

# 基于视频聚类的关键帧提取算法

刘华咏, 郝会芬, 李 涛

(华中师范大学 计算机学院, 湖北武汉 430079)

**摘 要** :关键帧可以有效减少视频索引的数据量,是分析和检索视频的关键。在提取关键帧过程中,为了解决传统聚类算法对初始参数敏感的问题,提出了一种改进的基于视频聚类的关键帧提取算法。首先,提取视频帧的特征,依据帧间相似度,对视频帧进行层次聚类,并得到初始聚类结果;接着使用K-means 算法对初始聚类结果进行优化,最后提取聚类的中心作为视频的关键帧。实验结果表明该方法可以大幅提高关键帧的准确率和查全率,能较好地表达视频的主要内容。

**关键词** :关键帧;特征提取;层次聚类;K-means 算法

**中图分类号** :TP391.4

**文献标识码** :A

**文章编号** :2095-1302(2014)08-0059-03

## 0 引 言

随着互联网和多媒体技术的飞速发展,形象生动的数字视频已经逐渐取代单调的文本信息,成为了人们网络生活中传播信息的重要方式之一。面对互联网上大量的视频,能否在较短的时间内找到需要的视频片段,已经成为了人们越来越关注的问题。在视频帧序列中,包含视频重要内容的帧可以简单地概括视频的主要内容,称为视频的关键帧。关键帧的提取技术在基于内容的视频检索中有着举足轻重的地位。在实际应用中,关键帧的提取技术可以分为以下4大类:

(1) 基于运动分析的关键帧提取技术。运动分析一般是基于流光运算的,通过分析和计算光流得出视频序列的运动量。然后比较运动量的值,并选取局部最小值处的帧为关键帧。这种方法提取关键帧的最大优点是:针对不同结构的镜头,可以根据实际情况提取数量合适的帧。但这种方法计算复杂,时间开销大,而且由局部最小值得到的帧不一定能准确描述视频内容。

(2) 基于镜头边界的关键帧提取技术<sup>[2-3]</sup>。这种方法首先将视频分割成若干个镜头,然后在每个镜头内部分别提取第一帧、中间帧和最后一帧作为关键帧。这种方法容易设计,计算简单,适合视频内容简单或场景固定的情况,但当镜头变换频繁且变换方式多样时,有可能导致提取的关键帧不能准确地描述视频的内容。

(3) 基于视觉内容的关键帧<sup>[4-6]</sup>提取技术。该方法根据每一帧图像的纹理、颜色等视频信息的变化来提取关键帧。这种方法可以根据视频内容变化的显著程度灵活的确

定要选取的关键帧的数目,但缺点是:当镜头变化频繁时容易导致选取的关键帧过多,造成信息冗余。

(4) 基于聚类分析的关键帧<sup>[7-9]</sup>提取技术。该方法充分考虑了镜头内以及镜头间的相关性,依据帧图像间相似度的大小,将视频帧序列进行聚类,然后依次从每类中选取一帧作为关键帧。大多数情况下,基于聚类分析提取关键帧能准确的描述视频主要的内容。但该方法最大的不足之处在于:需要在聚类前设定好聚类的数目和聚类中心。在视频内容不确定的情况下,提前设定聚类的数目和中心是十分困难的,且运算时间较长,这一缺陷极大的制约了这类方法的进一步发展。

层次聚类算法无需提前设定聚类中心和聚类数目,但该算法收敛速度相对较慢。K-means 算法设计简单、容易实现,且收敛速度较快,但其对初始聚类中心较敏感,容易陷入局部最优解<sup>[10]</sup>。针对这两种算法在提取关键帧过程中出现的不足,本文提出一种改进的基于视频聚类的关键帧提取算法。该算法过程为:首先,依次提取帧图像的信息熵,并使用欧式距离公式计算帧间相似度;然后,运用层次聚类算法对所有帧进行聚类,得到初始聚类结果;运用 K-means 算法优化并完成最终聚类;最后,将距离聚类中心最近的帧作为关键帧输出。

