

基于互信息熵的全景视频关键帧 数据实时提取系统设计

张大禹, 吴诗帆

(中国人民解放军 91550 部队, 辽宁 大连 116000)

摘要: 关键帧数据提取可以降低全景视频检索中产生的数据量, 为了提高全景视频镜头边缘的检测与关键帧的提取的处理能力, 提出基于互信息熵的全景视频关键帧数据实时提取系统设计; 根据全景视频关键帧数据提取系统的硬件结构, 分析视频播放器和镜头边缘检测器的工作原理; 在系统的软件设计中, 将待检测的两帧图像随机划分为子图像块, 通过计算子图像块之间的互信息熵, 获取全景视频图像的突变帧, 将关键图像帧的特征差值曲线作为全景视频关键帧的时序特征, 完成全景视频中关键帧数据的特征匹配, 选择一个能够体现全景视频图像属性的特征, 描述视频中的主要信息, 利用数值化分析的方式, 将全景视频帧的特征转换成数组形式的特征向量, 通过计算互信息熵值, 提取全景视频关键帧数; 系统测试结果表明, 文中系统将关键帧数据提取的查全率和查准率分别提高到 95% 和 98% 以上; 能够准确提取出全景监控视频的关键帧数据, 具有更好的全景视频处理能力。

关键词: 互信息熵; 关键帧数据; 特征提取; 实时提取; 全景视频; 镜头边缘检测

Full-motion Video Key Frame Data Real-time Extraction System Design Based on Mutual Information Entropy

ZHANG Dayu, WU Shifan

(Unit 91550, PLA, Dalian 116000, China)

Abstract: Key frame data extraction can reduce the amount of data generated in panoramic video retrieval. In order to improve the processing capability of panoramic video shot edge detection and key frame extraction, a real-time extraction system design of panoramic video key frame data based on mutual information entropy is proposed. According to the hardware structure of the panoramic video key frame data extraction system, the working principle of the video player and shot edge detector is analyzed; in the system software design, the two images to be detected are randomly divided into sub-image blocks, The mutual information entropy between the blocks is used to obtain the mutation frame of the panoramic video image, and the feature difference curve of the key image frame is used as the time series feature of the key frame of the panoramic video, and the feature matching of the key frame data in the panoramic video is completed. The characteristics of the video image are selected to describe the main information in the video. By means of numerical analysis, the features of the panoramic video frames are converted into the feature vectors in the form of arrays, and the number of key frames of the panoramic video is extracted by calculating the mutual information entropy value. The system test results show that the recall rate and precision rate of key frame data extraction in the system are improved by more than 95% and 98%, respectively. It can accurately extract the key frame data of the panoramic surveillance video, it has better panoramic video processing capability.

Keywords: mutual information entropy; key frame data; feature extraction; real-time extraction; full-motion video; lens edge detection

0 引言

视频信息作为网络资源的重要载体, 如何对其进行有效的处理和管理, 是目前国内外在视频处理方面研究的热点^[1]。目前, 针对传统的视频数据处理技术, 已经无法满足现有的需求, 于是, 基于内容的视频数据分析与检索技术应运而生。在基于内容的视频检索中, 镜头边缘的检测与关键帧的提取是其重要的技术^[2]。但是, 随着计算机技术和互联网技术的飞速发展, 只依靠人工获取关键帧的方

法已不能满足实际应用的需要, 因此, 利用计算机技术实现视频关键帧数据的实时提取具有重要的理论意义和现实意义^[3]。

梁建胜等^[4]为了解决传统的视频处理技术存在的问题, 本文采用了基于深度学习的方法, 对视频中的关键帧进行了提取, 提出了一种基于小波测距技术的图像检索方案, 并对各个场景进行了分类, 选择了特征最突出的图像作为主要图像。然后在卷积神经网络的基础上, 采用无监督、半监督和监督 3 种方法, 对卷积神经网络进行了训练, 并

收稿日期: 2022-09-08; 修回日期: 2022-10-31。

作者简介: 张大禹(1977-), 男, 黑龙江哈尔滨人, 大学本科, 高级工程师, 主要从事数据处理, 数字音视频方向的研究。

引用格式: 张大禹, 吴诗帆. 基于互信息熵的全景视频关键帧数据实时提取系统设计[J]. 计算机测量与控制, 2023, 31(1): 252-256, 263.

