文章编号: 1001-2486 (2000) 05-0103-06

一种自动镜头聚类方法*

熊华[1], 胡晓峰[2], 老松杨[1]

(1. 国防科技大学多媒体研究开发中心, 湖南 长沙 410073; 2. 国防大学模拟中心, 北京 100091)

摘 要: 镜头聚类是视频内容分析的重要途径。镜头聚类的基本任务是基于镜头的物理特征对镜头进行分类。本文设计和实现了一种新的镜头聚类方法, 这种方法从一个初始分割开始, 经多次聚类分裂与合并的迭代, 自动地进行误差校正。这种方法既不需要通过人工交互来解决试探聚类方法的误差调节问题, 也不需要迭代聚类算法中难以确定的经验参数和经验阈值的设定, 克服了普通聚类方法的缺点, 在实际应用系统中取得了较好的效果。

关键词: 镜头聚类; 合并和分裂; 视频内容分析; 基于内容检索
中图分类号: TP311 文献标识码: A

An Automatic Shot Clustering Method

XIONG Hua[1], HU Xiao-feng[2], LAO Song-yang[1]
(1. Multimedia R & D Center, National Univ. of Defense Technology, Changsha 410073, China;
2. Simulation Center of National Defense University, Beijing 10009, China)

Abstract: Shot clustering is an important aspect of video content analysis. The basic task of shot clustering is to classify shots based on their low-level features. This paper describes a novel shot-clustering technique. Beginning with an initial classification of the shot set, our algorithm proceeds with merging and splitting iteration alternatively to reduce the errors in the initial results. The main advantage of this algorithm is that it does not need any experiential parameters or thresholds, nor does it need any manual interaction. In this way, our algorithm overcomes shortcomings of traditional clustering algorithm and works well in practical systems.

Key words: shot clustering; merge and split; video content analysis; content-based video retrieval

镜头聚类是视频内容分析领域的一个重要问题, 其基本思想是用聚类方法对长视频的各组成镜头进行自动分类, 以期发现视频内容结构。许多研究小组涉足过镜头聚类[1-3]这一课题, 但都存在一定的问题, 主要表现为在聚类过程中需要人工交互[1]或需要在算法开始之前设置主观参数或阈值[2,3]。本文根据视频镜头特征数据的特殊性, 设计和实现了一种新的镜头聚类方法。这个镜头聚类方法有两个重要特点: 一个是相对于迭代聚类方法[4]而言, 不需设置经验参数; 另一个是相对于试探聚类方法[4]而言, 聚类方法提供了相当的精度, 这个精度不是通过人工交互得到的, 而是通过分裂和合并自动校准的。

1 镜头聚类算法的设计

我们的研究目标是要根据由低层特征计算得到的距离矩阵的指示把相似的镜头聚在一起, 得到场景层次上的视频内容单元。整个聚类算法从一个初始分割开始, 经多次聚类分裂与合并的迭代, 自适应地达到"最好的"聚类效果。初始迭代的结果聚类集(称为初始聚类集)是后续聚类迭代的基础。初始迭代结束后, 算法按照合并与分裂的准则对初始聚类集进行交替迭代的合并、分裂处理, 直到满足算法终止的条件为止。算法的设计任务包括初始迭代算法的设计, 合并与分裂的准则设计, 聚类合并与分裂的迭代设计, 以及整个算法的终止条件的设计。后续各小节将分别讨论这些问题。流程见图 1, 图中, 判断 1 判断本次迭代是否实际上未做任何操作, 判断 2 判断前次迭代是否实际上未做任何操作, 当连续两次迭代未执行任何操作时迭代终止。☆表示各次迭代的输出结果聚类集, 是紧跟其后的迭代的输入。

