	0.582 687
MSSIM	0.687 145	0.613 312
VSNR	0.667 632	0.666 803
VIF	0.800 633	0.543 079
VIFP	0.799 614	0.460 188
UQI	0.777 796	0.517 762
IFC	0.918 623	0.481 829
NQM	0.685 751	0.574 929
WSNR	0.693 149	0.508 371
SNR	0.693 256	0.477 746
VSI	0.651 71	0.490 813
FSIM	0.670 687	0.597 795
IWSSIM	0.663 611	0.625 153
GSM	0.689 182	0.599 428
MAD	0.749 518	0.644 753
SRSIM	0.751 581	0.571 254
RFSIM	0.726 493	0.601 062
MMF (无分组)	0.928 602	0.676 194
MMF (分组后)	0.929 236	0.794 227

5 结束语

本文提出了融合多特征的视频帧间篡改检测算法。首先计算视频帧间的 TI 和 SI，使用 TI 值和 SI 值对视频进行分组；分组后提取视频帧间的 VQA 特征；对提取的 VQA 特征使用特征递归消除法对特征进行排序；然后再使用顺序前向选择算法和 Adaboost 二元分类器对排好序的 VQA 特征逐个融合，最终得到每组融合方法的得分，最后取两组得分的平均值得到整个算法的帧间篡改分类得分。实验对比表明，融合算法的篡改检测效果优于单一特征，且对视频进行分组能够提高融合算法的得分。下一步工作，将做更多的实验来寻找更合适的视频分组策略，进一步提算法的融合性能。未来也会尝试使用深度学习的方法来自动提取特征及融合。

参考文献:

- [1] INDEX C V N. Forecast and methodology, 2016–2021[R]. White Paper, 2017, 6.
- [2] IULIANI M, SHULLANI D, FONTANI M, et al. A video forensic framework for the unsupervised analysis of MP4-like file container[J]. IEEE Transactions on Information Forensics and Security, 2018, 14(3): 635–645.
- [3] AFCHAR D, NOZICK V, YAMAGISHI J, et al. Mesonet: a compact facial video forgery detection network[C]//2018 IEEE International Workshop on Information Forensics and Security (WIFS). 2018: 1–7.
- [4] MATERN F, RIESS C, STAMMINGER M. Exploiting visual artifacts to expose deepfakes and face manipulations[C]//2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). 2019: 83–92.
- [5] SINGH R D, AGGARWAL N. Video content authentication techniques: a comprehensive survey[J]. Multimedia Systems, 2018, 24(2): 211–240.
- [6] CHAO J, JIANG X, SUN T. A novel video inter-frame forgery model detection scheme based on optical flow consistency[C]//International Workshop on Digital Watermarking. 2012: 267–281.
- [7] WU Y, JIANG X, SUN T, et al. Exposing video inter-frame forgery based on velocity field consistency[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2014: 2674–2678.
- [8] LIU Y, HUANG T. Exposing video inter-frame forgery by Zernike opponent chromaticity moments and coarseness analysis[J]. Multimedia Systems, 2017, 23(2): 223–238.
- [9] WANG W, FARID H. Exposing digital forgeries in video by detecting double MPEG compression[C]//The 8th Workshop on Multimedia and Security. 2006: 37–47.
- [10] AGHAMALEKI J A, BEHRAD A. Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding[J]. Signal Processing: Image Communication, 2016, 47: 289–302.
- [11] LI F, HUANG T. Video copy-move forgery detection and localization based on structural similarity[C]//The 3rd International Conference on Multimedia Technology (ICMT 2013). 2014: 63–76.
- [12] 张珍珍, 侯建军, 李赵红, 等. 基于 MSSIM 商一致性的视频插帧和删帧篡改检测[J]. 北京邮电大学学报, 2015, 38(4): 84–88.
ZHANG Z Z, HOU J J, LI Z H, et al. Video-frame insertion and deflection detection based on consistency of quotients of MSSIM[J]. Journal of Beijing University of Posts and Telecommunications, 2015, 38(4): 84–88.
- [13] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600–612.
- [14] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. 2003, 2: 1398–1402.
- [15] CHANDLER D M, HEMAMI S S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images[J]. IEEE Transactions on Image Processing, 2007, 16(9): 2284–2298.

