

基于关键帧和指示符运动模型的教学视频压缩算法

孟春宁*, 陈梓铭, 冯明奎, 赵强

公安海警学院电子技术系, 浙江 宁波 315801

摘要 为进一步提高教学视频的压缩比及其制作效率, 针对教学视频中有效信息暂留时间持续较长、信息展示区域固定等特点, 提出一种基于关键帧检测和指示符运动建模的智能教学视频压缩算法。首先检测投影区域作为每帧图像的有效区域, 减少知识冗余; 然后通过变化检测确定视频帧的类型, 并对光标和激光笔投影点建立指示符运动模型, 进一步减少教学视频特有的知识冗余; 最后针对关键帧编码及指示符运动模型设计了相应的播放算法用于视频回放。实验结果表明, 针对以幻灯片投影区域为有效信息的教学视频, 与 H.264 标准相比, 本文算法在相同峰值信噪比下, 可使码率平均降低约 88%, 且编解码过程满足实时要求, 无需额外人工剪辑, 可大幅提高在线教学视频的制作与传输效率。

关键词 图像处理; 教学视频; 视频压缩; 变化检测; 关键帧检测; 指示符建模

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP55.101005

Teaching Video Compression Algorithm Based on Key Frame and Indicator Movement Model

Meng Chunning*, Chen Ziming, Feng Mingkui, Zhao Qiang

Department of Electronic Technology, China Maritime Police Academy, Ningbo, Zhejiang 315801, China

Abstract To improve teaching video compression ratio and production efficiency, considering that effective information of teaching video normally remains for a long time and locates in fixed area, a teaching video compression algorithm based on key frame detection and indicator movement modeling is proposed. Firstly, the projection area is detected as the effective region of each frame to reduce the knowledge redundancy. Then, the type of each frame is determined using change detection, and the indicator motion model is established for the mouse cursor and the laser point to reduce more knowledge redundancy. Finally, a corresponding play algorithm based on the key frame coding and indicator motion model is designed to replay the videos by use of OpenCV and MFC. The experimental results showed that, for the teaching videos with effective information in the projection area, the proposed method achieved an average of 88% bitrate reduction than that of H.264 under the same peak signal to noise ratio value. Besides, the encoding and decoding efficiency of this algorithm could meet the real-time requirement. Without extra manual editing, the production and transmission efficiency of online courses could be significantly increased.

Key words image processing; teaching video; video compression; key frame detection; change detection; indicator motion modeling

OCIS codes 100.2000; 100.3005; 100.5010

1 引言

计算机辅助的多媒体教学方式能够展示更丰富更形象的教学内容。随着计算机技术的高速发展及个人计算机的普及, 多媒体教学迅速发展成为当今

主流的课堂教学方式。近年来, 伴随网络技术的发展兴起的慕课、微课等在线课程正在引发新的教育变革, 相比传统课堂, 在线课程没有时空局限, 极大地扩展了知识的传播速度和范围。在多媒体教学过程中, 除了 PPT 课件的展示, 还可能会涉及软件操

收稿日期: 2018-04-24; 修回日期: 2018-05-03; 录用日期: 2018-05-25

基金项目: 全国教育科学规划教育部重点课题(DIA150308)

* E-mail: mengchunning123@163.com

作、程序调试、教学指点等演示过程,只拷贝 PPT 课件或录制音频均无法重现完整的教学过程。因此,完整真实地记录多媒体教学过程的教学视频成为在线课程的核心资源。采用普通商用数码相机录制一段分辨率为 $1280 \text{ pixel} \times 720 \text{ pixel}$ 、帧率为 30 frame/s 、持续时间为 45 min 的教学视频,其文件大小约为 2G ;在保证画面清晰度(特别是 PPT 区域文字的辨识度)前提下,使用常规视频压缩算法取得的压缩比不高。因此,大幅提高压缩比、简化教学视频的处理过程将给在线课程的制作、分享与更新带来更大的便利。

目前,考虑到教学视频具有场景固定、有效信息集中在投影区域且变化速度慢等显著特点,未充分利用这些特点的通用压缩算法对教学视频的压缩效果不尽如人意,而专门针对教学视频的压缩研究相对较少。本文将多媒体教学授课场景中的幻灯投影区域视为教学视频的有效区域,充分利用有效区域图像帧间变化速度慢的特点,同时将鼠标光标和激光笔投影点的运动作为典型的教学指点过程建模,依据帧间变化情况将视频帧分为关键帧、有效帧和冗余帧,利用区域分割、变化检测、指示符检测及建模等技术实现了基于关键帧和指示符运动模型的教学视频压缩算法。该方法只保留原始视频帧中的投影区域,去除帧间变化小的冗余帧和包含指示符运动的有效帧,在大幅减少数据量的同时,还可保护授课教师的个人隐私。在压缩编码过程中,针对持续较长时间静止的关键帧采取无损压缩策略,保证关键帧中有效信息(如每页 PPT 课件等)的清晰度;采用常用的基于运动估计的帧间预测编码技术对有效帧进行压缩编码,在保证效率的同时尽量提高压缩比;针对包含指示符运动的有效帧,利用指示符运动模型代替普通的时域压缩技术,进一步减少数码率。并且设计了一种专门的播放器播放处理后的视频码流。实验结果表明,相比现有通用方法,在同等画面清晰度情况下可大幅降低码率。这种方法可为慕课、微课等在线课程视频网站的视频制作及分享提供高效的解决方案,还可为数码相机提供专门的教学视频录制功能。

2 相关工作

通用视频压缩编码标准的发展可分为三个阶段:以 H.261、H.262、MPEG-1、MPEG-2 等为代表的第一代视频压缩编码标准,以 H.264、MPEG-4、AVS 等^[1]为代表的第二代视频压缩编码标准,以及

以 H.265/HEVC^[2]为代表的新一代压缩编码标准。基于帧间预测编码及变换编码技术的 MPEG-4 和 H.264 仍然是当前最常用的视频压缩编码标准^[3-4]。H.265 标准在 H.264 基础上对帧内预测、帧间预测、变换量化以及熵编码等关键技术进行改进,进一步提高了视频压缩效率,平均比特率可节省 49.3% ^[2],但同时也增加了算法的复杂度,存在实时编码难度大、解码难度大、实现硬件要求较高等问题,因此目前尚未全面普及。很多研究工作致力于研究提高第二代压缩标准的压缩比或降低第三代标准的时间复杂度的方法。例如, Schwarz 等^[5]提出使用拉格朗日编码控制的方法提高 MPEG-4 编码效率,该算法在不同光强和比特率下压缩后的画面峰值信噪比(PSNR)均高于原算法。Manjanaik 等^[6]在 H.264 基础上提出了基于高斯脉冲沿对角线左下方向的帧内预测模式,在图像质量不变的前提下,平均可降低视频文件 30% 的比特率。Sharabayko 等^[7]提出了 H.265/HEVC 视频压缩的帧内预测模式的快速搜索算法,在平均只增加 1.9% 比特率的前提下可节省 41% 的运算时间。