## 1 基于视频聚类的关键帧提取过程

### 1.1 特征提取

1948年,信息论创始人 Shannon 首次提出信息熵的概念,用以描述随机变量  $X$  所包含的平均信息量。 $X$  的信息熵定义为:

$$H(X) = -\sum_{i=1}^n p(x_i) \log_2 p(x_i)$$

其中,  $X = \{x_i \mid i=1, 2, 3, \dots, n\}$  表示随机变量;  $p(x_i)$  表示  $x_i$  出现的概率。

对于一段包含  $N$  帧的视频序列:  $F = \{f_1, f_2, \dots, f_n\}$ , 假设

收稿日期:2014-04-28

基金项目:中央高校自主科研项目(No. CCNU10A01012)

该视频序列的每一帧均为 256 级的灰度图像。本文首先将图像分割为  $B$  块, 每块图像的信息熵定义为:

$$H = - \sum_{i=1}^{255} p(x_i) \log_2 p(x_i)$$

其中,  $p(x_i)$  表示图像中灰度值为  $i$  的像素出现的概率。图像的特征向量定义为:

$$F = \{H_j \mid j=1, 2, 3, \dots, B\}$$

其中,  $H_j$  表示第  $j$  块图像的信息熵。

## 1.2 计算帧间相似度

我们用欧式距离表示帧间的相似度。欧式距离越小, 帧间相似度就越高。帧间的欧式距离定义为:

$$d(F_a, F_b) = \sqrt{\sum_{r=0}^{N(N-1)} (F_a[r] - F_b[r])^2}$$

其中,  $N$  为视频中帧的数量,  $F_a$  和  $F_b$  分别表示图像  $a$  和图像  $b$  的特征向量,  $d(F_a, F_b)$  为帧间的相似度。

## 1.3 利用层次聚类得到初始聚类结果

本文使用凝聚型的层次聚类算法对视频序列进行初始聚类, 其主要思想是: 先将每个待聚类的样本做为一类, 然后依据帧间相似度的大小, 合并相似的类使之成为一类, 直到满足终止条件。

假设视频序列有  $N$  个待聚类的样本, 基本步骤为:

(1) 确定终止条件。设要提取的关键帧的数量为  $K$ 。计算特征向量的均值  $M$  和方差  $V$ :

$$M = \frac{1}{N} \sum_{i=1}^{N-1} D_i \quad V = \frac{1}{N} \sum_{i=1}^{N-1} (M - D_i)^2$$

其中,  $D = \{d_i \mid i=1, 2, 3, \dots, N\}$  表示距离向量。本文中,  $K$  的值与帧间距离  $d(F_a, F_b) > M + 2 \cdot \sqrt{V}$  的帧的数量相等。

(2) 初始化: 首先, 将每个样本单独归为一类, 依次计算每两类之间的相似度;

(3) 寻找距离最近的两个类, 并把他们归为一类;

(4) 重新计算新生成的这个类与各个旧类之间的相似度;

(5) 重复 3 和 4, 直到聚类数目为  $K$ , 结束。

## 1.4 利用 K-means 聚类算法优化聚类结果

本文将层次聚类产生的结果作为 K-means 算法的输入, 将层次聚类产生的每一类的中心, 作为 K-means 算法的初始聚类中心, 这样可以避免随机产生初始聚类中心对 K-means 算法的聚类结果产生的影响。由于层次聚类后, 样本已经基本聚类成功, 故 K-means 算法迭代的次数将大大减少。

具体步骤如下:

(1) 计算层次聚类产生的聚类的中心, 即每个聚类对象的平均值;

(2) 计算每一帧与聚类中心的距离, 并根据最小距离准则对相应帧重新进行分类;

(3) 重新计算每个类的聚类中心;

(4) 重复 2 和 3, 直到每个聚类中的对象不再发生变化;

