

文章编号:1007-2780(2019)05-0521-09

基于 SURF 和 SIFT 特征的视频镜头分割算法

张昊骅¹, 朱晓龙¹, 胡新洲¹, 任洪娥^{1,2*}

(1. 东北林业大学 信息与计算机工程学院, 黑龙江 哈尔滨 150040;

2. 黑龙江省林业智能装备工程研究中心, 黑龙江 哈尔滨 150040)

摘要:在视频镜头分割处理中, SIFT(Scale-invariant feature transform) 特征由于其具有尺度、旋转不变性等诸多优点而被广泛应用, 但是 SIFT 特征提取复杂, 计算量大, 导致程序效率低下, 很难满足实时性的要求。本文提出了一种基于 SURF(Speeded Up Robust Features)和 SIFT 特征的视频镜头分割算法。首先对相关帧进行颜色直方图特征提取, 然后通过一个自适应阈值判断是否为候选帧, 如果不是, 则融合 SURF 特征再进行复检。最后筛选出的候选帧再通过 SIFT 算法进行复选, 去除误检帧, 有效解决了 SIFT 特征低效率的问题。实验结果表明, 对于切变镜头查全率为 95%, 查准率为 93%, 对于渐变镜头查全率为 90%, 查准率为 76%, 平均单帧运行时间为 0.26 s, 本算法与 SIFT、SURF 等单一特征的算法相比, 在高效性和准确性之间取得了平衡, 并且它是鲁棒的。

关键词:镜头分割; SIFT; SURF; 颜色直方图

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.3788/YJYXS20193405.0521

Shot segmentation technology based on SURF features and SIFT features

ZHANG Hao-su¹, ZHU Xiao-long¹, HU Xin-zhou¹, REN Hong-e^{1,2*}

(1. College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China;

2. Heilongjiang Forestry Intelligent Equipment Engineering Research Center, Harbin 150040, China)

Abstract: In the field of video shot segmentation processing, SIFT features are widely used due to their many advantages such as scale and rotation invariance. However, SIFT feature extraction is complex and computationally intensive, resulting in low program efficiency and hard to achieve the requirement of real-time. In this case, an algorithm based on SURF features and SIFT features is proposed in this paper. The algorithm first performs color histogram feature extraction on the frame, determines whether it is a candidate frame through an adaptive threshold, and if not, fuses the SURF feature for re-detection. The candidate frames that are filtered out are then checked by the SIFT algo-

收稿日期: 2018-12-25; **修订日期:** 2019-03-12.

基金项目: 中央高校基本科研业务费专项资金资助项目(No.2572017PZ10, No.2572018BH09)

Supported by Central University Basic Research Funding For Specialized Projects Funded (No. 2572017PZ10, No.2572018BH09)

* 通信联系人, E-mail: nefu_rhe@163.com

rithm. This can effectively solve the problem of SIFT feature inefficiency. Experimental results show that the recall rate is 95% for the shear shots, 93% for the precision, 90% for the gradual shots, 76% for the precision and 0.26 s for the average single frame running time and compared with SIFT, SURF and other algorithms with a single feature, this algorithm achieves a balance between efficiency and accuracy, and has better robustness.

Key words: shot split ;sift; surf; color histogram

1 引言

视频镜头分割是视频处理中的第一步,它的任务是准确找出镜头之间的边界,这对视频的后续处理如关键帧提取等起着至关重要的作用。视频镜头边界主要分为两种^[1],一种是切变镜头,指的是前一个镜头直接切入到下一个镜头,镜头之间变化明显,也比较容易检测。另一种是渐变镜头,指的是前一个镜头通过几帧甚至几十帧缓慢过度到下一个镜头,镜头之间变化缓慢,检测比较困难。近些年来,为了解决镜头边界检测的问题,很多种用于镜头边界检测的方法被提出,并取得了显著效果。Kikukawa 等^[2]提出了像素帧差法,它采用前后相邻两帧的对应像素的色度并比较其差值来查找镜头边界。Zhang 等^[3]提出了直方图帧差法,采用度量临近帧之间直方图的差异来查找镜头的边界。杨瑞琴等^[4]提取 Hu 不变矩特征进行初检选出候选边界,然后利用三维颜色直方图进行复检。于俊清等^[5]通过构造回放因子特征来改进情感激励模型从而实现了镜头边界检测。李小琛等^[6]利用 SIFT 特征融合 CSIFT 颜色不变量特征进行镜头边界检测,有效地弥补了 SIFT 算法丢失色彩信息的缺陷。蔡铁珩等^[7]提出一种融合了色彩特征和改进的 SURF 特征的镜头分割算法,通过减少匹配点数量,加快了算法运算速度,通过使用自适应阈值,以及融入颜色特征等方式,提高了算法准确率。除此之外,还有很多类似的方法。简而言之,这些方法从不同的角度实现了对镜头识别分割这一过程。然而,在上述算法中,有些算法仅仅利用了全局特征,虽然速度较快但是准确率较差。有些算法虽然融合了局部特征和全局特征,但是融合方式不紧密导致可能会出现漏检。

