利用等价关系理论进行视频片段检索的方法

赵亚琴[1] 周献中[2] 何 新[1]
[1](南京理工大学自动化学院，南京 210094) [2](南京大学工程管理学院，南京 210093)

摘 要 视频片段检索是基于内容的视频检索的主要方式，可是现有的片段检索方法大多只是对预先分割好的片段进行检索。为了从连续的视频节目中自动分割出多个相似的片段，提出了一种新的有效的视频片段检索方法，并首次尝试将等价关系理论应用于视频片段的检索。该方法首先用等价关系理论定义了片段匹配函数，同时采用滑动镜头窗自动分割出多个真正相似的片段；然后把等价类映射为矩阵表达形式，再通过矩阵的特性来度量影响片段相似度的不同因子，实现了相似片段的排序。实验结果表明，该方法能够一次性快速准确地从连续视频库中自动分割出与查询片段相似的多个片段。

关键词 视频片段 基于内容的视频检索 等价关系 匹配函数 滑动镜头窗
中图法分类号：TP391.3 TN941.1 **文献标识码**：A **文章编号**：1006-8961(2007)01-0127-08

An Efficient Method for Video Clip Retrieval Using Equivalence Relation Theory

ZHAO Ya-qin[1], ZHOU Xian-zhong[2], HE Xin[1]
[1](*School of Automation, Nanjing University of Science & Technology, Nanjing* 210094)
[2](*School of Management and Engineering, Nanjing University, Nanjing* 210093)

Abstract Video clip retrieval plays a critical role in the content-based video retrieval. However, existing clip retrieval methods mostly focused on retrieving similar video clips from pre-segmented clips. To automatically segment and retrieve similar video clips from a continuous video database, this paper presents an efficient method for video clip retrieval in which equivalence relation theory is applied to video clip retrieval for the first time. Matching function is defined in terms of equivalence relations corresponding with shots of given video clip. Several similar video clips are automatically segmented from video database using sliding shot window according to matching degree between two video clips. Afterwards, equivalence classes are mapped into matrix representation. Therefore various factors are computed to rank the similarity of the selected video clips by characters of the matrix. Experimental results showed that the method could segment similar video clips from continuous video database quickly, exactly and automatically.

Keywords video clip, content-based video retrieval, equivalence relation, matching function, sliding shot window

1 引 言

随着多媒体技术的发展和视频数据库的普及，有越来越多的学者投入到基于内容的视频检索的研究中。视频检索一般分为镜头检索和片段检索。镜头一般是由摄像机一次摄像的开始和结束的所有帧构成，其表示一个物理概念，而片段则是由一连串语义相关的连续镜头构成，其表示一个语义概念。目前基于内容的视频检索的多数研究均集中在镜头检索上[1~3]，而片段检索方面的研究则相对较少。实际上，从用户的角度分析，他们对视频数据库的查询通常是一个视频片段，而很少会是单个的物理镜头。从信息量的角度分析，由几个镜头组成的视频片段比单个镜头有更多的语义，由于它可以表示用户感兴趣的事件，因此，查询的结果也比较有意义。例

基金项目：国家自然科学基金项目(70571032)；江苏省自然科学基金项目(BK2004137)
收稿日期：2005-09-10；**改回日期**：2005-11-30
第一作者简介：赵亚琴(1973～)，女。1997 年于山东科技大学获得学士学位，2003 年于南京工业大学获得硕士学位，目前在南京理工大学攻读博士学位。主要研究方向为基于内容的信息检索、多媒体技术。E-mail：yaqinzhao@163.com

如,从新闻视频中检索出感兴趣的事件、从体育视频中检索出喜爱的体育视频片段以及电视台检索某条广告是否播出等。

目前视频片段检索方法可分为以下几类:(1)把视频片段(以下简称片段)分为片段-帧两层考虑,片段的相似性利用组成它的帧的相似性来直接度量[4,5],这种方法要求相似的片段必须遵守同样的时间顺序,而实际的视频节目并不遵守这种约束,因为后期编辑的结果使得相似的片段完全可能具有不同的镜头顺序,如同一个广告的不同编辑,同时,这种基于每帧的比较,也使得检索速度比较慢;(2)片段检索是通过子样本的帧匹配进行[6,7],虽然这种方法的检索速度快,但是由于没有很好地考虑视频队列的时序关系,所以在连续视频片段中确定相似视频片段的位置不准确;(3)把视频片段分为片段-镜头-帧 3 层考虑,片段的相似性可通过组成它的镜头的相似性来度量,而镜头的相似性则通过它的关键帧[8]或所有帧[9]的相似性来度量。在上述这些方法中,方法(3)的思想虽比较合理,但大多数研究是对预先分割好的片段进行检索,并没有解决怎样在连续的视频节目里自动分割出多个相似片段的问题[8,9]。文献[9]虽然对连续视频检索进行了研究,但是由于两个片段的相似度仅仅取决于它们相似镜头的数量,因此即使某一个片段 Y 的所有镜头仅仅和另一个片段 X 的一个镜头相似,Y 也会被认为和 X 相似。文献[10]是采用基于图论的最大匹配算法来确定真正相似的片段,以排除伪相似片段,虽取得了较高的检索精度,但查找可能相似的片段和确定真正相似的片段需要两次进行,而且由于采用了最优匹配法计算视觉因子,因此检索速度仍然较慢。