由于教学视频的特殊性及其广泛的应用,专门针对教学视频的特点设计的非通用压缩算法可取得更好的压缩效果。例如, Purcell 等^[8]采用直接获取教学平台中 PPT 画面的方式录制教学视频,无需摄像机拍摄,视频清晰度高,但需要使用信号转换设备连接教学平台和录制视频的主机,且操作相对繁琐。Mittal 等^[9]利用相位相关及内容自适应技术实现对教学视频的压缩,通过对视频内容进行分类标记、分割,根据不同内容选择不同的压缩策略,该方法对视频部分的压缩比最高可达到 $1244:1$,但压缩后 PSNR 值最高只有 34.8 dB ,画面失真现象明显。古佳等^[10]利用帧间压缩和基于小波变换的帧内压缩方法对教学视频中的幻灯片内容进行压缩,该方法只能处理单色视频,且压缩后视频分辨率明显降低。

近些年来,基于内容分析的视频编码技术成为提高编码性能的重要方法,这种方法通过分析视觉相关性消除视觉冗余及知识冗余。针对教学视频,通常需要对投影区域内容进行变化检测,检测视频中是否存在有效的运动目标,并以此判断是否存在冗余。因此,变化检测技术是基于内容分析的编码技术的关键。背景差分法广泛应用于变化检测,例如, Barnich 等^[11]提出的 ViBe 非参数化前景检测算法采用随机背景更新策略,具有背景建模速度快、计算量小的优点,但存在无法消除阴影、检测结果中目

标不完整且存在鬼影等问题。陈海永等^[12]提出的基于超像素信息反馈的视觉背景提取算法,将原始图像分割为若干个超像素区域,在 ViBe 算法的初始阶段实现鬼影信息的准确检测,并更新鬼影区域像素点的背景模型,在一定程度上改善鬼影现象的同时增加了计算量,实时性较差。莫邵文等^[13]在 ViBe 算法基础上通过建立一个闪烁程度矩阵,判断背景高频扰动程度来设置自适应匹配阈值,并加入小目标丢弃和空洞填充策略,提高了算法稳健性,但鬼影现象在 40 帧后才能完全消除。另一种常用的变化检测方法是 Stauffer 等^[14]提出的混合高斯背景建模(GMM)法,该方法通过背景模型实时更新以适应场景的变化,提高背景学习效果。这种方法检测精度较高,能够克服阴影对检测效果的影响,在对变化检测的精度和对快速运动的小目标检测方面均优于 ViBe 算法,同时对光照变化和复杂背景具有很好的适应性。很多研究工作改善了混合高斯模型的性能,例如,Wang 等^[15]提出了结合三帧差法和改进更新策略的 GMM 算法,提高了混合高斯建模法的计算效率和精度。陈强等^[16]提出了结合边缘混合高斯模型与帧间差分的红外光照突变下的运动目标检测算法,提高了混合高斯模型在光照突变情况下的检测效果。近年来,St-Charles 等^[17]提出的基于局部二值相似性模式(LBSP)的运动目标检测算法(LOBSTER)和基于局部自适应灵敏度的通用运动目标检测(SuBSENSE)算法^[18]取得了较好的检测效果,LOBSTER 算法采用具有空间信息的 LBSP 描述子建立像素特征模型进行像素间的比较,算法稳定性较高。SuBSENSE 算法结合像素的 RGB 值和 LBSP 特征进行像素级建模,并运用一种新的反馈机制动态调控算法的敏感度和更新速度,该算法在复杂场景下更具灵活性,对动态背景的适应能力较强。但测试实验表明,这两种算法计算复杂度较大,难以保证对常见教学视频的实时处理。

3 基于关键帧和指示符运动模型的压缩算法

所提出的压缩编码方法主要包括以下部分:投影区域分割、投影区域中的变化检测、鼠标光标和激光笔投影点检测与建模、基于关键帧的编码算法和压缩码流播放算法。

3.1 投影区域分割

在多媒体教学过程中,投影区域是展示信息的

主要区域,课堂展示的多媒体信息内容通过投影仪投影在屏幕上。对国内各大慕课网站的调查显示,绝大多数在线教学视频都是以投影区域为有效区域,其中部分视频同时录制投影区域和教师讲课过程,增加后期制作过程实现适时切换,或者将课件与教师动作录像编辑合成教学视频,这两种教学视频的教学信息同样集中在投影区域,且需要人工进行后期剪辑合成。为减少后期视频制作的工作量同时保护授课人员的隐私,首先分割出投影区域并将其视为教学视频中的唯一有效区域。

为适应后续分割及检测算法的输入要求并尽量消除噪声对后期图像处理的影响,首先使用均值滤波、灰度化及形态学开运算等方法对原始视频帧进行预处理。

一般地,幻灯片投影区域会被投影仪光源照得更亮,而不同的拍摄角度可能导致投影区域发生形变,使得投影区域的形状非方正的矩形。针对这些特点,采取类矩形区域检测结合亮度值统计的方法检测投影区域,具体过程的伪代码如下。

算法 1 投影区域检测算法

输入:教学视频帧第一帧图像 frame

输出:投影区域感兴趣区域(ROI)

```

1) frame ← preprocess(frame)
   /* 图像预处理 */
2) edge ← Canny(frame);
   /* 使用 Canny 算子检测边缘 */
3) HoughLinesP(edge, line);
   /* 累计概率霍夫变换检测边缘中的直线,
   line 为检测到直线的集合 */
4) Find(line, Rect, 80, 100);
   /* 在所有直线围成的四边形中,找出内角在
   80°和 100°之间的候选矩形区域,Rect 为候选区域
   的集合 */
5) FOR i=1:|Rect|
   /* i 的取值从 1 到 |Rect|, |Rect| 为集合 Rect
   的大小 */
6) IF Rect.area > 0.04 * frame.area /* X.area
   为 X 的面积,剔除候选区域中面积较小的部分 */
7) THEN AvgB ← AvgBrightness(Rect[i],
   frame); /* 计算各候选矩形区域平均亮度,
   AvgB 为候选区域平均亮度的集合 */
8) END IF
9) END
10) ROI ← Max(AvgB, Area)

