

基于超图聚类的故事单元的抽取与分析*

樊 昀⁺, 王润生

(国防科学技术大学 ATR 国防科技重点实验室, 湖南 长沙 410073)

Extraction and Analysis of Story Unit Based on Clustering of Hyper-Graph

FAN Yun⁺, WANG Run-Sheng

(ATR National Defense Technology Key Laboratory, National University of Defence Technology, Changsha 410073, China)

+ Corresponding author: Phn: 86-10-66748997, E-mail: fanyun_cn@yahoo.com

<http://www.nudt.edu.cn>

Received 2001-09-10; Accepted 2001-10-18

Fan Y, Wang RS. Extraction and analysis of story unit based on clustering of hyper-graph. *Journal of Software*, 2003,14(4):857~863.

Abstract: It is the important precondition for browsing, retrieval and indexing videos that effectively and rationally organizing the video structure. In this paper, for hierarchically organizing the video structure, a method for extracting story units is proposed. k-nearest neighbor hyper-graph is used to represent the content relations among shots, and shots are clustered based on hyper-graph model. By analyzing time projection relations among shot clusters, story units are extracted, and represented by the 1D strings. A frame using specific domain knowledge for identifying the type of story unit is also proposed, and is applied to identify the dialogs in the videos. The new algorithm is applied to multiple test videos, and the experiments results are satisfying.

Key words: video retrieval; shot; story unit; k-nearest neighbor hyper-graph; hierarchical clustering

摘 要: 合理而有效地结构化组织视频数据是浏览、检索和管理视频的重要前提。提出了一种新的层次化的结构化组织视频数据的提取故事单元的方法。它采用 k-近邻超图描述镜头间的内容相似关系,利用超图模型对镜头聚类,并通过分析镜头类别间的时间投影关系提取故事单元,并采用一维字符串描述故事单元。此外,还提出了结合领域知识确定故事单元的类型算法框架,并将其用于对话场景的检测。将算法用于测试多段视频,实验结果令人满意。

关键词: 视频检索;镜头;故事单元;k-近邻超图;层次聚类

中图法分类号: TP319 文献标识码: A

随着因特网的快速增长,网上多媒体信息(multimedia information),如图像、视频、语音等迅速膨胀,传统的信息管理方法已不能满足索引、浏览、检索海量多媒体信息的需要,基于内容的检索方法应运而生,正得到日益广泛的应用与发展,视频数据的基于内容的处理与分析尤其引人注目^[1-3]。由于大部分原始视频数据是无结构化的,这给后续按内容的浏览、检索带来了极大的不便,目前视频数据的结构化研究受到了极大的关注。视频数据的结构化可以狭义地理解为将视频分割成镜头,再将镜头组织为故事单元的过程。

* 第一作者简介: 樊昀(1976—),男,甘肃古浪人,博士生,工程师,主要研究领域为图像理解,海量数据处理,内容检索。

镜头是摄像机一次连续起停操作拍摄的视频序列,在影片制作过程中,导演会根据剧本的内容,按照叙述的顺序并根据总的创作意图,合乎逻辑地、富有表现力地、有节奏地把镜头连接起来,构成完整的银幕形象(蒙太奇),形成比镜头更高一级的结构——故事单元^[4]。故事单元由一系列时间上连续的镜头组成,往往描述发生在一个地点的一个事件,是携带语义信息的比镜头更高一个层次的视频结构。目前对于故事单元的探测的研究相对较少,比较有代表性的故事单元探测方法有 Young^[5]和 Hanjalic^[6]提出的算法,这些方法取得了很好的效果,但也存在一定的局限性,比如:(1) 需要预先设定一系列门限,自适应较差,需要较多的人工干预;(2) 没有显式地描述故事单元内镜头间的彼此时间和内容之间的关系,缺乏描述故事单元的通用框架;(3) 不便于结合领域知识判断故事单元的类型。

