

# 提取多场景视频关键帧的复合HOG特征聚类方法

魏英姿, 尹苏渝, 张宇恒

(沈阳理工大学信息科学与工程学院, 辽宁 沈阳 110159)

**摘要:** 由于直接利用帧差数据提取动态多场景视频关键帧往往会产生过多冗余帧, 方向梯度直方图(HOG)特征对图像亮度、场景变化具有较好的稳定性。为此, 提出了用于提取多场景视频关键帧的复合HOG特征聚类方法来提高关键帧提取效率。首先, 通过提取视频帧的HOG特征引入图像信息熵构成复合特征矢量, 以保持数据特征相关性。其次, 根据复合特征矢量统计视频帧间差异数据确定视频分割镜头、关键帧提取个数; 再次, 分别考虑镜头内帧集合和完整视频帧集合, 无重复地将信息熵较大的视频帧选为初始聚类中心以引导聚类算法搜索方向, 并通过K均值聚类抽取视频关键帧。与传统K均值聚类方法比较后发现, 所提算法冗余度降低0.003~0.015, 查准率提高了0.14~0.21, 聚类时间得到下降, 精度和效率较优。

**关键词:** 关键帧提取; 视频分割; HOG特征; 复合特征矢量; K均值聚类; 图像熵

DOI: 10.11907/rjdk.231929

开放科学(资源服务)标识码(OSID):



中图分类号: TP391.4

文献标识码: A

文章编号: 1672-7800(2024)009-0187-06

## Composite HOG Feature Clustering Method for Extracting Key Frames of Multi-Scene Video

WEI Yingzi, YIN Suyu, ZHANG Yuheng

(College of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China)

**Abstract:** Due to the fact that extracting dynamic multi scene video keyframes directly from frame difference data often results in excessive redundant frames, the directional gradient histogram (HOG) feature has good stability for image brightness and scene changes. Therefore, a composite HOG feature clustering method for extracting keyframes from multi scene videos has been proposed to improve the efficiency of keyframe extraction. Firstly, by extracting the HOG features of video frames and introducing image information entropy, a composite feature vector is constructed to maintain the correlation of data features. Secondly, based on the composite feature vector, the difference data between video frames is calculated to determine the number of video segmentation shots and keyframe extractions; Again, considering both the intra shot frame set and the complete video frame set, select the video frames with high information entropy as the initial clustering centers without repetition to guide the clustering algorithm search direction, and extract video keyframes through K-means clustering. Compared with the traditional K-means clustering method, it was found that the proposed algorithm reduces redundancy by 0.003~0.015, improves precision by 0.14~0.21, reduces clustering time, and has better accuracy and efficiency.

**Key Words:** key frame extraction; video segmentation; HOG feature; composite feature vector; K-means clustering; image entropy

## 0 引言

视频已逐渐成为获取信息的重要渠道, 但海量视频给视频数据存储和浏览造成了极大压力, 抽取其中的关键信

息将会加快视频检索、视频描述、异常检测等领域数据处理效率。视频关键帧是一幅或多幅帧图像, 能反映镜头中最重要的、有代表性的信息。依据镜头内容复杂程度, 一个镜头中能提取出一个或多个关键帧, 如何快速、恰当地提取关键帧, 以尽量少的数据承载尽量多的信息量, 是关键

收稿日期: 2023-09-01

扫描二维码阅读全文:



基金项目: 辽宁省自然科学基金项目(2022-KF-12-08)

**作者简介:** 魏英姿(1973-), 女, 博士, 沈阳理工大学信息科学与工程学院教授, 研究方向为计算机视觉、机器人学等; 尹苏渝(1999-), 女, 沈阳理工大学信息科学与工程学院硕士研究生, 研究方向为信号与信息处理; 张宇恒(1996-), 男, 沈阳理工大学信息科学与工程学院硕士研究生, 研究方向为图像处理。