将其用于关键帧的提取中。利用公开的数据集进行验证, 结果显示, 该方法在关键帧提取中的准确率更高, 根据关键帧提取结果可以检索出相关视频。周丹等^[5]利用手语识别的方式, 提出一种关键帧提取算法, 利用卷积自编码算法对图像进行 K-means 聚类, 通过图像的清晰度过滤, 得到最清晰的图像, 并将其作为初始抽取的关键帧, 通过对初始关键帧的二次优化, 获得了用于手势识别的关键帧, 结果表明, 该方法在很大程度上消除了视频中的冗余帧, 提高了手语识别的准确性和有效性。

基于以上研究背景, 本文利用互信息熵设计了全景视频关键帧数据的实时提取系统, 从而保证全景视频的质量。

1 全景视频关键帧硬件结构及原理

在全景视频中, 关键帧数据的实时提取能够准确描述视频的详细内容, 系统的硬件由视频播放器、镜头边界检测器等主要设备组成, 具体结构如图 1 所示。

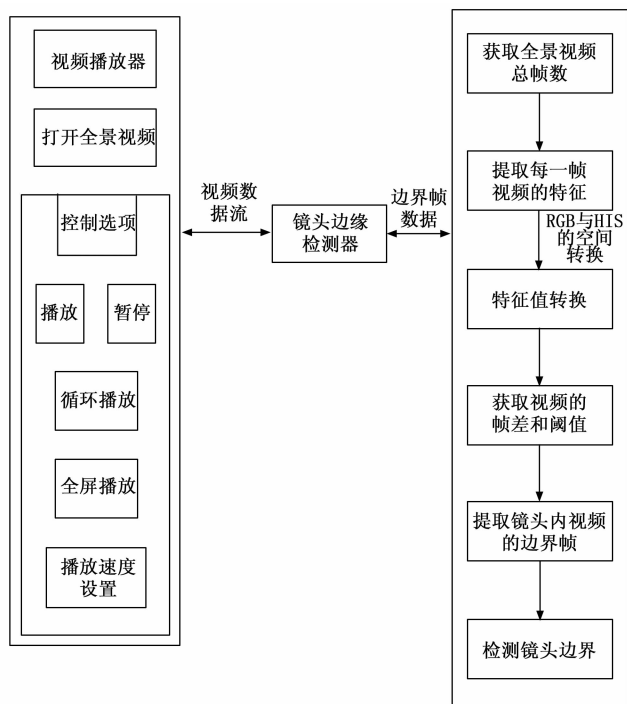


图 1 全景视频关键帧数据提取系统硬件结构

如图 1 所示, 首先, 获取全景视频与相关属性信息, 步骤如下:

1) 数据支撑层。该层由数据采集模块、数据存储模块、数据处理模块组成。其中数据采集模块包含摄像设备, 例如, 可以利用无人机采集全景视频, 将其传输至存储模块中。数据处理模块包含图像拼接单元, 主要职责是对全景视频进行融合处理, 获取全景大图, 并保存于全景数据库中。数据存储模块中包含 4 种类型数据库, 系统管理数据库用于存储系统管理文件、运行参数等信息; 业务数据库的职责是保存与全景视频关键帧相关的业务数据、功能数据等。全景数据库用于存储采集的视频信息及全景大图

信息。热点数据库用于保存各热点数据及其位置坐标, 能够对设备信息进行管理。通过热点可显示影像的具体属性, 并可呈现所有相关信息, 为用户全面掌握虚拟摄像机镜头拍摄的影像全景信息提供数据支持。在全景重建单元的 Flash Player 内可实现热点增加、修改以及热点位置坐标获取等, 并采用 XML 格式存储, 方便用户查阅。