```

/ * 取候选区域中平均亮度最大的作为投影区域 * /

考虑到教学视频录制过程中,摄像机有可能小幅移动,采用变化检测算法进行实时检测,结果表明,当画面变化较大时,重复上述分割方法进行投影区域的再检测,若发生连续的大幅度变化,则每两次检测之间间隔不少于 1 s,并加入必要的图像矫正以适应前次检测的大小。在确定每帧图像中的投影区域后,将其设置为有效区域,后续操作仅对视频帧的有效区域进行处理。

3.2 投影区域变化检测

在教学过程中,切换软件、幻灯片翻页以及每页中新条目的出现等情况是投影区域中常见的有效变化,除此之外每一页教学演示文稿往往长时间处于静止状态,对大量教学视频的统计表明,这个过程通常持续数分钟,即可能连续数千帧中投影区域展示的内容保持不变,存在大量冗余信息。采用教鞭、激光笔以及鼠标指点演示文稿展示的内容是教学视频中另一种常见且独有的现象,虽然激光笔或鼠标的指点运动也是一种包含有效信息的变化,但这种变化中运动目标面积基本固定,且变化部分面积不大,对背景图像的影响很小。

考虑到教学视频中常见变化现象的特点,对 Zivkovic 等^[19]提出的自适应高斯混合建模算法进行改进以用于检测投影区域中的各种变化。该方法根据视频中的每个像素在时域上的分布情况构建图像中各个像素的灰度分布模型,利用大量像素样本值的概率密度统计信息(如模式数量、每个模式的均值和标准差)表示背景,使用统计差分估计目标像素值。高斯混合模型的背景是动态更新的,只有当由前景像素匹配生成的混合高斯模型被频繁匹配时,才将该模型判定为背景模型,因此这种方法对光照变化具有较好的适应能力,以及抗噪能力强、检测精度高的优点。

相比自适应高斯混合建模法、LOBSTER 算法和 SuBSENSE 算法,ViBe 算法检测精度较低,特别对于激光笔投影点这样快速运动的小目标,会频繁出现漏检的情况,并且由于 ViBe 算法是利用第一帧初始化背景模型,当画面内容发生较大面积突变而后迅速进入静止状态时,背景模型的更新速率难以与背景变化同步,容易导致检测出虚假目标,即鬼影现象,如图 1 所示。图 1(a)是原视频帧,图 1(b)是混合高斯背景模型的变化检测结果,图 1(b)是

ViBe 算法的检测结果,图 1(d)是 LOBSTER 算法的检测结果,图 1(e)是 SuBSENSE 算法的检测结果,第 1 列是一段视频帧的第 325 帧,第 2 列是第 354 帧,第 3 列是第 368 帧,第 4 列是第 380 帧。LOBSTER 算法原理是基于像素间的比较,其背景模型的初始化与 ViBe 算法类似,但该算法采用具有空间信息 LBSP 描述子进行背景建模,提高了算法的检测精度和鬼影消除的速度。SuBSENSE 算法结合色彩信息和 LBSP 描述子进行像素级建模,通过样本一致性分割前景和背景,并使用反馈机制调控算法的灵敏度和模型更新速度,相比于 LOBSTER 算法,该算法的鬼影消除速度更快,但计算量也更大,处理速度较慢。而混合高斯背景模型使用多个高斯分布表征图像中各像素点的特征,在获得新一帧图像后,根据高斯分布的方差和均值匹配情况更新背景模型,背景模型更新及时,能够很好地消除鬼影现象。实验结果表明,LOBSTER 算法和 SuBSENSE 算法在处理速度方面较其余两种算法要慢很多,实时性较差,不适用于授课视频的变化检测。高斯混合模型计算量大于 ViBe 算法,为提高其运行效率,针对幻灯片投影区域出现阴影的几率小且阴影部分通常不包含有效信息的特点,本文舍弃了经典高斯混合模型中的阴影检测过程,提高了算法执行效率,测试结果表明减少阴影检测过程后平均运行速度约提高 5%。

根据自适应高斯混合模型的前景检测结果,将前景为空的无变化帧定义为冗余帧,有变化的帧(即前景不为空)定义为有效帧。将每段连续有效帧中的第一帧和最后一帧定义为关键帧。以演示文稿展示过程为例,切换或新条目出现的过程中涉及到的视频帧是有效帧,切换前或切换后的静止时间内的视频帧是冗余帧,有效帧的首尾帧是需要持续较长时间展示的关键帧。针对不同类型视频帧采用不同的压缩方法,其中关键帧采用无损压缩,尽可能地保证有效信息的清晰度,关键帧以外的有效帧采用有损压缩,舍弃所有冗余帧及包含指示符运动的有效帧。教学视频中包含大量冗余帧和指示符运动帧,传统视频压缩算法对于这些视频帧一般采用通用的帧间预测编码方法,数据量减少程度有限,而所提出的算法则直接舍弃冗余帧,使用指示符模型和位置信息代替包含指示符的有效帧,相比较传统视频压缩算法可以更大程度地减少视频中存在的时间冗余和知识冗余。

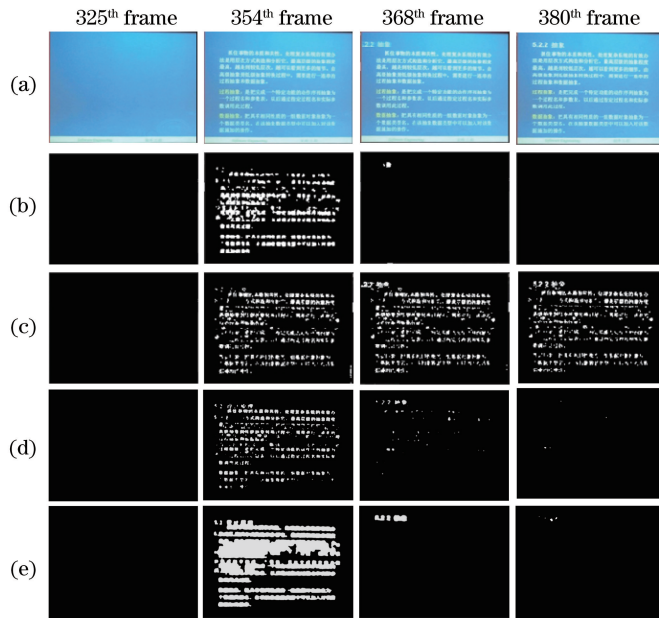


图 1 部分视频帧的变化检测结果。(a)原视频帧;(b)混合高斯背景模型的变化检测结果;(c) ViBe 算法的检测结果;
(d) LOBSTER 算法的检测结果;(e) SuBSENSE 算法的检测结果