为克服上述缺点,本文汲取了文献[7]中的超图层次聚类的思想,提出一种探测故事单元的新算法。该方法采用 k -近邻超图模型对镜头间的时间和内容的相互关系建模,并采用基于超图的层次聚类方法将镜头汇聚成若干个类别,并自动确定类别的数目。通过分析镜头类的时间投影关系确定故事单元所覆盖的时间范围,并采用一维字符串描述故事单元。本文还给出了结合领域知识确定故事单元类型的算法框架,并将其用于对话场景的检测。利用多段视频测试提出的算法,实验结果较为令人满意。

本文第 1 节简要介绍基于超图的层次聚类算法,第 2 节介绍基于超图聚类故事单元探测算法,第 3 节介绍描述故事单元的基本方法,第 4 节介绍结合领域知识的故事单元类型的判定,第 5 节给出实验结果并对全文进行总结。

1 基于超图的层次聚类算法

超图层次聚类的基本工作流程可以用图 1 表示。它的基本思想可以表述如下:首先建立超图模型,超图的顶点表示参加聚类的数据项,加权边则表示数据项之间的相似性;然后采用图分割算法获得初始分割子类集合,并采用类别的连接性(connectivity)和邻近性(closeness)指标度量两个子类间的相似性;接着采用层次合并算法合并最相似的子类,最终获得符合真实情况的聚类结果。下面我们分数据建模、类别相似性建模和聚类算法描述 3 个部分简要介绍文献[7]中提出的超图层次聚类算法。

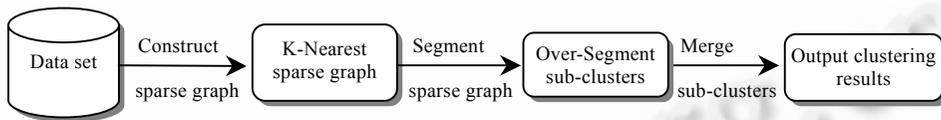


Fig.1 Flow chart of the hierarchical clustering algorithm based on hyper-graph

图 1 超图层次聚类算法流程图

1.1 数据建模

在构造稀疏图(sparse graph)时采用了 k -近邻方法,即当且仅当两个顶点中有一个是另外一个顶点的 k -近邻时,图中两个顶点间有边相连。采用 k -近邻建立超图具有如下优点:(1) 彼此远离的数据项间没有边相连;(2) 动态捕捉了邻域的概念,密集区域的邻域半径小,权值大,稀疏区域的邻域半径大,权值小;(3) 为后续处理带来了计算上的便利。

1.2 类别相似性建模

类别 C_i 和 C_j 间的相似性由连接性(connectivity)和邻近性(closeness)描述,分别记为 $RI(C_i, C_j)$ 和 $RC(C_i, C_j)$,下面简要描述两个指标的计算过程。

记 $EC_{[C_i, C_j]}$ 为连接 C_i 和 C_j 中顶点的边的权值之和,定义类内连接性为将该类别划分为两个子图的割边集的权值之和,记为 $EC_{|C_i|}$,同理可得 $EC_{|C_j|}$,定义:

$$RI(C_i, C_j) = \frac{|EC_{[C_i, C_j]}|}{0.5 * (|EC_{[C_i]}| + |EC_{[C_j]}|)}$$

邻近性可以由 C_i 和 C_j 间有边相连的顶点间的平均相似性来描述,注意到这种平均相似性实际上就是连接顶点的边的权值的平均值,记为 $S_{EC_{C_i, C_j}}$,同样可以定义类内邻近性为将 C_i 二分的割边集的权值的平均值,记为 $S_{EC_{C_i}}$,同理可得 $S_{EC_{C_j}}$,定义邻近性指标:

$$RC(C_i, C_j) = \frac{(|C_i| + |C_j|) * S_{EC_{[C_i, C_j]}}}{|C_i| * S_{EC_{C_i}} + |C_j| * S_{EC_{C_j}}}$$

1.3 聚类算法描述

聚类算法可以分解为两个步骤,第 1 步将数据项聚集为子类,第 2 步是层次合并子类.下面对这两步分别加以简单介绍.