2) Web 服务层。该层可提供编程接口, 并为数据传输选择合适路由, 具备请求接听等服务功能。

3) 应用逻辑层。该层由业务处理单元、全景重建单元、全景浏览与交互单元构成。全景重建单元调用获取的全景大图, 使之贴合于球面模型的表面, 将虚拟摄像机嵌入在球面模型中心, 并通过鼠标对其进行控制来进行角度调整, 以此实现全景视频关键帧场景的三维展示^[6-8]。同时, 该单元可从热点数据库中获取热点、位置数据, 并提供属性编辑功能。业务处理单元提供影像全景图像预处理、影像监测预警、运营维护、故障分析等功能。其中影像全景图像预处理可通过检测摄影镜头拍摄的影像的边缘, 获取影像边缘特征图像, 为利用改进 YOLOv3 网络完成影像监测预警提供数据支撑。

4) 用户层。通过该层实现虚拟摄像机镜头拍摄的影像全景监控的可视化呈现, 向用户展示影像全景视图以及业务处理结果。

然后, 视频播放器通过打开全景视频, 利用控制选项对视频进行播放、暂停、播放速度设置以及循环播放和全屏播放等功能, 将全景视频暂停可以分析该帧视频的数据, 通过判断关键帧数据的质量, 保留全景视频中高质量关键帧数据; 放慢视频播放速度、循环播放和全屏播放可以让用户清楚观察到任意一帧视频的质量, 为镜头边缘检测器对全景视频关键帧数据的检测和分割提供数据流支持。

最后, 镜头边缘检测器的主要任务是对全景视频的播放镜头进行检测和分割, 工作原理是先获取视频播放器中的全景视频总帧数, 提取出每一帧图像的颜色特征和纹理特征, 利用 RGB 与 HIS 的空间转换^[9], 对每一帧图像的颜色特征进行转换, 获取全景视频的帧差和阈值, 通过提取出镜头内视频的边界帧, 检测出镜头边界。

2 全景视频关键帧数据实时提取系统软件设计

全景视频关键帧数据提取是视频检索中的另一重要环节, 其关键帧是镜头的代表性帧, 利用关键帧技术实现对图像的快速检索, 从而极大地降低了检索所需的信息量, 提高了视频检索的准确性。

2.1 检测全景视频图像的突变帧

在提取全景视频关键帧数据的过程中, 需要先对全景视频进行预处理, 全景视频经过预处理之后, 在全景视频关键帧处理时, 由于邻近关键帧的不相关(非交迭区)大约占整个图像的 40% (此处对邻近关键帧的交迭需要 60%), 会存在含有镜头边界的片段^[10-12], 因此, 若将两个相邻的帧图像进行关联运算, 必然会提高误判率, 从而降低处理结果的精确性。检测全景视频图像的突变帧时, 只

需要对镜头的突变边界进行检测即可。

如果直接在全景视频中计算相邻两帧图像的互信息熵,就会存在一定偏差,因此本文将待检测的两帧图像随机划分为子图像块,通过计算子图像块之间的互信息熵,获取全景视频图像的突变帧。

通过对全景视频进行预处理^[13],筛选出包含镜头边界的候选视频片段,将其表示为:

$$F = \{F_1, F_2, \dots, F_k, \dots\} (k \leq n) \quad (1)$$

其中: n 代表全景视频的片段数量。

将 F 中第一帧到最后一帧图像的所有图像进行相邻两帧组合,划分成 $a \times a$ 个块,计算出子图像块之间的互信息熵 $I_{k,k+1}$ 。

对子图像块之间的互信息熵 $I_{k,k+1}$ 进行反运算,得到:

$$W_i = \frac{1}{I_{k,k+1} + 1} \quad (2)$$

将各个子图像块的熵值相加,计算全景视频中,相邻两帧图像的差值,公式为:

$$C_{k,k+1} = \sum_{i=1}^N W_i \quad (3)$$

其中: N 表示子图像块的数量。

利用相邻两帧图像的差值^[14-15],计算与 $C_{k,k+1}$ 相邻帧图像的局部极大值,那么 $C_{k,k+1}$ 对应的视频帧就是镜头的突变边界。

求解 $C_{k,k+1}$ 的方差,分析其显著性,如果 $|C_{k,k+1}| \geq \zeta I_{k,k+1}$,说明全景视频中相邻两帧图像的相似性超过了用户设定的阈值,将其作为镜头突变帧,如果 $|C_{k,k+1}| < \zeta I_{k,k+1}$,说明全景视频中相邻两帧图像的互信息熵没有明显变化,即不存在突变帧。