Fig. 1 Results of change detection of some frames. (a) Original frames; (b) results of GMM;
(c) results of ViBe; (d) results of LOBSTER; (e) results of SuBSENSE

3.3 鼠标光标和激光笔投影点检测与建模

鼠标指针和激光笔等指示符经常用于辅助教师指点当前授课的内容,是教学视频中的重要指示信息。因此,在连续的有效帧中,通常包含大量激光笔或鼠标的指点运动。提出一种指示符运动模型,通过检测激光笔投影点及鼠标光标的位置,通过预存的指示符模板图像将指示符再现在固定的背景图像中。在包含指示符运动的有效帧中,投影区域会检测到微小的变化,此时背景内容并未发生改变,检测并记录指示符位置,将预存的指示符图像在关键帧上的检测位置处再现,这样就可将一段连续的有效帧用一幅关键帧、预存的指示符模板图像以及每帧中指示符的位置信息取代,可大幅减少有效帧序列的码率。

由于摄像机帧率和曝光时间等因素的影响,教学视频中运动的鼠标光标在速度较快时会发生形

变,而运动的激光笔投影点则带有不同程度的拖影,如图 2 所示。对包含指示符运动的大量视频帧的统计分析显示,虽然指示符形状不固定,但其面积与投影区域的面积之比在一定范围内变化。随机选取 8000 帧包含指示符运动的视频帧进行测试,统计结果表明,指示符占有有效区域面积的 $1.5\% \sim 4.0\%$ 。然而在投影区域检测到的运动目标中,面积占比在此范围内的除了指示符外还有其他变化较小的过渡动画效果等。指示符运动区别于其他变化的主要特点是:指示符在画面中是一块连通的区域,且区域内像素之间的颜色信息差别较小。因此提出通过计算前景不同像素块颜色分量的欧氏距离,设置阈值对前景进行筛选。经过对 20000 帧光照条件不同的指示符运动视频帧检测实验结果的统计,阈值取 15 时检测准确率可达到 87% 。

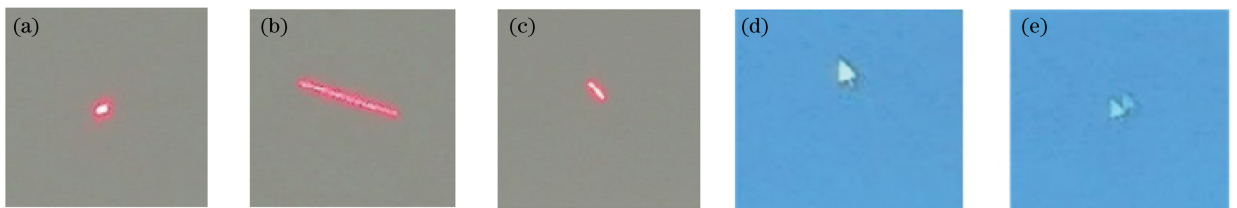


图 2 视频帧中不同状态的指示符。(a)(d)静止;(b)(e)快速运动;(c)慢速运动

Fig. 2 Different states of the indicator in the video. (a)(d) Static; (b)(e) moving fast; (c) moving slowly

由于鼠标光标和激光笔投影点在颜色上存在明显区别,激光笔投影点颜色是红色,而鼠标光标是白色,在 HSV 颜色空间容易分辨鼠标光标和激光笔投影点。统计不同光照条件的指示符视频帧,可以发现 H (Hue 色调)分量对光照不敏感,而当投影区域背景颜色改变时,指示符的颜色会发生相应偏色,在进行指示符检测前对投影区域图像进行白平衡处理可有效抑制背景颜色对指示符检测的影响,白平衡处理过程是通过计算图像三通道的平均灰度得到三通道的增益系数,然后利用原始灰度值乘以相应增益系数来调整原始图像色差。通过对不同光照和不同背景颜色指示符视频帧的测试和统计,鼠标光标 H 分量平均值在 75~110 之间,而激光笔投影点 H 分量平均值在 165~180 之间,两者数值差距明显。针对上述特点,提出了基于灰度及颜色信息的指示符检测与识别算法,该算法具体执行过程的伪代码如下。

算法 2 指示符检测算法

输入:前景图像 front,投影区域 ROI

输出:指示符类型 Ptype,指示符坐标 p

```

1) AWB(ROI);
   /* 对投影区域进行白平衡处理 */
2)  $x \leftarrow \text{Area}(\text{Front}) / \text{Area}(\text{ROI});$  /* 计算前景与投影区域面积之比 */
3) IF( $x > 1.5\%$  AND  $x < 4.0\%$ )
4)   THEN  $y = \text{CDcount}(\text{front});$  /*  $y$  为前景图像 front 中连通区域个数 */
5) IF( $y = 1$ )
6)   THEN Random(front, P, Q); /* P, Q 为前景图像中两个不重合  $3 \times 3$  的像素块 */
7)  $d \leftarrow \sqrt{(P_h - Q_h)^2 + (P_s - Q_s)^2 + (P_v - Q_v)^2}$ 
   /*  $d$  为像素块 P, Q 的 HSV 平均值的欧氏距离 */
8) IF( $d < t$ )
   /*  $t$  为阈值 */
9)   THEN IF( $P_h > 75$  AND  $P_h < 110$ )
   /*  $P_h$  为像素块 P 的 h 分量平均值 */
10)     THEN Ptype  $\leftarrow$  “鼠标”
   /* 记录指示符类型为“鼠标” */
11)      $p \leftarrow \text{position}(\text{front})$ 
   /* position(front) 为前景图像中心坐标 */
12)     ELSEIF( $P_h > 165$  AND  $P_h < 180$ )
13)     THEN Ptype  $\leftarrow$  “激光笔”

```

14) $p \leftarrow \text{position}(\text{front})$

为进一步提高指示符的检测正确率,提出一种数组运算法纠正部分误检情况。将连续有效帧中检测出来的指示符类型存放在一维数组中,其中检测为鼠标光标的标记为 A,检测为激光笔投影点的标记为 B,未检测到指示符的帧标记为 C。首先,取一维数组中任意连续的 10 个标记,若 C 的个数少于或等于 3,则将 C 置为其前一个非 C 的标记,同时位置也采用前一个非 C 标记的位置。然后,取一维数组中任意连续的三个标记 a_1, a_2, a_3 ,若 $a_1 = a_3$ 且 $a_1 \neq a_2$,则令 $a_2 = a_1$ 。实验结果表明,该方法可使平均检测准确率提高 2.4%。

例如:

原标记序列为

...ACABAAABAAABBCBBABBBAAA...