帧提取技术研究的重点。

关键帧提取的途径主要通过运动分析、图像特征、光流和聚类等方法<sup>[1]</sup>。运动分析方法通常容易实现和调整,但对复杂场景可能会发生误判。其中,聚类方法以距离为相似性评价指标处理复杂场景。Cheng等<sup>[2]</sup>提出视频关键帧提取框架分析运动视频,解决现有方法忽略的运动信息表示。Zhang等<sup>[3]</sup>考虑到全局和局部运动状态信息,通过分数阶傅里叶变换得到相位谱信息,利用帧间的相位谱信息表征目标运动状态变化再提取关键帧。Ping<sup>[4]</sup>通过计算视频的图像序列光流提取的关键帧冗余度低。Yuan等<sup>[5]</sup>在聚类基础上增加了时空一致性,提取的关键帧具有较好的代表性。李梦静等<sup>[6]</sup>选取适合的特征提升关键帧提取效果,将行人步行周期作为聚类特征提取视频关键帧,有效解决了行人重识别物体遮挡问题。曾志贤等<sup>[7]</sup>采用多级编码提取、表征视频和文本的全局、局部及时序特征提取关键帧,可减少视频信息丢失。Li等<sup>[8]</sup>提出具有组对数正则化器的高效结构化稀疏编码算法,并且组对数正则化器有助于剔除目标变量的冗余特征。

对于机器人同步定位与建图(Simultaneous Localization and Mapping, SLAM)渐变动态场景的关键帧提取技术<sup>[9-10]</sup>。Klein等<sup>[11]</sup>在PTAM中提出基于关键帧的SLAM算法,但关键帧选取规则太简单,在大型环境中效果不佳。视觉SLAM方法在光照不均匀场景下表现不佳,其根本原因是在相机的不同视角下同一物体上的像素点灰度值不一致。顾恺琦等<sup>[12]</sup>通过光度标定帧间位姿估计方法对原始输入图像进行光度校正。

然而,传统帧差法提取关键帧往往局限于固定背景或相对静态的视频。深度神经网络可提取图像特征,但算法结构相对复杂,需要一定的算力,因此花费较大。对于场景、亮度动态不确定的视频,本文利用方向梯度直方图(Histogram of Oriented Gradient, HOG)特征降维数据,以减少光照影响。同时,线性降维可缓解计算量大的问题,但样本间的非线性相关性可能丢失,导致图片信息缺失。为此,本文引入图像信息熵指标构建复合特征矢量划分场景镜头,通过聚类方法提升算法的即时性和准确性。

## 1 视频帧的HOG特征提取

HOG特征提取算法通过计算、统计视频帧图片中局部区域的梯度方向直方图构成HOG特征,然后对物体进行边缘检测<sup>[13]</sup>。该方法相较于其他图像特征描述方法具有以下优势:①HOG基于图像角度方向信息,与光线和色彩无关,在光线和色彩变化较大场景下也具有较好的鲁棒性;②将图像分割为若干个局部区域,通过提取每个区域内部的梯度方向信息以捕捉图像的局部特征,降低了聚类计算量;③HOG特征提取速度相对较快;④对尺寸变化具有一定的不变性。HOG特征提取过程主要包括以下6步:

