

基于关键帧提取与 DBSCAN 的视频 内容篡改识别技术

任 飞

(国家信息中心, 北京 100045)

摘 要: 针对篡改视频内容审查和筛选导致大量人力资源消耗问题, 本文提出了一种识别视频篡改内容的技术。首先, 该技术通过计算帧间差异提取关键帧; 其次, 通过对关键帧与原始帧进行差值分析获取篡改位置; 再次, 基于位置关联和语义聚合现象, 利用密度聚类算法 (DBSCAN) 算法解决因篡改位置和内容结构的随机性造成的位置分散和语义离散问题; 最后, 通过光学字符识别技术解析篡改的具体内容。本文方法实现了视频篡改内容的时空定位和内容识别, 为公共安全、传媒、商务等领域的视频内容检测提供了坚实的技术基础。

关键词: 关键帧抽取; SSIM; DBSCAN; 形态学变换; 光学字符识别

中图分类号: TP309.2 **文献标志码:** A **文章编号:** 1671-5497(2026)02-0555-09

DOI: 10.13229/j.cnki.jdxbgxb.20241261

Video content tampering identification technology based on key frame extraction and DBSCAN

REN Fei

(State Information Center, Beijing 100045, China)

Abstract: To solve the problem of a large amount of human resource consumption caused by tampering with video content review and screening, a technique for identifying tampered content in videos is proposed in this paper. First, key frames are extracted by calculating inter-frame differences. Then, the locations of tampering are determined through differential analysis between the key frames and the original frames. Subsequently, based on the phenomenon of location association and semantic aggregation, the problems of location dispersion and semantic dispersion are solved by the DBSCAN algorithm, which is caused by the randomness of tampering with location and content structure. Finally, optical character recognition (OCR) technology is applied to decipher the specific content that has been altered. Spatio-temporal position and content identification of tampered video content are achieved by the proposed method, providing a solid technical foundation for video content inspection in fields such as public safety, media, and business.

Key words: keyframe extraction; SSIM; DBSCAN; morphological transformation; optical character recognition

收稿日期: 2024-11-22.

基金项目: 国家重点研发计划项目 (2023YFB4503200).

作者简介: 任飞 (1986-), 女, 高级工程师, 博士. 研究方向: 网络 and 信息安全, 数据安全. E-mail: lypp_1116@163.com

0 引言

视频作为一种信息载体和表现形式,能够携带大量丰富的图像、文字等信息^[1],在公共安全、信息交流等场景中发挥着不可替代的作用。然而,视频在传输过程中易被截取并在数据流中掺入其他信息,如文字、数字、logo等,对视频数据传递的内容和含义造成不同程度的扭曲和破坏,导致内容失真甚至影响社会治安稳定。因此,为了防止原始视频被恶意篡改后广泛传播,对社会公共安全造成重大隐患,需要一种精准、高效、可行的视频内容篡改识别技术,检测视频在传播的过程中是否被恶意篡改,并定位到篡改位置和识别到篡改内容,维护网络安全。视频篡改内容识别技术在保障视频的真实性和完整性、维护公众利益、维护社会秩序以及促进技术创新和发展等方面都具有非常重要的意义。

近年来,视频内容篡改识别技术成为备受关注的研究方向,国内外众多研究人员持续致力于探索更为先进的识别技术。基于光流一致性的视频帧间篡改检测(A novel video inter-frame forgery model detection scheme based on optical flow consistency)^[2],通过窗口移动的方式计算帧之间的光流特征。当视频遭受篡改时,相邻帧之间的光流高度不一致,据此可查找出篡改帧的所在位置。光流通过跟踪图像中的特征点或像素估计运动的速度和方向,在计算机视觉领域常用于物体运动分析等方面。基于结构相似度均值(Mean structural similarity index, MSSIM)商的一致性检测^[3,4],提出相邻帧之间的MSSIM商具有连续性,而被篡改的视频其篡改位置的MSSIM商会发生突变,从而实现视频内容篡改识别。该算法的准确性较高,但其篡改识别的效果对专家经验具有较高的依赖性,并不具备自适应能力。基于聚类^[5]的篡改检测算法会利用多种方法对提取到的视频帧特征进行聚类,然后定位到篡改区域的位置,从而实现视频篡改识别。使用到的图像特征包括颜色特征(颜色直方图、颜色矩)、统计特征(Hu矩)、轮廓特征(SIFT、SURF)、空间关系特征等。基于聚类的篡改检测算法泛用性强,可以根据不同的视频内容有针对性地选择视频帧特征,从而获得较高准确度的检测结果。