纠正后的序列为

...AAAAAAAAAAAAABBBBBBBBBBAAA...

上述指示符检测与识别方法计算复杂度低,且较低的误检和漏检率均不会对压缩后视频质量和大小造成很大影响,误检为指示符的情况只是导致某一帧错加了指示符,对压缩后的视频质量影响甚微,少数漏检情况会使部分包含指示符的有效帧采用通用有损压缩算法,对比特数的影响也较小。

3.4 基于关键帧和有效帧的编码算法

针对不同类型视频帧采取不同的处理方法,根据前文对关键帧的定义可知,在视频画面中停留时间较长的每页演示文稿、重要图表和图片都将被确定为关键帧,在编码时要尽可能地保证关键帧中文字及图片信息的清晰度,最大限度减少图像信息损失,因此针对关键帧采用 JPEG2000 无损压缩方法。对于幻灯片切换过程、过渡动画、教鞭指点等连续变化的有效帧,变化过程的清晰度要求并不高,采用 H.264 编码标准中基于帧间预测的有损压缩算法尽量提高其压缩比。针对包含鼠标或激光笔投影点运动的有效帧,对指示符运动模型参数进行编码,将指示符出现的连续有效帧的首帧视为关键帧,其余有效帧舍弃,同时反复调用该关键帧作为指示符运动模型的背景,依次调用指示符模型图片叠加在相应位置处。编码算法处理过程如图 3 所示。

对音频部分不做压缩处理,将原视频的音频单独分离出来,保存为 MP3 格式。最后,在完成对所有有效帧和关键帧的压缩编码后,将压缩编码后的视频码流封装到 MP4 格式文件中。

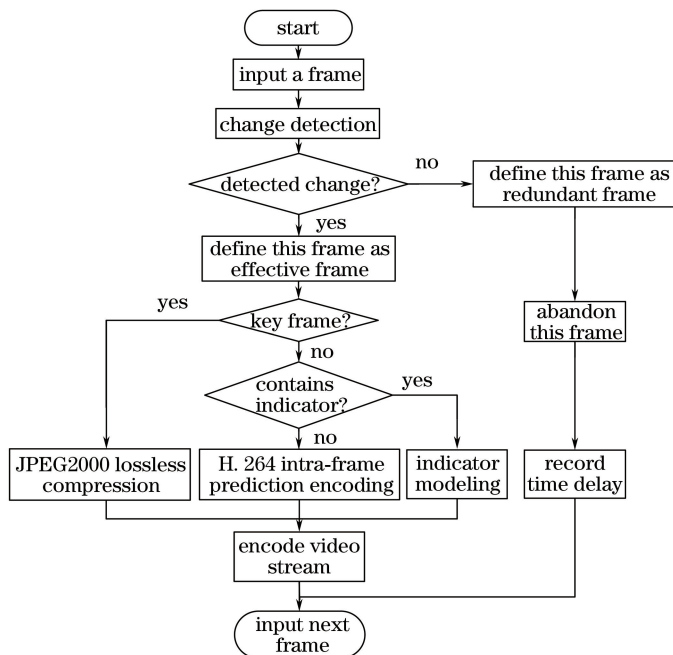


图 3 编码算法流程图

Fig. 3 Diagram of the encoding algorithm

3.5 压缩码流播放算法

由于压缩后的码流中只包含原视频中的关键帧和部分有效帧,剔除了冗余帧和包含指示符运动的有效帧,因此视频总帧数远少于原视频,导致在相同帧率下视频总时长大幅减小。若使用普通播放器进行播放将出现画面快速变化,音画不同步现象。为实现音画同步,重现包含指示符运动轨迹的完整教学视

