

基于仿射传播聚类的自适应关键帧提取

许文竹 徐立鸿

(同济大学电信学院 上海 200092)

摘 要 关键帧提取技术,是基于内容的视频检索的一个重要组成部分。为了能从不同类型的视频里有效地提取关键帧,提出了基于仿射传播聚类的自适应关键帧提取算法。首先通过图像的颜色特征获取视频镜头的相似性矩阵,然后通过仿射传播聚类自适应地提取视频关键帧。该算法从视频的本身信息分布出发,自适应地搜索出视频最优关键帧,且运算速度快。实验表明,该算法能有效地提取出视频最优关键帧,且算法快速稳健。

关键词 视频检索,关键帧,视频镜头,仿射传播聚类

中图法分类号 TP391 文献标识码 A

Adaptive Key-frame Extraction Based on Affinity Propagation Clustering

XU Wen-zhu XU Li-hong

(School of Electronics and Information, Tongji University, Shanghai 200092, China)

Abstract Key-frame Extraction is important to content-based video analysis. For extracting key frames efficiently from different video, this paper presented an efficient method for key-frame extraction in which affinity propagation clustering is applied to key-frame extraction. First get the shot comparability matrix by histogram differences, extract key frames adaptively through affinity propagation clustering. The algorithm which this paper proposed can get the key frames adaptively according to the content of video shot and fast. The experiment indicates the algorithm is efficient in key-frame extraction and has enough robustness.

Keywords Video indexing, Key-frame, Video shot, Affinity propagation clustering

1 引言

随着多媒体技术和网络技术的高速发展,产生了大量的视频文档,视频数据信息越来越多。传统的基于关键词描述的视频检索由于其描述能力有限、工作劳动力大、主观性强等很多客观因素,已经不能适应海量视频检索的需要。关键帧提取技术能对视频数据库进行快速的检索、访问,已经成为基于内容的视频检索的一个重要组成部分,越来越受到人们的重视。

2 关键帧提取方法

目前常用的关键帧提取技术大致可以分为以下几类^[1]:

(1)基于镜头边界的方法,把镜头的第一帧和最后一帧或者中间帧作为关键帧,这类方法简单,但不能很好地反映镜头内容的复杂度,当镜头较长或者摄像机在运动时,该方法就不适合;(2)基于视觉内容分析的方法,利用视频每一帧的颜色、纹理、边缘等视觉信息的改变实现关键帧的提取,它可以根据镜头内容的变化程度选择相应数目的关键帧,但所选的关键帧不一定具有代表意义,而且有镜头运动时,容易选取过多的关键帧;(3)基于运动分析的方法^[2],通过光流分析来计算镜头中的运动量,在运动量局部最小值处选取关键帧,这种算法的

缺点是计算量大。

尽管在关键帧的提取技术上已有了许多研究成果,但寻求能涵盖整个镜头内容,自动快速地提取出视频关键帧仍是一个难题。因此本文提出了一种新的视频关键帧提取算法,该算法能够根据视频本身内容的分布,自适应地快速提取出视频关键帧。该算法首先通过视频帧的颜色特征计算出视频镜头的相似性矩阵,然后通过仿射传播聚类自适应地提取视频关键帧。该算法不需要用户提供聚类个数或者聚类中心等任何参数,克服了运用 K 均值算法时用户难以确定聚类个数和初始聚类中心以及因初始聚类中心的随机性而带来结果的不稳定性等缺陷。该算法能根据视频内容复杂度和视频内容信息分布,从视频数据本身的结构出发,自动确定最优的聚类中心,快速找出视频关键帧,同时其运算量小,运算速度快且算法稳健。

3 特征提取

颜色是图像内容最基本的元素,也是图像最直观的特征。选择一个符合人眼视觉特性的颜色空间对于视频检索是至关重要的^[3]。HSV 颜色空间不但反映了人眼视觉观察彩色的方式,同时也有利于图像处理。HSV 颜色空间有两个重要的特点:第一,V 分量(即亮度)与彩色信息无关;第二,H 分量

到稿日期:2009-02-25 返修日期:2009-04-29 本文受国家高技术研究发展计划(863 项目)(No. 2008AA10Z227)资助。

许文竹(1979—),女,博士生,主要研究方向为基于内容的视频检索,E-mail: xuwenzhu@163.com;徐立鸿(1960—),男,教授,博士生导师,主要研究方向为图像处理。