当前国内外在视频内容篡改识别方面已取得显著进展。然而,对于视频中插入了具有歧义性

的文字,从而导致视频传达的信息出现误导的情况,现有方法识别仍存在困难。传统的篡改检测算法通常能够定位到被篡改内容在视频中的时间位置,但对于篡改的具体内容,常需要依赖人工审查才能做出最终判断。此外,随着视频文件体积的增大,逐帧检测的方法逐渐成为效率瓶颈,在实际应用中难以满足快速响应的需求。

针对以上问题,本文提出一种基于关键帧提取与DBSCAN的视频内容篡改识别技术。该算法的流程如下:首先,通过分析篡改视频的图像特征^[6,7],获取能够体现视频主要内容的关键帧;其次,对篡改视频的关键帧和源视频相同位置的原始帧进行差值分析,获得篡改内容的空间位置信息;再次,根据语义近邻现象和空间位置信息,利用DBSCAN算法聚合内容相关的篡改区域,获取具有语义完整性的篡改内容的空间位置;最后,依据空间位置信息截取篡改区域,利用光学字符识别技术(Optical character recognition, OCR)^[8,9]提取篡改内容。

一方面,该算法基于关键帧提取技术,能够在实现应检尽检的同时只检应检,从而降低后续内容识别过程中的计算消耗。另一方面,利用位置关联和语义近邻现象,基于DBSCAN的语义聚合技术能够提高单区域篡改内容的完整性和多区域篡改内容之间的独立性,从而提高对篡改内容的识别能力,有利于发挥篡改检测的自动化优势。

1 相关工作

1.1 关键帧

处理视频前需要对视频进行数据预处理,以有效提取视频中的关键信息。视频作为信息传递的媒介,由连续或非连续的图像序列构成。随着技术的进步,分辨率、码率和帧率等参数不断提升,以提高视觉质量,这也导致视频中存在大量冗余信息。为了有效识别篡改内容,必须对这些冗余信息进行过滤和处理。

如图1所示,视频是由帧、镜头、场景,通过不断堆叠,并按时间顺序拼接而成。帧是视频的最小组成单元,也是内容识别的主要对象,数据预处理模块的主要任务就是从完整的视频流中提取到具有关键信息的帧,即关键帧,也称为代表帧,它通常会反映一个镜头的主要内容^[10,11]。在一个镜头中,相似度最高的帧被选为关键帧,关键帧的数

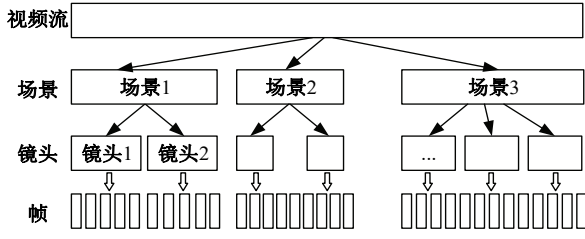


图1 视频构成

Fig. 1 Video composition

量取决于镜头的复杂程度。

1.2 帧间差异性算法

帧间差异性衡量视频内容变化程度的重要指标,是基于场景检测的关键帧提取技术的核心。常见的帧间差异性算法有峰值信噪比(Peak signal-to-noise ratio, PSNR)^[12]和结构相似性(Structural similarity, SSIM)^[13,14]。

PSNR用于评估图像信号与噪声之间的比值,数值越高表示图像的质量越好。在图像处理领域,PSNR通常采用对数分贝单位表示,30~50 dB的值被认为是理想的。对于彩色图像,PSNR可以针对每个颜色通道进行计算,最终取平均值以评估整体质量。尽管PSNR广泛使用,但由于其基于像素点误差,未考虑人眼视觉特性,因此有时评价结果与主观感受不一致。

PSNR的公式假设一张大小为 $H \times W$ 的图像由清晰的原始图像 X 和覆盖在原始图像上的噪声 Y 组成,该图像的均方误差(MSE)定义为:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2 \quad (1)$$

该图像的PSNR定义为:

$$PSNR = 10 \lg \left(\frac{(2^n - 1)^2}{MSE} \right) \quad (2)$$

式中: $(2^n - 1)^2$ 表示这张图像的可能最大像素值; n 表示这张图像的每个像素需要多少个字节(bit)。

SSIM认为自然图像具有显著的结构性,尤其在空间相似情况下,像素间存在强烈的关联性。这些关联性在视觉场景中承载了关于物体结构的重要信息。其假设人类视觉系统(Human visual system, HVS)主要从可视区域内获取结构信息^[13]。因此,通过探测结构信息是否发生变化,可以感知图像失真的近似信息。

