

融合空间约束和梯度结构信息的视频篡改检测算法

普茜^{1,2,3}, 黄添强^{1,2,3}, 翁彬^{1,2,3}, 肖辉^{1,2,3}, 黄维^{1,2,3}

(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. 数字福建大数据安全技术研究所, 福建 福州 350007)

摘要: 相邻帧间相似性原理的传统视频被动取证方法会对画面运动剧烈的视频发生大量误检测, 针对这个问题, 提出了一种融合空间约束和梯度结构信息的视频篡改检测方法。首先, 利用空间约束准则, 提取低运动区域和高纹理区域, 并将两个区域进行融合, 获取顽健的量化相关性丰富区域用于提取视频最优相似性特征; 然后, 改进原有特征的提取和描述方法, 运用符合人类视觉系统特性的梯度结构相似性 GSSIM 来计算空间约束相关性值, 最后, 利用切比雪夫不等式对篡改点进行定位。实验证明, 针对画面运动剧烈的视频, 所提算法误检率更低, 精确度更高。

关键词: 空间约束; 量化相关性丰富区域; 梯度结构相似性; 画面运动剧烈的视频

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2019052

Video tampering detection algorithm based on spatial constraint and gradient structure information

PU Han^{1,2,3}, HUANG Tianqiang^{1,2,3}, WENG Bin^{1,2,3}, XIAO Hui^{1,2,3}, HUANG Wei^{1,2,3}

1. Mathematics and Informatics, Fujian Normal University, Fuzhou 350007, China

2. Fujian Provincial Engineering Research Center of Big Data Analysis and Application, Fuzhou 350007, China

3. Digital Fujian Big Data Security Technology Institute, Fuzhou 350007, China

Abstract: The traditional video passive forensics method using only the principle of similarity between adjacent frames will cause a lot of false detection for the video with severe motion. Aiming at this problem, a video tamper detection method combining spatial constraints and gradient structure information was proposed. Firstly, the low motion region and the high texture region were extracted by using spatial constraint criteria. The two regions were

收稿日期: 2019-04-14; 修回日期: 2019-06-06

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实验室基金资助项目 (No.SX201803)

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

论文引用格式: 普茜, 黄添强, 翁彬, 等. 融合空间约束和梯度结构信息的视频篡改检测算法[J]. 网络与信息安全学报, 2019, 5(5): 64-79.

PU H, HUANG T Q, WENG B, et al. Video tampering detection algorithm based on spatial constraint and gradient structure information[J]. Chinese Journal of Network and Information Security, 2019, 5(5): 64-79.

merged to obtain the robust quantitative correlation rich regions for extracting video optimal similarity features. Then improving the extraction and description methods of the original features, and using the similarity of the gradient structure in accordance with the characteristics of the human visual system to calculate the spatial constraint correlation value. Finally, the tampering points were located by the Chebyshev inequality. Experiments show that the proposed algorithm has lower false detection rate and higher accuracy.

Key words: spatial constraints, the quantitative correlation rich regions, GSSIM (gradient structure similarity), videos with severe motion

1 引言

由于数字视频的篡改检测在司法取证、新闻媒体等方面有重要的意义,因此近年来国内外对该领域展开了大量研究^[1]。但现有的数字视频取证技术大多针对画面运动平缓的视频,对画面运动剧烈视频的取证会造成大量误检测^[2]。因此,针对画面运动剧烈视频的篡改取证是一个亟待解决的问题,具有重要的研究意义。

目前,国内外的一些研究团队关于视频帧删除篡改的研究已经取得一定进展。Zhao等^[3]首先通过比较视频相邻帧之间的HSV颜色直方图的相似性进行帧间篡改的粗检测,然后结合SURF特征提取和FLANN进行细检测,进一步确认篡改点,但由于HSV颜色直方图的基本原理是利用不同色彩在视频帧中所占的比例,并没有考虑不同色彩在视频帧中的空间位置信息,故该特征不能很好地代表帧图像信息,所以该方法的稳健性有待提高;Sowmya等^[4]通过时空三元特征关系(STTFR, spatiotemporal triad feature relationship)技术对任何给定视频生成一个128 bit的信息数字,作为该视频唯一的指纹,利用视频时域上的篡改会破坏这种指纹的原理进行篡改取证,但该方法属于主动取证,需要预先获得原始视频,所以其实用性受到一定限制。Wang等^[5]提出一种基于非负张量分解的视频篡改检测方法,但该方法对细微的帧删除篡改和同源帧插入篡改会造成漏检测,具有一定的局限性。Lin等^[6]提出一种融合音频的多通道视频帧间篡改检测方法,通过融

合音频通道和视频帧序列通道的检测结果进行检测,然后结合QDCT特征进行细检测定位,但对于无声视频和画面运动剧烈的视频,该方法不能体现出它的优越性。Wang等^[5]根据篡改视频重压缩保存后会导致离散余弦变换系数分布直方图出现周期性伪影,以及运动估计误差会显现周期性尖峰进行篡改取证,但该方法在检测含有噪声的视频时性能显著下降,而且无法检测到删除帧数是GOP整数倍时的篡改,同样不适用于使用恒定比特率编码模型的视频,所以该方法的适用局限性比较大。Liu等^[6]提出了专门为H.264编码视频设计的帧删除检测方案,证明了在视频帧删除篡改的情况下P帧序列平均残差会在时域上表现出周期性,但该方法只适用于特定编码的视频。

由于视频本身的时域冗余性,相邻帧之间存在很大的相关性,所以基于视频内容连续性的检测方法对于大部分遭受帧删除篡改的视频有良好的性能。传统的帧删除篡改检测方法的主要步骤是:首先提取视频帧特征,然后计算相邻帧特征的相关性,最后通过相关性值出现的异常点进行篡改点的定位。然而,由于画面运动比较剧烈的视频相邻帧之间的相关性波动比较大,传统的检测方法会对其造成大量误检测。因此,本文提出一种融合空间约束和梯度结构信息的视频篡改检测方法,在更加精准地检测到篡改位置的同时,大大降低了误检率。鉴于视频的运动剧烈区域是影响检测精确率的主要因素,故本文算法首先采用自适应阈值最佳划分法对待检测视频进行空间约束,提取出低运动和高纹理区域,然后把两者

进行融合, 获取最优量化相关性丰富区域, 并结合梯度结构相似性特征梯度结构相似性(GSSIM, gradient structure similarity)和 GSSIM 熵等方法, 提高了特征的顽健性表达和检测的准确率。

由于在视频的量化相关性丰富区域提取的特征变化相对一致, 所以该算法能够适用于不同画面运动剧烈程度的视频; 同时利用基于人类视觉特性的 GSSIM 特征的顽健性表达, 该方法不限制视频清晰程度以及视频类型, 能检测到仅仅几帧的细微篡改。使用空间约束和梯度结构信息的结合方法, 提高了检测方法的顽健性, 针对画面运动剧烈的视频, 与现有算法^[5-7]相比, 误检率更低, 精确率更高。

2 视频空间约束

现有算法大多基于视频相邻帧之间的极大相关性原理进行篡改检测。对于画面运动剧烈的视频, 鉴于相邻帧之间的相关性变化比较剧烈, 仅利用传统的检测算法会造成大量误检测, 因此可对视频进行空间约束, 获取有利于篡改检测的空间区域, 提高检测的精准率。视频空间约束是结合视频时空域上的相关信息对视频帧进行空域上的约束, 通过定义一定的约束条件, 过滤掉一部分影响篡改检测效果的区域, 保留有利于篡改检测的空间区域。

本文主要采用自适应阈值最佳区域划分法对视频进行空间约束, 主要流程如图 1 所示, 主要分为两部分的内容: ① 利用自适应低运动区域划分法获取视频的低运动区域; ② 首先通过自适应