步骤1:在检测口对普通视频帧进行灰度化处理以降低数据维度。

步骤2:采用Gamma校正方法进行归一化处理,以降低图片中光照和阴影变化影响,抑制噪音干扰。

步骤3:计算视频帧图片中局部像素的梯度幅值 $G(x, y)$ 、方向 $G_\theta(x, y)$ ,如式(1)、式(2)所示。

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (1)$$

$$G_\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (2)$$

式中: $G_x(x, y)$ 、 $G_y(x, y)$ 为水平和垂直方向梯度值,可弱化光照影响得到图片中的轮廓信息。

步骤4:逐个将图像划分为多个大小为 $8 \times 8$ 的细胞单元,每 $2 \times 2$ 个细胞单元组成一个块。

步骤5:计算每一个细胞单元的特征梯度,将每4个细胞单元的特征描述组成一个块的HOG特征描述。

步骤6:将全图所有块依次连接形成特征矢量,得到第 $i$ 帧图像的HOG特征矢量 $f_i = [x_1, x_2, \dots, x_n]$ 。

采用HOG算法提取特征视频帧的效果如图1所示。由此可见,不同明暗背景下HOG可有效保留人物轮廓信息,降低数据维度,适于处理动态多场景视频数据。

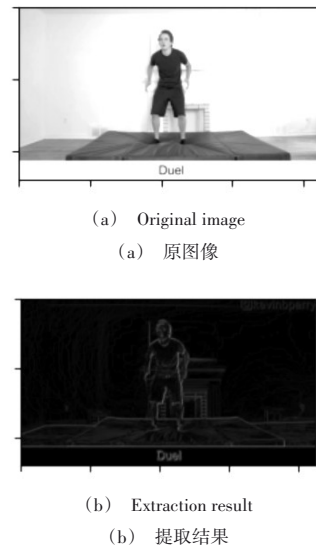


Fig. 1 HOG feature extraction results

图1 HOG特征提取结果

## 2 HOG复合特征矢量构建

为了在提取关键帧时尽量保留呈现较多信息量的帧,本文提出HOG复合特征矢量计算视频帧一元灰度熵 $x_h$ ,如式(3)所示。

$$x_h = -\sum_{j=0}^{255} P_j \log_2 P_j \quad (3)$$

式中: $P_j$ 表示灰度值为 $j$ 的像素比例。

将灰度熵 $x_h$ 与HOG特征矢量 $f_i$ 融合,构建HOG复合

特征矢量 $f'_i$ ,如式(4)所示。

$$f'_i = [f_i, x_H] = [x_1, x_2, \dots, x_n, x_H] \quad (4)$$

式中: $x_n \in [6, 8]$ ;  $x_n$ 列矢量中的元素值 $\in [0, 1]$ ; 设 $b$ 为熵特征 $x_H$ 的权重, 即将 $x_H$ 扩展为 $b$ 维列矢量 $x_H$ 。融合后的复合特征矢量 $f'_i$ 将用于K均值聚类方法抽取关键帧。

### 3 多场景视频镜头分割

针对动态场景视频, 视频分割可避免多镜头条件下的关键帧漏帧。目前, 镜头分割方法主要有基于颜色特征、运动特征、边缘特征<sup>[14]</sup>、训练模型<sup>[15-16]</sup>和统计学习<sup>[17]</sup>等方法。同时, 基于帧间差分原理分割视频比较视频帧图像颜色直方图特征差异, 例如取最大值, 因此镜头边缘帧常被选为分割帧。

然而, 上述方法往往适用于场景颜色变化明显的视频, 对处理渐变场景、移动场景效果不佳, 特征表达能力较弱。HOG特征可提取景物轮廓特征, 提出了基于视频帧间HOG特征差的视频分割法。首先将视频帧的HOG特征矢量转化为一维矢量, 计算HOG特征差值 $dh_i = |f_i - f_{i+1}|$ ; 其次构建HOG特征差值 $dh_i$ , 由大到小进行排列, 将排在前的 $S_F$ 个视频帧取出作为分割帧,  $S_F = \frac{N}{100}$ ,  $N$ 为视频总帧数, 视频被分割的镜头个数则为 $M = S_F + 1$ 。

### 4 基于HOG复合特征的K均值聚类方法

传统K均值聚类首先将原始数据划分为 $K$ 个类, 随机选取 $K$ 个点为起始聚类中心, 计算所有样本与各中心的距离, 距离近的归为一类; 其次重新计算每个聚类中心, 重复上述步骤直到聚类中心不再移动<sup>[18]</sup>; 再次根据样本的HOG复合特征矢量测度计算各样本差异, 以信息熵最大化为依据确定K均值聚类算法的初始聚类中心; 最后聚类视频帧样本以提取视频关键帧。主要步骤如下:

步骤1: 输入视频。逐帧计算HOG复合特征矢量 $f'_i$ , 计算HOG特征帧间差值 $dh_i$ , 将 $dh_i$ 较大的 $S_F$ 个帧作为分割镜头的帧, 得到镜头分割后各个帧的集合 $R_m$ 。