(5) 计算每一类的聚类中心, 将距离聚类中心最近的帧作为关键帧输出。

本文将每个视频帧分为  $B$  块并提取每块的信息熵, 然后组成特征向量, 充分考虑了每个视频帧的特征。同时, 该算法利用了层次聚类能在短时间内得到聚类结果的优势, 大大提高了工作效率。此外, 本文将已经初步分类成功的层次聚类的结果作为 K-means 算法的输入, 避免了人为定义初始聚类中心对 K-means 聚类结果的影响。

## 2 实验结果与分析

本文在 Intel i7, 2.4 GHz CPU, 4 GB 内存, Windows 8(64 位) 环境下和 Matlab 平台上实现该算法, 以检验该算法的有效性。本文使用的实验视频长度从几百帧到几千帧, 类型丰富全面, 包括广告、电影、新闻、卡通等。同时, 人工对数据集进行关键帧的识别统计作为算法的比较标准。

本文采用查全率和准确率这两个参量来检验算法的有效性<sup>[11]</sup>。其中, 查全率和准确率的定义如下公式所示:

查全率 = 正确检测数 / (正确检测数 + 漏检数)

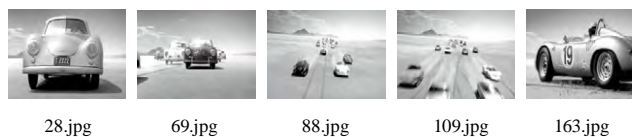
准确率 = 正确检测数 / (正确检测数 + 误检数)

由于传统的直方图法<sup>[12]</sup> 广泛应用于关键帧的提取, 且该方法具有较好的检测性能和较快的运算速度, 故本文采用其作为对比实验。提取关键帧的结果如表 1 所列。

表 1 实验结果

数据集	总帧数	关键帧数目	视频镜头检测方法	查全率	准确率
广告	1 235	21	直方图	65.5%	62.1%
			本文	89.2%	90.5%
电影	1 087	31	直方图	64.1%	63.4%
			本文	87.8%	86.5%
新闻	796	18	直方图	74.8%	77.4%
			本文	81.4%	84.4%
卡通	1 348	23	直方图	70.4%	68.6%
			本文	86.9%	85.7%

由表 1 可知, 对于广告、电影、新闻、卡通等不同类型的视频片段, 使用本文的算法提取关键帧具有较大优势。例如, 在随机选取的卡通类视频中, 本文提出的算法查全率为 86.9% 和准确率为 85.7%, 比直方图法的查全率和准确率高很多。由此可见, 本文提出的算法能很好的描述视频的主要内容, 且查全率和准确率较高。本文选取了一段包含 1 348 帧的卡通视频序列, 以验证该算法提取的关键帧具有代表性, 部分关键帧如图 1 所示。



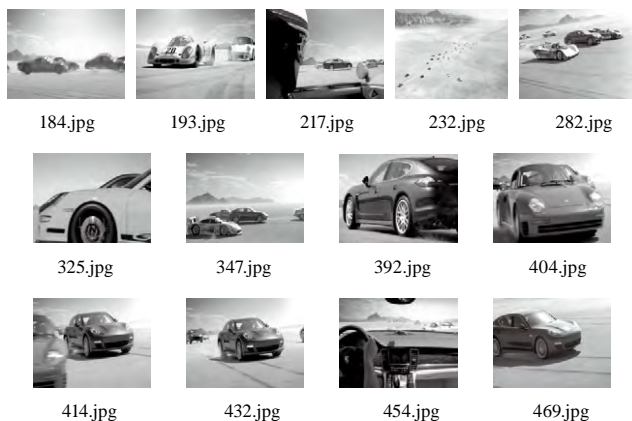


图1 新闻视频的关键帧

图1列举了从视频序列中提取的部分关键帧。其中,前10个关键帧内容变换较大,准确反映了汽车在空旷的平原上静止、加速、超车等一系列过程,因此提取为关键帧。后8个关键帧内容变化不大,关键帧序列特写了黑色轿车超过红色轿车的过程。从以上实验结果可以看出,本文提出的基于视频聚类的关键帧提取算法,有较高的查全率和准确率,提取的关键帧能较为准确的描述视频的主要内容。