高纹理区域划分法实现对视频高纹理区域的空间约束; 然后融合低运动和高纹理区域, 获取有利于篡改检测的量化相关性丰富区域。

2.1 自适应低运动区域划分法

2.1.1 背景模型的初始化

步骤 1 令 P_i 、 P_{i+1} 为视频第 i 帧和第 $i+1$ 帧的灰度像素值, 以相邻两帧灰度像素值的绝对差值来构造差图像, 然后通过阈值 Th_m 判断得到差图像对应的二进制掩码 DM_i 。其中, 阈值 Th_m 的自适应选取方法在下文详细介绍。

$$DM_i(x, y) = \begin{cases} 0, & |P_i(x, y) - P_{i+1}(x, y)| < Th_m \\ 255, & |P_i(x, y) - P_{i+1}(x, y)| > Th_m \end{cases} \quad (1)$$

步骤 2 构建平均背景模型。首先计算每帧背景区域的对应像素点灰度值的累加值 $Sum(x, y)$, 并且计算背景区域像素点的累加次数, 用 $Num(x, y)$ 表示。

$$Sum(x, y) = \begin{cases} Sum(x, y) + P_i(x, y), & DM_i(x, y) = 0 \\ 0, & DM_i(x, y) = 255 \end{cases} \quad (2)$$

$$Num(x, y) = \begin{cases} Num(x, y) + 1, & DM_i(x, y) = 0 \\ 0, & DM_i(x, y) = 255 \end{cases} \quad (3)$$

步骤 3 令 $i = i + 1$, 返回步骤 1 迭代, 当迭代到视频的最后一帧时, 停止迭代。求得平均背景模型为

$$B(x, y) = \frac{Sum(x, y)}{Num(x, y)} \quad (4)$$

2.1.2 低运动区域的检测

首先通过当前帧与背景模型绝对差值的计算

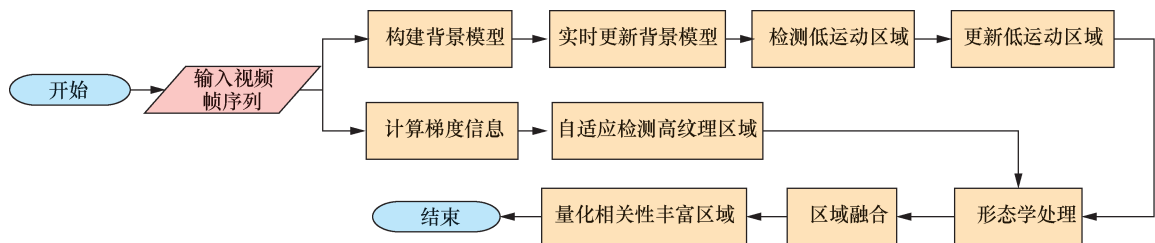


图 1 视频空间约束的流程

得到差图像，然后利用自适应阈值 Th_m 判断得到差图像对应的二维掩码，以此确定低运动区域。

$$LM_i = \begin{cases} 0, |P_i - B_i| < Th_m \\ 255, |P_i - B_i| > Th_m \end{cases} \quad (5)$$

对得到的二进制掩码 LM_i 进行形态学操作来填补空隙和移除含有噪声的小面积区域。

$$LM_i = (LM_i \cdot SE) \circ SE \quad (6)$$

其中， $\cdot$ 是形态学操作的闭操作， $\circ$ 是形态学操作的开操作，SE 是开运算与闭运算的结构化元素。

2.1.3 低运动区域的更新

由于视频的画面一直在变化，低运动区域也在实时变化，所以需要定期对低运动区域进行更新，主要包含两种更新方法：① 当视频的场景微小变化时，进行像素级更新；② 当视频的场景变化大时，进行帧级更新。具体更新的方法如下。

方法1 像素级更新

当 $P(x, y)$ 连续 k 次没有发生变化时，则把该像素点更新为低运动区域；同时考虑到视频光线环境变化的影响，低运动区域并不只是静止区域，所以对超过 k 次没有发生变化时的像素点采用对低运动区域和整个帧图像加权的方法来更新低运动区域的像素点。

$$LM_i(x, y) = \begin{cases} P_i(x, y), P_i(x, y) \text{ 连续 } k \text{ 次没有发生变化} \\ w1 * P_i(x, y) + w2 * LM_i(x, y), P_i(x, y) \text{ 超过 } k \text{ 次没有发生变化} \end{cases} \quad (7)$$

当低运动区域发生较大范围变化时，仅使用像素级更新，较难达到很好的效果，需要利用接下来介绍的帧级更新方法。

方法2 帧级更新

当低运动区域的面积小于整个帧图像面积超过一定百分比时，说明低运动区域发生了较大范围的变化。当连续多帧图像都出现这种情况时，需要利用上文低运动区域的检测方法对该区域进行重新获取。

2.1.4 空间约束阈值 Th_m 的自适应选取

若 Th_m 的取值太大，会引入大量噪声；若 Th_m 的取值太小，会造成低运动区域的碎片化，所以 Th_m 的取值对视频低运动区域的空间约束至关重要。通过观察 2.1.2 节描述的差图像的直方图可得，直方图大体上呈递减趋势。为了使直方图呈现完全递减趋势，将该直方图按照纵坐标的频数值进行排序，得到完全递减的直方图^[7]，构建差图像直方图的简化模型如图 2 所示，通过下面的步骤进行自适应最佳阈值 Th_m 的选取。

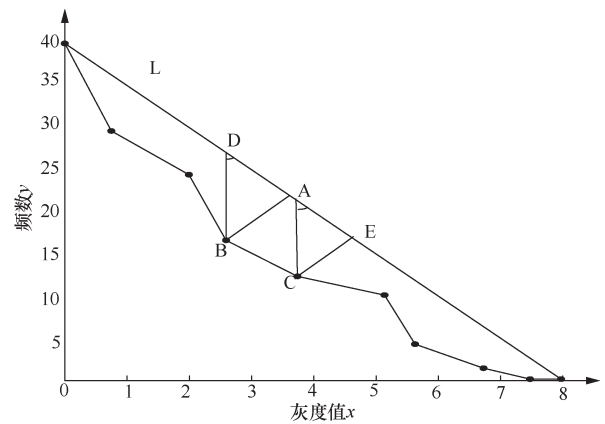


图2 差图像的直方图的简化模型

步骤1 把直方图的频数最大值点和频数最小值点进行连线，得到直线 L。

步骤2 计算直方图的每个点到该直线 L 的距离。如图 2 所示，点 B 到直线 L 的距离为 BA，点 C 到直线 L 的距离为 CE，即 $BA \perp L$ ， $CE \perp L$ ，其中 BD 和 CA 均垂直于 x 轴。

步骤3 由于需要计算直方图上的每个点到直线 L 的距离，为了节省时间和减少计算量，可以考虑快速算法。快速算法的过程是：由于 $\triangle BAD$ 和 $\triangle CEA$ 均为直角三角形，所以 $AB = BD \cdot \sin \angle BDA$ ， $CE = CA \cdot \sin \angle CAE$ 。又因为 $BD \parallel CA$ ， $\sin \angle BDA = \sin \angle CAE$ ，所以比较 AB 和 CE 的大小转换为求解 BD 和 CA 的大小，而后的计算量相对较小，如 $CA = y_A - y_C$ 。

步骤 4 选取最大距离的点对应的横坐标 (灰度值) 作为二值掩码式(5)的阈值。

2.2 量化相关性丰富区域

2.2.1 自适应高纹理区域划分法

计算视频帧的梯度结构信息, 对梯度图像通过自适应阈值 Th_i 的判断筛选得到对应的二值掩码, 以此确定视频的高纹理区域, 分别如式(8)和式(9)所示。

$$\|\Delta P_i\| = \sqrt{P_{ix}^2 + P_{iy}^2} \quad (8)$$

$$TM_i = \begin{cases} 255, \|\Delta P_i\| > Th_i \\ 0, \|\Delta P_i\| \leq Th_i \end{cases} \quad (9)$$