视频片段检索需要解决以下两个问题:一是从视频库里自动分割出与查询片段相似的多个片段;二是按照相似度从高到低的顺序排列这些相似片段。目前从连续的视频中自动分割出多个相似的片段的研究较少,而且现有的检索方法都比较复杂,通常需要很大的计算量和运算时间,为此本文提出了一种新的从连续的视频库中自动检索视频片段的方法,不同于文献[10]的基于图论的方法。该方法首先通过等价关系理论来定义片段匹配函数,然后采用滑动镜头窗来确定真正相似片段的位置,并从连续视频片段中自动分割出与查询片段相似的多个片段。这不仅消除了文献[9]由于视频片段的自相似性而引起的一对多、多对一、多对多的影响,而且可一次性地得到与查询片段真正相似的片段,而不必要像文献[10]那样分两次进行。为了排列相似片段,本文还考虑了片段相似度度量的不同因子,把等价类映射为矩阵表达形式,并通过矩阵的特性来度量影响片段相似度的不同因子。

2 相似片段的自动分割

本文把视频片段分为片段-镜头-帧 3 层考虑,即在计算镜头的相似度的基础上,定义了片段匹配函数,并采用滑动镜头窗自动分割出与查询片段真正相似和匹配的多个片段。

2.1 镜头的分割

为了充分表示镜头内部的变化,一般可用一个镜头的几个关键帧来表示该镜头。与用一帧来表示一个镜头的方法相比,这种方法确实可以有更好的查准率和查全率,但检索速度比单帧表示要慢,甚至慢很多。因此,片段检索的准确性和检索时间是存在的一对矛盾。本文运用了文献[3,10]提出的基于相机运动的关键帧表示方法,这种方法不仅能很好地检测出镜头突变,还能检测出镜头缓变,如画像、淡入淡出,而且可直接在 MPEG 流上通过分析 DC 图像来实现,并具有很快的检测速度。镜头分割时,根据相机的运动信息,首先把一个内容变化的镜头划分为几个内容一致的子镜头,如果一个镜头是静止后变焦,那么该镜头可分为两个子镜头,如果是静止、扫描、静止,那么就分为 3 个子镜头;然后针对不同运动的子镜头来构造不同的关键帧表示。设镜头 G_i 关键帧的集合为 $\{f_{i,1}, f_{i,2}, \cdots, f_{i,k}\}$,则镜头 G_i 和 G_j 的相似度定义为

$$S(G_i,G_j) = \frac{1}{2}\{M(G_i,G_j) + \tilde{M}(G_i,G_j)\} \quad (1)$$

其中,

$$M(G_i,G_j) = \max_{p=\{1,2,\cdots\}} \max_{q=\{1,2,\cdots\}} \{I(f_{i,p}, f_{j,q})\}$$

$$\tilde{M}(G_i,G_j) = \tilde{\max}_{p=\{1,2,\cdots\}} \max_{q=\{1,2,\cdots\}} \{I(f_{i,p}, f_{j,q})\}$$

$$I(f_{i,p}, f_{j,q}) = \frac{1}{A(f_{i,p}, f_{j,q})} \times \sum_h \sum_s \sum_v \min\{H_i(h,s,v), H_j(h,s,v)\}$$

$$A(f_{i,p}, f_{j,q}) = \min\left\{ \sum_h \sum_s \sum_v H_i(h,s,v), \sum_h \sum_s \sum_v H_j(h,s,v) \right\}。$$

这里,$I(f_{i,p}, f_{j,q})$表示镜头 G_i 的关键帧 $f_{i,p}$ 与镜头 G_j 的关键帧 $f_{j,q}$ 的直方图的交,用来判段两个关键帧的相似度。$H_i(h,s,v)$是 HSV 颜色空间的直方图,本文使用 H,S,V 分量在 18×3×3 的 3 维空间中的统计直方图,并以归一化后的 162 个数值作为颜色特征值。max 表示第 2 大的值,使用 M 和 $\tilde{M}$ 的平均值之所以可以加强算法的鲁棒性,是因为 HSV 颜色直方图表示的是关键帧,如果两个关键帧有相似的颜色分布,那么即使它们的内容不一样,也会认为这两个关键帧相似[10]。