步骤2: 依据视频的平均帧间距确定聚类中心个数 $K$ , 作为提取关键帧的数目。根据复合特征数据矢量, 采用欧式距离计算第 $i$ 帧与 $i+1$ 帧的帧间距, 然后求出整个视频的平均帧间距。

$$d_i = |f'_{i+1} - f'_i| \quad (1 < i < N - 1) \quad (5)$$

$$\bar{d} = \frac{\sum_{i=1}^{N-1} d_i}{N-1} \quad (6)$$

式中: $d_i$ 为复合特征矢量 $f'_i$ 差值的模;  $\bar{d}$ 的大小反映视频对象的变化程度。

根据平均帧间差 $\bar{d}$ 大小确定关键帧数目, 如果 $\bar{d} < Th\_d$ 表明视频变化缓慢, 可降低关键帧提取数目 $K = \frac{N}{T}$ , 应设置

较大的系数 $T$ ,  $Th\_d$ 一般取值为 $50^{[19]}$ 。如果 $\bar{d} > Th\_d$ 表示视频变化剧烈,  $T$ 值应该设置小一些。对于多场景视频, 关键帧的个数应大于镜头个数, 即 $K > S_F + 1$ ,  $T < 100 \times \frac{N}{N+1}$ 。

步骤3: 分批筛选信息熵较大的视频帧作为初始聚类中心。计算所有视频帧的图像信息熵, 镜头分割后的帧集合为 $R_m, m \in [1, M]$ , 通过比较镜头内的帧数据, 优先选出 $R_m$ 内熵值最大的视频帧, 共取出 $M$ 个帧, 再按熵值由大到小排序所有视频帧, 无重复地取排在前面的 $K-M$ 帧, 共计 $K$ 个帧被选为初始聚类中心。

步骤4: 迭代更新聚类中心。计算各帧到对应聚类中心的距离, 判断是否满足平均距离最小, 若不满足则中心继续移动; 若满足则进行近邻冗余帧复核和删除, 根据记录的原始帧位置检查是否存在原视频相邻帧同时被选入聚类中心, 若存在则丢弃熵值较小的部分, 从未曾被选为初始聚类中心的帧集合中按照熵大小, 概率性选择熵值较大的帧替换聚类中心, 直到聚类中心继续移动到最终结果没有相邻帧。

步骤5: 停止移动, 聚类结束。各聚类中心即为提取关键帧。

## 5 实验结果与分析

### 5.1 实验结果

本文实验环境为AMD Ryzen 7 6800 H, 3.2 GHz CPU, 16 GB内存, Windows 10。实验数据集为自采集数据集, 包括移动渐变场景、人体摔倒、徘徊、人物跑酷等多段人体运动视频, 通过提取不同类别视频的关键帧来分析算法。在数据集中截取部分样本片段, 测试集总帧数分别为{移动渐变 819, 摔倒 1 500, 跑酷 840}, 采用HOG复合特征K均值聚类方法提取上述视频关键帧。图2为单个人体摔倒镜头下提取的关键帧。



Fig. 2 Key frames of human fall video

图2 人体摔倒视频关键帧

图3为多场景跑酷视频的HOG特征数据帧间差值曲线, 横坐标表示帧号。纵坐标表示归一化后的HOG特征帧间差 $dh_i$ , 帧间差突变处对应多个镜头间转换的视频帧, 该视频镜头转换帧为86帧, 155帧, 398帧, 447帧, 527帧,

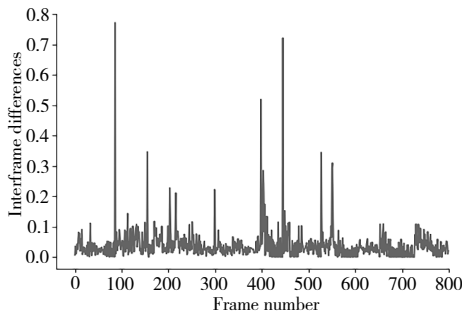


Fig. 3 Video HOG feature value frame difference variation curve

图3 视频HOG特征值帧间差变化曲线

554帧,因此视频被分为7个镜头。

图4为多场景人体跑酷视频未作镜头分割提取的关键帧。图5为跑酷视频经过镜头分割处理后提取的关键帧。由图4、图5可见,未能提取到中间多个场景的关键帧,加入镜头分割后避免了多场景视频关键帧漏帧现象。图6为算法提取的人物移动渐变视频关键帧。