### 3 结 语

本文提出了基于视频聚类的关键帧提取算法,有效地改进了传统聚类算法在提取关键帧过程中出现的不足。该算法将层次聚类和 K-means 算法的优势充分结合,并扬长避短,使用 K-means 算法对层次聚类的结果再次优化,最终得到了视频的关键帧。从大量的实验数据可以看出,该算法有很好的

适应性,可以针对不同类型的视频提取适当数目的关键帧,且提取关键帧的查全率和准确率高于传统方法。

### 参考文献

- [1] WOLFW. Key frame selection by motion analysis [C]//Proceedings of IEEE International Conference on Acoustic, Speech and signal processing. Washington D C, USA :IEEE Computer Society, 1996 :1228-1231.
- [2] MENDI E, BAYRAK C. Shot boundary detection and key frame extraction using salient region detection and structural similarity[C]//Proceedings of the 48th Annual Southeast Regional Conference. ACM, 2010 :66-68.
- [3] 方勇, 戚飞虎. 一种新的视频镜头边界检测及关键帧提取方法 [J]. 华南理工大学学报 :自然科学版, 2004( S1 ) :18-23.
- [4] 陆燕, 陈福生. 基于内容的视频检索技术 [J]. 计算机应用研究, 2003( 11 ) :1-4.
- [5] 丁洪丽, 陈怀新. 基于镜头内容变化率的关键帧提取算法 [J]. 计算机工程, 2009( 13 ) :225-227, 231.
- [6] 王松, 韩永国, 吴亚东, 等. 基于图像主色彩的视频关键帧提取方法 [J]. 计算机应用, 2013( 9 ) :2631-2635.
- [7] 朱映映, 周润汝. 一种基于视频聚类的关键帧提取方法 [J]. 计算机工程, 2004( 4 ) :12-13, 121.
- [8] 王方石, 须德, 吴伟鑫. 基于自适应阈值的自动提取关键帧的聚类算法 [J]. 计算机研究与发展, 2005( 10 ) :109-114.
- [9] 印勇, 蒋海娜. 优化初始聚类中心的关键帧提取 [J]. 计算机工程与应用, 2007( 21 ) :165-167.
- [10] 孙淑敏, 张建明, 孙春梅. 基于改进 K-means 算法的关键帧提取 [J]. 计算机工程, 2012( 23 ) :169-172.
- [11] 周祥东, 李国辉, 涂丹, 等. 一种新的视频镜头分割算法 [J]. 计算机工程与科学, 2004, 25( 6 ) :5-8.
- [12] LO C, WANG S J. A histogram-based moment-preserving clustering algorithm for video segmentation [J]. Pattern recognition letters, 2003, 24( 14 ) :2209-2218.

作者简介:刘华咏(1978—),男,湖北武汉,副教授,博士。研究方向为图像处理、语音信号处理及多媒体技术。

郝会芬(1988—),女,湖北武汉,硕士。研究方向为图像视频处理及多媒体技术。

李涛(1989—),男,湖北武汉,硕士。研究方向为图像处理、视频摘要分析。

## Key Frame Extraction Algorithm Based on Video Clustering

LIU Hua-yong, HAO Hui-fen, LI Tao

(Department of Computer Science, Central China Normal University, Wuhan 430079, China)

**Abstract:** Key frame can dramatically reduce the data of video indexing, and it is the fundamental processes in video analysis and video retrieval. In order to solve the problems that the traditional clustering algorithm is sensitive to the initial parameter in key frame extraction process, we propose an improved key frame extraction algorithm based on video clustering. Firstly, we extract the features of video frames. And the hierarchical clustering algorithm is used to obtain an initial clustering result, according to the similarity between two video frames. Then, K-means algorithm is conducted to optimize the initial clustering result and obtain the final clustering result. Finally, the center frame of each clustering is extracted as key frame. Experimental results show that the precision and recall ratio of our proposed algorithm are greatly improved. The key frames extracted by our algorithm are better to express the primary content of video.

**Keywords:** keyframe; feature extraction; hierarchical clustering; K-means algorithm