根据以上步骤,即可检测全景视频中的图像是否存在的突变帧,为关键帧数据的特征匹配提供了质量保证。

2.2 匹配全景视频中关键帧数据的特征

利用计算机直接读出图像的灰度矩阵,避免了点特征的提取和匹配,降低了计算量,提高了处理速度。匹配全景视频中关键帧数据的特征是关键帧提取的关键,根据全景视频的图像特征,获取每一帧图像的子区域灰度均值,给出具有 m 帧图像的序列 X_c ,表示为:

$$X_c = \frac{c \oplus i}{[G_{c_1}, \dots, G_{c_i}]} \quad (4)$$

其中: G_{c_i} 代表全景视频 c 中第 i 帧关键图像。

利用全景视频图像序列 X_c 将 $[G_{c_1}, G_{c_i}, G_{c_i}]$ 转化为对应的灰度值^[16],提取出全景视频中关键图像的时序特征,即:

$$X_f = n \times \{Af(1,k), \dots, Af(i,k)\} \quad (5)$$

其中: $Af(i,k)$ 代表全景视频 f 中,第 i 帧图像在 k 范围内的灰度均值。

在进行相关系数的计算时,如果将当前关键帧内的交叠区域看作是运动物体,而图像被看作是静态物体,则会发现,重叠区域会随着拍摄时间的推移,逐渐地向图像的后面移动,直到图像完全消失。如果将 a 帧全景视频 U 与 b 帧目标视频 V 进行匹配^[17],先计算出 U 和 V 在关键帧时序

特征曲线的相邻点差值,公式为:

$$\begin{cases} \Delta AU(i,j) = |AU(i+1,k) - AU(i,k)| \\ \Delta AV(i,j) = |AV(i+1,k) - AV(i,k)| \end{cases} \quad (6)$$

将式(10)中关键图像帧的特征差值曲线作为全景视频关键帧的时序特征,对全景视频中关键帧数据的特征进行匹配,为了消除突变因素对匹配效果的影响,引入异常因子 ξ 判断关键帧数据的特征能否成功匹配^[18],即:

$$\xi = \left(1 - \Delta SF_v \times \frac{M}{K}\right) \quad (7)$$

其中: ΔSF_v 代表全景视频中关键图像帧的时域特征差值曲线, M 代表全景视频中包含的合格图像帧数, K 代表关键图像帧的像素差值。

当异常因子在容忍范围内,并且关键图像帧的像素差值比用户设置的阈值小,说明全景视频中关键帧数据的特征成功匹配,通过对大量特征进行多项式拟合,可以近似拟合图像中的相关系数的变化趋势,特别是在数据量大、有一定规则的情况下,通过选取最好的多项式和抽样时间,可以精确地反映出结果的微小差异(即,曲线拟合函数所得到的最大值和实际位置之间的差异很小),从而快速地确定数据的极值位置,在对图像序列进行颜色、纹理、形状、运动等方面进行分析的基础上,选择出能够反映镜头整体内容的图像序列。能够快速、真实地反映出图像的动态变化,计算效率高,拟合精度高,定位精度高。

2.3 基于互信息熵设计全景视频关键帧数据提取算法

关键帧数据提取过程可以降低视频索引所需的信息量,并为视频的查询与检索提供了一个有效的框架。其中,镜头边缘检测和透镜边缘探测是图像处理中的两大主要环节。

一支镜头的关键帧长度随镜头长度及镜头内容的强弱而不同,随着镜头长度的增加,镜头内容的改变速度会加快,该镜头的关键帧序列也会随之延长,反之则会缩短。由于视频本身就是一个庞大的数据流,如果算法太过复杂,就很难实现实时的处理。在全景视频中提取关键帧数据时,需要选择一个能够体现全景视频图像属性的特征,对视频中的主要信息进行描述。利用数值化分析的方式^[19],将全景视频帧的特征转换成数组形式的特征向量,该特征向量可以利用计算机识别得到,实现镜头的边缘检测和帧间相似度计算。