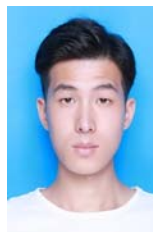
- [16] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing, 2006, 15(2): 430-444.
- [17] WANG Z, BOVIK A C. A universal image quality index[J]. IEEE Signal Processing Letters, 2002, 9(3): 81-84.
- [18] SHEIKH H R, BOVIK A C, DE VECIANA G. An information fidelity criterion for image quality assessment using natural scene statistics[J]. IEEE Transactions on Image Processing, 2005, 14(12): 2117-2128.
- [19] DAMERA-VENKATA N, KITE T D, GEISLER W S, et al. Image quality assessment based on a degradation model[J]. IEEE Transactions on Image Processing, 2000, 9(4): 636-650.
- [20] MANNOS J, SAKRISON D. The effects of a visual fidelity criterion of the encoding of images[J]. IEEE Transactions on Information Theory, 1974, 20(4): 525-536.
- [21] ZHANG L, SHEN Y, LI H. VSI: a visual saliency-induced index for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2014, 23(10): 4270-4281.
- [22] ZHANG L, ZHANG L, MOU X, et al. FSIM: a feature similarity index for image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(8): 2378-2386.
- [23] WANG Z, LI Q. Information content weighting for perceptual image quality assessment[J]. IEEE Transactions on Image Processing, 2011, 20(5): 1185-1198.
- [24] LIU A, LIN W, NARWARIA M. Image quality assessment based on gradient similarity[J]. IEEE Transactions on Image Processing, 2012, 21(4): 1500-1512.
- [25] LARSON E C, CHANDLER D M. Most apparent distortion: full-reference image quality assessment and the role of strategy[J]. Journal of Electronic Imaging, 2010, 19(1): 011006.
- [26] ZHANG L, LI H. SR-SIM: a fast and high performance IQA index based on spectral residual[C]//2012 19th IEEE International Conference on Image Processing. 2012: 1473-1476.
- [27] ZHANG L, ZHANG L, MOU X. RFSIM: a feature based image quality assessment metric using Riesz transforms[C]//2010 IEEE International Conference on Image Processing. 2010: 321-324.
- [28] LIU T J, LIN W, KUO C C J. Image quality assessment using multi-method fusion[J]. IEEE Transactions on Image Processing, 2012, 22(5): 1793-1807.
- [29] LIN J Y, LIU T J, WU E C H, et al. A fusion-based video quality assessment (FVQA) index[C]//Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. 2014: 1-5.
- [30] HUANG T, ZHANG X, HUANG W, et al. A multi-channel approach through fusion of audio for detecting video inter-frame forgery[J]. Computers & Security, 2018, 77: 412-426.
- [31] SHANABLEH T. Detection of frame deletion for digital video forensics[J]. Digital Investigation, 2013, 10(4): 350-360.
- [32] LIU T J, LIN W, KUO C C J. Image quality assessment using multi-method fusion[J]. IEEE Transactions on Image Processing, 2012, 22(5): 1793-1807.
- [33] INSTALLATIONS T, LINE L. Subjective video quality assessment methods for multimedia applications[J]. Networks, 1999, 910(37): 5.
- [34] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002, 46(1-3): 389-422.
- [35] HU Y J, SALMAN A H, WANG Y F, et al. Construction and evaluation of video forgery detection database[J]. Journal of South China

University of Technology (Natural Science), 2017, 45(12):57-64.

- [36] QADIR G, YAHAYA S, HO A T, Surrey university library for forensic analysis (SULFA) of video content[C]//IET Conference on Image Processing. 2012: 1-6.

- [37] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, et al. Scikit-learn: machine learning in Python[J]. Journal of Machine Learning Research, 2011, 12: 2825-2830.

[作者简介]



肖辉(1991-), 男, 福建建瓯人, 福建师范大学硕士生, 主要研究方向为信息安全、数字多媒体取证。



翁彬(1981-), 男, 福建福州人, 博士, 福建师范大学讲师, 主要研究方向为机器学习及应用。



黄添强(1971-), 男, 福建仙游人, 博士, 福建师范大学教授、博士生导师, 主要研究方向为机器学习、数字多媒体取证。



普茵(1995-), 女, 河南平舆人, 福建师范大学硕士生, 主要研究方向为信息安全、数字多媒体取证。



黄则晖(1999-), 女, 福建仙游人, 主要研究方向为多媒体编辑与传播。