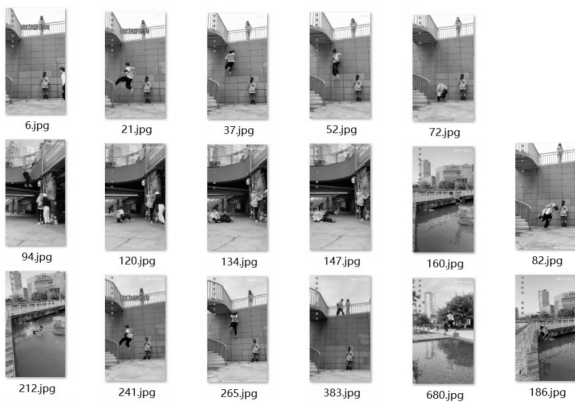


Fig. 4 Key frames of multi scene parkour videos without shot segmentation

图4 未作镜头分割的多场景跑酷视频关键帧

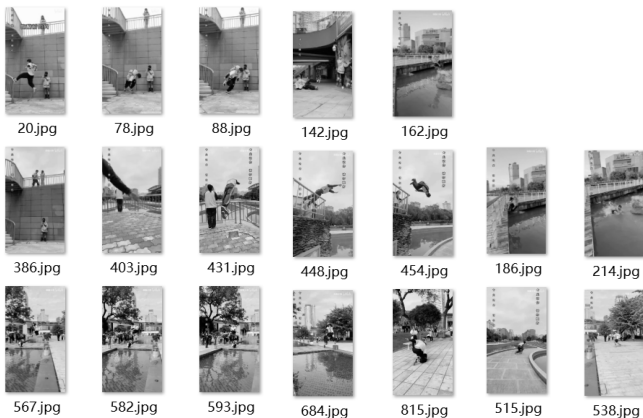


Fig. 5 Multi-scene parkour video key frames with shot segmentation

图5 加入镜头分割的多场景跑酷视频关键帧

### 5.2 策略浸润试验

本文为测试复合HOG特征对聚类算法性能的影响,采用策略浸润试验并利用图像信息熵评价关键帧提取效

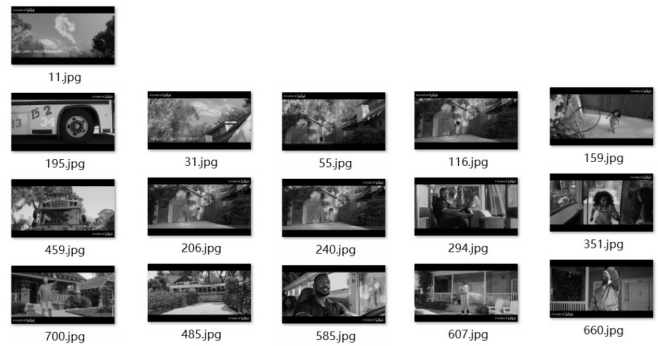


Fig. 6 Key frames of moving gradient video

图6 移动渐变视频关键帧

果。将跑酷原视频分割为多段视频,按步骤3确定初始聚类中心,通过聚类提取关键帧。

图7为视频分割段的熵平均值及熵平均差值的比较曲线,熵平均值体现了不同视频段抽取关键帧的信息量平均水平。由此可见,7段跑酷视频段总的熵平均值分别为{原视频 50.596, HOG 特征法 50.550, 复合 HOG 特征法 51.008},这7段视频总的熵平均差值分别为{原视频 0.700, HOG 特征法 0.588, 复合 HOG 特征法 0.681};4段移动渐变视频总的熵平均值分别为{原视频 26.301, HOG 特征法 26.241, 复合 HOG 特征法 26.588},这4段视频总的熵平均差值分别为{原视频 0.982, HOG 特征法 0.351, 复合 HOG 特征法 0.814}。

通过比较原视频帧、HOG 特征K均值聚类提取的关键帧、复合 HOG 特征K均值聚类提取的关键帧可知,本文聚类方法具有较视频段内熵平均值和熵平均差值较大,说明提取的关键帧具有较高的信息量和多样性。此外,本文方法抽取的关键帧与原视频帧熵值、熵差异曲线形状相似度更高,证明了HOG+熵构建复合特征后聚类可较好保留原视频帧数据特点。