频内容,基于开源计算机视觉库(OpenCV)与微软基础类库(MFC)针对关键帧编码及指示符运动模型设计了一种“LVPlayer”播放器进行视频回放。

播放器的输入信息包括分离出来的音频文件、压缩后的视频码流、指示符坐标和类型、指示符模型图片、视频帧类型标记以及视频帧率等信息。具体播放处理流程如图 4 所示。

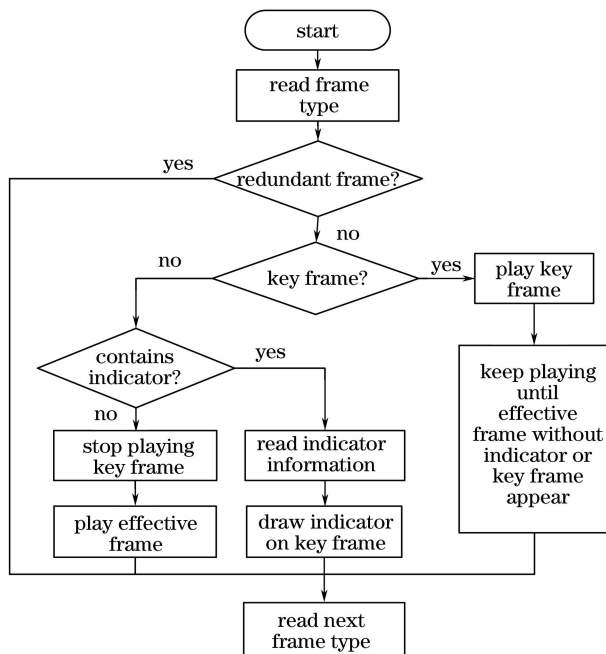


图 4 播放算法流程图

Fig. 4 Diagram of the paly back algorithm

4 实验结果与分析

实验基于 Windows10 环境下的 Visual Studio 2015 结合 OpenCV 的软件环境,计算机配置为 Intel(R)Core(TM)i5-4200H CPU 2.80 GHz,8 GB 内存。采用三脚架放置 JVC GC-P100BAC 数码摄像机拍摄教学视频,拍摄过程中镜头可发生小范围位移或变焦,选择 1920 pixel \times 1080 pixel 和 1280 pixel \times 720 pixel 两种分辨率拍摄测试视频数据,视频帧率选择 50 frame/s 和 25 frame/s 两种,视频保存格式为 MP4。数据采集的光照条件共有 4 种:白天晴天、白天阴天、傍晚、夜晚有光源(日光灯)。教学视频拍摄场景除了传统的 PPT 演示外,还包括演示 flash 动画、演示 word 文档、操作应用软件、播放视频资料等教学场景以及各场景的融合过程。取教学视频中包含投影区域的首帧作为初始背景图像建立混合高斯背景模型。为确定初始高斯分布个数,取 4 种光照条件的教学视频各 3000 帧进行实验测试,对比不同初始高斯分布个数的检测效果发现,当初始高斯分布个数设置为 5 时检测效果最好,高斯分布个数大于 5 时检测效果提升不明显而耗时增加明显。将用于训练混合高斯模型中背景的帧数设置为 500 帧,学习速率为 0.002,实验结果能够较好地保证背景更新频率的实时性,同时尽量避免漏检。为验证算法有效性,分别从变化检测、指示符检测与识别、视频压缩效果三个方面分析实验结果。

4.1 变化检测实验

为测试不同算法对教学视频中的变化检测效果,分别采用混合高斯背景模型、ViBe 算法以及

LOBSTER 算法和 SuBSENSE 算法进行对比实验。采用单线程运行评估算法的计算效率,结果如表 1 所示,LOBSTER 算法和 SuBSENSE 算法的计算耗时远高于另外两种算法,实时性较差。这是因为这两种算法均使用 LBSP 描述子进行像素级背景建模,增大了运算量,SuBSENSE 算法还结合了色彩信息,进一步增加了运算负担。对包含指示符快速运动、慢速运动或静止状态的 3228 帧教学视频进行对比测试,检测结果如表 2 所示。从表 2 中可看出,混合高斯背景模型对指示符的检测准确率最高,且漏检率也明显低于其余三种算法。这主要得益于自适应混合高斯背景模型的动态背景更新机制,而 ViBe 算法背景更新速度慢,当画面中的目标运动状态发生变化时,背景更新速度跟不上画面的变化,容易将已经静止的目标误检为前景或造成对快速运动小目标的漏检。LOBSTER 算法和 SuBSENSE 算法虽然比 ViBe 对鬼影的消除速度更快,但由于同样基于样本的相似性,易将从运动转入静止的目标误检为前景,特别是针对小的指示符运动情况检测效果较差,所以整体表现差于 GMM 算法。部分检测结果如图 1 所示,针对某段测试视频,从第 325 帧到第 354 帧,各算法均能检测出投影区域变化,但从第 354 帧到第 368 帧,ViBe 算法背景模型未更新,误将第 368 帧已经静止的部分检测为前景,LOBSTER 算法也将部分静止区域检测为前景;直到第 380 帧时,画面已保持静止,GMM 算法没有检测到运动目标,但 ViBe 算法仍检测到有运动目标存在,出现误检,LOBSTER 算法和 SuBSENSE 算法均出现较小面积的误检情况。

表 1 4 种算法处理速度对比实验结果

Table 1 Comparison of processing speed of the four algorithms

Image resolution / (pixel \times pixel)	Frames	Average time cost for per frame /ms			
		GMM	ViBe	LOBSTER	SuBSENSE
640 \times 320	7256	13.31	10.84	166.95	220.99
1280 \times 720	9138	16.29	14.13	658.01	683.09

表 2 4 种算法检测效果对比实验结果

Table 2 Comparison of detection results of the four algorithms

Type	Frames	False negatives rate /%				Precision rate /%			
		ViBe	GMM	LOBSTER	SuBSENSE	ViBe	GMM	LOBSTER	SuBSENSE
Laser point	1646	33.60	10.70	27.36	22.55	62.50	88.40	70.88	75.79
Mouse cursor	1582	67.50	9.20	52.32	48.82	23.50	89.00	46.65	50.09

4.2 指示符检测与识别

为测试所提出的指示符检测与识别算法的性能,随机选取 26000 帧进行测试检测效果,其中

9080 帧为鼠标光标运动帧,9065 帧为激光笔投影点运动帧,7855 帧为无指示符运动的有效帧。实验结果表明,该算法对鼠标光标检测准确率为 88.52%,

召回率为 95.88%， F -measure 值为 0.93；对激光笔投影点检测准确率为 87.55%，召回率为 95.78%， F -measure 值为 0.91。虽然检测准确率不足 90%，但指示符的误检和漏检对压缩后的视频质量不会产生太大影响。然而使用指示符建模的方法后，对包含指示符运动的连续有效帧的压缩比可提高为原来的百倍以上，对包含指示符运动的有效帧采用指示符模型及采用通用压缩后的比较如表 3 所示，其中 IM 为采用指示符建模的压缩方法，non-IM 为不采用指示符建模的标准压缩方法。由表 3 可知，采用

表 3 采用指示符模型与采用 H.264 的压缩效果对比

Table 3 Comparison between compressing algorithms with IM and H.264

Frame per second	Image resolution / (pixel×pixel)	Frames contain indicator	Size of frames contain indicator /MB	Size after compression		PSNR /dB		Decrease on per frame /kB
				Non-IM/MB	IM/kB	Non-IM	IM	
50	1920×1080	5049	433	36.81	132	37	39	7.40
50	1920×1080	1652	141	11.51	53	37	39	7.10
50	1920×1080	934	80	7.56	42	37	39	8.24
50	1280×720	1329	58	4.45	44	37	39	3.40
50	1280×720	1148	56	3.35	39	37	39	2.95
50	1280×720	605	30	1.92	28	37	39	3.20
25	1920×1080	3843	354	32.64	53.8	37	39	8.68
25	1920×1080	1215	104	10.45	27.5	37	39	8.79
25	1920×1080	632	54	5.47	19	37	39	8.84
25	1280×720	883	54	3.21	37.7	37	39	3.70
25	1280×720	507	26	1.70	27.82	37	39	3.42
25	1280×720	348	15	1.04	16.77	37	39	3.04