其中, P_{ix} 是 P_i 帧在 x 方向的偏导数, P_{iy} 是 P_i 帧在 y 方向的偏导数, $\|\Delta P_i\|$ 是 P_i 帧的梯度结构信息。

对得到的二进制掩码 TM_i 进行形态学操作来填补空隙和移除含有噪声的小面积区域。

$$TM_i = (TM_i \cdot SE) \circ SE \quad (10)$$

其中, $\cdot$ 是形态学操作的闭操作, $\circ$ 是形态学操作的开操作, SE 是开运算与闭运算的结构化元素。

2.2.2 阈值 Th_i 的自适应选取

视频帧是动态变化的, 其每一帧的纹理信息都会发生变化, 因此阈值 Th_i 的选取应根据视频的画面变化进行动态自适应选取, 阈值的自适应选取可以通过梯度图像的灰度值来实现。

步骤 1 首先计算当前帧灰度图像对应的梯度图像的最小像素值和最大像素值, 然后计算二者的平均值作为阈值的初始值 T 。

步骤 2 通过初始阈值 T 把当前帧划分为高纹理区域和低纹理区域, 然后分别计算两区域梯度图像的平均灰度值 A_1 和 A_2 , 及两区域梯度图像的灰度分别占整个帧梯度图像灰度的百分比 b_1 和 b_2 。

$$A_1 = \frac{\sum_{\|\Delta P(i,j)\| > T} \|\Delta P(i,j)\|}{\sum_{\|\Delta P(i,j)\| > T} N} \quad (11)$$

$$A_2 = \frac{\sum_{\|\Delta P(i,j)\| \leq T} \|\Delta P(i,j)\|}{\sum_{\|\Delta P(i,j)\| \leq T} N} \quad (12)$$

$$b_1 = \frac{\sum_{\|\Delta P(i,j)\| > T} \|\Delta P(i,j)\|}{\sum_{\|\Delta P(i,j)\| > T} \|\Delta P(i,j)\|} \quad (13)$$

$$b_2 = \frac{\sum_{\|\Delta P(i,j)\| \leq T} \|\Delta P(i,j)\|}{\sum_{\|\Delta P(i,j)\| \leq T} \|\Delta P(i,j)\|} \quad (14)$$

步骤 3 高纹理区域检测的阈值 Th_i 的选取为

$$Th_i = b_1 b_2 (A_1 + A_2) \quad (15)$$

2.2.3 区域融合

对空间约束得到的低运动和高纹理区域进行融合, 取两个区域的交集, 得到视频的量化相关性丰富区域 MR_i 。

$$MR_i = \bigcap_{j=1}^n LM_j \cap TM_j \quad (16)$$

对得到的量化相关性丰富区域 MR_i 进行形态学操作来填补空隙和移除含有噪声的小面积区域。

$$MR_i = (MR_i \cdot SE) \circ SE \quad (17)$$

其中, $\cdot$ 是形态学操作的闭操作, $\circ$ 是形态学操作的开操作, SE 是开运算与闭运算的结构化元素。

3 篡改检测算法

本文算法融合空间约束和梯度结构信息对给定视频进行篡改检测。首先, 通过自适应低运动区域划分法获取视频的低运动区域, 并结合自适应高纹理区域划分法获取视频的高纹理区域; 然后, 把两者进行融合, 获取有利于基于相邻帧之间极大相关性原理检测方法的量化相关性丰富区域; 其次, 基于人类视觉特性改进原有的特征提取和描述方法, 在此区域上提取梯度结构相似性特征 $GSSIM$ 得到空间约束相关性值, 并利用 $GSSIM$ 商对其进行后处理; 最后, 利用切比雪夫不等式对篡改点进行定位。算法的具体流程如图 3 所示。

3.1 视频空间约束

考虑到视频本身的特性，视频的空间约束过程主要分为3部分：① 首先对视频进行预处理，然后对视频的低运动区域进行空间约束；② 视频高纹理区域的空间约束；③ 视频量化相关性丰富区域的空间约束。

步骤1 对视频低运动区域的空间约束。

为了减少计算复杂度，把待检测视频转换成帧序列图像，并对得到的帧序列图像进行灰度化的预处理。

$$P = \{P_1, P_2, \dots, P_i \dots P_n\} \quad (18)$$

画面变化剧烈的视频，由于原始视频的高运动区域使相邻帧之间相关性值比较小，使视频帧间篡改点的相邻两帧之间相关性值也会变小^[8]，因此为了减少误检率，可以定义一定的约束条件，

过滤掉视频帧的高运动区域，获取视频的低运动区域。利用本文2.1节提出的自适应低运动区域划分法，对视频低运动区域进行空间约束的主要步骤如图4所示。首先，根据视频场景信息构建视频的背景模型；其次，当前帧和背景模型作差得到对应的差图像；然后，通过自适应阈值判断是否为低运动区域^[9]，并对低运动区域进行形态学操作来填补空隙和移除含有噪声的小面积区域；最后，根据视频场景的变化剧烈程度进行低运动区域的更新。

步骤2 对视频高纹理区域的空间约束。

由于视频帧的高纹理区域包含帧图像大量的梯度信息和边缘信息，所以对视频高纹理区域进行空间约束可以确保特征提取的有效性和精准性^[10]。利用2.2节提出的自适应高纹理区域

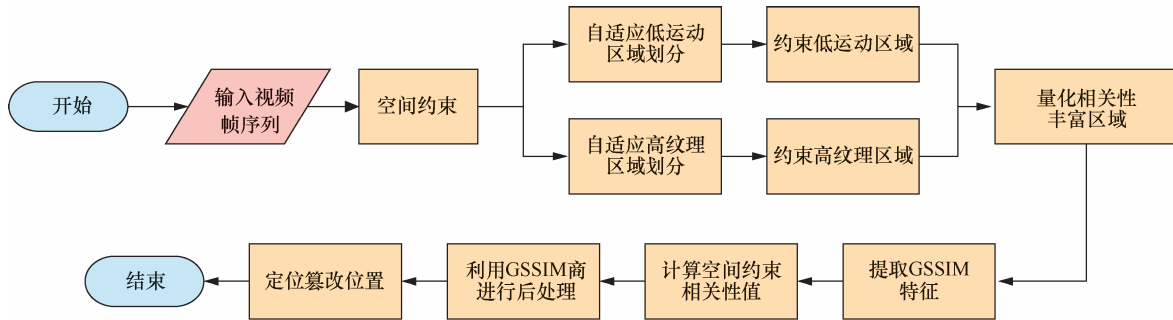


图3 视频篡改检测流程

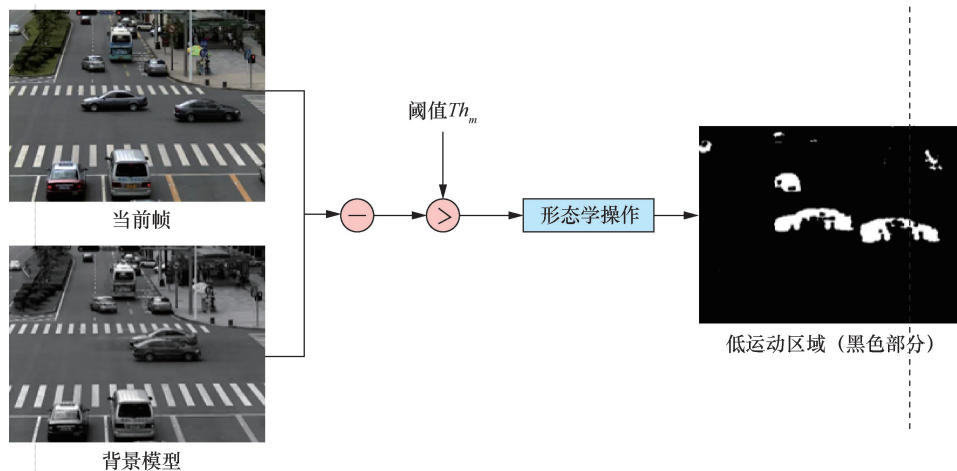


图4 视频低运动区域的空间约束

划分法,对视频高纹理区域进行空间约束的主要步骤如图 5 所示。首先,计算视频帧的梯度信息;然后,通过自适应阈值的判断得到视频的高纹理区域;最后,通过形态学操作来填补空隙和移除含有噪声的小面积区域。