提取全景视频图像特征是全景视频关键帧数据提取的重要环节,而透镜边缘探测的目标是把视频分割成一系列的透镜,透镜是检索、检索和浏览的最小单元,利用互信息熵的属性提取全景视频的图像特征,具体步骤如下:

Step1: 将全景视频的片段分割成 n 段,记作 p_1, p_2, \dots, p_n ,由于每一个图像帧在计算像素点的过程是独立完成的,因此将 GPU 线程块设置为 32×32 ,每一个线程可以计算唯一一个视频图像区域的像素^[20],还可以在 1 个宏模块中计算 4 个 16×16 块的互信息,最后得到 3 种类型分别为 32×32 、 32×16 和 16×32 的互信息值,表示为:

$$I_n = \{I_{1,2}, I_{2,3}, I_{3,4}, \dots, I_{i,i+1}\} \quad (8)$$

Step2: 根据互信息值, 消除全景视频帧相似度较高的帧图像, 利用 $I_{i,j}^R$ 、 $I_{i,j}^G$ 和 $I_{i,j}^B$ 代表 R、G、B 三个通道的互信息熵值, 计算公式为:

$$I_{i,j}^R = \sum_{x=0}^h \sum_{y=0}^h p_{i,j}^R(x,y) \log \frac{p_{i,j}^R(x,y)}{p_i^R(x) p_j^R(y)} \quad (9)$$

$$I_{i,j}^G = \sum_{x=0}^h \sum_{y=0}^h p_{i,j}^G(x,y) \log \frac{p_{i,j}^G(x,y)}{p_i^G(x) p_j^G(y)} \quad (10)$$

$$I_{i,j}^B = \sum_{x=0}^h \sum_{y=0}^h p_{i,j}^B(x,y) \log \frac{p_{i,j}^B(x,y)}{p_i^B(x) p_j^B(y)} \quad (11)$$

其中: h 代表全景视频帧的灰度等级, $p_i(x)$ 代表视频 i 的图像灰度概率, $p_j(y)$ 代表视频 j 的图像灰度概率, $p_{i,j}(x, y)$ 代表两帧图像的联合灰度概率。

获取 R、G、B 三个通道的互信息熵值之后, 利用下式计算总的互信息熵值, 即:

$$I_{i,j} = I_{i,j}^R + I_{i,j}^G + I_{i,j}^B \quad (12)$$

Step3: 利用 R、G、B 三个通道总的互信息熵值, 计算每段全景视频中相邻两帧图像的互信息标准差:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - \bar{s})^2} \quad (13)$$

其中: S_i 代表相邻两帧图像的互信息熵值, \bar{s} 代表互信息熵的均值, n 代表互信息总量。

Step4: 将 Step3 中的标准差 σ 与阈值 λ 作对比, 如果 $\sigma < \lambda$, 说明全景视频片段中的图像内容不显著, 帧间相似度较大, 反之执行 Step5;

Step5: 获取最大信息熵值对应的图像帧作为参考^[21], 并与其他信息熵比较, 即:

$$E = 2I_x - I_{\max} \quad (14)$$

当 $E > 0$ 时, 说明全景视频中两帧图像的相似度较高, 将其删除, 当 $E \leq 0$ 时, 说明全景视频中两帧图像的相似度较低, 将全景视频中该帧图像的位置进行标记, 提取出被标记的视频帧, 将其作为全景视频的关键帧, 为后续的检索、查询、检索和浏览提供了精确的依据。

3 测试分析

3.1 测试环境

为了验证文中系统在提取全景视频关键帧数据时的性能, 搭建了如下实验环境:

- 操作系统: Windows 7
- CPU 型号: Intel Core i9-4890
- CPU 频率: 4.8 GHz
- RAM 容量: 8 GB
- 运行环境: Matlab 2018b
- 软件开发包: OpenCV3.2