4.3 视频压缩效果

为测试本文算法对教学视频的压缩性能，采用 PSNR^[20] 衡量压缩后的视频画质。PSNR 是原图像与被处理图像之间的均方误差相对于 $(2^n - 1)^2$ 的对数值（即信号最大值的平方，其中 n 是每个采样值的比特数），单位是 dB。采用本文算法、H.264 以及 MPEG-4 三种方法对 50 段教学视频编码的率失真(RD)曲线如图 5 所示。由图 5 可知，本文方法在不同码率下的性能均远超 H.264 和 MPEG-4。在相同 PSNR 值的条件下，本文算法压缩后的视频码率相比 H.264 平均降低了 88.12%，相比 MPEG-4 平均降低了 90.3%。H.264 和 MPEG-4 算法在码率低于 400 kb/s 时画面的 PSNR 值均达不到 30 dB，画面中有效信息丢失较多，而本文算法压缩后的视频在码率大于 800 kb/s 时画面 PSNR 值便可达到 38 dB，即使在码率为 300 kb/s 的情况下压缩后画面 PSNR 值也能达到 36 dB 以上。这主要是由于本文算法针对关键帧采用无损压缩，最大限度

指示符模型后对原视频的压缩比相比通用 H.264 压缩算法提高了百倍以上，同时画面也拥有更高的 PSNR 值，压缩比的大幅提高得益于采用指示符模板图像和位置信息代替删除的连续有效帧，较高的画面质量则是由于调用了采用无损压缩的关键帧作为背景。另外，连续的包含指示符的有效帧数越多、图像帧的分辨率越高，采用基于指示符模型的方法压缩比越大，这主要因为帧数的增加对于基于指示符模型的压缩方法来说只是增加了位置坐标信息，而分辨率的提高对小面积的指示符模板图像影响甚微。

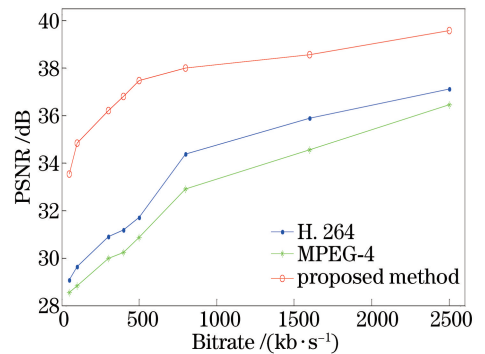


图 5 率失真曲线

Fig. 5 Rate distortion curves

地保证了关键信息的清晰度，而另外两种方法采用统一的有损压缩算法，即使在码率达到 2500 kb/s 时 PSNR 值仍不能达到 38 dB，远低于本文算法的 39.57 dB。同时，指示符运动模型的引入以及去除冗余帧和部分有效帧的方法在基本不影响画面质量的同时，大幅降低了码率。部分压缩后的截图如图 6 所示，可以看出，MPEG-4 及 H.264 算法在码率

为 500 kb/s 时,部分图像信息丢失严重,导致画面失真,字迹模糊不清,甚至出现“马赛克”现象,关键的图片及文字无法识别的情况将严重影响教学视频发挥其应有的作用。本文算法分别对关键帧和部分有效帧编码,在较低的数码率下依然能保持较高的视频质量,特别是对持续时间较长的每页演示文稿或关键图片信息采取无损编码策略,最大限度地保证了关键信息的画面质量。

为评估提出的方法在不同场景下的综合性能,录制了不同环境下、不同场景、不同分辨率的教学视频共 49.25 G,并在同一 PSNR 值下与 H.264 标准进行了比较,对比实验结果如表 4 所示。对于 PPT 演示和 word 文档阅读等画面变化不频繁的场景,在 PSNR 值为 39 时,本文算法压缩比超过 40:1;而

对于包含有大量指示符运动的软件操作场景,压缩比也能达到 30:1;在综合场景中,本文算法取得的压缩比超过 20:1,均远超采用 H.264 标准取得的压缩比。这是由于这些场景中均存在大量冗余帧或包含指示符运动的有效帧。但对于视频资料和 FLASH 动画演示等画面变化较频繁,少有单纯的指示符运动场景时,教学视频中大部分为一般有效帧,压缩过程大多由 H.264 压缩编码完成,因而取得的压缩比与 H.264 标准相当。

对分辨率为 1280 pixel×720 pixel 的视频帧,本文算法处理每帧平均耗时 0.015 s(每秒可处理 66 帧),对分辨率为 1920 pixel×1080 pixel 的视频帧,本文算法处理每帧平均耗时 0.028 s(每秒可处理 35 帧),基本可达到实时处理的要求。

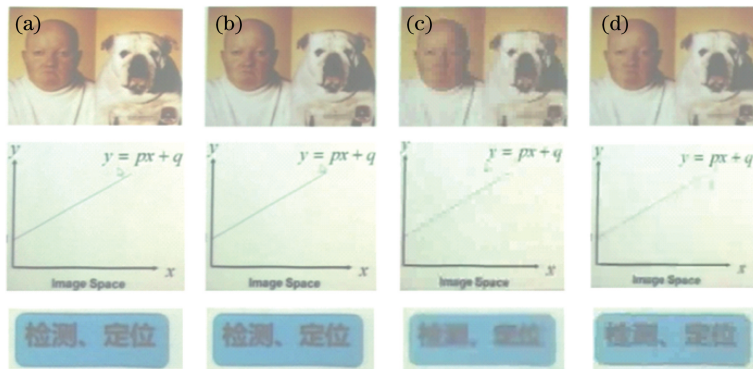


图 6 不同算法在码率为 500 kb/s 时的压缩效果示例图。(a)原视频帧;(b)本文算法;(c) MPEG-4;(d) H.264

Fig. 6 Compression graphs of different algorithms at the bitrate of 500 kb/s. (a) Original frame; (b) proposed algorithm; (c) MPEG-4; (d) H.264

表 4 对不同场景教学视频压缩效果对比

Table 4 Comparison of compression ratio in different scenes

Type	Image resolution / (pixel×pixel)	PSNR	Size /GB	Size after compression		Compression ratio	
				H.264 /GB	Proposed method /GB	H.264	Proposed method
Word	1920×1080	39	5.12	0.648	0.124	8.1:1	42:1
	1280×720	39	1.79	0.230	0.045	7.9:1	40:1
Video	1920×1080	39	6.54	1.88	1.860	3.4:1	3.5:1
	1280×720	39	2.39	1.03	0.874	2.8:1	2.7:1
Software Operation	1920×1080	39	7.03	0.999	0.212	7.2:1	34:1
	1280×720	39	2.01	0.294	0.063	7:1	32:1
FLASH	1920×1080	39	5.22	1.36	1.350	3.8:1	3.8:1
	1280×720	39	2.22	0.757	0.721	3:1	3.1:1
PPT	1920×1080	39	6.72	782	0.092	8.8:1	75:1
	1280×720	39	2.55	0.302	0.038	8.6:1	69:1
Combination	1920×1080	39	5.71	1.24	0.268	4.6:1	22:1
	1280×720	39	1.95	0.453	0.094	4.4:1	21:1