### 5.3 评价指标

本文根据冗余度、聚类时间、查准率、查全率等4个指标验证算法的有效性。其中,冗余度  $R_r = \text{冗余帧数 } I_r / \text{视频总帧数 } N$ ;查准率  $R_a = \text{查对帧数 } I_a / (\text{查错帧数 } I_f + \text{查对帧数 } I_a)$ ;查全率  $R_w = \text{查对帧数 } I_a / (\text{查对帧数 } I_a + \text{漏查帧数 } I_2)$ 。本文 HOG 复合特征 K 均值聚类法(Ours)、基于视频帧数据 K 均值聚类方法(Fm\_data)、基于运动目标尺度不变特征转换特征<sup>[20]</sup>(Scale Invariant Feature Transform, SIFT)<sup>[21]</sup>提取不同视频的关键帧评价指标如表1所示。

由此可知,本文复合 HOG 特征聚类法性能指标均优于基于运动目标特征的关键帧提取算法,相较于传统视频帧数据聚类方法(Fm\_data)而言准确率有所提高。其中, HOG 特征聚类法有效降低了聚类处理的数据量,减少了计算时间花费;采用的相邻视频帧剔除策略降低了关键帧的冗余度。此外,对于不同人物动作类视频,复合 HOG 特征聚类法准确率均能保持在0.83以上,在跑酷视频上进行的消融实验结果如表2所示。