3.2 实验数据

在 3.1 测试环境的基础上, 通过全景监控视频、全景动画视频、全景体育视频、全景电影视频和全景新闻视频进行实验测试, 各视频数据如表 1 所示。

实验过程中使用的所用视频都采用 AVI 格式, 每一段视频中图像帧的总长度为 2 000 帧~5 000 帧。

表 1 实验测试视频数据

全景视频类型	全景视频总帧数	镜头划分的视频片段
全景监控视频	5 068	86
全景动画视频	5 947	47
全景体育视频	3 642	68
全景电影视频	4 257	88
全景新闻视频	2 674	62

3.3 设置性能指标

为了检验文中系统的优越性, 在相同的实验条件下, 采用基于深度学习的提取系统、面向手语识别的提取系统和文中提取系统对表 1 中视频数据的关键帧进行提取, 测试了全景视频关键帧数据提取的查全率和查准率, 计算公式为:

$$Q = \frac{U_c}{U_c + U_f} \times 100\% \quad (15)$$

$$Z = \frac{U_c}{U_c + U_m} \times 100\% \quad (16)$$

式中, Q 代表全景视频关键帧数据提取的查全率, Z 代表查准率, U_c 代表正确提取的关键帧数据量, U_m 代表漏选的关键帧数据量, U_f 代表错误选择的关键帧数据量。

3.4 提取全景视频的关键帧数据

将表 1 中的视频数据导入到基于互信息熵的全景视频关键帧数据提取系统中, 以全景监控视频为例, 提取出全景监控视频中的关键帧数据, 结果如图 2 所示。



图 2 关键帧数据提取结果

由图 2 可知, 文中系统能够准确提取出全景监控视频的关键帧数据, 而且关键帧数据提取的响应频率也比较稳定。

3.5 性能分析

在性能测试中, 采用基于深度学习的提取系统、面向手语识别的提取系统和文中提取系统, 分别提取表 1 中 5 个类型视频的关键帧数据, 全景视频关键帧数据提取的查全率和查准率测试结果如图 3、图 4 所示。

由图 3 可知, 采用基于深度学习的提取系统和面向手语识别的提取系统时, 提取全景视频关键帧数据时的查全率在 60%~80% 之间, 而文中系统提取全景视频关键帧数