* 收稿日期: 2000-04-10
基金项目: 国家部委基金项目资助(15.8.3)
作者简介: 熊华(1973-), 女, 博士生。

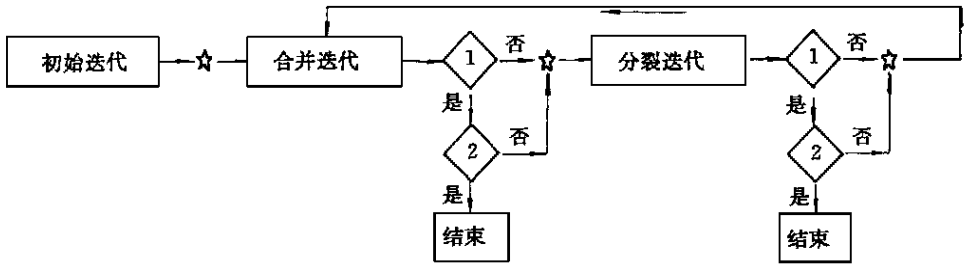

图 1 镜头聚类算法的整体流程

Fig. 1 Shot clustering scheme

1.1 初始迭代算法设计

初始迭代的任务是对被聚类镜头集进行一个大致的划分, 尽量把相似的镜头放在同一个聚类中, 为后续的分裂合并微调打好基础。

距离矩阵是初始迭代的唯一输入, 距离矩阵的元素描述了待聚类样品镜头集中两两镜头之间的距离。这个距离可由镜头的任何低层特征或低层特征的组合的相似性测度计算得到, 距离越近的两个镜头(按照某种低层特征) 越相似。距离矩阵是个对称矩阵。

初始迭代算法的基本原则是距离最近的镜头最有资格被聚为一类, 依据这一原则, 初始迭代算法按照距离矩阵中最小距离排名依次聚类相应的镜头。首先将距离最近的两个镜头聚为一类, 然后将次近的镜头聚为一类, 依此类推, 直到所有镜头都被有所归属。在次小距离(大于等于当前最小距离的最小距离) 搜索过程中, 若构成次小距离的两个镜头之一已被聚入某类 A, 则将另一镜头也聚入类 A, 不另外成立新类。

1.2 后续迭代算法设计

初始迭代的初步聚类结果还存在一些可修正的误差, 后续迭代的任务就是对初始迭代的聚类结果进行“微调”, 修正这些误差。有两种误差需要修正: 一种是有些类分得过小, 需要通过合并操作来修正; 另一种是有些类中包含噪声点, 使得这些类的半径过大, 需要通过分裂操作来修正。合并操作和分裂操作应该交替进行, 直到再没有需要合并的聚类对和需要分裂的聚类。

1.2.1 合并迭代设计

借用几何上关于两圆相交的定义, 将聚类合并准则定义为: 当两个聚类的质心距离小于两个聚类半径之和时, 两个聚类可以合并。只有通过合并判断, 满足合并准则的两个聚类才 能被合并。见图 2, 两个椭圆分别表示两个聚类, 实线两端圆点分别为两类质心, 实线长度为两类质心距离, 两条带箭头线分别表示两类半径。该准则中的质心距离指两个聚类的质心镜头之间的距离, 可由距离矩阵中查到。质心是一个聚类中所有样品点的中心, 一般由所有样品矢量的平均值计算而得。由于初始迭代算法的唯一输入是距离矩阵, 样品矢量的值不可见, 用另外一种方法间接得到尽可能精确的质心。定义一个类中的质心为该类中的一个镜头, 满足: 该类中其它所有镜头离该镜头距离之和最小。算法实现中, 通过计算由该类所包含的镜头所构成的缩减距离矩阵中的最小行和来得到相应的质心镜头。聚类半径等于离质心最远的镜头距质心镜头的距离。

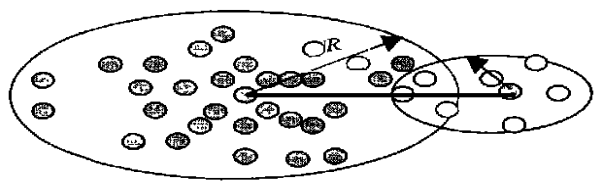

图 2 应合并的两类: 质心距离大于半径的例子

Fig. 2 A cluster pair need to merge

合并判断受噪声点影响较大。若聚类 C_n 包含噪声镜头 S_n, 则 C_n 的半径 d_n 会明显增大。这时 C_n 的质心镜头离其它聚类质心的距离很容易小于半径 d_n, 从而小于 d_n 与其它聚类半径的和。事实上, 一个强烈的噪声点会导致某个聚类的半径奇大, 其结果会导致多个其它类(常常是毫不相干的类) 被并入这