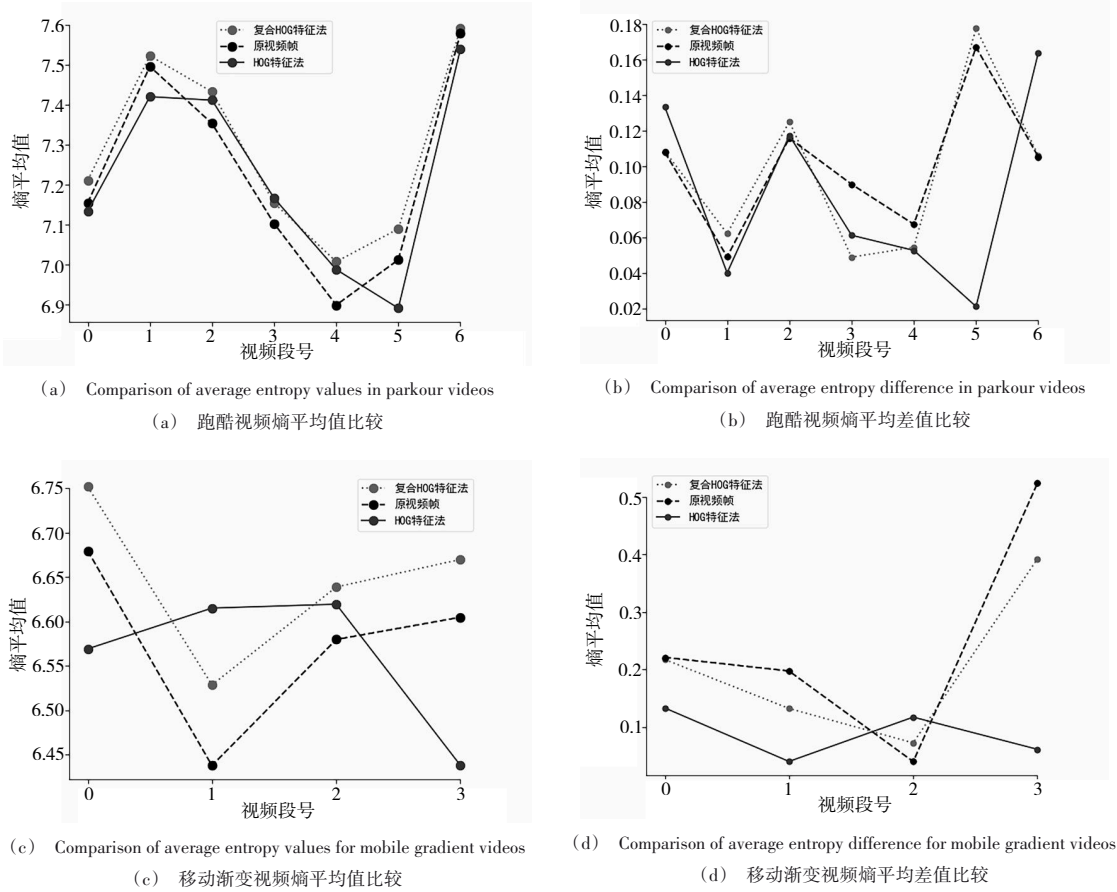


Fig. 7 Comparison of entropy values between original video frames and extracted keyframes

图 7 原视频帧、提取关键帧熵值比较

Table 1 Performance indicators of K-Means, SIFT, and the composite HOG feature clustering method proposed in this paper

表 1 K-Means、SIFT 与本文复合 HOG 特征聚类法的性能指标

视频	总帧数	冗余度			时间花费(s)			查准率			查全率			关键帧数
		Ours	Fm_data	SIFT	Ours	Fm_data	SIFT	Ours	Fm_data	SIFT	Ours	Fm_data	SIFT	
摔倒	1500	0.005	0.020	0.010	34.480	330.50	240.50	0.870	0.670	0.670	0.900	0.670	0.800	31.000
跑酷	840	0.001	0.004	0.003	37.670	66.570	180.30	0.940	0.730	0.800	1.000	0.730	0.830	19.000
移动渐变	819	0.001	0.014	0.003	18.850	36.610	298.20	0.930	0.850	0.620	0.880	0.750	0.500	16.000

Table 2 Ablation experiment

表 2 消融实验

方法	查准率	查全率
HOG	0.79	0.88
熵	0.53	0.63
HOG+熵	0.94	1.00

## 6 结语

由于视频中相近的 HOG 特征通常对应视觉上相似的帧,而聚类可寻找具有代表性的关键帧;对于场景、光照渐变的视频,HOG 特征可有效降低帧间差分数据计算量。为此,本文通过提取多场景视频关键帧的复合 HOG 特征聚类方法来提升关键帧提取效率,将融合信息熵作为特征参

数,使 K 均值聚类方法较好地保留了原视频数据的熵信息特点,在一定程度上弥补了 HOG 数据降维造成的特征损失。

目前,本文采用分割镜头方法避免关键信息遗漏,未来也可通过设置密集的时间窗大小进行聚类迭代运算,例如每隔 5~10 帧聚类一次,或通过降低视频帧图像特征维度来减少聚类计算量。虽然,本文算法对硬件平台成本要求较低,可为动态构建地图技术提供较好的参考,但无法提取图片深层特征,未来可融合深度学习模型进一步优化关键帧提取方法。

## 参考文献:

[1] LIU Q P, SHENG Y Q, GAO R X, et al. Sign language recognition algorithm based on key frame and attention residual network[J]. Computer En-