步骤 1. 目前超图的二分问题已经得到了很好的解决,尤其是基于多级模式分解的算法可以快速、有效地二分超图,在这里本文采用文献[7]中设计的 hM_ETiS 软件包中的算法,在此不再赘述.本文采用的是带有平衡限制的二分图算法.将 C_i 分解为两个子类 C_i^A 和 C_i^B ,保证每个子类至少含有 C_i 中 25% 的顶点,分割的基本过程为:(1) 初始化,将所有数据项归于一类;(2) 选择当前顶点数最多的子类做平衡限制二分;(3) 反复迭代,直至最大子类的数据项数目小于预先设定的门限,通常门限取为数据项总数的 1%~5%.

步骤 2. 利用动态框架合并子类.计算子类间的 $RC(C_i, C_j)$ 和 $RI(C_i, C_j)$,并综合两者得到类别相似性指标 $Sim(C_i, C_j) = RI(C_i, C_j) \cdot RC(C_i, C_j)^a$,若 $a > 1$,则 $RC(C_i, C_j)$ 中在合并中所起的作用较大,若 $a < 1$,则 $RI(C_i, C_j)$ 在合并中所起的作用较大.每一次合并当前最相似类别对,反复迭代直至相似性指标小于某一门限,或者迭代到仅有一个类为止,可以得到层次聚类的树状图.实验中, $a = 1$.

2 基于超图聚类故事单元探测算法描述

故事单元由彼此关联的多个镜头组成,是介于镜头和整个视频的中层结构单位.它通常与一个特定的事件、某一个特定的场景或一个特定的角色相关.故事单元通常可以区分为交叠型故事单元和渐进型故事单元两种类型^[4],在文献[5,6]中检测的都是交叠型故事单元.交叠型故事单元在视频中广泛存在,它可以概括地定义为由多个具有相似内容元素的镜头构成的集合所交叠而成的视频序列,通常在它的内部同一人物或同一布景的镜头会反复与其他内容的镜头交替出现.而渐进型故事单元则由内容互不相似的一组短镜头构成,通常表现紧张激烈的场面和简要有力的论述,它没有很明确的结构特征,不便于直接检测,但若在检测交叠型故事单元之后,再加以分析,则往往可以得到较好的效果.本文重点讨论交叠型故事单元的检测,将两个交叠故事单元之间的视频片段判定为渐进型故事单元.

故事单元探测的基本流程如图 2 所示.它主要包括 3 个步骤:(1) 视频分镜头,并提取描述镜头的关键帧;(2) 建立描述镜头间相似性的 k-近邻超图,并依据超图对镜头聚类;(3) 分析不同镜头类别间的时间投影关系,以抽取故事单元.

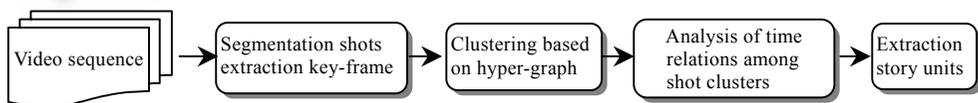


Fig.2 Flow chart of algorithm for detecting story units

图 2 故事单元探测算法流程图

2.1 分镜头和提取关键帧

本文采用文献[8]提出的基于颜色直方图的双重比较法,关键帧选镜头的首帧.双重比较法的基本思想如下:

首先用一个较低的阈值确定出渐变切换序列的可能起始帧.一旦确定了该帧,就用它与后续的帧进行比较,用积累的差值取代帧间的差值.如果渐变的确存在,则积累差值会单调增长,当增长过程中止时,将积累的差值与第 2 个预先设定的较高的门限比较,若超过该门限,则认为积累差值单调增长的序列为渐变序列.将 R, G 和 B 分别量化为 8 个格子,帧 i 的 R, G 和 B 分量归一化直方图分别为 $H_i^R(l), H_i^G(l), H_i^B(l), l=1,2,\dots,8$, 帧 i 和帧 j 的直方图差异定义为

$$Diff(i, j) = 1 - \frac{1}{3} \sum_{k=R, G, B} \sum_{l=1}^8 \min(H_i^k(l), H_j^k(l)).$$