个含强烈噪声点的类, 这样很快会使聚类效果恶化, 得到荒唐的结果, 使前面的初始聚类成果化为乌有。

为了防止噪声的影响, 给上述合并准则加了如下合并约束: 聚类半径大于等于整个聚类集平均半径的聚类不能参加合并。通过实验, 发现这个约束能够很好地避开那些不该合并的聚类对, 在一定程度上抑制了噪声对合并判断的影响。不过这个约束也把那些虽含噪声镜头, 但其它元素相距很近的、确实需合并的聚类对“拒之门外”。所以在聚类合并准则和合并约束下的合并操作仍需分裂操作的辅助来进一步降低噪声的影响。

另外为了防止同一个镜头不停地被相邻的合并、分裂迭代操作, 产生所谓的“振荡”效应, 我们规定若被合并聚类中包含单镜头聚类, 则该单镜头不能是最近才从候选聚类对中的另外一个聚类中剔除的镜头。

1.2.2 分裂迭代设计

聚类分裂是指将一个聚类分裂为两个子类, 以使各类的凝聚度更高, 改善整体聚类效果。在任何被聚类镜头集中都会存在一些零散镜头, 不应该归于任何类, 如新闻片、广告片的片头镜头。但在初始迭代算法中却未考虑零散镜头的情况, 所有的镜头都按照就近原则被硬性分到各个聚类中。初始迭代算法中的这种强制分类使得零散镜头成为其所在聚类的噪声点, 需要通过分裂操作来剔除。

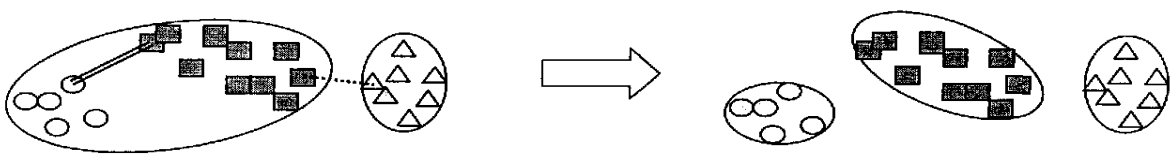

图 3 应分裂的类

Fig. 3 Cluster need to split

我们认为只有当一个聚类的类内空洞大于类间空隙时, 该聚类才能被分裂, 见图 3, 左侧为分裂前的情况, 右侧为分裂后的情况。由圆点和方点构成的椭圆为应分裂的类。双实线表示类内空洞, 虚线表示类间空隙。由于类内空洞大于类间空隙, 应该执行分裂操作。两个聚类的类间空隙指这两个聚类所含元素之间的最近距离。任何一个聚类都可看作由两个或多个子类构成, 类内空洞就是指这些子类之间的空隙。当某聚类的类内空洞大于类间空隙时, 说明该聚类的子聚类与其它已存在的聚类实际上位于相同的颗粒层次, 应该通过分裂把该聚类的子聚类释放出来。将聚类分裂准则定义为: 当某聚类的最大类内空洞大于等于整个聚类集的最小类间空隙时, 该聚类可以被分裂。只有通过分裂判断, 满足分裂准则的聚类才能被分裂。注意采用这种分裂准则的前提是被分类样品集所包含的各类密度较相似, 不存在有的类特密, 而有的类特疏的情况。或者毋宁说, 采用这种分裂准则分出来的聚类集, 各聚类的密度相仿。这实际上正是我们所希望的情况。

给定两个聚类 $C_i = \{S_1^i, S_2^i, \cdots, S_k^i, \cdots, S_I^i\}$, $C_j = \{S_1^j, S_2^j, \cdots, S_k^j, \cdots, S_J^j\}$, 其中 S_k^i, S_k^j 分别表示 C_i, C_j 的第 k 个镜头, I 和 J 分别表示 C_i, C_j 的镜头个数。令 d_{m_i}, n_j 表示 C_i 中镜头 S_m^i 与 C_j 中镜头 S_n^j 之间的距离。则两类的类间空隙为:

$$Space(C_i, C_j) = \min_{1 \leqslant m_i \leqslant I,\ 1 \leqslant n_j \leqslant J} (d_{m_i}, n_j) \tag{1}$$