步骤 3 对视频量化相关性丰富区域的空间约束。

为了避免造成误检测,可以获取视频的量化相关性丰富区域^[11]。根据 2.2 节的内容,对视频的量化相关性丰富区域进行空间约束的主要步骤如图 6 所示,首先,对低运动和高纹理区域进行融合,取两个区域的交集;然后,通过形态学操作来填补空隙和移除含有噪声的小面积区域,获取视频的最优量化相关性丰富区域,实现对视频的空间约束。

3.2 特征提取

由于人类视觉系统具有对图像的边缘梯度信息比较敏感的特性,而梯度信息本身可以很好地反映出图像中微小的细节反差和纹理特征变化,因此可以将梯度作为重要的结构信息^[12]。基于这一特性,改进原有结构相似性(SSIM, structural similarity)特征的提取和描述方法,在量化相关性丰富区域上采用梯度结构相似性(GSSIM),度量相邻帧的相似性。梯度结构相似性主要把从原始图像和梯度图像提取到的局部信息融合^[13],从而提升传统 SSIM 在模糊图像上的性能,其中,梯度图像获取的具体步骤是:首先把原始帧图像进行 Sobel 边缘检测,然后进一步对其进行逐像素动态范围压缩,凸显图像的重要显著特征^[14]。

GSSIM 的定义如下

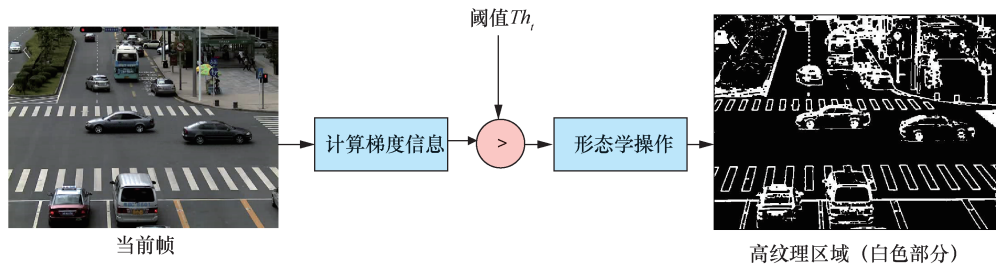


图 5 视频高纹理区域的空间约束

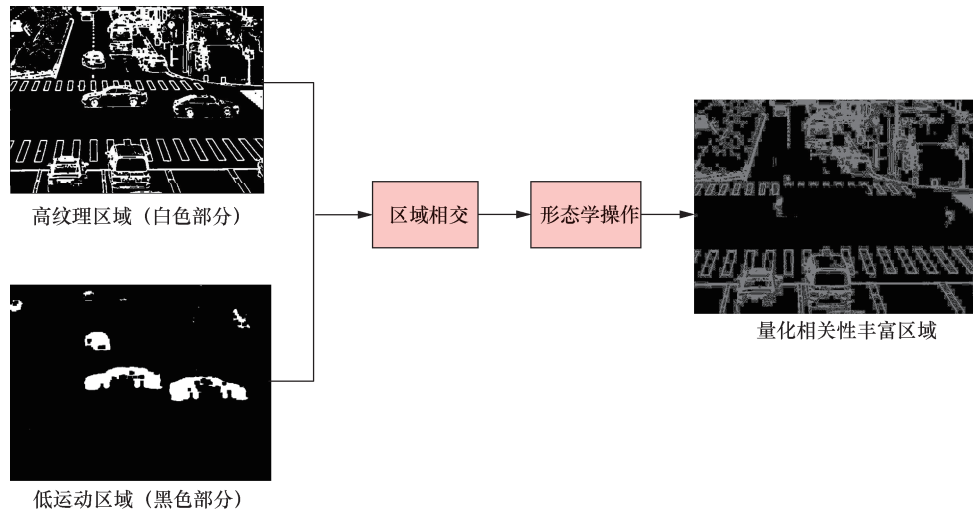


图 6 视频量化相关性丰富区域的空间约束

$$GSSIM(x, y) = [l(x, y)]^\alpha [c(x'_{Drc}, y'_{Drc})]^\beta [s(x'_{Drc}, y'_{Drc})]^\gamma \quad (19)$$

其中, $l(x, y)$ 是图像 X 和 Y 的对应图像块 x 和 y 上的亮度比较函数, X' 和 Y' 是图像 X 和 Y 利用 Sobel 算子计算得到的图像, 并通过式(20)、式(21)对 X' 和 Y' 进行动态范围压缩得到梯度图像 X'_{Drc} 和 Y'_{Drc} 。 $c(x'_{Drc}, y'_{Drc})$ 和 $s(x'_{Drc}, y'_{Drc})$ 分别是 X'_{Drc} 和 Y'_{Drc} 的对应图像块 x'_{Drc} 和 y'_{Drc} 上计算得到的对比度比较函数和结构比较函数。亮度比较函数 $l(x, y)$ 、对比度比较函数 $c(x'_{Drc}, y'_{Drc})$ 和结构比较函数 $s(x'_{Drc}, y'_{Drc})$ 分别定义为

$$X'_{Drc} = \log \log(X' + 1) \quad (20)$$

$$Y'_{Drc} = \log \log(Y' + 1) \quad (21)$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (22)$$

$$c(x'_{Drc}, y'_{Drc}) = \frac{2\sigma_{x'_{Drc}}\sigma_{y'_{Drc}} + C_2}{\sigma_{x'_{Drc}}^2 + \sigma_{y'_{Drc}}^2 + C_2} \quad (23)$$

$$s(x'_{Drc}, y'_{Drc}) = \frac{\sigma_{x'_{Drc}y'_{Drc}} + C_3}{\sigma_{x'_{Drc}}\sigma_{y'_{Drc}} + C_3} \quad (24)$$

其中, $\sigma_{x'_{Drc}}\sigma_{y'_{Drc}}$ 表示 x'_{Drc} 和 y'_{Drc} 对应图像块的方差, $\sigma_{x'_{Drc}y'_{Drc}}$ 表示 x'_{Drc} 和 y'_{Drc} 的协方差, C_1 、 C_2 、 C_3 是为了避免分母为 0, 所以公式加上一个值很小的常数, 其中 $C_1 = C_2, C_3 = \frac{C_1}{2}$ 。由式(19)~式(24)可以推出式(25)。

$$GSSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x'_{Drc}}\sigma_{y'_{Drc}} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_{x'_{Drc}}^2 + \sigma_{y'_{Drc}}^2 + C_2)} \quad (25)$$

最后将图像块之间的梯度结构相似度取平均值 (MGSSIM, mean gradient structural similarity), MGSSIM 衡量的是两幅图像帧之间的相似度。两幅图像对应的所有图像块 GSSIM 的均值可以由式(26)得出。

$$MGSSIM(x, y) = \frac{1}{M} \sum_{i=1}^M GSSIM(x_i, y_i) \quad (26)$$

对于得到的长度为 n 的帧序列, 根据式(26)计算相邻帧图像的梯度结构相似性的均值 MGSSIM, 可以得到 $n-1$ 个前后相邻帧间的空间约束相关性值 r_i , 将产生一个相关系数序列 r 。

$$r = \{r_1, r_2, \dots, r_i, \dots, r_{n-1}\} \quad (27)$$

其中, r_i 代表了 i 帧和 $i+1$ 帧之间的 MGSSIM。

3.3 特征处理

根据人类视觉系统的视觉停留现象, 为了人类在视觉上能够感知连续的画面, 视频的帧率一般是 24~25 fp/s, 所以相邻帧之间的相关性非常大, 而经过人为帧删除恶意篡改之后, 篡改点的相邻两帧之间相关性变小。但不能简单地通过分析相关性值的大小来判断视频是否被篡改^[1], 因为对于一个内容非静止的待检测视频, 它的内容变化程度是未知的, 同一视频的帧间相关性值也会出现变化。为了进一步消除视频内容运动变化对检测结果造成的影响, 本文用 MGSSIM 商替换 MGSSIM, MGSSIM 商的定义如式(28)所示。

$$\Delta r_i = \begin{cases} \frac{r_i}{r_{i+1}}, & r_i \geq r_{i+1} \\ \frac{r_{i+1}}{r_i}, & r_i < r_{i+1} \end{cases} \quad (28)$$