2.2　基于滑动镜头窗的相似片段自动分割方法

设 V_q(下角 q 代表 query,下同)表示查询片段,V_d(下角 d 代表 datebase)表示视频数据库,相似片段分割的目的是在连续视频库 V_d 中查找与查询片段 V_q 相似的多个片段。滑动镜头窗是一个以镜头为单位的滑动的窗口(如图 1 所示)。

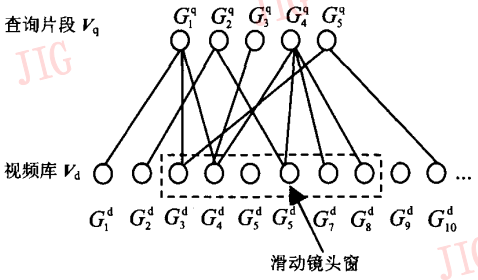

图 1　查询片段与视频库镜头的对应示意图

Fig. 1　Correspondence between query & database shots

由于视频片段是由表示同一语义的连续镜头组成,因此一个视频片段的内部镜头本身就会相似,这称为视频片段的自相似性。由于这种自相似性的存在,使得查询片段中的镜头和视频库中的镜头之间出现普遍的一对多、多对一、多对多的情况,如图 2~4所示,这些情况会导致得到一些伪相似片段而影响检索的精度。实际上,两个片段中相似镜头一一对应的数目才能真正反映片段的相似程度。基

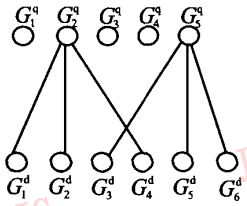

图 2　两个伪相似的视频片段(一对多)

Fig. 2　Two false similar video clips(one to many)

于上述原因,本文定义了视频片段的匹配函数,以便通过过滤掉伪相似片段来得到真正相似的多个片段。

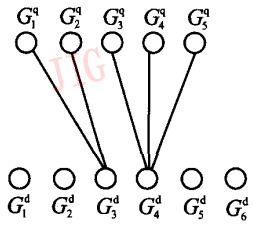

图 3　两个伪相似的视频片段(多对一)

Fig. 3　Two false similar video clips(many to one)

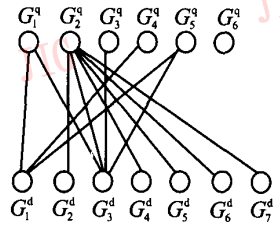

图 4　两个伪相似的视频片段(多对多)

Fig. 4　Two false similar video clips(many to many)

定义 1(镜头对应的等价关系)　设查询片段 V_q 中的镜头数为 K,滑动镜头窗中的镜头集合为 G_W(下角 W 代表 window,下同),窗口中的镜头数目为 L,对于查询片段 V_q 中的任一镜头 G_i^q,定义滑动镜头窗中与 G_i^q 相应的镜头的等价关系为 $R_i = \{\{[G_j^W]_{R_i}\}, \{G_W - [G_j^W]_{R_i}\}\}$,其中,镜头等价类 $\{[G_j^W]_{R_i}\} = \{G_j^W \mid S(G_i^q, G_j^W) \geqslant \beta_1\}$,$i = 1,2,\cdots,K$;$j = a, a+1, \cdots, a+L-1$,$a$ 为当前滑动镜头窗中的第 1 个镜头的序号,$S(G_i^q, G_j^W)$为镜头 G_i^q 和 G_j^W 的相似度,β_1 为阈值。显然,$\{[G_j^W]_{R_i}\} \cap \{G_W - [G_j^W]_{R_i}\} = \varnothing$,且 $\{[G_j^W]_{R_i}\} \cup \{G_W - [G_j^W]_{R_i}\} = G_W$。

这里的等价关系与通常意义的等价关系有一些不同,通常的等价关系是就某一个论域而言的,其某一对象的等价类中必然包含该对象本身,而本文定义的等价关系则是针对两个视频片段中的镜头而言的,其与查询片段中某一镜头对应的等价关系中的对象就是镜头窗中的镜头(另一片段中的镜头),对于查询片段中的任一镜头 G_i^q,从相似性的角度看,镜头窗中与 G_i^q 相似的所有镜头都可以看作是等价的,这就构成了一个等价类,可称其为相似等价类。同样镜头窗中与 G_i^q 不相似的镜头也可以看作是等

价的，这就构成了另一个等价类，可称其为不相似等价类。相似等价类$[G_i^q]_{R_i}$中对象的个数表征了镜头G_i^q一对多的镜头数目，而相似等价类相交的镜头数目则反映了查询片段镜头多对一的情况。