SIFT 特征是一种经典的特征点提取描述算

法。并且它已被大量应用于图像、视频分析等领域中,它具有对尺度变换、旋转变换、仿射变换、视角变换、噪声的不变性等诸多优点,然而由于其特征点提取过程比较复杂,导致效率低下,不适用于数据量大和具有实时性要求的场合,而且由于算法使用灰度图像进行特征点的检测,缺乏图像色彩信息,所以对于视频检测来说容易造成误检。

相比之下,SURF 特征比 SIFT 特征的速度快很多,并且对亮度等变化不太敏感。所以,本文针对以上文献中存在的高效性与准确性不平衡的问题,从 SIFT 特征效率较低的特点入手,提出了基于 SURF 和 SIFT 特征的镜头分割技术。首先判断两帧的颜色特征距离,如果大于设定的自适应阈值,则加入候选帧,否则提取出 SURF 特征并与颜色特征进行特征融合,然后使用滑动窗口技术计算当前帧与后续帧的特征距离并使用阈值筛选出候选切变镜头和渐变镜头。最后采用 SIFT 特征用来进行复选,从而去掉误检镜头。

2 基于多特征融合的镜头分割算法

由于 SIFT 特征效率低下,显然不能直接使用 SIFT 进行镜头分割,所以在镜头边界帧初选阶段我们引入 SURF 特征来替代 SIFT 特征以实现高效分割。

2.1 SURF 特征点提取

SURF 特征即加速稳健特征,是一种局部特征点的提取和描述算法^[8],它基于 SIFT 特征并改良了其特征点提取和描述方式,所以 SURF 也具有 SIFT 特征的对尺度、旋转等不变性的特性,而且 SURF 由于采用了 Harr 小波特征和积分图像的方法,运行效率相较 SIFT 有了大幅度的提升。

SURF 特征点的提取过程如下:

(1) 假设图像用函数 $f(x, y)$ 来表示。首先利用黑塞矩阵获取特征点, 则图像中相应点的黑塞矩阵为:

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}, \quad (1)$$

从而有判别式:

$$\det(H) = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2, \quad (2)$$

经过高斯滤波的对应矩阵为:

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}, \quad (3)$$

在计算黑塞矩阵之前, 应该使用二阶标准高斯函数进行滤波, 这是因为它可以使求得的特征点具备尺度无关特性。用 $I(x, y)$ 代替点 (x, y) , $L(x, t)$ 为高斯函数 $G(t)$ 在点 (x, y) 处卷积的结果:

$$L(x, t) = G(t) \cdot I(x, y), \quad (4)$$

其中, 高斯核 $G(t)$ 为:

$$G(t) = \frac{\partial^2 g(t)}{\partial x^2}. \quad (5)$$

为了加快计算速度, SURF 算法改变了以往先进行高斯滤波, 然后求二阶导数, 进而求取黑塞矩阵的方式, 转而采用了盒式滤波来近似高斯滤波。设在经过盒式滤波之后, 用 P_{xx} 、 P_{yy} 、 P_{xy} 来代表 L_{xx} 、 L_{yy} 、 L_{xy} , 所以有判别式:

$$\det(H) = P_{xx} P_{yy} - (0.9 P_{xy})^2, \quad (6)$$

其中, 0.9 为经验值。根据判定结果的符号, 若 $\det(H) > 0$, 则此点为极值点。

(2) 生成尺度空间, 如图 1 所示。

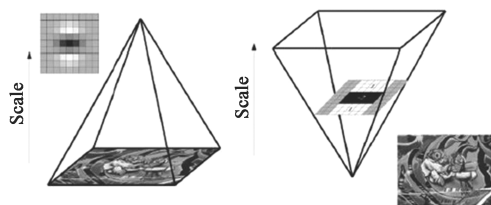


图1 尺度空间的建立

Fig.1 Building of a scale space

在 SURF 特征中, 图像尺度是不变的, 通过对图像进行不同尺度的盒式滤波来生成尺度空间

的每一层。

(3) 利用非最大抑制来定位特征点, 如图 2 所示。

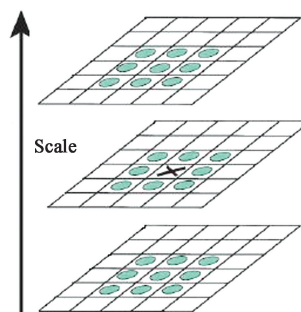


图2 定位特征点

Fig.2 Location feature points

经过黑塞矩阵处理过的每个特征点与相邻的 26 个点进行大小比较, 如果最大或最小, 则保留下来做备选点。然后, 采用 3 维线性插值法进一步得到亚像素级的特征点, 同时也去掉一些较弱或错误的特征点, 通过控制极值大小来控制特征点数量, 最终仅留下一些较强的特征点。

(4) 为特征点选取主方向, 如图 3 所示。

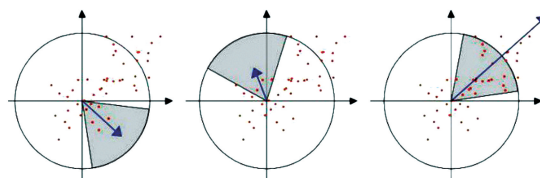


图3 选取特征点主方向

Fig.3 Selection of the main direction of feature points

SURF 特征统计了以特征点为中心, 在固定半径大小的圆形区域内以一定间隔转动的扇形内全部点的 Harr 小波特征, 并将取得最大值的扇形方向设定为主方向。这样, 通过计算每一个特征点, 得到每一个特征点的主方向。

(5) 构成特征点的描述算子, 如图 4 所示。在特征点邻域选取一个矩形框并将其分成 4×4 个子区域, 然后统计每个区域中共 25 个像素的 Haar 小波特征来得到描述子。

2.2 改进的 SURF 特征匹配及特征融合算法

特征点匹配一般采用 FLANN 算法。FLANN 算法基于 KD 树, 可以根据数据集的分布特点, 来推荐索引类型和检索参数^[9]。该算法首先根据得到的特征点, 计算图像中某一特征点

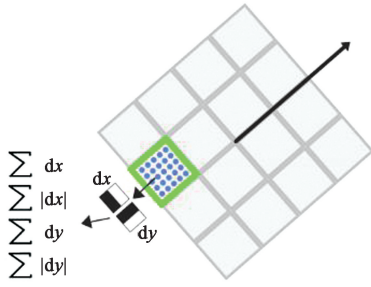


图 4 构成特征点描述子

Fig.4 Constitution of the descriptor of feature points

x 与另一图像中特征点 y 的欧式距离,接着采用欧式距离比值法进行匹配^[10],最后筛选经过配对的特征点,仅保留符合要求的点。

为了提升 SURF 特征的准确性,本文改进了特征点的匹配方式。传统的匹配方式通常为使用固定阈值或者设定阈值为最小匹配距离 $D_{\min}$ 的 N 倍来进行匹配,但是,如果使用固定阈值将可能无法满足不同类型视频匹配的要求,如果设定阈值为最小匹配距离的 N 倍,在某些情况下,最小匹配距离可能为 0,而其他匹配点的匹配距离可能并不为 0,从而导致误检。为解决这个问题,使用最大匹配距离 $D_{\max}$ 和颜色特征距离 D_c 并采用线性结合权重的自适应阈值的方法,如式 7 所示。

$$D_t = \rho \cdot D + (1 - \rho) \cdot D_c, \quad (7)$$

其中: D_c 为颜色特征距离, D 为特征点距离, D_t 为融合后的距离。

对于两幅图像而言,式(7)中融合了相同的颜色特征信息,只有 D 是可变的,因此经过融合后的最大距离 $D_{\max}$ 与最小距离 $D_{\min}$ 的比值应该与 $D_{\max}$ 和 $D_{\min}$ 的比值相等。设 $D_{\min} = \rho \cdot D_{\min} + (1 - \rho) \cdot D_c$, 所以有式(8):

$$\begin{aligned} D_{\max} &= \frac{D_{\max}}{D_{\min}} \cdot D_{\min} = \\ &= \frac{D_{\max}}{D_{\min}} \cdot (\rho \cdot D_{\min} + (1 - \rho) \cdot D_c) = \\ &= \rho \cdot D_{\max} + (1 - \rho) \cdot \frac{D_{\max}}{D_{\min}} \cdot D_c, \quad (8) \end{aligned}$$