(即色调)和 S 分量(即饱和度)与人感受彩色的方式紧密相连。这些特点使得 HSV 颜色空间非常适合于借助人眼视觉系统来感知彩色特性的图像处理方法。

在确定采用 HSV 空间模型后,将色调 H 空间、饱和度 S 空间、亮度 V 空间 3 个分量按照人的颜色感知进行非等间隔的量化。其中色调 H 空间被分成 8 份,饱和度 S 空间被分成 3 份,亮度 V 空间被分成 3 份,因此整个 HSV 空间被分为 72 个子空间($8 \times 3 \times 3$),把这 3 个颜色分量(H, S, V)通过式(1)合成一维特征矢量:

$$l = 9H + 3S + V \quad (1)$$

通过统计帧图像在各个子空间上的像素个数,即可得到该帧的 HSV 颜色直方图矢量,该矢量有 72 个分量,即 0 到 71。在得到每个帧图像的 HSV 直方图后,需要选择合适的距离度量来表示帧图像之间直方图的差异^[4],这可以简单地用绝对值距离来表示,计算简单,且满足计算速度要求。

$$D(H_i, H_j) = \sum_{k=0}^{71} |H_i(k) - H_j(k)| \quad (2)$$

H_i 是第 i 帧图像的直方图矢量, H_j 是第 j 帧图像的直方图矢量, $D(H_i, H_j)$ 是第 i 帧图像和第 j 帧图像之间的直方图距离。

这里定义视频帧之间的相似性矩阵 S 为视频帧直方图之间距离的负数,即 2 帧图像直方图距离越大,其相似性越小,2 帧图像直方图距离越小,其相似性越大。

$$S(H_i, H_j) = -D(H_i, H_j) \quad (3)$$

$S(H_i, H_j)$ 表示第 i 帧图像和第 j 帧图像之间的相似性。

4 关键帧提取

仿射传播聚类是由 Frey 等人提出的一种新的聚类算法^[5]。其优势是算法快速、有效。它与 K 均值算法都是属于 K 中心聚类方法。然而 K 均值算法需要用户指定聚类个数以及初始聚类中心,且对初始聚类中心的选择敏感,不同的初始聚类中心会导致不同的聚类结果。仿射传播聚类算法则克服了这些缺点,其迭代过程不断搜索合适的聚类中心,同时也使得聚类的适应度函数最大化,能给出比较准确的聚类结果,且运算速度快。

仿射传播聚类算法以所有样本的相似度组成的样本相似性矩阵 $S(N \times N)$ 为基础, N 为样本数。首先将数据集的全部样本点都视为候选的聚类中心;然后在循环迭代过程中,各个样本点相互竞争最终的聚类中心^[6]。处于聚类中心的样本点对于其他样本点的吸引力之和比较大,在竞争中胜出的可能性也较大;反之,处在边缘位置的样本点,其吸引度之和较小且胜出的可能性也较小。用 $R(i, k)$ 表示样本点 k 对样本点 i 的吸引度,即样本点 k 适合做为样本点 i 的类代表的度, $A(i, k)$ 表示样本点 i 对样本点 k 的归属感,即表示样本点 i 选择样本点 k 做为其类代表的适合程度。 $R(i, k)$ 与 $A(i, k)$ 越大,说明样本点 k 做为最终聚类中心的可能性越大。这样,通过每次的迭代循环进行消息的传递,不断更新样本点的吸引度和归属感,最终形成高质量的类代表和对应的聚类。

样本吸引度和归属度的计算公式为:

$$R(i, k) = S(i, k) - \max\{A(i, j) + S(i, j)\} \quad (4)$$

$(j \in \{1, 2, \dots, N\}, \text{但是 } j \neq k)$

$$A(i, k) = \min\{0, R(k, k) + \sum_j \{\max(0, R(j, k))\}\} \quad (5)$$

$(j \in \{1, 2, \dots, N\}, \text{但是 } j \neq k \text{ 且 } j \neq i)$

仿射传播聚类最主要的 2 个参数是偏向参数 $p(i)$ 和阻尼因子 l , 偏向参数 $p(i)$ 表示样本点 i 被选做聚类中心的倾向性,是置于相似性矩阵 S 对角线上的,因此当 $i=k$ 时:

$$R(k, k) = p(k) - \max\{A(k, j) + S(k, j)\} \quad (6)$$

从式(6)可以看出 $p(k)$ 较大时,得到的 $R(k, k)$ 吸引度也较大,同时 $A(i, k)$ 归属感也较大,从而类代表 k 做为最终聚类中心的可能性也就越大。因此增大或者减少偏向参数 p 的值可以增加或者减少聚类个数。通常在没有先验知识的情况下,将所有的偏向参数 $p(k) (k \in \{1, 2, \dots, N\})$,全都设定为相似性矩阵 S 元素的中值,然而在视频检索中对于内容变化缓慢的视频来说,这样做常常会造成冗余帧,即聚类个数太多,同时也会影响运算速度,针对这种情况我们设计了一个偏移量 b 来改变偏移参数 p ,首先把相似性矩阵 S 中的元素除去对角线上的元素即 $S(i, i)$ 后按照从小到大的顺序排列成 1 列, S_{sort} 就是排序后的序列,取 S_{sort} 中间位置 $p_{\text{media}} = \text{round}g(\text{size}(S_{\text{sort}})/2)$, round 是四舍五入取整函数, size 返回 S_{sort} 元素的个数。偏向参数 $p = S_{\text{sort}}(p_{\text{media}} + b)$, p_{media} 是排序后的相似性矩阵 S_{sort} 的中间位置,偏移量 b 由视频平均帧间差即 $\text{mean}(-S(i, i+1))$ 以及视频镜头长度 N 共同决定,当平均帧间差较大且镜头长度较大时,说明视频内容变化比较剧烈, b 就适当取大一点的正值,以保证偏向参数 p 比较大,当平均帧间差较小时,说明视频内容变化比较缓慢, b 就取负值,以保证偏向参数 p 比较小。

在每次循环迭代过程中,都以式(4)~式(6)更新一次样本点的吸引度 R_i 和归属感 A_i ,此时有一个重要的参数即阻尼因子 l ,阻尼因子反映了本次迭代后的样本吸引度 R_i 和归属感 A_i 与上次的样本吸引度 R_{i-1} 和归属感 A_{i-1} 之间的关系,即:

$$R_i = R_i * (1-l) + R_{i-1} * l \quad (7)$$

$$A_i = A_i * (1-l) + A_{i-1} * l \quad (8)$$

从式(7)和式(8)可以看出,阻尼因子 l 也可以改进算法的收敛性,在本文中取值为 0.5。

仿射传播聚类算法输出的聚类数目很大程度上依赖于偏向参数 p ,经过加入偏移量 b ,使得 p 在很大程度上得到了优化,加速了运算速度,而且使得聚类结果更为准确。为了保证结果是最优的(为了使结果更加可靠),从聚类有效性指标即在运行过程中以初始给定的 p 值出发,循环过程中每次更新吸引度 R_i 和归属感 A_i ,当循环过程中收敛到某一个类数 K ,减小参数 p 后,继续循环迭代,从而获得不同的聚类个数和聚类结果,通过聚类有效性指标来评价聚类结果的质量,从而得到最优的聚类结果,即最优的关键帧。关于聚类有效性指标本文使用了应用广泛、计算简单的 Silhouette 指标,对于每一个样本 i ,其中 $a(i)$ 是样本 i 到同类其他所有样本的平均距离,而 $b(i)$ 是样本 i 到最近的类中所有样本的平均距离。样本 i 的 Silhouette 指标为:

$$\text{sil}(i) = [b(i) - a(i)] / \max(b(i), a(i)) \quad (9)$$

一个聚类中所有样本的 sil 平均值反映了聚类结果的质量^[7],平均值越大表示聚类结果越好,其最大值对应的聚类个数和聚类结果对应着最优的聚类个数和聚类结果。本算法主要流程是:

(1)首先判断视频镜头的平均帧间差 $\text{mean}(-S(i, i+1))$ 是否小于 T , T 是一个很小的数,本文取值为 0.0001,如果

视频镜头的平均帧间差小于 T , 则说明该镜头内容变化非常缓慢, 为了计算方便就取镜头的中间帧作为关键帧, 算法结束。若视频的平均帧间差大于 T , 就进入流程(2)。

(2) 初始化偏向参数 $p = S_{\text{err}}(p_{\text{media}} + b)$, 定义最大循环次数 $mcount = 50000$, 收敛条件为聚类中心 50 次循环无变化, 初始化吸引度 R_i 和归属度 A_i 为 0, 每次偏向参数 p 下调步伐 $p_{\text{sep}} = 0.1 * p_{\text{min}}$ (p_{min} 是偏向参数 p 的最小值)。

(3) 按照式(4)一式(8)进行循环迭代, 不断更新吸引度 R_i 和归属度 A_i , 算法产生 k 个聚类中心。

(4) 判断是否收敛, 若收敛或者是趋于收敛, 则转入流程(5), 若聚类个数不下降且聚类指标值下降, 则 $p = p + p_{\text{sep}}$, 转入流程(5), 否则转入流程(3)。