定义 2（匹配函数） 设查询片段V_q的镜头数为K，滑动镜头窗中的片段为V_W，片段V_W的镜头数为L，对于$\forall G_i^q \in V_q$，G_i^q相应的等价关系为R_i，则视频片段V_q和V_W之间的匹配函数定义为

$$r_q^W = \frac{\sum_{i=1}^{K} \delta_i}{L} \tag{2}$$

其中，

$$\delta_i = \begin{cases} 1 & \text{if } \{[G_j^W]_{R_i}\} \notin \{[G_j^W]_{R_n}\} \text{ 且 } \{[G_j^W]_{R_i}\} \neq \{[G_j^W]_{R_n}\} \\ 0 & \text{otherwise} \end{cases}$$

其中，$i=1,2,\cdots,K$，$n=1,2,\cdots,i-1$。定义 2 中的$\sum_{i=1}^{K}\delta_i$表征了两个视频片段一一对应的相似镜头的数目，假设查询片段和滑动镜头窗中的片段分别如图 2 中的两个片段所示，若根据定义 1 得到的与查询片段的各镜头相应的滑动镜头窗中的镜头的等价关系分别为$R_1=R_3=R_4=\{\varnothing,\{G_1^d G_2^d, G_3^d, G_4^d, G_5^d, G_6^d\}\}$，$R_2=\{\{G_1^d, G_2^d, G_4^d\},\{G_3^d, G_5^d, G_6^d\}\}$，$R_5=\{\{G_3^d, G_5^d, G_6^d\},\{G_1^d, G_2^d, G_4^d\}\}$，则查询片段和滑动镜头窗中的片段的匹配度为$r_q^W=2/6$，图 3 中两个片段的匹配度为$r_q^W=2/6$，图 4 中两个片段的匹配度为$r_q^W=3/7$，图 5 中两个片段的匹配度为$r_q^W=5/7$。下面介绍如何应用定义 2 给出的匹配函数，采用滑动镜头窗确定相似片段的位置，从视频库中自动分割出与查询片段真正相似的片段的方法。

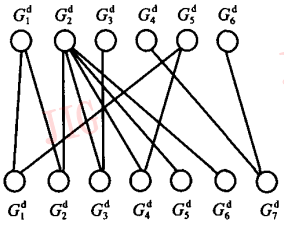

图 5 两个相似的视频片段

Fig. 5 Two similar video clips

算法 1 基于滑动镜头窗的相似片段自动分割方法

输入：查询片段$V_q=\{G_1^q, G_2^q, \cdots, G_K^q\}$及视频库$V_d=\{G_1^d, G_2^d, \cdots, G_N^d\}$

输出：m 个相似片段的镜头序列

（1）初始化，输入查询片段V_q及视频库V_d，并设定滑动窗口中镜头的数目为L，L的取值范围在$1.1K\sim1.2K$；

（2）如果视频库V_d为空，则算法停止，否则转第（3）步；

（3）对于视频库V_d中的每一个镜头G_i^d，如果G_i^d至少与查询片段V_q中的一个镜头相似，则把G_i^d记为G_j^S（上角 S 代表 similar），这样就得到一组相似镜头集合$G_{shot}^S=\{G_1^S, G_2^S, \cdots, G_J^S\}$；

（4）计算匹配度

①如果相似镜头集合为空，则算法停止，否则在当前的相似镜头G_j^S（$1\leqslant j\leqslant J$）位置使用滑动镜头窗；

②计算查询片段V_q与滑动镜头窗中片段V_W的匹配度r_q^W，如果$r_q^W\geqslant\beta_M$（下角 M 代表 match），则转第③步。否则$j=j+1$，转第①步；

③以一个镜头的增量移动滑动镜头窗，并计算每个镜头窗中片段V_W与查询片段V_q的匹配度r_q^W；

（5）确定相似片段

①以每一个窗口的起始位置的镜头序号为横坐标，绘制匹配度r_q^W的变化曲线（如图 6 所示）；

②设定阈值β_2，在与曲线上大于β_2的局部最大值对应的镜头中，当属于相似镜头序列G_{shot}^S的最小镜头序号对应于相似片段的起始位置$j=j+1$，则转第①步。

其中，β_M是匹配度阈值，只有当相似镜头对应的镜头窗中的片段与查询片段的匹配度大于β_M时，也就是说，只有与当前相似镜头G_j^S对应的镜头窗中与查询片段相似的镜头个数达到一定数目时，