令 $D_{\max}/D_{\min} = M$, 设优秀匹配点阈值 D_m 为融合后最大距离 $D_{\max}$ 的 K 倍 ($K \in (0, 1)$), 如式(9)所示。

$$D_m = K \cdot (\rho \cdot D_{\max} + (1 - \rho) \cdot M \cdot D_c). \quad (9)$$

经过大量实验证明,在 SURF 检测器阈值为

3 000, K 值为 0.4 时,这种方法尤其对切变镜头有较高的匹配率。在本文中的 SIFT 特征匹配方式也是按照相似的方式进行的。

对于特征融合算法,传统的特征融合方式包括串行融合和并行融合两种^[11]。串行融合是指新的特征向量是由两组特征向量通过串行连接成的。其优点是构造方法简单,几乎不需要进行运算,缺点是特征向量 γ 的维数可能急剧增加。并行融合是利用复向量,将特征向量在复向量空间进行融合。如果两个向量维数不同,则通过补 0 来对齐。其优点是对比串行融合方式,有效地降低了特征向量的维数,缺点是计算量可能较大。

为了尽可能减小计算量,避免维度剧烈增加,本文采用式 7 的方式通过赋予每种特征的特征向量之间的距离以相应的权重,再求其和进而表示特征融合后两图之间的距离。通过此公式融入 FLANN 匹配算法来获得优秀的匹配点,实验证明,权重 ρ 在取值为 0.6 时准确率最高。

2.3 改进的颜色特征的检测算法

SURF 特征是一种局部特征,它反映了图像的部份像素间的关系,然而仅仅使用局部特征还不能对图像精确地描述。全局特征可以表达出整张图像的信息,而且它的速度比局部特征要快,所以为了增加速度和准确性,本文使用了全局特征及全局特征融合局部特征的方式进行镜头检测。

颜色直方图是一种常用的全局特征,主要因为其算法简单,对匹配图像的旋转、变形、模糊变换具有不变性的优势,并且对噪声、镜头位置、分辨率等也并不敏感,所以在图像、视频处理系统中被广泛应用。而在这其中 HSV 空间是直方图最常用的颜色空间,它包含色相 H 、饱和度 S 和亮度 V 3 个分量,同时,它也是最适合人类的视觉系统。

计算颜色直方图首先需要量化颜色。颜色量化即是将颜色空间分为若干区间(bin),然后计算处于每个区间的像素个数即可得到颜色直方图。

为了减少计算复杂度,首先将直方图分为 64 个 bin, H 范围为 (0, 44), (45, 89), (90, 134), (135, 179), S 范围为 (0, 63), (64, 127), (128, 191), (191, 255), V 范围为 (0, 63), (64, 127), (128, 191), (191, 255), 并将其归一化,即将每个 bin 值除以图像像素数量。最后得到颜色特征向量:

$$V = \left\{ \frac{M_1}{S}, \frac{M_2}{S}, \frac{M_3}{S}, \frac{M_4}{S} \dots \right\}, \quad (10)$$

其中: S 为图像像素数量, M_n 代表 bin 元素数量。

为了后续的特征融合阶段,本文中采用欧式距离作为 HSV 直方图的距离对比方式:

$$d(M, N) = \sqrt{(M_1 - N_1)^2 + \dots + (M_i - N_i)^2} \\ = \sqrt{\sum_{i=1}^n (M_i - N_i)^2}, \quad (11)$$