2.2 基于k-近邻超图的镜头聚类

2.2.1 建立反映镜头时间和内容相互关系的超图

本文采用 k-近邻超图模型,描述镜头间的时间临近性和内容相似性.设镜头探测算法检测到了 N 个镜头,将这 N 个镜头按时间排序得到镜头序列 Sh_1, Sh_2, \dots, Sh_N , 任取两个镜头 Sh_i 和 Sh_j , 序号之差 $|i-j|$ 反映了镜头 Sh_i 和 Sh_j 之间的时间关系.定义函数 f_1 为 $|i-j|$ 的单调减函数,归一化使得 $f_1 \in [0,1]$.由于钟形函数具有中部平稳、两端下降较快的特点,本文采用该形式定义 $f_1 = \exp\left(-\frac{(x-1)^2}{T^2}\right)$, 由 T 控制可容忍的镜头跨度,由上式可见,当两个镜头彼此相邻时 $f_1 = 1$, 随着 x 的增大, f_1 的值单调减小,实验中 $T = 10$.

镜头的内容由镜头的特征所描述,设描述镜头内容的特征矢量为 $\vec{Fv} = (fv_1, fv_2, \dots, fv_m)$, 其中 fv_i 为特征,对应的相似性度量为 S_{fv_i} , 设 $S_{\vec{Fv}}(Sh_i, Sh_j)$ 为镜头 i 和 j 之间基于特征描述的内容的相似性度量, $S_{\vec{Fv}}$ 是诸 S_{fv_i} 的函数,归一化 $S_{\vec{Fv}}$, 使得 $S_{\vec{Fv}} \in [0,1]$. 本文采用镜头关键帧的颜色直方图描述镜头内容,定义两个镜头间的相似性为对应的关键帧直方图的交,即 $S_{\vec{Fv}}(Sh_i, Sh_j) = \sum_{k=1}^{hn} \min(h_k^i, h_k^j)$, 其中 hn 为直方图的格子数目, $(h_k^i)_{k=1, \dots, hn}$ 和 $(h_k^j)_{k=1, \dots, hn}$ 为镜头 Sh_i 和 Sh_j 各自关键帧分别对应的归一化直方图.

综合时间的临近性和内容的相似性度量定义镜头间的相似性度量 $S_{im}(Sh_i, Sh_j)$, 注意到只有当时间临近且内容相似时,镜头间的相似度才会高,定义镜头间的相似性度量为 $S_{im}(Sh_i, Sh_j) = f_1 \times S_{\vec{Fv}}^a$. 其中 a 反映两种度量在相似性测度中所占的比重.

在建立了镜头间的相似性描述模型后,采用 k-近邻超图 $G = (V, E)$ 来描述镜头集合内部的相似性关系.顶点集合 V 代表镜头集合,当两个顶点中的某一个为另一个顶点的 k-近邻,而且相似性测度大于门限 $\min S_{im}$ 时,两个顶点间才有边相连.边的权为所连接的两个镜头间的相似性测度.实验中取 $k = 5$, 显然由上述方法生成的超图是一个稀疏图.

2.2.2 基于超图的镜头层次聚类

由前面的叙述可知,通过基于超图的层次聚类,可以得到镜头聚类树状结构图.根节点包含所有的镜头,而叶子节点则是由超图过分割得到的子类.对于一段特定的视频来说,聚类树中存在一个最优层次,在该层上聚类结果与实际的镜头的真实类别最接近.本文通过衡量不同层次的聚类的有效性加以自动选择最优聚类层次.好的聚类应该具有如下两条性质:(1) 类别间彼此远离,即具有较好的孤立性;(2) 类内的数据项间应该彼此联系紧密,即具有较高的类内连接强度.综合上述两个指标可以定义聚类的有效性度量.