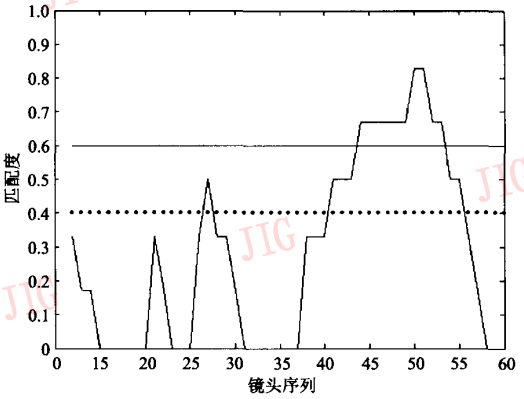

图 6 匹配度r_q^W的变化曲线

Fig. 6 Variation of matching degree r_q^W

算法才继续进行,否则忽略该相似镜头 G_j^S,这样处理是为了减少计算量和消除个别相似镜头的干扰。因为 G_j^S 是镜头窗的起始镜头,如果它对应的镜头窗的相似镜头很少,则说明它与相邻的下一个相似镜头 G_{j+1}^S 之间不存在相似片段。所以 β_M 的选择不易太大,通常选择在 0.15 ~ 0.2 之间,以确保不遗漏相似片段。β_2 用来限制相似片段的个数,若选择 β_2 的值越大,则所得到的相似片段的数目越少,这样那些与查询片段相似度低的片段就会被忽略(如图 6 所示),当设定 $\beta_2 = 0.4$ 时,曲线上有两个局部最大值大于 β_2,则可以得到两个相似片段;当设定 $\beta_2 = 0.6$ 时,曲线上只有一个局部最大值大于 β_2,则只能得到一个相似片段。为了避免那些因匹配镜头数很少而不足以判定相似片段的局部最优值的影响,β_2 的初值可先选择为 0.4 ~ 0.5,然后逐渐增大 β_2 的值,每次忽略一个最小的局部最优值,也就是忽略一个相似片段。这样,通过动态地调整阈值 β_2,就可以得到满足不同匹配度要求的相似片段。

对于上述算法,计算查询片段 V_q 的每个镜头与视频库 V_d 中的所有镜头的相似度,时间复杂度是 $O(KM)$,当某一个镜头窗中的片段 V_W 与查询片段 V_q 的匹配度计算完成后,下一个镜头窗只要计算最后一个镜头与查询片段的 δ_i 即可,因为前面($L-1$)个镜头已经在上一个镜头窗计算过,其时间复杂度最高为 $O(KM)$。由此可见,算法的时间复杂度是 $O(KM)$,其中,K 是查询片段 V_q 中的镜头数,M 是视频库 V_d 中的镜头数。

3　视频片段的相似度度量因子

由于镜头的相似性主要是基于视觉信息的,而视频片段的相似性,则除了视觉信息以外,还依赖于组成片段的镜头之间的内部关系,如粒度、时间顺序和干扰因子等,因此,按照相似度从高到低的顺序排列,如果已经得到与查询片段相似的多个片段,则必须考虑影响片段相似性的多个因子。为了计算用于度量片段相似度的不同因子,本文先把定义 1 中给出的相似等价类映射为一个矩阵,然后再由矩阵的特性计算这些影响因子。如图 5 所示的相似片段,可根据定义 1,计算与查询片段 V_q 的每个镜头对应的等价关系 $\boldsymbol{R}_1 = \{\{G_1^d, G_2^d\}, \{G_3^d, G_4^d, G_5^d, G_6^d, G_7^d\}\}$,$\boldsymbol{R}_2 = \{\{G_2^d, G_3^d, G_4^d, G_5^d, G_6^d\}, \{G_1^d, G_7^d\}\}$,$\boldsymbol{R}_3 = \{\{G_3^d\}, \{G_1^d, G_2^d, G_4^d, G_5^d, G_6^d, G_7^d\}\}$,$\boldsymbol{R}_4 = \boldsymbol{R}_6 = \{\{G_7^d\}, \{G_1^d, G_2^d, G_3^d, G_4^d, G_5^d, G_5^d, G_6^d\}\}$,$\boldsymbol{R}_5 = \{\{G_1^d, G_4^d\}, \{G_2^d, G_3^d, G_5^d, G_6^d, G_7^d\}\}$。映射后得到的矩阵为 $\boldsymbol{R} = \begin{bmatrix} s_{1,1} & s_{1,2} & 0 & 0 & 0 & 0 & 0 \\ 0 & s_{2,2} & s_{2,3} & s_{2,4} & s_{2,5} & s_{2,6} & 0 \\ 0 & 0 & s_{3,3} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s_{4,7} \\ s_{5,1} & 0 & 0 & s_{5,4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s_{6,7} \end{bmatrix}$,其中矩阵的行表示查询片段 V_q 的镜头,列表示相似片段 V_S 的镜头,$s_{i,j}$ 表示查询片段 V_q 中的镜头 G_i^q 与相似片段 V_S 的镜头 G_j^S 之间的相似度,0 表示两个镜头不相似,也就是说并不考虑不相似镜头的相似度。