其中, r_i 代表了 i 帧和 $i+1$ 帧之间的 MGSSIM。

用式(28)根据上一步得到的相关系数集合 r 计算 GSSIM 商, 来衡量相邻帧之间相关性的变化情况, 得到长度为 $n-2$ 的相关系数集合 Δr , 如式(29)所示。

$$\Delta r = \{\Delta r_1, \Delta r_2, \dots, \Delta r_i, \dots, \Delta r_{n-2}\} \quad (29)$$

3.4 篡改点的定位

本文使用切比雪夫不等式^[15]定位异常点。切比雪夫不等式衡量的是随机变量与均值之间的偏离程度, 随机变量越远离平均值, 概率越低。基于切比雪夫不等式的基本原理, 可以用其进行离群点检测。切比雪夫不等式证明, 对于任何一个

随机变量 Z ，它的任何一个取值与均值的距离都遵循一定的概率 P ， P 的取值只与它的方差有关^[5]。切比雪夫不等式定义如式(30)和式(31)所示。

$$P(|Z - \mu| \geq \varepsilon) < \frac{\sigma^2}{\varepsilon^2} \quad (30)$$

$$P(|Z - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2} \quad (31)$$

其中， ε 是 σ 的整数倍。由上式可知，对于随机变量 $Z = \{z_1, z_2, \dots, z_i, \dots, z_n\}$ ，其取值分布在 $(\mu - \varepsilon, \mu + \varepsilon)$ 的概率为 $1 - \frac{\sigma^2}{\varepsilon^2}$ 。

当 $\varepsilon = 3\sigma$ 时， $P(|Z - \mu| < 3\sigma) \geq \frac{8}{9} \approx 88.89\%$ 。

当 $\varepsilon = 4\sigma$ 时， $P(|Z - \mu| < 4\sigma) \geq \frac{15}{16} \approx 93.75\%$ 。

当 $\varepsilon = 5\sigma$ 时， $P(|Z - \mu| < 5\sigma) \geq \frac{24}{25} \approx 96\%$ 。

由上可知，可由切比雪夫不等式剔除异常点，只要随机变量 z_i 满足 $|z_i - \mu| > k\sigma$ ，则认为 z_i 为异常点。

根据概率论中的切比雪夫不等式进行异常点的检测，需要用到上一步得到的相关系数变化情况的集合 Δr 的均值 μ 和方差 σ^2 ，其定义分别如式(32)和式(33)所示。

$$\mu = \frac{\sum_{i=1}^{n-2} \Delta r_i}{n-2} \quad (32)$$

$$\sigma^2 = \frac{\sum_{i=1}^{n-2} (\Delta r_i - \mu)^2}{n-2} \quad (33)$$

以 $\varepsilon = k\sigma$ 为界限，当 Δr_i 与 Δr 的均值 μ 距离 $|\Delta r_i - \mu|$ 超过此界限时，即可把该离散变量定义为篡改点，本文通过大量实验发现 $k=5$ 时可以很好地检测到篡改点。

4 实验结果及分析

4.1 实验设置

为了评估算法的检测效果，本文在 3 个公开

数据集上进行了实验，分别是 SULFA 视频库 (surrey university library for forensic analysis surrey university library for forensic analysis)^[15]、CDNET 视频库 (a video database for testing change detection algorithms)^[16]，华南理工视频篡改检测数据库 VFDD (Video Forgery Detection Database Version1.0)^[17]，共约 200 个视频。使用 Adobe Premiere Pro CC 软件对所有待检测视频进行不同数量程度的帧删除篡改，分别是 5、10、20、25、50、75、100。本文的算法由 Python 和 Matlab 编程实现。以下为实验所使用的计算机配置：中央处理器 Intel(R) Xeon(R) CPU，16 GB 内存以及操作系统为 Microsoft Windows 10PC。

4.2 实验结果

为了验证本文算法的有效性，分别针对不同运动剧烈程度的视频进行不同程度的帧删除篡改，具体步骤是：对所有视频运行检测算法并分析实验结果，同时和现有方法的检测结果进行比较和量化分析。为了验证本算法的顽健性，分别从空间约束、梯度结构相似度 GSSIM 两个角度进行对比实验，同时和现有的较优算法进行实验效果的比较分析。实验的具体流程如下。① 验证空间约束对检测算法准确率的影响，具体步骤是：对于运动变化剧烈和平缓的两类视频，在同等条件下，分别对这两类视频进行空间约束和不进行空间约束进行篡改检测，对比实验效果。② 验证 GSSIM 特征的顽健性，具体步骤是：在同等条件下，对画面模糊和清晰的两类视频分别提取特征 SSIM 和 GSSIM 进行对比实验。

4.2.1 对视频进行空间约束的实验结果

选取视频 Camera Road 01.avi 进行空间约束实验结果的展示说明。首先，运用自适应的运动区域检测算法对低运动区域进行空间约束。实验结果如图 7 所示，选取第 8 帧图像构建的原始背景模型如图 7(a)所示，第 8 帧的背景模型如图 7(b)

所示,第 8 帧图像如图 7(c)所示,图 7(b)和图 7(c)相减得到第 8 帧的差图像如图 7(d)所示。对得到的差图像进行直方图统计如图 7(e)所示,利用 2.1 节低运动区域阈值的自适应选取算法进行阈值的选取,该帧图像的阈值选取为 20,进而得到第 8 帧的二值化掩码,如图 7(f)所示,其中黑色区域是检测到的第 8 帧的低运动区域。然后,利用 2.2 节的方法对视频的高纹理区域进行空间约束,阈值为 150,实验结果如图 8 所示,白色区域即为高纹理区域。最后,对视频帧的高纹理区域和低运动区域进行融合,取两区域的交集,得到第 8 帧的量化相关性丰富区域,如图 9 所示,实现对

视频的空间约束。

4.2.2 篡改检测结果展示

针对视频的帧删除篡改,可以通过观察检测

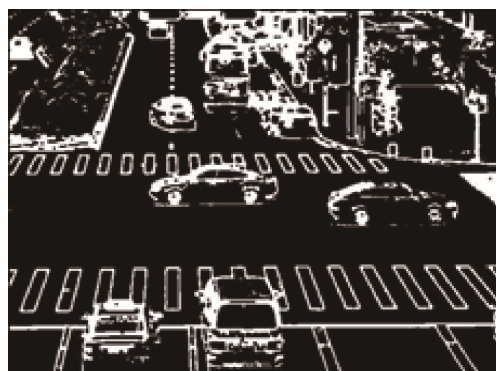


图 8 视频高纹理区域的空间约束



(a) 原始背景模型



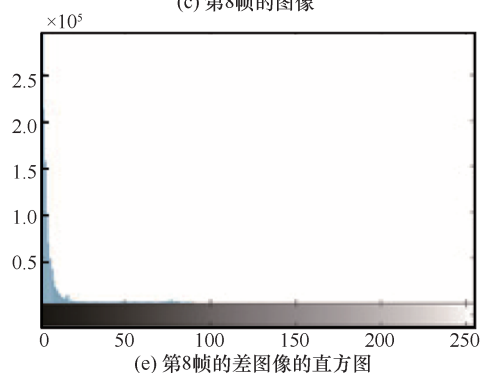
(b) 第8帧的背景模型



(c) 第8帧的图像



(d) 第8帧的差图像



(e) 第8帧的差图像的直方图



(f) 第8帧的低运动区域 (黑色部分)