这里 M, N 是两帧图像的 HSV 颜色直方图 M_i, N_i 是颜色直方图第 i 个 bin 的元素个数。

算法主要是通过比较前后两帧颜色直方图距离与某一阈值的大小来寻找镜头边界。传统的判别方法使用固定阈值,然而固定阈值对于不同类型的视频来说阈值设定不尽相同,且需要大量的实验来确定,很容易发生漏检、误检现象,所以本文通过使用自适应阈值来判断切变镜头。对于渐变镜头,其变化过程多样,单一颜色特征检测渐变镜头较难判定,容易造成误检。所以并未采用单一颜色特征来判定渐变镜头。本节算法过程为:

首先提取视频的相邻两帧图像,然后分别把这两帧图像从 RGB 空间转换成 HSV 空间并将进行颜色量化,接着分别计算这两帧图像的颜色直方图,最后计算它们颜色特征距离,如果颜色距离大于自适应阈值 T ,则判定切变镜头。根据实验得到的经验,对于一般情形,前后两帧的信息具有一定连续性,颜色变化程度较小,所以相应颜色特征距离不会超过其与零向量的欧式距离。所以,将自适应阈值选择为两帧中第一帧的颜色特征向量与零向量欧氏距离的 0.8 倍(式(12))。经过实验检验,本节算法对检测切变镜头有较高的准确性。

$$T = 0.8 \times \sqrt{(M_1 - 0)^2 + \dots + (M_i - 0)^2}, \quad (12)$$

2.4 基于颜色和 SURF 特征的镜头分割算法

本节算法过程为:首先提取颜色特征,使用上述基于颜色特征的检测算法判断当前帧是否为切变镜头,如果是则进入下一帧,否则提取 SURF 特征,然后将改进的融合模式用于特征融合,并筛选优秀的匹配点,接着计算相邻两帧之间的匹配度,如果帧间匹配度小于阈值 T_1 ,则将其判定为切变镜头。如果匹配度大于阈值 T_2 ,则判定为无变化。如果匹配度处于 T_1 与 T_2 之间,则有可能存在渐变镜头。针对切变和渐变两种镜头边界的差别,我们使用了滑动窗口的检测手段。维持一

个窗口,设定它的最初大小 w 为 0,它的起始位置为当前帧 c ,并不断向右侧拓展,并不断比较帧 F_c 和 F_{c-w} 的匹配度,如果两帧匹配度小于阈值 T_1 ,那么将其认定为渐变镜头,并加入候选帧,然后恢复最初设定。若在此窗口扩展中检测到相邻两帧存在切变镜头或无变化,则直接初始化设定。重复如此过程,直到检测完所有帧。经过大量实验,确定阈值 $T_1=0.3, T_2=0.8$ 时准确率最高。算法流程如图 5 所示。

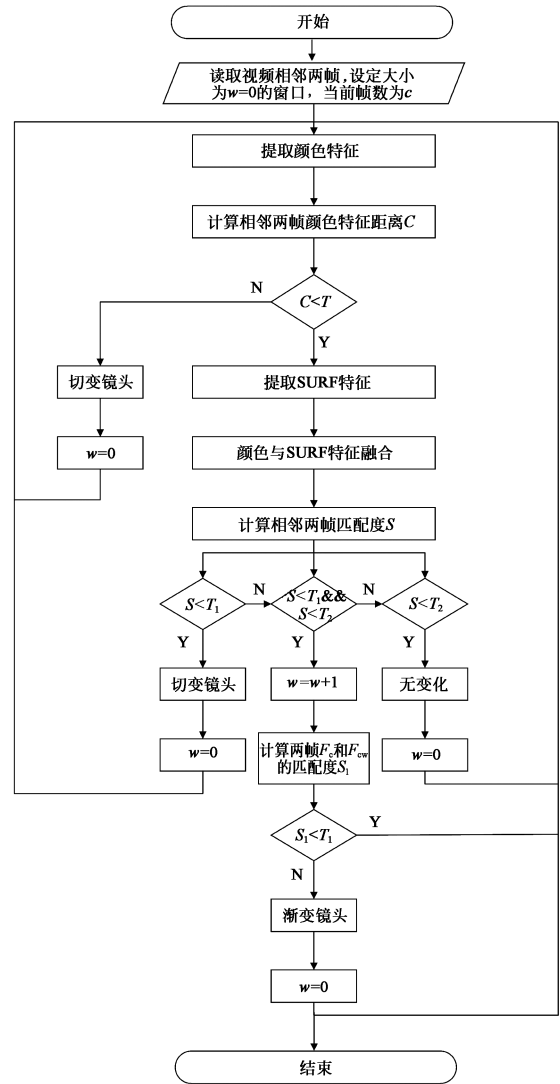


图 5 基于颜色特征和 SURF 特征的镜头分割算法流程图

Fig.5 Flowchart of shot segmentation algorithm based on color feature and SURF feature