采用式 $Fitness(C) = \frac{\sum_{e \in C} Weight(e)}{\sum_{|e \cap C| > 0} Weight(e)}$ 衡量某一层类 C 的孤立性.设总的类别数为 N , 定义该层聚类的孤立性为所有包含类别的孤立性之和,有 $Isol(N) = \sum_{i=1}^N Fitness(C_i)$, 该指标有随着 N 的减小而增大的趋势.

定义类别 C 的连接强度为 $Strong(C) = \frac{\sum_{e \in C} Weight(e)}{|e \in C|}$, 定义该层聚类的平均连接强度为

$S(N) = \sum_{i=1}^N \text{Strong}(C_i)$. 容易看到,当类别数目较多时,聚类的平均连接强度倾向于较大.

为了便于比较不同层次间的上述指标,本文对上述指标做了稳健归一化处理.若 $\xi = \{\xi_1, \xi_2, \dots, \xi_l\}$ 为数据集,定义 $Z(\xi_i) = \frac{\xi_i - \text{median}(\xi)}{\text{MAD}(\xi)}$,其中 $\text{median}(\xi)$ 为数据集 ξ 的中值, $\text{MAD}(\xi) = \text{median}\{|\xi_i - \text{median}(\xi)|, i = 1, 2, \dots, l\}$,综上所述,定义聚类有效性度量指标为 $Va(N) = Z(\text{Isol}(N)) + Z(S(N))$.

由层次聚类可以得到类别数目由 $1 \sim N_{\max}$ 的聚类树,分别计算各层的 $Va(N)$, $2 \leq N \leq N_{\max}$,选择使得 $Va(N)$ 最大的聚类层次作为最终聚类结果输出.

2.3 镜头类别间的时间投影关系分析

镜头聚类以后,需要通过分析镜头类别间的时间投影关系,以确定故事单元的边界.设得到镜头类别 c_i ,它的镜头序号边界区间为 $[L_{c_i}, H_{c_i}]$,由交叠型故事单元的定义可知,在该区间内的所有镜头都应属于同一个故事单元.将所有的镜头类别的序号边界区间都投影到序号轴上,采用以下公式来分析彼此交叠关系:

$$F\{[L_{c_i}, H_{c_i}], [L_{c_j}, H_{c_j}]\} = \begin{cases} [\min(L_{c_i}, L_{c_j}), \max(H_{c_j}, H_{c_i})] & \text{if } [L_{c_i}, H_{c_i}] \cap [L_{c_j}, H_{c_j}] \neq \text{NULL} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

对镜头类别的序号边界区间反复作上述分析,直到区间关系不再变换时,划分镜头序号轴所得到的互不交叠的区间,就是所求的故事单元.分析过程示意如图 3 所示.

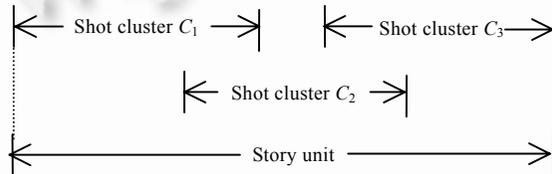


Fig.3 Schematic of analysis of the relations of time intervals projection
图 3 时间区间投影关系分析

在获取所有交叠型故事单元之后,将夹在两个交叠型故事单元之间的镜头集合视为渐进型故事单元.

3 故事单元描述

设当前考察的故事单元 S_i 包含 M 个镜头,按时间先后关系依次为 $Sh_{i1}, Sh_{i2}, \dots, Sh_{iM}$,这些镜头分属于 N 个类,若将每一个类用一个字符表示,记为字符集 $C = \{c_1, c_2, \dots, c_N\}$,则 S_i 可由 C 中字符所构成的字符串描述.利用字符串描述故事单元,可以清晰地显示故事单元的基本结构

4 基于领域知识确定故事单元类型

确定故事单元的类型对于正确理解视频内容具有非常重要的作用.故事单元可以划分为对话场景、动作场景等类型.具体类型的确定与视频涉及的领域知识有关,例如,电视节目的谈话节目的确定就与故事片中对话场景的确定有所不同.