图 7 视频低运动区域的空间约束

结果中被篡改的位置是否出现尖锐峰值, 及是否会定位到异常点来判断检测效果。根据 GSSIM 商公式的计算, 在一个篡改点的位置产生两个异常点, 相应地会导致一对相邻的峰值出现, 篡改点的位置用黑色圆圈标记。

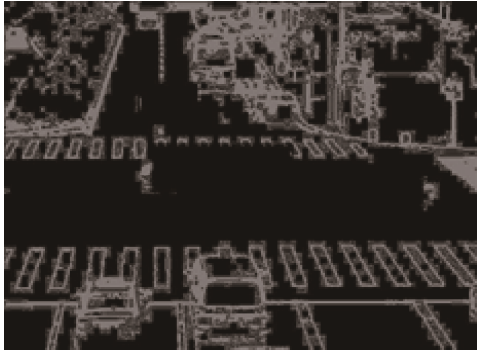


图9 视频量化相关性丰富区域的空间约束

(1) 验证本文算法对运动剧烈视频的有效性

为了验证本文方法的有效性, 针对不同运动剧烈程度的视频均进行了篡改检测, 检测结果如

图 10 所示。对于画面运动较为剧烈的视频, 选取的实验视频是 SULFA 视频库中 fuji_2800_road(1).avi, 该视频中记录的是在高速公路上行驶的汽车, 并对其在 95 帧处进行了 10 帧的帧删除篡改(95~105), 检测结果如图 10(a)所示, 可以观察到在 95 帧处有尖锐明显的峰值出现, 用黑色圆圈标记; 如图 10(b)所示, 实验视频是 CDNET 视频库中 turbulence3.avi, turbulence 分组的视频画面运动较为剧烈, 并在 238 帧处进行了 75 帧的帧删除篡改(238~313), 可以观察到在 238 帧处有尖锐明显的峰值出现, 同时篡改点用黑色圆圈标记。对于画面运动相对剧烈的视频, 选取的实验视频是 HUANAN 视频库中 walkman.avi, 该视频记录的是奔跑的行人, 同时对该视频在 200 帧处进行了 25 帧的帧删除篡改(200~225), 检测结果如图 10(c)所示, 可以观察到在 221 帧处有尖锐明显的峰值出现, 而且用黑色圆圈标记。如图 10(d)

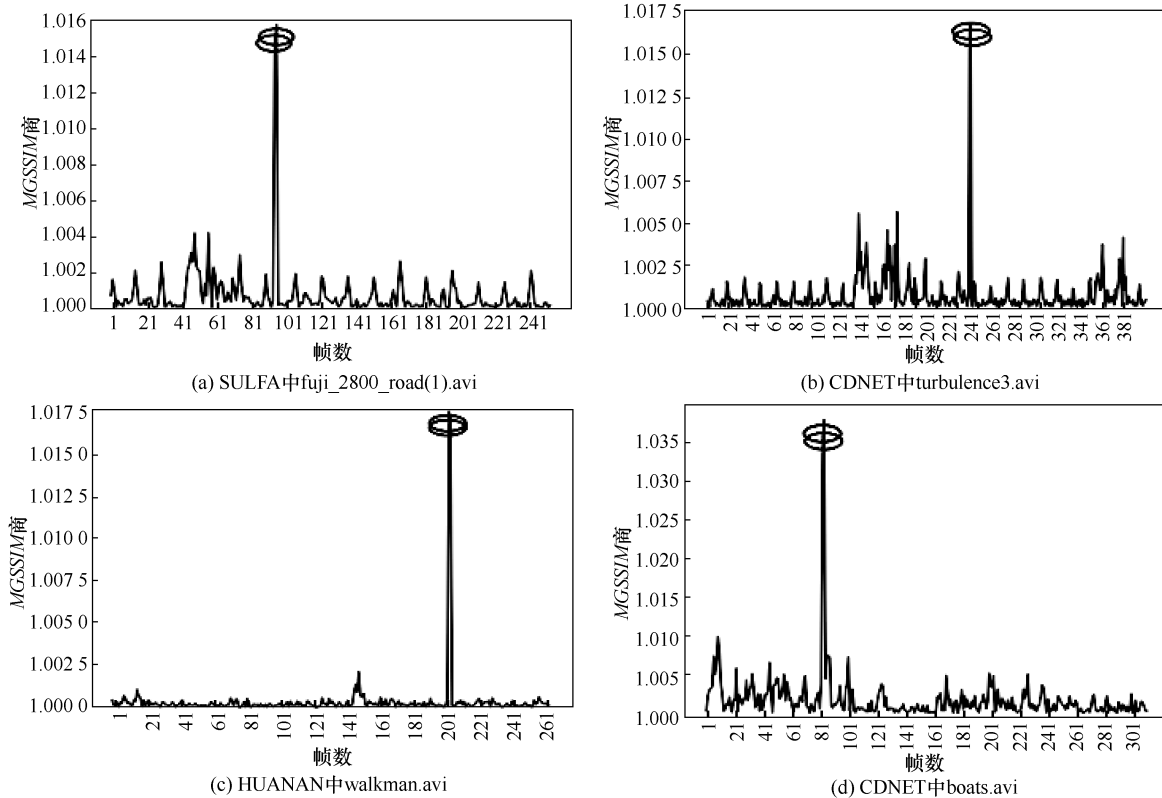


图10 运动变化剧烈视频的篡改检测结果

所示,实验视频是CDNET视频库中dynamicBackground的boats.avi,由于它是动态背景的视频,所以整体画面运动相对剧烈,同时在82帧处进行了20帧的帧删除篡改(82~102),可以观察到在82帧处有尖锐明显的峰值出现,用黑色圆圈标记。

通过实验证明,本文算法对不同运动剧烈程度的视频检测性能都比较好,均能精准地把篡改点定位出来。

(2) 验证空间约束效果的对比实验

为了验证本文算法所提的空间约束是否能提升检测的准确率,在同等条件下,针对运动变化剧烈和运动变化平缓的两类视频,对视频进行空间约束和不进行空间约束分别做对比实验。

① 运动变化剧烈的视频

针对运动变化剧烈视频的对比实验结果如图11所示,实验视频是CDNET视频库中Night Video的Video1.avi,并在36帧处进行了5帧的帧删除篡改(36~51)。首先,对待检测视频不进行空间约束,篡改检测的效果如图11(a)所示。可以明显看到被篡改的位置36帧处没有出现尖锐的峰值,且并没有定位到异常点,却在92帧和151帧处标记了篡改点,存在严重的误检测。造成误检测的主要原因是该视频的运动变化较为剧烈,从图11(a)中可以观察到该视频相邻帧的相关性波动很大,因为该视频中存在高速运动的汽车,所以该视频的运动变化非常剧烈。然后,对视频进行空间约束,同样的条件下,对得到的量化相关性丰富区域进行视频篡改检测,检测效果如图11(b)所示,可以观察到在36帧处有尖锐明显的峰值出现,用黑色圆圈标记。

通过对比实验,证明了对视频进行适当的空间约束可以提高视频篡改检测的准确率,减少误检率。而且受视频内容变化的影响,现有算法对仅仅几帧的删除篡改往往检测不到,但

本文算法却可以实现精确检测。说明本文算法不仅能适应于运动变化剧烈的视频篡改,还同样能检测到几帧的细微帧删除篡改,具有很强的顽健性。

② 运动变化平缓的视频

针对运动变化平缓视频的实验结果如图12所示,实验视频是HUNAN视频库中的Yvmaoqiu.avi,并在53帧处进行了25帧的帧删除篡改(53~78)。首先,对视频不进行空间约束,篡改检测的结果如图12(a)所示。可以看到被篡改的位置53帧处出现尖锐的峰值,以及黑色圆圈标记的异常点。从图像中可以观察到该视频相邻帧的相关性波动很小,因为视频中不存在高速运动的物体,所以该视频的运动变化较为平缓。然后,对视频进行空间约束,同样的条件下,对空间约束区域进行篡改检测,检测效果如图12(b)所示。可以观察到在53帧有尖锐明显的峰值出现,用黑色圆圈标记。通过对比实验,说明本文算法同样适用于运动变化平缓的视频。