2.5 SIFT 特征匹配

虽然颜色特征结合 SURF 特征可以满足要求,但是根据 Juan 等人对 SURF 和 SIFT 等特征

的实验比较来看, SIFT 特征可以获得数量更多的特征点, 而且 SURF 特征对于尺度和旋转变换等方面相比检测效果略差^[12], 综合来看 SIFT 特征的检测效果要强于 SURF 特征, 此外, 在渐变镜头检测中由于滑动窗口未限定尺寸, 有可能产生误检, 为了提高算法准确性, 我们通过构造 SIFT 特征来去除一些由于尺度、旋转等方面误检的候选帧, 如图 6 所示。

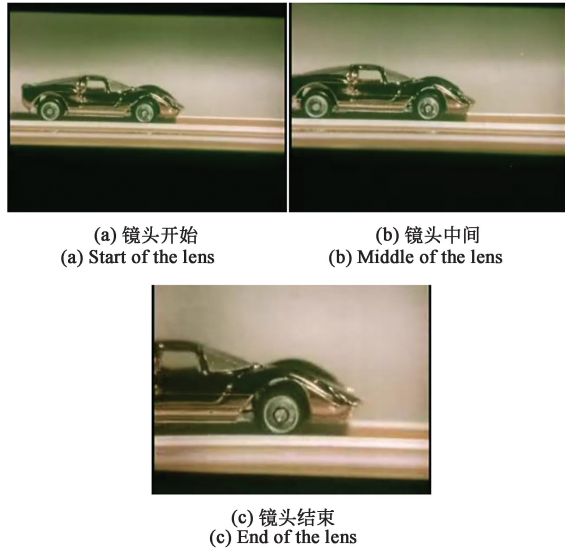


图 6 镜头放大造成的误检

Fig.6 Misdetection caused by lens enlargement

SIFT 特征的全称是尺度不变的特征变换^[13], 它具有许多优点, 例如对旋转、尺度、视角变化、仿射、噪声等变换的不变性。SIFT 特征与 SURF 特征的主要区别在于:

(1) SIFT 是在高斯差分图像金字塔中对每一层图像进行高斯滤波求取图像差来求取特征点。

(2) SIFT 算法通过计算梯度直方图, 然后查找其最大方向或大于 80% 处的最大 bin 方向来确定相应特征点的主方向。

(3) 在确定特征描述子方面, SIFT 算法采用了在 16×16 的邻域内分成 4×4 的区域, 每个区域取 8 个方向, 即特征向量维数 $= 4 \times 4 \times 8 = 128$ 维。

SIFT 特征复选的算法如下: 对候选帧首先计算其前一帧和后一帧的 SIFT 距离, 如果匹配度小于阈值 T_3 , 则判定为切变镜头, 反之计算其前后 4 帧的 SIFT 距离, 如果匹配度小于阈值 T_4 , 则判定为渐变镜头, 若匹配度大于 T_4 , 则说

明既不是切变镜头也不是渐变镜头, 去除即可。经过大量实验, 确定阈值 $T_3 = 0.3$, $T_4 = 0.5$ 时准确率最高。

3 实验流程和结果对比评价

3.1 实验环境

本实验硬件环境是 CPU 为 Intel Core I7 6700, 内存为 4 GB, 软件环境是操作系统为 Windows 10 64 位, 编程环境为 Visual Studio.NET 2015, OpenCV 2.4.13。测试视频数量为 70 个, 分辨率均为 320×240 , 码率均为 326 kb/s。

3.2 实验数据

本文视频数据源来自 OPENVIDEO 视频库, OPENVIDEO 视频库包含历史、纪录片、教育、航天等多类视频。该视频包括了切变镜头、渐变镜头, 其中渐变镜头还包含了多种特效镜头。

3.3 实验流程

本文算法实施流程如下:

(1) 读取视频文件。

(2) 读取成功进入下一步, 失败则退出。

(3) 读取下一帧, 计算其 HSV 颜色特征, 通过是否大于自适应阈值判断是否为切变镜头, 若是, 则将此帧加入候选帧, 直接进入下一帧, 否则进入下一步。

(4) 结合 SURF 特征与 HSV 颜色特征, 通过设定滑动窗口来判断此帧是否为切变镜头或渐变镜头, 如果是, 则加入候选帧。然后回到第二步。

(5) 使用 SIFT 特征, 通过对其前后帧及前后四帧的特征距离进行判断, 筛选候选帧, 剔除伪渐变镜头和伪切变镜头。

(6) 保存并输出分割结果。

3.4 实验结果及评价

本文使用查准率和查全率作为评价指标, 如式(13)、式(14)所示。

$$\text{查全率 } R_r = \frac{\text{正确检测数目}}{\text{正确检测数目} + \text{漏检数目}}, \quad (13)$$

$$\text{查准率 } R_p = \frac{\text{正确检测数目}}{\text{正确检测数目} + \text{误检数目}}, \quad (14)$$

将视频数据通过图 5 算法流程进行分割, 将得到的视频帧内容及帧数输出, 与该视频数据人工标注帧对比, 并计算其查准率和查全率, 进而得到本

文算法的综合指标。本节选取了几个视频片段在本文算法下的部份分割结果:

视频1长度为59 s,总帧数为1 798 帧。视频2长度为7 min 43 s,总帧数为13 888 帧。由于篇幅有限,仅选部份结果进行展示。视频片段1和视频片段2的镜头分割输出结果如图7、8所示。



(a) 切变镜头输出结果

(a) Output of shear shot



(b) 渐变镜头输出结果

(b) Output of gradual shot

图7 视频片段1 输出结果

Fig.7 Output of shot segmentation in video clip 1



(a) 切变镜头输出结果

(a) Output of shear shot



(b) 渐变镜头输出结果

(b) Output of gradual shot

图8 视频片段2 输出结果

Fig.8 Output of shot segmentation in video clip 2

本文算法与其他算法的精准度比较如表1所示。

由图7、图8可以看出,算法能够对切变镜头和渐变镜头进行有效的识别,为了说明本文算法的优势,我们通过使用70个视频数据测试本文算

法与其它几种单一特征算法或文章中提到的一些算法进行比对,通过比较其平均查准率和平均查全率,显示了本文的方法无论在查准率还是查全率上都要稍高于其他算法。另外,通过和单一特征算法的单帧平均运行时间比较表明,融合后的特征比单一特征更为均衡,兼顾性能和准确性。

表 1 本文算法与其他算法的精准度比较
Tab.1 Comparion of the accuracy of the proposed algorithm with that of other algorithms

	查全率/%		查准率/%	
	切变镜头	渐变镜头	切变镜头	渐变镜头
融合多种度量方法 ^[1]	89.5	78.9	92.4	87.2
像素帧差法 ^[2]	69	69	67	13
颜色矩法 ^[3]	81	26	48	34
基于颜色直方图和 SURF ^[7]	96	60	80	70
基于 HSV 直方图	80	70	90	87
基于 SIFT	92	87	81	63
基于 SURF	90	83	68	59
本文算法	95	90	93	76

本文算法与其他算法的平均单帧运行时间比较如表 2 所示。

表 2 本文算法与其他算法的平均单帧运行时间比较
Tab.2 Comparison of the average single frame run time between the proposed algorithm and other algorithms

算法	单帧平均运行时间/s
SIFT	1.46
SURF	0.14
HSV 直方图	0.05
本文算法	0.26

4 结 论

本文实现了一种基于 SURF 和 SIFT 特征的

视频镜头分割算法.通过在初选阶段引入颜色和 SURF 特征并进行了特征融合,精巧地解决了 SIFT 特征效率不高的难题,同时增加了算法的准确性。在复选阶段使用 SIFT 进行筛选,又利用了 SIFT 特征的优势,进一步加强了算法的准确性。该算法有效地结合了局部特征和全局特征,可以精准快速地识别镜头边界,对切变镜头的检测查全率达到 95%,查准率达到 93%,对渐变镜头的检测查全率达到 90%,查准率达到 76%,并且平均单帧运行时间为 0.26s。实验结果表明,通过特征融合,融合后的特征同时具备了两种特征的优点,而弥补了相应的缺点,起到了扬长补短的作用。此外,对于算法中 SURF、SIFT 特征检测器的阈值以及两帧匹配度的判定阈值均为实验所得的固定阈值,可以通过加入无监督学习来进一步优化。