- gineering, 2023, 49(12):224-230, 242.  
刘群坡, 盛月琴, 高如新, 等. 基于关键帧和注意力残差网络的手语识别算法[J]. 计算机工程, 2023, 49(12):224-230, 242.
- [2] CHENG H, WANG H M. A novel key-frames selection frame work for comprehensive video summarization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(2):577-589.
- [3] ZHANG Y Z, ZHANG J Y, TAO R. Key frame extraction of surveillance video based on fractional fourier transform [J]. Journal of Beijing Institute of Technology: English Edition, 2021, 30(3):311-321.
- [4] PING Y. Key frame extraction method of music and dance video based on multicore learning feature fusion [J]. Scientific Programming, 2022 (7): 1-8.
- [5] YUAN Y, ZHE L, ZHOU Y, et al. Key frame extraction based on global motion statistics for team-sport videos [J]. Multi-media Systems, 2022, 28 (2):387-401.
- [6] LI M J, JI G L, ZHAO B. Key frame extraction algorithm for video-based person re-identification based on walking cycle clustering [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53 (5): 780-788.  
李梦静, 吉根林, 赵斌. 基于步行周期聚类的视频行人重识别关键帧提取算法[J]. 南京航空航天大学学报, 2021, 53(5):780-788.
- [7] ZENG Z X, CAO J J, WENG N F, et al. Dual encoding integrating key frame extraction for video-text cross-modal entity resolution [J]. Acta Armentarii, 2022, 43(5):1107-1116.  
曾志贤, 曹建军, 翁年凤, 等. 结合关键帧提取的视频-文本跨模态实体分辨双重编码方法[J]. 兵工学报, 2022, 43(5):1107-1116.
- [8] LI Z N, LI Y J, TAN B Y, et al. Structured sparse coding with the group log-regularizer for key frame extraction [J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(10):1818-1830.
- [9] ZENG S Y, ZHANG L, XU F, et al. Laser semantic SLAM system fused with poisson reconstruction [J]. Computer Engineering and Design, 2023, 44(1):84-91.  
曾圣尧, 张雷, 徐方, 等. 融合泊松重建的激光语义SLAM系统[J]. 计算机工程与设计, 2023, 44(1):84-91.
- [10] CUI Y X, LIU G H, YU D Y, et al. Lidar-mono-inertial SLAM system with fusion of point-line features [J]. Computer Engineering, 2022, 48 (7):254-263.  
崔云轩, 刘桂华, 余东应, 等. 点线特征融合的激光雷达单目惯导SLAM系统[J]. 计算机工程, 2022, 48(7):254-263.
- [11] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C]// 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007:225-234.
- [12] GU K Q, LIU X P, WANG G, et al. Semi-direct visual slam algorithm based on online photometric calibration [J], Robot, 2022, 44 (6): 672-681.  
顾恺琦, 刘晓平, 王刚, 等. 基于在线光度标定的半直接视觉SLAM算法[J]. 机器人, 2022, 44(6):672-681.
- [13] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005:886-893.
- [14] ZHANG H S, ZHU X L, HU X Z, et al. Shot segmentation technology based on SURF features and SIFT features [J]. Chinese Journal of Liquid Crystals and Displays, 2019, 34(5):74-82.  
张昊骅, 朱晓龙, 胡新洲, 等. 基于SURF和SIFT特征的视频镜头分割算法[J]. 液晶与显示, 2019, 34(5):74-82.
- [15] LI Q J, LI Q N. A lens segmentation method based on mutual information and self-regression model [J]. Computer Technology and Development, 2019, 29(1):35-39.  
李强军, 李启南. 基于互信息量和自回归模型的镜头分割方法[J]. 计算机技术与发展, 2019, 29(1):35-39.
- [16] HUANG Q, FENG H C, LIU L. Multi-modal video scene segmentation optimization algorithm based on convolutional neural network [J]. Application Research of Computers, 2022(5):1595-1600.  
黄清, 丰洪才, 刘立. 基于卷积神经网络的多模态视频场景分割优化算法[J]. 计算机应用研究, 2022(5):1595-1600.
- [17] XU Y C, FENG H C, LIU L. Shot segmentation algorithm of agrometeorological popular science video based on double detection model [J]. Software Guide, 2022, 21(9):159-166.  
徐迎春, 丰洪才, 刘立. 基于双重检测模型的农业气象科普视频镜头分割算法[J]. 软件导刊, 2022, 21(9):159-166.
- [18] MACQUEEN J. Some methods for classification and analysis of multivariate observations [C]// Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1965:281-297.
- [19] XU W Z, XU L H. Combining principal components analysis and clustering to extract key-frame [J]. Computer Engineering and Applications, 2009, 45(15):9-23.  
许文竹, 徐立鸿. 结合主成分分析和聚类的关键帧提取[J]. 计算机工程与应用, 2009, 45(15):9-23.
- [20] TIAN L H, ZHANG M, LI C. Key frame extraction algorithm based on feature of moving target [J]. Application Research of Computers, 2019, 36(10):3183-3186.  
田丽华, 张咪, 李晨. 基于运动目标特征的关键帧提取算法[J]. 计算机应用研究, 2019, 36(10):3183-3186.
- [21] LOWE D. Distinctive image features from scale-invariant key points [J]. International Journal of Computer Vision, 2004, 60(2):91-110.

(责任编辑:刘嘉文)