4.1 确定故事单元类型的算法框架

确定故事单元类型的算法的基本过程为:首先用户确定要查找的故事单元类型,再参考领域知识库确定该类型故事单元的基本模式和组成要素,接着匹配引擎在视频故事单元库中查找,并将检索的结果返回给用户.

4.2 对话场景的确定

本文通过对故事单元描述字符串做相关分析,结合人脸和字幕伪语义特征确定对话场景,检测算法流程如图 4 所示.

4.2.1 故事单元描述字符串分析

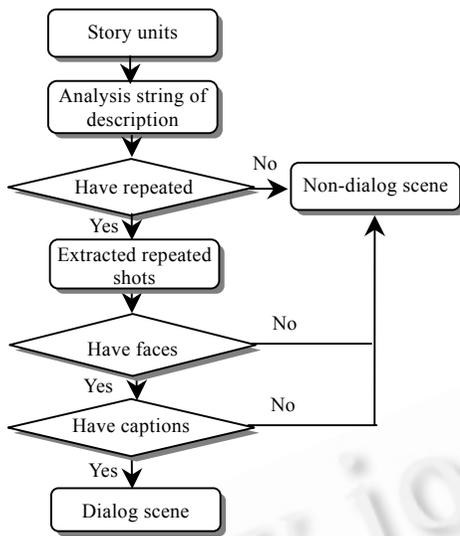


Fig.4 Flow chart of detecting dialog scene of movies
图4 电影中对话场景检测流程

对话场景一般由多个重复出现的特写镜头和间或穿插的中景和全景镜头所构成,在描述字符串中会重复出现特定的结构,判断检测这种结构是否存在,可以为对话场景的检测提供重要线索.本文采用直方图统计方法分析描述字符串.首先统计故事单元中出现的模式(字符)数目 N ,并统计每种模式的出现频率 h_i^c , $1 \leq i \leq N$,并对 h_i^c 归一化处理,然后判断是否出现高重复率模式,即是否存在 $h_i^c > T_h$,若满足上述条件的模式数目 M 大于 2,则作进一步分析,否则判定该故事单元为非对话场景.下一步判断重复出现的模式之间是否存在一定的结构关系.设两个重复出现的模式为 i_1, i_2 ,记镜头间的距离 $d(i_1, i_2)$ 为两镜头的序号之差,统计镜头间距离为 d 模式对的出现频率,记为 $h_d^D(i_1, i_2)$, $-L \leq d \leq L$,其中 L 为故事单元所包含的镜头总数,记 $h_{\max}^D(i_1, i_2) = \max_{-L \leq d \leq L} h_d^D(i_1, i_2)$,可以构造

$$\frac{M(M-1)}{2}$$

个上述的直方图,若存在 i_1, i_2 ,使得

$h_{\max}^D(i_1, i_2) > ThD$, $ThD = 2$,则该故事单元有高重复并有一定结构的模式对出现,判断该故事单元可能为对话场景,记满足条件的模式对为 (i_1^*, i_2^*) .下一步验证 i_1^* 和 i_2^* 中是否存在人脸和字幕,

4.2.2 人脸检测

本文采用了一种在 MPEG 流中检测人脸的新算法.它采用彩色图像分割结合色调证实区分人脸与背景;采用基于规则的区域合并解决可能出现的过分割;利用距离变换结合 watershed 分割的方法分解形状复杂的区域,解决了多个人脸交叠的问题;利用帧间冗余性,自适应调整肤色检测器;利用 MPEG 流中直接提取的运动矢量在一个 GOP 内跟踪人脸;并依据场景变化更新分割码本.