SSIM基于两张图像 X 与 Y 之间的亮度、对比度和结构3种关系的对比定义:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2\mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

式中: c_1, c_2 为两个常数; μ_x 和 μ_y 分别为图像 X 和 Y 的均值; σ_x^2 和 σ_y^2 分别为 X 和 Y 的方差; σ_{xy} 为 X 和 Y 的协方差。

1.3 图像增强

图像增强^[15]是一种数字图像处理技术,旨在改善图像的质量,以增强视觉效果,或者有选择地强调图像整体或局部特性。该技术通常包括除去图像中的噪声,使边缘清晰,以及突出图像中的某些性质等。图像增强还用于扩大图像中不同物体特征之间的差异,抑制不感兴趣的特征,以改善图像质量、丰富信息量、增强图像的解读和识别效果,从而满足特定分析的需求。在视频图像传输过程中,为避免因篡改而导致的图像质量下降,需要在关键帧提取后进行图像增强。本文采用的图像增强技术包括图像去噪、图像锐化和图像重构等。

1.3.1 图像去噪

图像去噪是消除图像噪声、提高图像质量的过程。噪声可能由传感器故障、环境干扰或图像传输错误等因素引起。常用的去噪算法包括均值滤波、中值滤波、维纳滤波和高斯低通滤波器等。均值滤波通过计算邻域像素的平均灰度值替代每个像素的灰度值,以达到去噪效果。中值滤波基于排序统计理论,是一种有效抑制噪声的非线性平滑滤波技术。维纳滤波通过最小化原始图像和恢复图像之间的均方误差实现去噪。高斯低通滤波器是一种线性平滑滤波器,适用于去除服从正态分布的噪声。

1.3.2 图像锐化

图像锐化技术用于增强图像的边缘和细节,主要通过提取和增强图像的高频分量实现。常见的锐化算法包括Laplac算子、Gauss-Laplac算子(LoG)和Sobel算子。Laplac算子为图像的二阶导数,但可能增加噪声和丢失边缘方向信息。LoG通过Gauss滤波器平滑图像,再用Laplac算子提取边缘,能够减少噪声同时保留更多边缘信息。Sobel算子提取图像的梯度信息,用于边缘检测和方向判别。

1.3.3 图像重构

图像重构通过利用图像中已有的像素信息,对图像中缺少的像素进行填补,以达到重建图像的目的。这种技术可以用于图像修复、超分辨率重建等领域。基于单张图像的重构方法可以分为

邻近插值、双线性插值、双立方插值等。这些算法复杂度低,易于实现,但是恢复效果有限,难以恢复高频的图像细节。基于深度学习的图像重构算法包括变分自编码器(Variational autoencoder, VAE)和生成对抗网络(Generative adversarial network, GAN)^[16]。VAE是一种基于概率模型的深度学习算法,它通过编码和解码的过程,将图像转换为低维的潜在空间表示,并从中生成新的图像。GAN则是一种基于竞争性训练的深度学习算法,它通过训练两个神经网络,一个生成器网络和一个判别器网络,生成新的图像。

1.4 形态学变换

形态学变换是图像处理中的一种重要技术,它基于图像形状进行操作,通常在二值化图像上执行^[12,17]。形态学变换包括腐蚀、膨胀、梯度等,具体效果如图2所示。腐蚀作用类似于土壤侵蚀,削减前景物体的边界,帮助去除小的白噪声或分离连接的对象。膨胀与腐蚀相反,扩展前景物体的边界,常用于连接对象的破碎部分。形态学梯度通过计算膨胀与腐蚀的差值,突出图像边缘。



图2 形态学变换示意图

Fig. 2 Schematic diagram of morphological transformation

1.5 DBSCAN

密度聚类算法(Density-based spatial clustering of applications with noise, DBSCAN)^[18]通过设置搜索半径(Eps)和最小邻域点数(MinPts)进行聚类,适用于不同形状的数据集,特别是流型数据。在数据分布较为稀疏或密度差距较大时,容易出现误判的情况。同时为了确保聚类精度,DBSCAN算法还需要提前分析数据集的密度设置这两个参数。DBSCAN算法流程中涉及3个重要概念。①核心点:如果任意样本点 p_i 的Eps内含有的数据点数量大于MinPts,则判断 p_i 为核心点。②非核心点:如果任意样本点 p_i 的搜索半径内的数据点少于最小邻域点数,但是 p_i 位于核

心点的邻域范围内,则判断 p_i 为非核心点。③异常点:除去核心点和非核心点的剩余样本点,被标记为异常点。

其算法具体执行步骤为:

(1)设定Eps和MinPts参数。

(2)对于每个样本点,若其Eps邻域内至少包含MinPts个其他样本点,则标记该为核心点。若样本点不是核心点但位于某个核心点的邻域内,则标记为边界点。既不是核心点也不是边界点的样本点被标记为噪声点(异常点)。

(3)利用密度可达性进行聚类。

图3所示为DBSCAN算法划分结果。其中,黑色圆心表示样本点,虚线表示该样本点的搜索半径Eps。最小搜索半径MinPts=2时,样本点A及其他搜索半径内有2个以上其他样本点的样本点,则被判别为核心点。样本点B、C被判别为边界点。样本点N被标记为噪声点。

DBSCAN算法的优点在于其能够自动确定类别数量,并能够自动检测到被篡改的位置。此外,其能够发现任意形状的簇,具备较高的计算效率和鲁棒性。

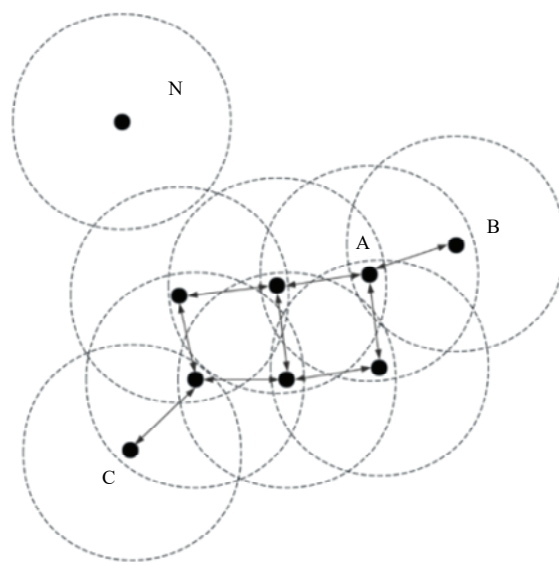


图3 DBSCAN 算法示意图

Fig. 3 DBSCAN algorithm diagram

1.6 OCR技术

OCR技术^[19]能够将文档中的手写或印刷文字转换为计算机可编辑的格式,从而显著提升信息处理和数据挖掘的效率。OCR技术通过先进的图像处理算法,精确分离图像中的文本区域,随后利用模式识别算法将这些文本转换为计算机可

编辑格式。OCR在文档数字化、车牌识别、银行支票处理等领域有广泛应用。然而,其准确性和可靠性受到诸多因素的影响,如文本的字体、大小、旋转角度以及图像质量等。

常见的OCR技术包括以下几种方法:

(1)基于规则的方法:通过编写规则识别字符,优点是速度快,但对于复杂字符的识别率较低,并且需要大量人工干预。

(2)基于统计的方法:采用机器学习算法,如支持向量机(Support Vector machine, SVM)和朴素贝叶斯等,进行模型训练。此方法对复杂字符有较好的识别率,并能自动学习规则,但需要大量

训练数据。

(3)基于深度学习的方法:利用深度神经网络(如CNN和RNN)进行字符识别。该方法对复杂字符具有较高的识别率,并能自动学习特征,但需要较多的计算资源和训练时间。

图4展示了基于CNN的训练方法。该方法首先通过多次应用卷积层和池化层提取输入图像的高维特征,随后利用全连接层训练这些特征与输出之间的权重。全连接层是建立完整识别模型的重要组成部分。在模型验证过程中,输入待检测的图像会再次经历特征提取,并根据全连接层的权重得出最终识别结果。

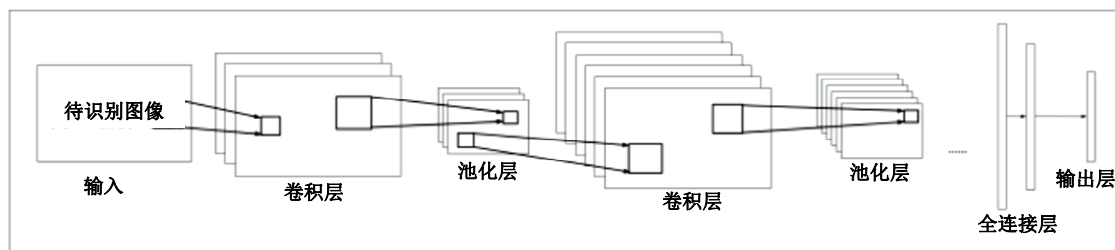


图4 基于CNN的神经网络

Fig. 4 Neural network based on CNN

2 技术路线

本文提出了一种基于关键帧提取与DBSCAN算法的视频内容篡改识别技术,旨在检测恶意篡改视频中的篡改内容。该技术的目标是通过时空定位与内容识别,实现对视频篡改内容的精确检测,从而为公共安全、传媒、商务等