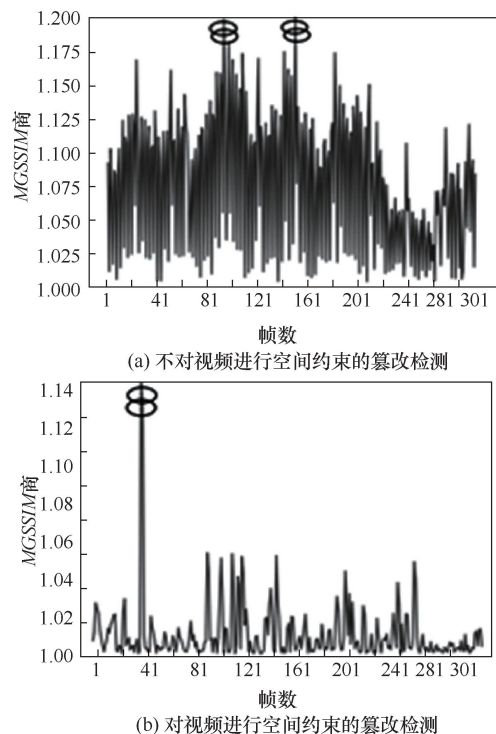


图11 运动变化剧烈视频的对比实验

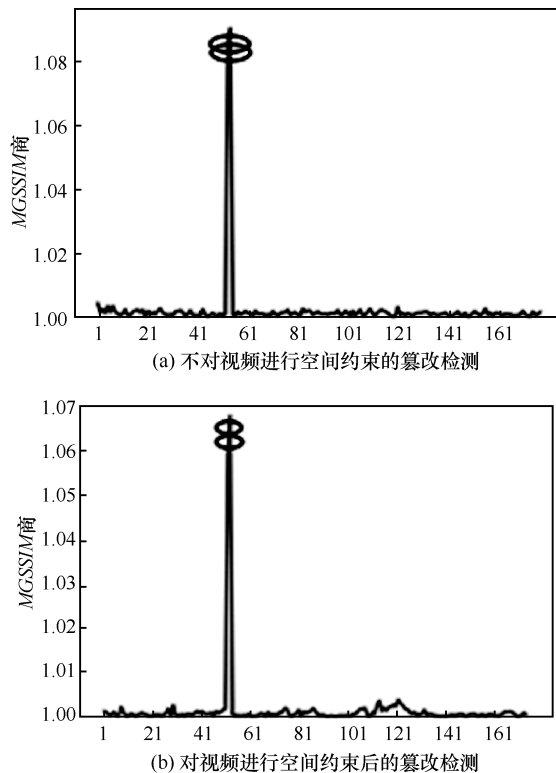


图 12 运动变化平缓视频的对比实验

(2) 验证特征 GSSIM 的顽健性

基于人类视觉特性的原理改进 SSIM 的特征提取和描述方法, 本文算法提出在量化相关性丰富区域上提取梯度结构相似性 GSSIM 进行篡改检测。为了验证 GSSIM 特征具有较强的顽健性, 能够适用于画面模糊视频的特征提取。在同等条件下, 针对画面模糊和画面清晰两种类型的视频, 分别提取 SSIM 和 GSSIM 进行对比实验。

① 画面模糊的视频

针对画面模糊视频的对比实验结果如图 13 所示, 实验视频是 CDNET 视频库中 Night Video 的 Video1.avi, 并在 75 帧处进行了 20 帧的帧删除篡改 (75~95)。首先, 用 Adobe Premiere Pro CC 软件对篡改后的视频进行高斯模糊处理, 其中模糊因子设为 10; 其次, 对空间约束后的视频提取 SSIM 特征进行篡改检测, 检测的效果如图 13(a) 所示, 可以看到被篡改的位置 75 帧处并没有出现尖锐的峰值, 同时也没有黑色圆圈标记的异常点, 但在 170

帧处标记到了异常点, 造成了误检测; 然后, 同样的条件下, 对视频提取 GSSIM 特征进行篡改检测, 检测效果如图 13(b) 所示, 可以观察到在 75 帧有尖锐明显的峰值出现, 用黑色圆圈标记。通过对比实验, 说明本文提出的 GSSIM 特征具有较强的顽健性, 能够适用于模糊视频的篡改。

② 画面清晰的视频

针对画面清晰视频的对比视频实验结果如图 14 所示, 实验视频是 SULFA 视频库中的 02_original.avi, 并在 61 帧处进行了 20 帧的帧删除篡改 (61~81)。首先, 对空间约束后的视频提取 SSIM 特征进行篡改检测, 检测效果如图 14(a) 所示, 可以看到被篡改的位置 61 帧处出现尖锐的峰值, 同时也有黑色圆圈标记的异常点; 其次, 同样的条件下, 对同样的视频提取 GSSIM 特征进行检测, 检测效果如图 14(b) 所示, 可以观察到在 61 帧处同样有尖锐明显的峰值出现, 而且用黑色圆圈进行了标记。通过对比实验, 说明本文算法提出的 GSSIM 特征适用于清晰视频的篡改。

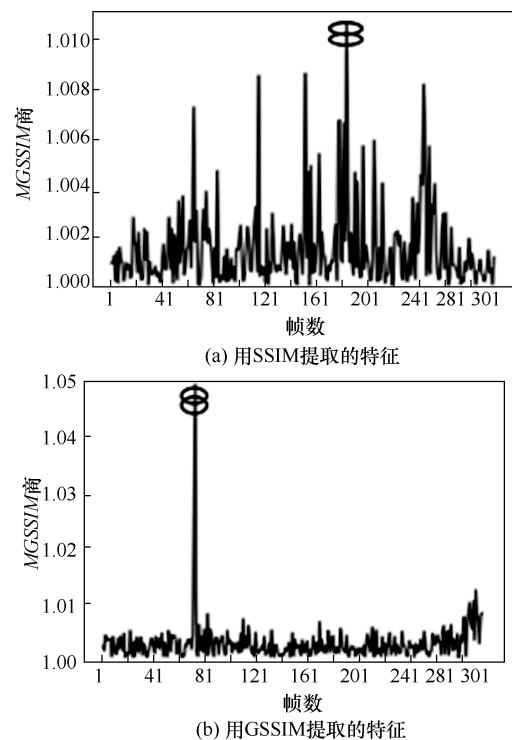
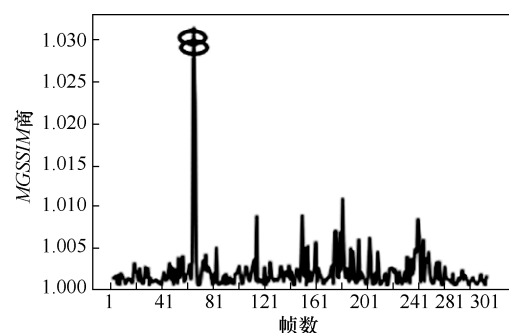
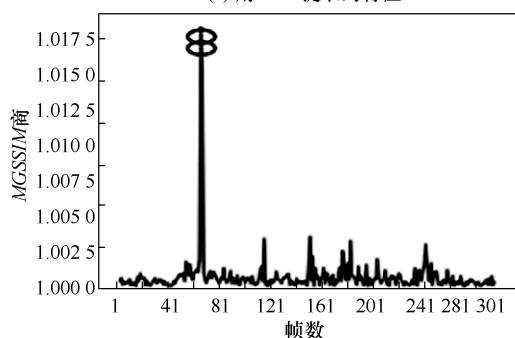