两类的类间空隙可以通过遍历两类所含的镜头之间的距离得到。整个聚类集中各聚类之间的空隙构成一个对称矩阵, 最小类间空隙是这个矩阵中的最小值。

如前所述, 任何一个聚类都可看作由两个或多个子类构成, 类内空洞就是指这些子类之间的空隙。各子类之间的空隙构成一个对称矩阵。最大类内空洞是这个矩阵中的最大值。由于对于任意给定的聚类, 我们无法得到它的子类构成, 只能设法得到最大可能类内空洞, 即我们要寻找该聚类的一个子类分割, 满足: 该子类集的最大类间空隙(也即其父类的最大类内空洞) 是所有可能的子类分割中最大的。然而, 遍历所有可能的子类分割开销过大。不过, 由初始迭代算法所构造的聚类的子类分割有个重要特

点: 即最多只有一个子类包含多于一个以上的元素。也即初始迭代算法的输出聚类中不会出现图 3 中那样的真正意义上的子类分割, 而只会出现由噪声点导致的一个个“孤岛”。聚类分裂实际上做的是剔除噪声点的工作。如果我们一次分裂只剔除一个噪声点, 那么对应的子类分割由两个子类构成, 其中一个子类只包含一个元素, 另外一个子类包含除该元素外的其它所有元素。

在这种情况下可以证明, 某类的最大可能类内空洞等于该类的最大最小距离。所谓最小距离是指某个元素距其所属类中的其它元素的最短距离。一个聚类的最大最小距离指该聚类中所有镜头的最小距离中的最大者, 持有最大最小距离的元素也是该类可能的噪声点。

于是, 分裂准则可以近似表述为: 当一个类的最大最小距离大于最小类间空隙时, 一个类应该分裂。当一个类需要分裂时, 应该分为两类, 其中一类只包含一个元素, 这个元素是该类最大最小距离的持有者。在算法实现中, 我们按照这个近似分裂准则将聚类分裂转化为单个噪声点的剔除, 将类内空洞的计算转化为最大最小距离的计算。值得注意的是, 最大最小距离持有者并非离质心最远的镜头(最大半径持有者) , 而是离所有的元素最远的镜头, 比最大半径持有者更有可能是噪声。因为最大半径持有者仅由质心一个元素判断, 而最大最小距离持有者则反映了类中所有其它元素的集体抉择。从这个意义上, 最大最小距离持有者也是一个好的噪声候选者。

分裂判断的“门槛”是最小类间空隙。若在聚类集中存在需合并但还未被合并的聚类对, 比如 C_i, C_j, 那么 Space(C_i, C_j) 会非常小, 用这个错误的“门槛”来指导分裂, 会使许多不需分裂的类被支解, 导致荒谬的结果。理想的情况是在分裂操作之前所有该合并的类都已经合并。

1.2.3 迭代方式设计

从以上分析知, 初始迭代的结果必须经过合并与分裂操作的校正才能达到较好的聚类效果。但是合并与分裂又互相希望对方先行, 为其铺平道路。合并操作需要分裂操作帮助把那些理应合并的聚类中的噪声点剔除, 分裂操作需要合并操作帮助消除错误的最小类间空隙以得到较为精确的“门槛”。我们采取合并与分裂交替迭代的方式来达到这种效果。即每次迭代只选几个最佳候选聚类作合并或分裂操作, 而把可能受噪声影响的操作推迟到对方操作降低噪声后再进行。

在每次合并迭代中选择两个候选聚类对。一是质心距最近的聚类对, 这是按合并准则最有资格合并的聚类对; 二是类间空隙最小的聚类对, 因为小的类间空隙也能在一定程度上指示需合并的两类。第二个候选聚类对的选择是为了帮助分裂操作得到尽可能不含水份的最小类间空隙。如果候选聚类对满足聚类合并准则和合并约束, 就将该聚类对合并。

在每次分裂迭代中选择两种候选聚类。首先选择上次合并迭代中想合而因为半径过大未能合并的聚类, 这是为了减少噪声对合并操作的影响; 其次选择那些类内空洞最大的聚类, 这是按照分裂准则最有资格被分裂的类。如果候选聚类满足分裂准则, 则分裂之。

每次迭代(合并或分裂) 的输出聚类集是下次迭代的输入, 如果连续两次迭代未有任何聚类被操作(被合并或被分裂) , 即整个聚类集中既没有可合的类, 也没有可分的类, 则整个聚类迭代算法终止(见图 1) 。