3.1　视觉因子

两个片段在低层视觉特征(如颜色)上的相似程度,是决定片段相似最重要的因子。该视觉因子表征了两个片段的所有镜头在视觉上的综合相似程度。若用 $S_{\max}^{q,S}(i) = \max\limits_{1 \leqslant j \leqslant N} S(G_i^q, G_j^S)$ 表示查询片段 V_q 的某一个镜头 G_i^q 与相似片段 V_S 中所有镜头的相似度最大值,其中,G_j^S 表示相似片段的镜头,N 是相似片段的镜头个数。用 $\overline{S}_{\max}^{q,S}$ 表示查询片段 V_q 中所有镜头的相似度 $S_{\max}^{q,S}(i)$ $(1 \leqslant i \leqslant K)$ 总和的平均值,而相似片段 V_S 的某一个镜头 G_j^S 与查询片段 V_q 中所有镜头的相似度最大值用 $S_{\max}^{S,q}(j) = \max\limits_{1 \leqslant i \leqslant K} S(G_j^S, G_i^q)$ 表示,其中,K 是查询片段的镜头个数,用 $\overline{S}_{\max}^{S,q}$ 表示相似片段 V_s 的所有镜头的 $S_{\max}^{S,q}(j)$ $(1 \leqslant j \leqslant N)$ 的总和的平均值,则视觉因子 F_V(下角 V 代表 visual)定义为

$$F_V = \frac{1}{2}(\overline{S}_{\max}^{q,S} + \overline{S}_{\max}^{S,q}) \tag{3}$$

仔细分析矩阵 $\boldsymbol{R}$,可以发现,矩阵 $\boldsymbol{R}$ 的行元素的最大值及列元素的最大值分别与 $S_{\max}^{q,S}(i)$ 和 $S_{\max}^{S,q}(j)$ 的值相对应,因此 $\overline{S}_{\max}^{q,S}$、$\overline{S}_{\max}^{q,S}$ 的值可分别用下式计算:

$$\overline{S}_{\max}^{q,S} = \frac{\sum\limits_{i=1}^{K} \max\limits_{1 \leqslant j \geqslant N}(c_{i,j})}{K} \tag{4}$$

$$\overline{S}_{\max}^{S,q} = \frac{\sum\limits_{j=1}^{N} \max\limits_{1 \leqslant i \leqslant K}(c_{i,j})}{N} \tag{5}$$

计算视觉因子的时间复杂度是 $O(2KN)$,K,N 分别为查询片段 V_q 与相似片段 V_S 的镜头数目,而文献[10]计算视觉因子的复杂度为 $O(t^3)$,其中 $t = \max(K, N)$。所以本文算法的检索速度更快。

3.2 顺序因子

与查询片段 $\boldsymbol{V}_q$ 视觉相似的多个片段，它们与 $\boldsymbol{V}_q$ 的时间顺序可能不一致。在这种情况下，与 $\boldsymbol{V}_q$ 时间顺序相似的片段应该被赋予更高的相似度。也就是要考察查询片段 $\boldsymbol{V}_q$ 与相似片段 $\boldsymbol{V}_S$ 中的镜头按照时间顺序的对应情况，可通过计算两个片段所包含的最长公共镜头序列来度量顺序因子。通过观察矩阵 $\boldsymbol{R}$，不难发现，矩阵 $\boldsymbol{R}$ 的对角线上连续不为 0 的元素个数就是两个片段所包含的公共镜头序列号。设初始每个对角线的最长公共镜头序列为 $Q_{\max}(i)=0$，对于矩阵 $\boldsymbol{R}$ 的每个对角线，可用以下递归式计算公共镜头序列：

$$Q=\begin{cases}0 & \text{if} \quad i,j=0\\ Q+1 & \text{if} \quad c_{i,j}\neq 0\\ Q & \text{if} \quad c_{i,j}=0\end{cases} \tag{6}$$

如果 $Q\geqslant Q_{\max}(i)$，且 $Q_{\max}(i)=Q$，则最大的 $Q_{\max}(i)$ 就是两个片段所包含的最长公共镜头序列。定义顺序因子 F_O（下角 O 代表 order）为

$$F_O=\frac{\max\limits_{1\leqslant i\leqslant(K+N-1)}(Q_{\max}(i))}{K} \tag{7}$$