4.2.3 字幕检测

本文采用了一种简单的方法检测字幕,该方法可以概括为两个主要步骤:(1) 利用文字的边缘突出和背景一般比较均匀的特点,对待检测图像高通滤波,并采用大尺度高斯滤波使得孤立的点相互作用形成区域,通过二值化检测出该区域,并把它作为可能存在文字的候选区;(2) 采用 Hough 变换分析候选内扫描线的灰度分布模式,判断该区域内的扫描线灰度的峰谷是否呈现明显的规则分布,并以此证实候选区,同时通过对相邻扫描线灰度分布相同模式的检测和定位,修改候选区的形状,使其尽可能的紧凑的包含文本.输出结果为包含文本的矩形区域.

5 实验结果及其讨论

5.1 实验结果

本文的测试数据取自有一定代表性的《阿甘正传》、《卧虎藏龙》和《神鬼传奇》3 部故事片,片长都在 100 分钟~130 分钟左右.出于方便考虑,本文从每部故事片截取了大约 45 分钟左右的视频段进行实验.表 1 为故事单元检测的统计结果,表 2 为对话场景检测的统计结果.

Table 1 Statistical results of detection of story units
表 1 故事单元检测的统计结果表

Sequence name	Total frame number	Number of the actual story units	Number of detected story units	Number of the false story units	Number of the missed story units
阿甘正传	63 582	14	15	3	2
卧虎藏龙	67 517	17	17	2	2
神鬼传奇	62 251	11	12	2	1

Table 2 Statistical results of detection of dialogs
表 2 对话场景检测统计结果表

Sequence name	Number of the actual dialog scenes	Number of the detected dialog scenes	Number of the false dialog scenes	Number of the missed dialog scenes
阿甘正传	6	9	4	1
卧虎藏龙	7	9	2	0
神鬼传奇	4	5	2	1

故事单元检测中的漏检和漏警主要是因为以下几个原因:(1) 分镜头错误,关键帧选择不当;(2) 聚类出现不符合人的视觉感知的局部错误.对话场景检测中,《阿甘正传》中的错误主要是由于没有对旁白字幕和对话字幕加以区分,而《卧虎藏龙》和《神鬼传奇》中的错误主要是由于字幕与图像没有时间同步、人脸检测错误和描述字符串分析错误所造成的.

5.2 总 结

本文提出了一种利用超图层次聚类方法检测、描述和鉴别视频中故事单元的算法.实验结果表明,该算法行之有效.在未来的工作中,我们将结合更好的分镜头技术、人脸识别、字符识别技术和语音识别技术,以进一步提高故事单元检测的准确率.

References:

[1] Bach J, Fuller C, Gupta A, Hampapur A, Horowitz B, Humphrey R, Jain R. The virage image search engine: an open framework for image management. In: Proceedings of the SPIE, Storage and Retrieval for Image and Video Databases IV. San Jose: SPIE Press, 1996. 76~87. <http://citeseer.nj.nec.com/context/565/0>.

[2] Gudivada VN, Raghavan VV. Content based image retrieval systems. IEEE Computer, 1995,28(9):18~22.

[3] Wactlar H, Kanade T, Smith MA, Stevens SM. Intelligent access to digital video: informedia project. IEEE Computer, 1996,29(3): 46~52.

[4] Ge D. Introduction to the Camera Work of Cinematics. Beijing: China Film Press, 1995 (in Chinese).

[5] Yeung M, Yeo BL, Liu B. Time constraint clustering for segmentation of video into story units. In: Proceedings of the International Conference on Pattern Recognition. 1996. 375~380. <http://citeseer.nj.nec.com>.

[6] Hanjalic A, Lagendijk RL, Biemond J. Automatically segmenting movies into logical story units. In: Proceedings of the 3rd International Conference on Visual Information Systems. Amsterdam, 1999. 229~236. <http://citeseer.nj.nec.com/context/1074914/179834>.

[7] Karypis G, Han EH, Kumar V. Chameleon: Hierarchical clustering using dynamic modeling. IEEE Computer, 1999,32(8):68~75.

[8] Zhang HJ, Kankanhalli A. Automatic partitioning of full-motion video. ACM Multimedia Systems, 1993,1(1):10~28.

附中文参考文献:

[4] 葛德.电影摄影艺术概论.北京:中国电影出版社,1995.