初始迭代后首先进行合并迭代, 然后分裂与合并交替进行。首先进行合并迭代基于两点认识: 一则合并迭代受到合并约束的保护, 只会“漏合”, 不会“误合”, 比分裂迭代受误差影响要小; 二则初始迭代可以看作一种分裂迭代(把一个大的聚类分割成数个小的聚类) , 从而与合并迭代有交替性。

另外, 每次合并迭代中的候选类对为两个。而分裂迭代中的候选聚类数由其上一次的合并迭代中实际操作的镜头数决定。因为分裂操作一次只操作一个镜头(剔除单噪声点) , 所操作的聚类数和所操作的镜头数相等, 而一次合并操作则会移动一个聚类中的所有镜头。把分裂迭代中的候选聚类数设计为合并迭代中实际操作的镜头数是可以使合并迭代与分裂迭代在操作幅度上大致对称, 加快算法收敛的速度。

2 实验结果与分析

共对 78 段视频进行了镜头聚类实验, 主观视觉效果非常好。实验证明, 虽然算法未使用任何人工

交互和经验参数, 但通过合并、分裂迭代, 初始迭代的聚类效果仍能得到自动的校正, 主观视觉效果非常好。我们发现这个算法对长视频的效果比对短视频的效果好, 这可能是由于被聚类样品集越多, 其统计特性就越明显的缘故。另外, 由于聚类算法的唯一输入是距离矩阵, 通过灵活采用不同低层特征来计算距离矩阵, 可以得到各种不同意义的聚类结果。在实验中分别采用基于镜头运动特征[2]的距离矩阵, 和基于镜头颜色特征[5]的距离矩阵测试我们的聚类算法, 发现根据前一种距离矩阵进行聚类可以把以相似运动方式变化的镜头聚在一起, 而以后一种距离矩阵进行聚类的结果则指示包含同一对象的镜头或发生在同一地理位置的镜头。利用这些特点, 在自行开发的两个原型系统中使用了这一方法: 在 VideoCAR 系统中[5], 使用这个算法构造视频场景; 在 NewsVideoCAR 系统中, 使用这个算法帮助探测新闻故事边界, 都得到了较好的效果。

由于对聚类效果的好坏不易给出量化评价, 在此给出一个典型结果, 代表大多数实验得到的印象, 并帮助进一步阐述算法过程。素材是一段长为 4min23s 的 MPEG－1 文件。首先对该视频进行镜头分割。先使用镜头探测软件进行自动分割, 然后再进行人工调整, 得到基本上无漏判, 误判的视频文件镜头集, 共 40 个镜头。代表帧取的是视频流中离镜头中间帧最近的 I 帧。此实验采用基于代表帧颜色特征[5]的距离矩阵进行聚类。

初始迭代共分出 12 个聚类。图 4 列出了部分聚类所含镜头的代表帧, 每一行图片是一个聚类, 各代表帧按距质心镜头远近排序, 行尾是该聚类的编号。初始迭代已有较好的效果。从图 4 可以看出视觉上相似的镜头被聚在一起。不过从图 4 也可以较明显的看出有些聚类需要合并, 如聚类 1 和聚类 2, 聚类 3 和聚类 4。从图 4 也能看出有些聚类含有明显的噪声点, 需要分裂, 如聚类 4 的最后一个镜头, 聚类 7 的最后一个镜头等。