(5) 计算 Silhouette 指标值, 判断 k 是否小于等于 2, 循环次数是否等于 $mcount$, 若聚类中心个数小于等于 2, 或者循环次数达到最大值, 则转入流程(6), 否则减小参数 $p = p + p_{\text{sep}}$, 转入流程(3)。

(6) 找出 sil 最大值所对应的聚类个数, 即最优的聚类结果, 其聚类中心就是最优的关键帧。

5 实验结果

为了验证上述算法的有效性, 采用了不同类型的实验视频, 包括: MTV、足球、广告、新闻不同类型的视频, 视频序列从几百帧到几万帧, 视频帧大小各不相同。其基本情况如表 1 所列。

表 1 实验数据集

样本	帧数	本算法检测出的关键帧数
MTV	449	34
足球	3525	85
广告	577	24
新闻	14862	71

以两段典型的视频为例, 在视频内容变化缓慢的广告视频第一个镜头中, 应用本算法找出的关键帧如图 1 所示。



图 1 提取的广告关键帧

镜头 1 有 168 帧, 应用本算法提取的关键帧有 3 帧, 反映了视频的运动信息, 包括昂头、平视、侧视这 3 个动作, 没有冗余。

在足球运动视频中, 以其中一个目标运动变化剧烈的镜头为例, 应用本算法找出的关键帧如图 2 所示。



图 2 提取的足球运动关键帧

我们可以看到通过这 3 帧反映了人物的移动, 表达了视频镜头的主要内容, 尤其是运动信息。

从以上的实验结果分析可以看到, 不管是对目标对象运动缓慢的视频, 还是对目标运动剧烈、镜头快速移动的视频, 本算法都能有效地找出代表视频主要内容和运动信息的最优关键帧, 而且算法稳健快速。

结束语 本文提出了基于仿射传播聚类的自适应关键帧提取算法, 克服了运用 K 均值算法时难以确定聚类个数以及初始聚类中心的随机性带来的结果的不稳定性等缺陷, 本算法能够根据视频本身的信息分布以及视频的复杂度, 自适应地提取出最优关键帧, 在一系列不同类型的视频片断上进行实验和分析, 实验的提取结果与人的视觉有良好的一致性, 证明了该算法能快速有效地提取出关键帧, 并且具有良好的鲁棒性。

参考文献

- [1] Zhang Hongjiang, Wang J Y A, Altunbasak Y. Content - based video retrieval and compression: A unified solution[A]// IEEE International Conference on Image Processing[C]. Washington, DC, 1997: 13-16
- [2] Worf W. Key frame selection by motion analysis[C]// Proceedings of the 1996 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Atlanta, 1996: 1228-1231
- [3] 伯晓晨, 刘建平. 基于颜色直方图的图象检索[J]. 中国图像图形学报, 1999, 4(1): 33-37
- [4] 邢伟利. 图像检索中颜色的特征提取及匹配算法[J]. 微机发展, 2002(2): 86-88
- [5] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976
- [6] 王开军, 张军英, 李丹, 等. 自适应仿射传播聚类[J]. 自动化学报, 2007, 12(33): 1242-1245
- [7] Velamuru P K, Renaut R A, Guo H B, et al. Robust clustering of positron emission tomography data[C]// Joint Interface CSNA. USA, 2005
- [8] Hettich S, Bay S D. The UCI KDD Archive[OL]. Irvine, CA: University of California, Department of Information and Computer Science. <http://kdd.ics.uci.edu/database/kddcup99/1999>
- [9] 施伟, 战守义, 盛思源. 基于 Rough Set 的数据预处理[J]. 计算机工程与应用, 2003, 39(22): 190-191
- [10] Forrest S, Allen L, Cherukuri R. Self-nonself discrimination in a computer[C]// Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy. IEEE Computer Society Press, 1994: 271-281
- [11] 徐宗本, 张讲社, 郑亚林. 计算智能中的仿生学: 理论与算法[M]. 北京: 科学出版社, 2003

(上接第 242 页)

- [5] Kim J. Towards an artificial immune system for network intrusion detection: an investigation of clonal selection with a negative selection operator[C]// Proc. of the Congress on Evolutionary Computation. Seoul, Korea, 2001: 1244-1252
- [6] 焦李成, 杜海峰, 刘芳, 等. 免疫优化计算、学习与识别[M]. 北京: 科学出版社, 2006
- [7] Fang Xianjin. An Artificial Immune Model with Vaccine operator for Network Intrusion detection[C]// Proc. of 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application. IEEE CS Press, 2008: 488-491