领域提供可靠的视频内容检测技术支持。具体解决方案的流程如图5所示,具体步骤总结如下:

(1)对原始视频和篡改视频进行解码,以加载视频信息。

(2)对篡改视频进行特征提取和分析,获取能够反映视频主要内容的关键帧。

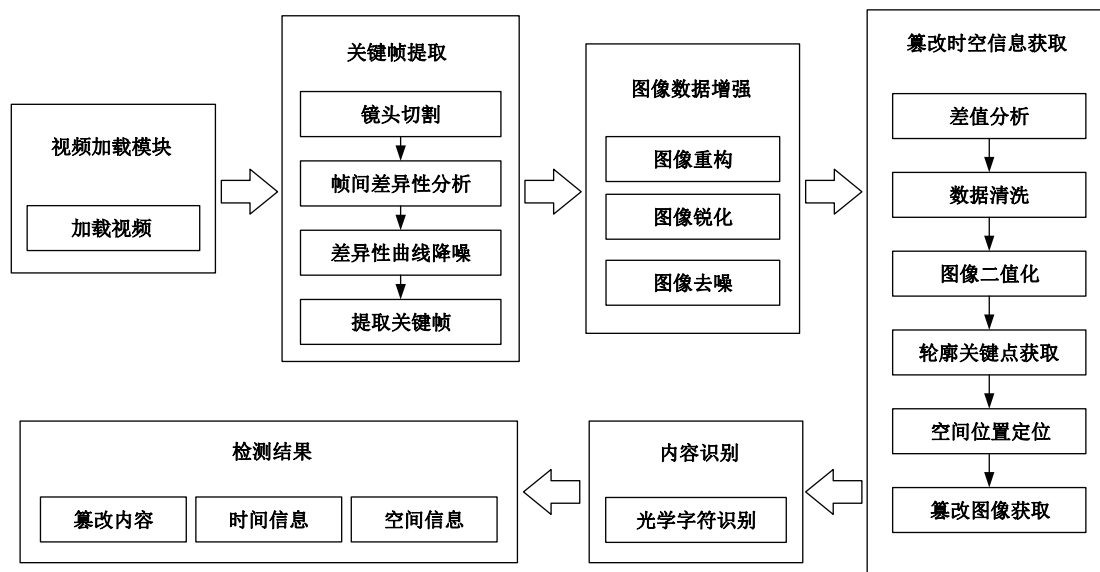


图5 解决方案流程图

Fig. 5 Solution flow diagramm

(3)对提取的关键帧进行图像增强处理,以提高图像质量。

(4)提取原始视频中相同位置的帧,并通过差值分析等算法截取视频篡改区段。

(5)利用基于深度学习的内容识别算法识别视频中的篡改内容。

(6)输出篡改内容信息及其时空信息。

篡改内容的时空信息包括时间信息和空间信息,其中时间信息指示篡改内容在视频中的时间位置,空间信息则为篡改内容在某一帧中的坐标位置。

在解决方案中,步骤(2)用于获取篡改视频的关键帧,步骤(4)用于获取视频篡改内容的时空信息,步骤(5)则负责内容识别,这也是本文的主要创新点。最终,所获得的信息将被转化为规范化、易于用户阅读的文件并输出。下文将对分点进行论述。

2.1 基于场景检测的关键帧提取

在处理包含多个镜头的复杂视频内容时,推荐采用基于场景检测的关键帧提取方法。此方法主要通过分析帧图像的颜色、纹理、轮廓等视觉特征的变化进行镜头切割,并在每个镜头内提取关键帧。

关键帧用于描绘镜头中的重要画面,其内容应清晰明了,能够充分体现镜头的主题。在确定关键帧时,通常遵循保守原则,即宁可选择并非完全准确的图像,也不能遗漏任何关键信息。此外,若图像特征不够明确,通常会去除重复或冗余的帧图像。具体步骤如下:

(1)提取当前镜头的所有图像信息,并计算每张图像的特征。

(2)将视频的第一帧设为当前帧,下一帧为对比帧。计算两帧图像视觉特征的差异分数和 SSIM。

(3)如果差异分数小于设定阈值,则继续比较下一帧,返回步骤(2)。若差异分数大于或等于阈值,或当前对比帧为视频的尾帧,则进入

步骤(4)。

(4)在当前帧位置进行镜头切割。将切割位置之前的所有 SSIM 传递至步骤(5)。切割位置之后返回步骤(2)。

(5)以当前镜头的帧序列为横坐标,SSIM 为纵坐标,绘制帧间差异性曲线。

(6)对帧间差异性变化曲线进行平滑和去噪处理,识别并提取关键帧。

图像视觉特征的差异分数和帧间差异性使用的算法不同。帧间差异性则使用 SSIM 算法,以衡量前后帧之间是否存在巨大差异,能否作为关键帧。而差异分数主要目的是进行镜头切割,关注的是图像中的颜色、轮廓等。具体流程如下:

(1)获取当前帧的 HSL 色域图像信息和轮廓图像信息。

(2)若上一帧的图像信息为空,则将当前帧的图像信息作为上一帧图像信息,返回差异分数为 0;否则,进入步骤(3)。

(3)计算上一帧与当前帧图像信息之间的平均像素距离,作为当前帧的图像视觉特征差异分数。

(4)将当前帧的图像信息覆盖上一帧图像信息,并返回差异分数。

2.2 视频篡改时空信息获取

在获取关键帧在视频中的相对位置后,能够回溯到原始视频并提取出相同位置的原始帧。通过对比原始帧和篡改帧,可以计算出篡改内容的时空信息。该过程具体步骤如图 6 所示。具体步骤如下:

(1)差值分析:计算原始帧与篡改帧的差值,识别二者的差异部分。

(2)数据清洗:去除数据传输过程中产生的噪声,以避免对后续计算的影响。

(3)图像二值化:将图像像素的灰度值设置为 0 或 255。

(4)轮廓关键点获取:识别二值图像中篡改内容的轮廓点位置。

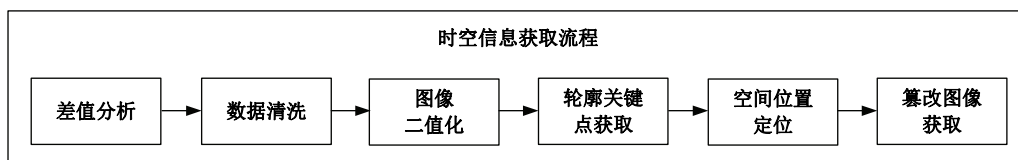


图 6 视频篡改时空信息获取流程图

Fig. 6 Flow chart of obtaining information of video tampering space-time

(5)空间位置定位:利用DBSCAN算法对轮廓关键点进行聚类,以获取篡改内容的空间信息。

(6)篡改图像获取:根据篡改内容的位置截取图像,用于后续内容识别。

在视频篡改时空信息获取的过程中,步骤(1)~(3)对篡改图像进行数据预处理,轮廓关键点获取和空间位置定位为主要算法,旨在获取篡改内容的空间信息。

轮廓关键点是指二值化后文字轮廓构成的散点。通过这些散点可以计算出篡改信息的详细坐标,从而截取更为精准的篡改内容图像。获取图像轮廓的方法多种多样,包括基于边缘检测和形态学操作的提取方法。由于任务目标是获取篡改内容的时空位置,因此,轮廓信息最好由散点构成。本文在二值化图像的基础上进行形态学操作以提取图像的轮廓散点。获取所有文字轮廓散点后,根据散点位置确定篡改信息的具体空间位置。在同一画面中,不同文字信息存在被篡改的可能,这些篡改位置可能分布在不同区域。直接根据所有轮廓核心点的位置获取篡改内容的空间位置,

可能会产生冗余信息。因此,本文提出了一种基于DBSCAN算法的关键点位置聚类方法。

DBSCAN算法是一种基于密度的聚类算法,它的优点是可以发现任意形状的聚类,并且可以避免噪声数据对聚类结果的影响。本文将轮廓关键点位置作为输入数据,利用DBSCAN算法进行聚类。得到轮廓关键点的聚类结果后,将每个簇视为一个整体,并根据簇的边缘位置找到篡改内容的空间信息。与传统的基于像素的方法相比,基于DBSCAN算法的空间位置定位能够更准确地定位篡改内容。同时,该方法可以有效处理多个篡改区域的情况,并避免冗余信息的产生。

2.3 基于OCR的篡改内容识别

经过视频篡改时空信息获取后,输出结果为被篡改位置的文字截图。接下来,将待识别图像输入基于OCR的篡改内容识别模块。如图7所示,本文中的OCR模块的主要流程如下。与传统OCR算法不同的是,本文中输入的图像已完成预处理。