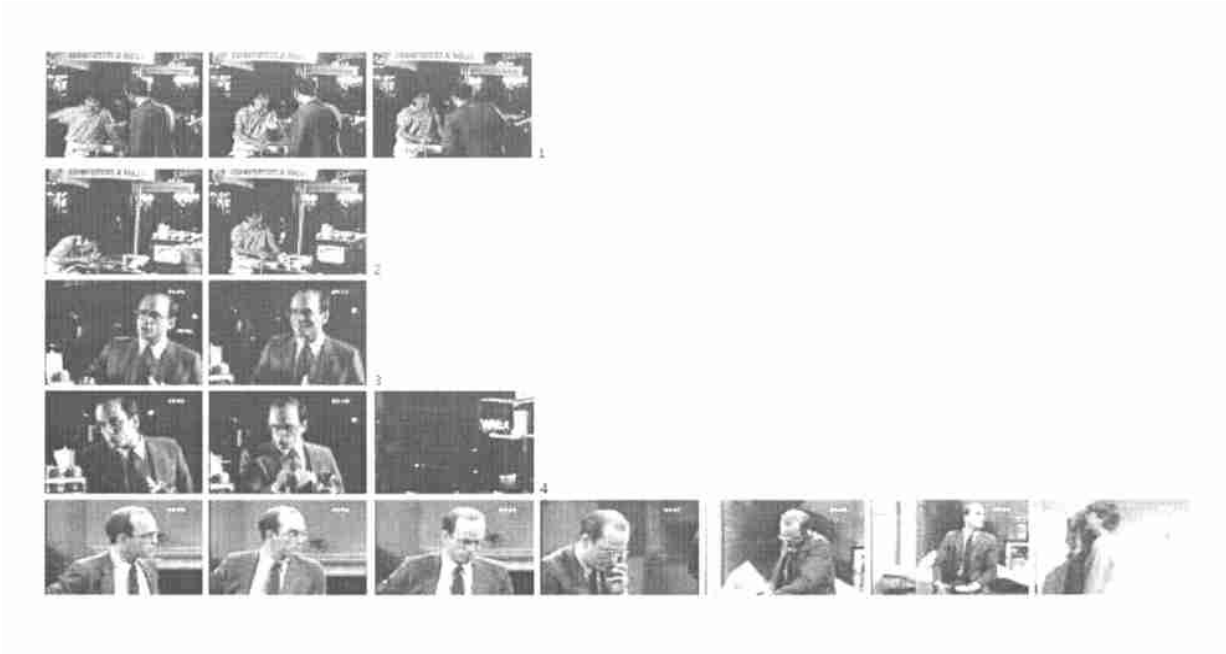

图 4　初始迭代结果: 聚类 1, 2, 3, 4, 7 所含镜头代表帧

Fig. 4　Results of initial iteration: R-frames of cluster 1, 2, 3, 4, 7

该实验共迭代 15 次, 最后共得到 9 个聚类和 1 个杂类。图 5 列出了部分聚类所含镜头的代表帧。将图 4, 图 5 相比较可以看出, 初始迭代结果中的聚类 1 和聚类 2, 以及聚类 3 和聚类 4 被正确地合并在一起, 而聚类 4、聚类 7 中的噪声点也被剔除, 初始迭代中的误差得到了较好的校正。从图 5 还可以看出, 镜头聚类的结果在一定程度上反映了语义内容。比如聚类 1 是关于街头小贩的镜头, 聚类 2、5 则是关于男主角分别在不同地方的镜头。

图 5 最终迭代结果: 聚类 1, 2, 5 所含镜头代表帧

Fig. 5 Final results: R-Frames of cluster1, 2, 5

3 结束语

为了避开经验阈值的设定, 我们设计了一种新的镜头自动聚类方法。这种方法以初始迭代算法对整个镜头集作大致分割, 用特别设计的合并和分裂算法对初始聚类结果作交替迭代的微调。整个过程不需要任何经验参数和人工交互。算法的实质是通过分裂准则和合并准则的使用, 在程序运行期间动态地由当前聚类集的统计特性隐式地设置控制收敛的阈值, 如: 聚类集平均半径, 最小类间空隙, 最大类内空洞等。

本文侧重于聚类方法的设计, 相似度准则的设计在实际应用中会至关重要, 需要下功夫研究。目前所用的测试视频仅包括教学视频和新闻视频, 在以后的研究中应该采用更宽范围内的实验素材。

参考文献:

[1] Yeung M M, Yeo B L, Wolf W et al. Video browsing using clustering and scene transitions on compressed sequences[C]. Proc. Of SPIE Conf. on Multimedia Computing and Networking, USA, 1995.

[2] Zhong D, Zhang H J, Chang S F. Clustering methods for video browsing and annotation[C]. Proc. Of SPIE Conf. Storage and Retrieval for Image and Video Database, USA, 1996.

[3] 薛峰. 基于内容检索的图象和视频存储结构和索引技术的研究和实现[D]. 硕士学位论文, 长沙:国防科技大学, 1999.

[4] 沈清, 汤霖. 模式识别导论[M]. 长沙: 国防科技大学出版社. 1990.

[5] 熊华, 胡晓峰等. 基于镜头的视频场景构造方法研究[J]. 小型微型计算机系统, 2000, 21(6).