图 13 画面模糊视频的对比实验



(a) 用SSIM提取的特征



(b) 用GSSIM提取的特征

图14 画面清晰视频的对比实验

(precision)、召回率(recall)对实验结果进行分析, 计算公式如下。

$$precision = \frac{N_c}{N_c + N_f} \quad (34)$$

$$recall = \frac{N_c}{N_c + N_m} \quad (35)$$

其中, N_c 是检测出的正确(correct)点数, N_f 是检测出的错误(false)点数, N_m 是未检测出(miss)的篡改点数, 即漏检的篡改点的数量。

部分实验结果如表1和表2所示, 从表1中可以看出本文算法对不同视频可以有效准确地检测到不同程度的帧删除篡改, 其中篡改帧数为5帧时的召回率最低, 原因是该篡改帧数较小, 不足以影响到相邻帧间的相关性值的变化, 造成了漏检的篡改点的数量较多; 篡改帧数为100的检测准确率最高, 因为该篡改帧数较大, 漏检的篡改点的数量较少, 同时误检率也比较低, 所以精确率较高。表2是本文算法与对比文献算法的性

4.2.3 量化分析与讨论

为了评价算法的性能, 本文使用准确率

表1

视频的帧删除篡改检测结果

篡改帧数	篡改点总数	检测出的正确点数	未检测出的篡改点数	检测出的错误点数	召回率	精确率
5	200	168	32	6	84.00%	96.55%
10	200	176	24	5	88.00%	97.24%
20	200	178	22	4	89.00%	97.80%
25	200	179	21	6	89.50%	96.76%
50	200	183	27	8	91.50%	95.81%
75	200	184	16	7	92.00%	96.34%
100	200	189	11	5	94.50%	97.42%

表2

各算法性能比较结果

数据集	本文算法		文献[5]算法		文献[6]算法	
	准确率	召回率	准确率	召回率	准确率	召回率
SULFA	99.95%	99.45%	82.43%	85.26%	89.02%	90.53%
CDNET	84.62%	91.11%	81.65%	83.31%	81.83%	83.71%
HUANAN	99.71%	86.25%	83.12%	84.21%	84.91%	85.72%

能比较结果,通过分析表 2 可以得出结论,本文通过空间约束对量化相关性丰富区域进行篡改检测,在不同公开视频数据库上的检测结果在准确率和召回率上总体均优于文献[5]和文献[6]的算法。本文算法对 CDNET 库中夜间视频的检测性能相对不太理想,但也优于对比算法,原因可能是夜间视频光线比较暗,对该类视频提取到的帧特征不能很好地代表帧图像内容,所以算法在夜间视频的检测性能略微下降。

5 结束语

本文提出了融合空间约束和梯度结构信息的视频篡改检测算法,首先对给定视频进行空间约束,分别自适应地提取出低运动和高纹理区域,把两区域进行融合,获取量化相关性丰富区域,通过在此区域上提取梯度结构相似性特征 GSSIM 得到空间约束相关性值,并利用 GSSIM 商对空间约束相关性值进行后处理,最后利用切比雪夫不等式实现对篡改点的检测和定位。通过对比实验证明,对视频进行空间约束可以过滤掉一些易造成误检的区域,使约束得到的量化相关性丰富区域的相邻帧之间相关性值大小较为稳定,使之能够适用于运动变化剧烈视频的篡改检测;同时基于人类视觉系统的特性改进了原有的特征提取和描述方法,提出了检测性能较好的梯度结构相似性特征 GSSIM,可以适用于模糊视频的篡改取证。实验证明本文算法具有很高的精确率、极低的误检率及较强的顽健性,不足之处在于对夜间视频的篡改检测性能不太好,这也是后继需要重点研究的工作。

参考文献:

- [1] 陈威兵,杨高波,陈日超等. 数字视频真实性和来源的被动取证[J]. 通信学报, 2011, 32(6):177-183.
CHEN W B, YANG G B, CHEN R C, et al. Digital video passive forensics for its authenticity and source[J]. Journal on Communications. 2011, 32(6):177-183.
- [2] SINGH, RAAHAT DEVENDE R, AGGARWAL N. Video content authentication techniques: a comprehensive survey[J]. Multimedia Systems, 2018:1-30.
- [3] ZHAO D N, WANG R K, LU Z M. Inter-frame passive-blind forgery detection for video shot based on similarity analysis[J]. Multimedia Tools & Applications, 2018(1):1-20.
- [4] SOWMYA K N, CHENNAMMA H R, LALITHA R. Video authentication using spatio temporal relationship for tampering detection[J]. Journal of Information Security and Applications, 2018, 41:159-169.
- [5] WANG W, FARID H. Exposing digital forgeries in video by detecting double MPEG compression[C]//8th Workshop on Multimedia and Security (MM&Sec'06). NY, 2016: 37-47.
- [6] LIU H, LI S, BIAN S. Detecting frame deletion in H.264 video[C]//10th International Conference ISPEC. 2017: 262-270.
- [7] 左建军, 吴有富. 一种基于直方图的实时自适应阈值分割方法[J]. 毕节学院学报, 2014, 32(4):53-56.
ZUO J J, WU Y F. A real-time adaptive threshold segmentation method based on histogram[J]. Journal of Bijie University, 2014, 32(4): 53-56.
- [8] 黄添强, 陈智文, 苏立超, 等. 利用内容连续性的数字视频篡改检测[J]. 南京大学学报(自然科学), 2011, 47(5):493-503.
HUANG T Q, CHEN Z W, SU L C, et al. Digital video forgeries detection based on content continuity[J]. Journal of Nanjing University(Natural Sciences), 2011, 47(5):493-503.
- [9] 薛丽霞, 罗艳丽, 王佐成. 基于帧间差分的自适应运动目标检测方法[J]. 计算机应用研究, 2011, 28(4):1551-1552.
XUE L X, LUO Y L, WANG Z C. Detection algorithm of adaptive moving objects based on frame difference method[J]. Application Research of Computers, 2011, 28(4):1551-1552.
- [10] 王珉, 毕秀丽, 马建峰, 等. 基于概率统计模型与图像主纹理方向分析的非线性滤波算法[J]. 中国图象图形学报, 2018, 13(5):858-864.
WANG X, BI X L, MA J F, et al. Nonlinear filtering algorithm using probability statistic and main texture direction analysis based on random transforms[J]. Journal of Image and Graphics, 2018, 13(5):858-864.
- [11] 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: 289-302.
- [12] 孙健钧, 赵岩, 王世刚. 基于图像梯度信息强化的 SIFT 特征匹配算法改进[J]. 吉林大学学报(理学版), 2018, 56(1):82-88.
SUN J J, ZHAO Y, WANG S G. Improvement of SIFT feature matching algorithm based on image gradient information enhancement[J]. Journal of Jilin University (Science Edition), 2018, 56(1): 82-88.
- [13] NERCESSIAN S, AGAIAN S S, PANETTA K. A. An image similarity measure using enhanced human visual system characteristics[C]//SPIE Defense, Security and Sensing. 2011.
- [14] QADIR G, YAHAYA S, HO A T. Surrey university library for forensic analysis (SULFA) of video content[C]// IET Conference on Image Processing (IPR 2012). 2012: 1-6.
- [15] HE Z. The data flow anomaly detection analysis based on LipChebyshev method[J]. Computer System Application, 2009, 18(10): 61-64.

- [16] GOYETTE N, JODOIN P M, PORIKLI F, et al. Changedetection.net: a new change detection benchmark dataset[C]// Computer Vision & Pattern Recognition Workshops. 2014: 1-8.
- [17] HU Y J, AL-HAMIDI S. Construction and evaluation of video forgery detection database[J]. Journal of South China University of Technology (Natural Science), 2017, 45(12): 57-64.

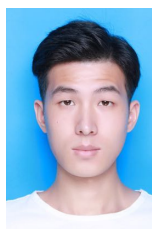


翁彬（1981- ），男，福建福州人，博士，福建师范大学讲师，主要研究方向为机器学习及应用。

[作者简介]



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。



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。



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



黄维（1994- ），女，福建莆田人，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。