5 结 论

针对教学视频的特点,提出一种基于关键帧检测和指示符运动建模的智能教学视频压缩算法。利用变化检测将视频帧分成有效帧、冗余帧及关键帧三种类型,针对不同类型的视频帧采用不同的压缩策略。针对包含指示符的有效帧,提出一种基于指示符运动模型的编码方法,提出一种指示符检测及识别算法,利用指示符位置将指示符模板图像再现在关键帧上,提高压缩比的同时保证了图像的清晰度。设计了专门的播放器播放包含延时信息、位置坐标、指示符类型、帧类型、指示符模板等信息的压缩码流。实验结果证明,对关键帧采用无损压缩最大限度地保证了关键信息的画面质量;去除冗余帧及包含指示符的有效帧可大幅提高压缩比;所提出的指示符运动模型是去除包含指示符的有效帧的完美替代,可在保证再现指点教学过程的同时再现更清晰的背景图像。对有效区域的分割不仅能降低码率,还能保护授课人的隐私。压缩后的码流可通过专门设计的LVPlayer播放器播放。压缩性能对比实验结果表明,本文算法在相同PSNR值下可大幅降低码率,且编解码过程满足实时要求,为提高在线教学视频的制作与传输效率提供了高效的解决方案,同时可为数码相机提供专门的教学视频录制功能,可大幅降低教学视频的录制对内存空间的占用。今后的工作将致力于提高变化检测算法的效率及指示符检测与识别算法的精度,同时研究针对更多更复杂应用场景的压缩编码及视频制作解决方案。

参 考 文 献

- [1] Abomhara M, Khalifa O O, Zakaria O, *et al.* Video compression techniques: an overview[J]. Journal of Applied Sciences, 2010, 10(16): 1834-1840.
- [2] Uhrina M, Frnda J, Sevcik L, *et al.* Impact of H.264/AVC and H.265/HEVC compression standards on the video quality for 4K resolution[J]. Advances in Electrical and Electronic Engineering, 2014, 12(4): 905-908.
- [3] Bauer S, Kneip J, Mlasko T, *et al.* The MPEG-4 multimedia coding standard: algorithms, architectures and applications[J]. Journal of VLSI signal processing systems for signal, image and video technology, 1999, 23(1): 7-26.
- [4] Kwon S K, Tamhankar A, Rao K R. Overview of H.264/MPEG-4 part 10 [J]. Journal of Visual Communication and Image Representation, 2006, 17(2): 186-216.
- [5] Schwarz, Heiko, Wiegand, Thomas. An improved MPEG-4 coder using lagrangian coder control[C] // Proceedings of 2001 Video Coding Experts Group Conference on New Tools for Video Compression Technology, 2001: 1-8.
- [6] Manjanaik N, Parameshachari B D, Hanumantappa S N, *et al.* Intra frame coding in advanced video coding standard (H.264) to obtain consistent PSNR and reduce bit rate for diagonal down left mode using Gaussian pulse[J]. IOP Conference Series: Materials Science and Engineering, 2017, 225(1): 012209.
- [7] Sharabayko M P, Markov N G. Fast search for intra prediction mode in H.265/HEVC video compression [J]. Key Engineering Materials, 2016, 685: 897-901.
- [8] Purcell D D, Hess C P, Durack J C, *et al.* Recording, editing, archiving, and distributing radiology lectures: a streamlined approach[J]. Radio Graphics, 2007, 27(6): 1839-1844.
- [9] Mittal A, Gupta S, Jain S, *et al.* Content-based adaptive compression of educational videos using phase correlation techniques [J]. Multimedia Systems, 2006, 11(3): 249-259.
- [10] Gu J. The study of compression technology of multimedia educational information based on wavelet [D]. Xi'an: Xidian University, 2009: 25-40.
古佳. 基于小波变换的多媒体教学资源压缩编码方法研究[D]. 西安: 西安电子科技大学, 2009: 25-40.
- [11] Barnich O, Droogenbroeck M V. ViBe: a universal background subtraction algorithm for video sequences [J]. IEEE Transactions on Image Processing, 2011, 20(6): 1709-1724.
- [12] Chen H Y, Qie L Z, Yang D D, *et al.* Visual background extraction algorithm based on superpixel information feedback[J]. Acta Optica Sinica, 2017, 37(7): 0715001.
陈海永, 鄯丽忠, 杨德东, 等. 基于超像素信息反馈的视觉背景提取算法[J]. 光学学报, 2017, 37(7): 0715001.
- [13] Mo S W, Deng X P, Wang S, *et al.* Moving object detection algorithm based on improved visual background extractor[J]. Acta Optica Sinica, 2016, 36(6): 0615001.
莫邵文, 邓新蒲, 王帅, 等. 基于改进视觉背景提取的运动目标检测算法[J]. 光学学报, 2016, 36(6): 0615001.

- [14] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, 2: 246-252.
- [15] Wang H L, Wang J Q, Ding H F, *et al.* Moving target detection based on the improved gaussian mixture model background difference method [C] // Advanced Materials Research, 2012, 482: 569-574.
- [16] Chen Q, Sheng H X, Zhang Z, *et al.* Moving object detection under infrared light mutation [J]. Laser & Optoelectronics Progress, 2016, 53(11): 111005.
陈强, 盛惠兴, 张卓, 等. 红外光照突变下的运动目标检测 [J]. 激光与光电子学进展, 2016, 53(11): 111005.
- [17] St-Charles P L, Bilodeau G A. Improving background subtraction using local binary similarity patterns [C] // IEEE Winter Conference on Applications of Computer Vision (WACV), 2014: 509-515.
- [18] St-Charles P L, Bilodeau G A, Bergevin R. SuBSENSE: a universal change detection method with local adaptive sensitivity [J]. IEEE Transactions on Image Processing, 2015, 24: 359-373.
- [19] Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction [C] // Proceedings of the 17th International Conference on Pattern Recognition, 2004, 2: 28-31.
- [20] Seshadrinathan K, Soundararajan R, Bovik A C, *et al.* Study of subjective and objective quality assessment of video [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1427-1441.