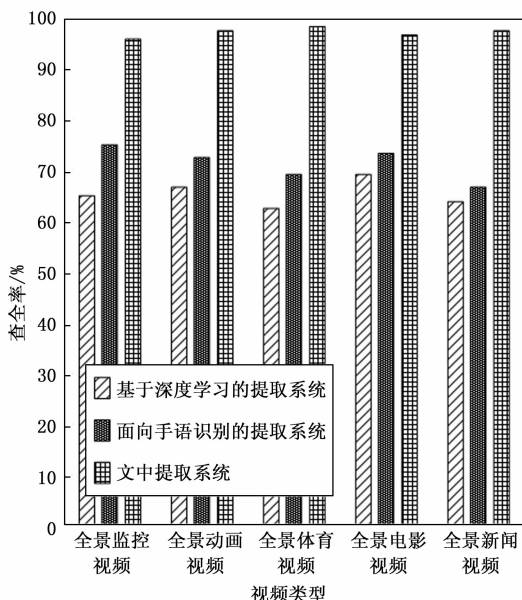


图 3 全景视频关键帧数据提取的查全率

据时的查全率在 95% 以上, 说明采用文中系统时, 全景视频关键帧数据提取错误的概率较低, 能够保证关键帧数据提取的准确性。

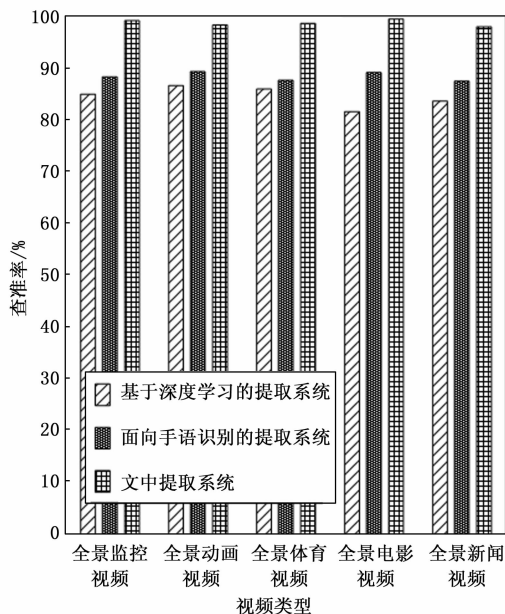


图 4 全景视频关键帧数据提取的查准率

由图 4 可知, 在查准率测试中, 3 个系统的性能较查全率都有所提高, 基于深度学习的提取系统和面向手语识别的提取系统将全景视频关键帧数据提取的查准率提升到 80% 以上, 文中系统对 5 种类型视频关键帧提取的查准率可以达到 98% 以上, 能够减少关键帧数据漏选的数量。

4 结束语

将互信息熵应用到全景视频关键帧数据实时提取系统

设计中。将待检测的两帧图像随机划分为子图像块, 通过计算子图像块之间的互信息熵, 获取全景视频图像的突变帧, 将关键图像帧的特征差值曲线作为全景视频关键帧的时序特征, 完成全景视频中关键帧数据的特征匹配, 选择一个能够体现全景视频图像属性的特征, 描述视频中的主要信息, 将全景视频帧的特征转换成数组形式的特征向量, 通过计算互信息熵值, 提取全景视频关键帧数。经过实验测试发现, 该系统能够准确提取出全景监控视频的关键帧数据, 并提高关键帧数据提取的查全率和查准率。

目前大多数视频都是经过压缩后的视频, 如果不经过解压直接提取关键帧数据, 就会加大关键帧数据提取的时间开销, 在今后的研究中, 希望可以先对视频进行解压处理, 保证关键帧提取的质量。

参考文献:

- [1] 朱树明, 王凤随, 程海鹰. HEVC 压缩域的视频摘要关键帧提取方法 [J]. 信号处理, 2019, 35 (3): 481-489.
- [2] 毋立芳, 赵宽, 简萌, 等. 级联手工特征与深度特征的视频关键帧检测方法 [J]. 信号处理, 2019, 35 (11): 1871-1879.
- [3] 罗志伟, 刘持标. 基于 H. 266/通用视频编码的帧内双模式的视频水印算法 [J]. 计算机应用, 2021, 41 (2): 456-460.
- [4] 梁建胜, 温贺平. 基于深度学习的视频关键帧提取与视频检索 [J]. 控制工程, 2019, 26 (5): 965-970.
- [5] 周舟, 韩芳, 王直杰. 面向手语识别的视频关键帧提取和优化算法 [J]. 华东理工大学学报 (自然科学版), 2021, 47 (1): 81-88.
- [6] 安利智, 何平, 张薇, 等. 动态卷积生成对抗网络的视频帧预测方法研究 [J]. 小型微型计算机系统, 2022, 43 (2): 278-284.
- [7] 张航, 卢小平, 张晓强, 等. 面向矿山监管的无人机视频关键帧影像动态提取方法 [J]. 遥感信息, 2020, 35 (1): 112-116.
- [8] 王俊玲, 卢新明. 基于语义相关的视频关键帧提取算法 [J]. 计算机工程与应用, 2021, 57 (4): 192-198.
- [9] 蔡冠蓝. 柔性姿态估计和时空特征结合的乒乓球动作视频片段关键帧提取 [J]. 科学技术与工程, 2019, 19 (25): 268-272.
- [10] 卞鹏程, 郑忠龙, 李明禄, 等. 基于注意力融合网络的视频超分辨率重建 [J]. 计算机应用, 2021, 41 (4): 1012-1019.
- [11] 熊炜, 王传胜, 管来福, 等. 基于特征跟踪和网格路径运动的视频稳像算法 [J]. 计算机工程与科学, 2020, 42 (5): 843-850.
- [12] 黄胜, 李萌芳, 张倩云. 基于背景帧的监控和会议视频帧间快速算法 [J]. 计算机工程与应用, 2020, 56 (22): 205-210.
- [13] 张德正, 翁理国, 夏旻, 等. 基于深度卷积长短时神经网络的视频帧预测 [J]. 计算机应用, 2019, 39 (6): 1657-1662.

(下转第 263 页)