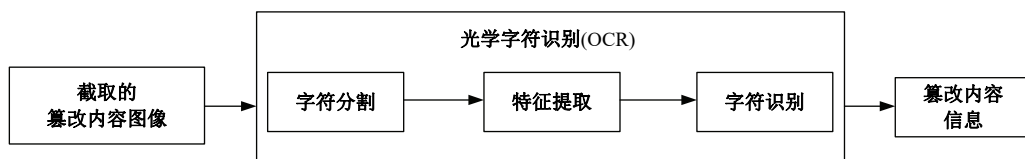


图7 基于OCR算法的视频篡改内容识别流程图

Fig. 7 Flow chart of video tampering content recognition based on OCR algorithm

(1)字符分割:将图像中的每个字符分割出来,以便独立识别。

(2)特征提取:对文本图像进行图像处理和特征提取。

(3)字符识别:利用基于卷积残差神经网络CRNN的光学字符识别技术,识别每个字符。

通过以上步骤,可以获取完整的篡改内容信息,并最终输出结果。

3 实验论证

为了验证本文论述的视频篡改内容识别技术的可行性和有效性,本节将展示本文方案的实验效果,并与基于光流一致性的视频帧间篡改检测(下文称“光流法”)、基于结构相似度均值MS-SIM商的一致性检测(下文称“MSSIM商”)、基于聚类的篡改检测(下文称“聚类法”)3种算法进行对比。

3.1 效果展示

本节展示本文提出的篡改检测方法的检测效果。图8所示为本文方法对篡改视频进行检测后的效果。其中左上图像是在篡改视频中提取到的关键帧,右上图像是在源视频中相同位置的帧,左



图8 识别效果展示图

Fig. 8 Recognition effect diagram

下图像是对原始帧与篡改帧的差值二值化后获得的差值关键点。最后,右下图像是当前篡改帧的识别结果显示,并用蓝色框将篡改位置标注出来。

对于检测到发生篡改的时间位置和空间位置,具体结果如表 1 所示。其中 snapshot 是篡改帧的文件名;time 是当前篡改帧在视频中的时间位置;text 是 OCR 检测识别到的篡改内容;location 是篡改内容在帧上的坐标点(如图 8 右下图像上蓝色框的四角坐标位置)。

表 1 检测结果对比表
Table 1 Table of test result

snapshot	time	tamper	
		text	location
12_95.png	12.095 s	点击链接查看详细视频	[90.0, 11.0]
			[390.0, 11.0]
			[90.0, 41.0]
			[390.0, 41.0]
		www.splj.com	[180.0, 26.0]
			[320.0, 69.0]
			[180.0, 26.0]
			[320.0, 69.0]

3.2 对比展示

本节展示与其他算法的对比结果。实验使用的硬件环境为: Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHZ, 16 GB 内存。实验数据方面, 本文从 Facebook AI Research 公开发布的 How-To100M 数据集中挑选了 100 个视频作为源视频。然后在视频的任意位置添加文字后作为篡改视频。

与其他 3 种算法对比结果如表 2 所示。如果检测到的篡改位置是在视频中添加文字的位置, 则记为正确检测, 反之记为误检测。检测时间为所有视频检测完毕的总耗时。

检测结果对比表证明本方法可以在保证检测正确率的情况下, 大幅缩短检测时间, 提高检测效率。

表 2 检测结果对比表
Table 2 Table of test result

方法	正确检测	误检测	准确率/%	召回率/%	检测时间/min
本文	1 630	165	96.55	90.21	56.84
光流法	2 419	1 701	65.12	54.20	171.67
MSSIM 商	3 103	869	91.87	74.25	185.36
聚类法	2 521	226	93.33	91.60	68.68

4 结束语

本文提出了一种基于关键帧提取与 DB-SCAN 的视频内容篡改识别技术。旨在解决在视频数据传输的过程中可能发生数据泄露, 导致视频被恶意篡改后广泛传播, 对社会公共安全造成重大隐患。本文算法采用关键帧检测替代逐帧检测, 从而降低算法的计算消耗。

基于篡改位置关联和语义聚合现象, 利用 DBSCAN 聚类算法提高单区域篡改内容的完整性和多区域篡改内容之间的独立性, 从而提高篡改检测的准确率。综上所述, 本文提出的方法实现了对视频篡改内容的时空定位与内容识别, 有利于针对安全隐患的快速实现和响应。

参考文献:

- [1] 崔雪冰, 冯巧娟, 崔平非. 基于内容特征的 MPEG 视频认证方案[J]. 计算机应用, 2010, 30(1): 214-216.
Cui Xue-bing, Feng Qiao-juan, Cui Ping-fei. MPEG video authentication scheme based on content feature[J]. Journal of Computer Applications, 2010, 30(1): 214-216.
- [2] Sun T F, Jiang X H, Chao J. A novel video inter-frame forgery model detection scheme based on optical flow consistency[C]//International Workshop on Digital Watermarking, Berlin, Germany, 2012: 261-281.
- [3] 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 Zhen-zhen, Hou Jian-jun, Li Zhao-hong, et al. Video-frame insertion and deletion detection based on consistency of quotients of MSSIM[J]. Journal of Beijing University of Posts and Telecommunications, 2015, 38(4): 84-88.
- [4] Lv C H, Huang Y. Effective keyframe extraction from personal video by using nearest neighbor clustering[C]//11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Beijing, China, 2018: 1-4.
- [5] Valognes J, Amer M A, Dastjerdi N S. Effective keyframe extraction from RGB and RGB-D video sequences[C]//Seventh International Conference on Image Processing Theory, Tools and Applications (IP-

- TA), Montreal, Canada, 2017: 1-5.
- [6] Duan L, Xiong D, Lee J, et al. A local density based spatial clustering algorithm with noise[C]//IEEE International Conference on Systems, Man and Cybernetics, Taipei, China, 2006: 4061-4066.
- [7] 赵黎, 祁卫, 李子青, 等. 基于关键帧提取的最近特征线(NFL)聚类算法的镜头检索方法[J]. 计算机学报, 2000, 23(12): 1292-1296.
- Zhao Li, Qi Wei, Li Zi-qing, et al. Key frame extraction based improved nearest feature line(NFL) classification algorithm[J]. Chinese Journal Computers, 2000, 23(12): 1292-1296.
- [8] 秦绪佳, 王慧玲, 杜轶诚, 等. HSV 色彩空间的 Retinex 结构光图像增强算法[J]. 计算机辅助设计与图形学学报, 2013, 25(4): 488-493.
- Qin Xu-jia, Wang Hui-ling, Du Yi-cheng, et al. Structured light image enhancement algorithm based on retinex in hsv color space[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(4): 488-493.
- [9] Sabu A M, Das A S. A survey on various optical character recognition techniques[C]//Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, India, 2018: 152-155.
- [10] Sarika N, Sirisala N, Velpuru M S. CNN based optical character recognition and applications[C]//6th International Conference on Inventive Computation Technologies, Coimbatore, India, 2021: 666-672.
- [11] Jun L. An improved DBSCAN clustering algorithm[J]. Computer and Communications, 2008, 8: 47468-47476.
- [12] 顾益军, 解易, 夏天. 基于内容代表性评价的关键帧抽取[J]. 计算机科学, 2014, 41(8): 286-288.
- Gu Yi-jun, Xie Yi, Xia Tian. Keyframe extraction based on representative evaluation of contents[J]. Computer Science, 2014, 41(8): 286-288.
- [13] Decombas M, Dufaux F, Renan E, et al. A new object based quality metric based on sift and SSIM[C]//19th IEEE International Conference on Image Processing, Orlando, USA, 2012: 1493-1496.
- [14] Gupta P, Srivastava P, Bhardwaj S, et al. A modified PSNR metric based on HVS for quality assessment of color images[C]//International Conference on Communication and Industrial Application, Kolkata, India, 2011: 1-4.
- [15] Alain H, Ziou D. Image quality metrics: PSNR vs. SSIM[C]//20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010: 2366-2369.
- [16] 王浩, 张叶, 沈宏海, 等. 图像增强算法综述[J]. 中国光学, 2017, 10(4): 438-448.
- Wang Hao, Zhang Ye, Shen Hong-hai, et al. Review of image enhancement algorithms[J]. Chinese Optics, 2017, 10(4): 438-448.
- [17] 金利娜, 于炯, 杜旭升, 等. 基于生成对抗网络和变分自编码器的离群点检测算法[J]. 计算机应用研究, 2022, 39(3): 774-779.
- Jin Li-na, Yu Jiong, Du Xu-sheng, et al. Generative adversarial network and variational auto-encoder based outlier detection[J]. Application Research of Computers, 2022, 39(3): 774-779.
- [18] Evans A N. Morphological gradient operators for colour images[C]//International Conference on Image Processing, Singapore, 2004: 3089-3092.
- [19] Smiti A, Eloudi Z. Soft DBSCAN: improving DBSCAN clustering method using fuzzy set theory[C]//6th International Conference on Human System Interactions, Sopot, Poland, 2013: 380-385.