参 考 文 献:

[1] 来毅,辛可嘉,刘颖. 多特征融合的视频镜头分割[J]. 电讯技术, 2018, 58(7): 792-797.
LAI Y, XIN K J, LIU Y. Video shot segmentation based on multi-feature fusion [J]. *Telecommunication Engineering*, 2018, 58(7): 792-797 . (in Chinese)

[2] KIKUKAWA T, KAWAFUCHI S. Development of an automatic summary editing system for the audio visual resources [J]. *Trans on Electronics and Information*. 1992, 75: 204 - 212.

- [3] ZHANG H J, KANKANHALLI A, SMOLIAR S W. Smoliar. Automatic partitioning of full-motion video [J]. *Multimedia Systems*, 1993, 1(1): 10-28.
- [4] 杨瑞琴, 吕进来. 基于双重检测的视频镜头分割方法[J]. 计算机工程与设计, 2018, 39(5): 1393-1398.
YANG R Q, LV J L. Video shot segmentation method based on double detection [J]. *Computer Engineering and Design*. 2018, 39(5): 1393-1398. (in Chinese)
- [5] 于俊清, 张强, 王赠凯, 何云峰. 利用回放场景和情感激励检测足球视频精彩镜头[J]. 计算机学报, 2014, 37(6): 1268-1280.
YU J Q, ZHANG Q, WANG Z K, HE Y F. Soccer highlight detection based on replay and affection arousal model [J]. *Chinese Journal of Computers*, 2014, 37(6): 1268-1280. (in Chinese)
- [6] 李小琛, 黄添强. 融合彩色信息与 SIFT 特征的帧内复制粘贴篡改检测[J]. 计算机系统应用, 2018, 27(7): 11-18.
LI X C, HUANG T Q. Forgery detection of copy-paste video based on fusion of color information and SIFT feature [J]. *Computer Systems & Applications*, 2018, 27(7): 11-18. (in Chinese)
- [7] 蔡轶珩, 胡朝阳, 崔益泽. 融合颜色信息与特征点的镜头边界检测算法[J]. 计算机应用, 2017, 37(S2): 95-111.
CAI Y H, HU Z Y, CUI Y Z. Shot boundary detection algorithm combining color information and feature points [J]. *Journal of Computer Applications*, 2017, 37(S2): 95-111. (in Chinese)
- [8] BAY H, ESS A, TUYTELAARS T. Speeded-up robust features [J]. *Computer Vision & Image Understanding*, 2008, 110(3): 346-359.
- [9] 冯亦东, 孙跃. 基于 SURF 特征提取和 FLANN 搜索的图像匹配算法[J]. 图学学报, 2015, 36(4): 650-654.
FENG Y D, SUN Y. Image Matching Algorithm Based on SURF Feature Extraction and FLANN Search [J]. *Journal of Graphics*. 2015, 36(4): 650-654. (in Chinese)
- [10] 任克强, 胡梦云. 基于改进 SURF 算子的彩色图像配准算法[J]. 电子测量与仪器学报, 2016, 30(5): 748-756.
REN K Q, HU M Y. Color image registration algorithm based on improved SURF [J]. *Journal of Electronic Measurement and Instrument*. 2016, 30(5): 748-756. (in Chinese)
- [11] 王蒙军. 唇读发声器中视觉信息的检测与处理[D]. 天津: 天津大学, 2016.
WANG M J. Detecting and processing visual information in speech synthesis system driven by visual-speech [D]. Tianjin: *Tianjin University*. 2016. (in Chinese)
- [12] JUAN L O. A Comparison of SIFT, PCA-SIFT and SURF [J]. *International Journal of Image Processing*. 2009, 3: 143-152.
- [13] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004: 91-110.

作者简介:



张昊骥(1991—),男,黑龙江哈尔滨人,硕士研究生,2014年于东北林业大学获得学士学位,主要从事视频分析、处理的研究。E-mail: 98774871@qq.com



任洪娥(1962—),女,吉林白山人,教授,博士,博士生导师,2009年于东北林业大学获得博士学位,主要从事图像识别与智能控制的研究。E-mail: nefu_rhe@163.com