3.3 干扰因子

查询片段 $\boldsymbol{V}_q$ 与相似片段 $\boldsymbol{V}_S$（下角 S 代表 similar）各有一些零星的镜头，其找不到对应的相似镜头。这体现了镜头相似对应存在不连续性，这些零散的镜头越多，两个视频片段的相似程度越差。如果 $\boldsymbol{V}_q$ 中的某个镜头 G_i^q 在 $\boldsymbol{V}_S$ 中找不到相似镜头，则矩阵 $\boldsymbol{R}$ 的第 i 行的所有元素必然全是 0，同样如果 $\boldsymbol{V}_S$ 中的某个镜头在 $\boldsymbol{V}_q$ 中找不到相似镜头，则与它相应的列元素全为 0。所以找不到相似镜头的镜头数为

$$N_{\text{dissimilar}}=N_{\text{row}}+N_{\text{colum}} \tag{8}$$

这里 N_{row}，N_{colum} 分别为矩阵 $\boldsymbol{R}$ 的全 0 行和全 0 列的个数。干扰因子 F_I（下角 I 代表 interference）定义为 $F_I=1-\frac{N_{\text{dissimilar}}}{K+N}$。这里 K，N 分别为查询片段 $\boldsymbol{V}_q$ 与相似片段 $\boldsymbol{V}_S$ 的镜头数目，F_I 表明了在两个片段中，能找到对应相似镜头的镜头比例。

3.4 视频的总体相似度

以上都是从视频角度考虑镜头相似关系对视频总体相似的影响。而从镜头层次相似出发，考虑镜头组成的视频片段的相似性，既保证了微观底层上的相似，又保证了宏观整体上的相似，且具有粒度性，则查询片段 $\boldsymbol{V}_q$ 与它的相似片段 $\boldsymbol{V}_S$ 的相似度可用下式计算：

$$S(\boldsymbol{V}_q,\boldsymbol{V}_S)=\omega_V\cdot F_V+\omega_O\cdot F_O+\omega_I\cdot F_I \tag{9}$$

这里，ω_V，ω_O，ω_I 是各个因子权重，表明了用户对视觉因子、顺序因子、干扰因子的重视程度，不同的用户可以根据自己的判段标准来调整。

4 实验结果

为验证本文方法的检索效果，从电视视频中录制了几天的节目作为实验的视频数据库，总长为 263min，共 5 939 个镜头，360 162 帧图像。视频数据库中包括新闻、广告、电视剧、体育、谈话等各种类型的节目。这里面有重复的相同片段，如广告、影视片段等，对于这些片段，可以进行精确的片段检索；有相似的视频片段，如体育节目中不同的排球比赛、不同时间长度和编辑的相同广告、相同影片预告等，对于这样的片段，也能够实现相似的片段检索，可见这个视频库非常具有挑战性。图 7 的第 1 行是查询片段，共从视频库中检索到 4 个相似片段，从第 2 行开始，从上到下按照相似度从高到低的顺序排列相似片段，即与查询片段最相似的片段排在首位，依次往下，排在最后的片段与查询片段的相似程度最低。最右边 1 列数据显示了每一个相似片段与查询片段之间的相似度。排列的相似片段体现了不同因子的作用，如前两个相似片段的排列体现了顺序因子的作用，而后两个相似片段则更多地体现了视觉因子和干扰因子的作用。

文献[10]的方法是采用基于图论的最大匹配算法和最优匹配算法实现的，其与本文提出的基于等价关系理论的滑动镜头窗方法并不相同，但是由于这两种方法都能够从连续视频库中自动分割出相似片段，同时考虑了影响片段的相似度的不同度量因子，并可按照相似度从大到小的顺序排列它们，而且文献[10]的方法是目前类似方法中检索效果最好的一种新的方法，因此，本文将本文提出的方法与文献[10]的方法进行实验比较，以验证本文方法的性能。并采用视频检索常用的两个评价指标——准确率和召回率来进行评价，表 1 所示为文献[10]的视频片段检索的实验结果，表 2 所示为本文的方法的视频片段检索的结果。表 1，表 2 的前 3 个片段是精确的片段检索，后 4 个片段是相似的片段检索。在表 2 中，对于排球比赛片段的检索，视频库中出现排球比赛的片段共 3 次，本文的检索方法漏掉了

图 7　相似视频片段的检索和排列结果

Fig. 7　Results of retrieval and ranking of similar clips

表 1　基于图论的视频片段检索的实验结果[10]

Tab. 1　Results of clip retrieval based on graph theory[10]

视频片段	帧的数目	准确率(%)	召回率(%)	检索时间(s)
雪碧广告	391	100	100	62
潘婷广告	455	100	100	84
足球新闻	809	100	100	53
排球比赛片段	614	100	66.7	65
人鱼小姐剧情预告片段	1 069	60	75	321
每日 C 广告	463	100	100	114
脑白金广告	374	100	60	129

表 2　本文方法的视频片段检索的实验结果

Tab. 2　Results of clip retrieval proposed in this paper

视频片段	帧的数目	准确率(%)	召回率(%)	检索时间(s)
雪碧广告	391	100	100	41
潘婷广告	455	100	100	50
足球新闻	809	100	100	39
排球比赛片段	614	100	66.7	44
人鱼小姐剧情预告片段	1 069	75	100	95
每日 C 广告	463	100	100	86
脑白金广告	374	100	100	82

1 个,原因是片段中主要是选手和观众镜头,反映比赛的镜头很少。同样文献[10]也漏掉了这个片段。片段 5 的语义较强,由于视频库中有 1 个剧情预告片段的颜色特征与它很相似,因此这个片段也被作为相似片段。比较表 1 和表 2 的实验结果可见,在准确率和召回率上本文算法能够达到等同或优于文献[10]的效果。但在检索速度上,本文的方法比文献[10]的方法快,特别是对于长的视频片段,本文方法的优点越明显。例如,对于查询片段 5,文献[10]算法的检索时间是本文算法的 3 倍多。

5　结　论

本文给出了查询片段的镜头对应的等价关系和片段匹配函数的定义,并实现了基于滑动镜头窗的视频片段自动分割,而且只要对视频库一次性浏览,就可以检索到真正与查询片段相似的多个片段。此外,影响片段相似度的各个因子的计算快速简单,因此,与同功能的算法[10]相比,本文算法的检索速度更快,特别是对于长的查询片段的检索,检索速度的优点更加明显。同时能够达到较高的检索精度,是一种快速有效的视频片段检索方法。但本文的方法仍然存在一些问题,例如上述实验中片段 4,5 的检索精度较低的问题;对于同一个体育比赛节目的连续视频中,片段的精确检索效果不是太好,今后将继

续致力于这些问题的研究。

参考文献(References)

1 Zhao Li, Qi Wei, Li Zi-qing, *et al.* Content-based retrieval of video shot using the improved nearest feature line method[J]. Journal of Software, 2002, **13**(4): 586 ~ 590. [赵黎,祁卫,李子青等. 利用改进 NFL 算法对镜头进行基于内容的检索[J]. 软件学报, 2002, **13**(4): 586 ~ 590.]

2 Lin Tong, Zhang Hong-jiang, Feng Ju-fu, *et al.* Shot content analysis for video retrieval application [J]. Journal of Software, 2002, **13**(8): 1577 ~ 1585. [林通,张宏江,封举富等. 镜头内容分析及其在视频检索中的应用[J]. 软件学报, 2002, **13**(8): 1577 ~ 1585.]

3 Ngo C W, Pong T C, Zhang H J. On clustering and retrieval of video shots through temporal slices analysis [J]. IEEE Transactions on Multimedia, 2002, **4**(4): 446 ~ 459.

4 Naphade M R. A novel scheme for fast and efficient video sequence matching using compact signatures [A]. In: Proceedings of SPIE Storage and Retrieval for Media Databases[C], San Jose, California, USA, 2000: 564 ~ 572.

5 Mohan R. Video sequence matching [A]. In: International Conference on Acoustics, Speech, and Signal Processing(ICASSP'98) [C], Washington, USA, 1998: 3679 ~ 3700.

6 Jain A K, Vailaya A, Wei X. Query by video clip[J]. Multimedia ystem, 1999, **7**(5): 369 ~ 384.

7 Kim Sang-Hyun, Park Rae-Hong. An efficient algorithm for video sequence matching using the modified Hausdorff distance and the directed divergence[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, **12**(7): 592 ~ 596.

8 Zhuang Yue-ting, Liu Xiao-ming, Wu Yi, *et al.* A new approach to retrieve video by example video clip [J]. Chinese Journal of Computers, 2000, **23**(3): 300 ~ 305. [庄越挺, 刘小明, 吴翌等. 通过例子视频进行视频检索的新方法[J]. 计算机学报, 2000, **23**(3): 300 ~ 305.]

9 Chen L P, Chua T S. A match and tiling approach to content-based video retrieval[C]. In: Proceedings of IEEE International Conference on Multimedia and Expo(ICME 2001)[C], Tokyo, Janpan, 2001: 417 ~ 420.

10 Peng Yu-xin, Ngo C W, Dong Qing-jie, *et al.* An approach for video retrieval by video clip [J]. Journal of Software, 2003, **14**(8): 1409 ~ 1417. [彭宇新, Ngo Chong-Wah, 董庆杰等. 一种通过视频片段进行视频检索的方法[J]. 软件学报, 2003, **14**(8): 1409 ~ 1417.]