第 33 卷 第 9 期 计 算 机 工 程 2007 年 5 月
Vol.33 No.9 Computer Engineering May 2007
·软件技术与数据库· 文章编号：1000—3428(2007)09—0095—02 文献标识码：A 中图分类号：TP391.4

基于向量空间模型的视频语义相关内容挖掘

谢晓能[1,2]，吴 飞[1]
(1. 浙江大学人工智能研究所，杭州 310027；2. 杭州广播电视大学，杭州 310009)

摘 要：对海量视频数据库中所蕴涵的语义相关内容进行挖掘分析，是视频摘要生成方法面临的难题。该文提出了一种基于向量空间模型的视频语义相关内容挖掘方法：对新闻视频进行预处理，将视频转化为向量形式的数据集，采用主题关键帧提取算法对视频聚类内容进行挖掘，保留蕴涵场景独特信息的关键帧，去除视频中冗余的内容，这些主题关键帧按原有的时间顺序排列生成视频的摘要。实验结果表明，使用该视频语义相关内容挖掘的算法生成的新闻视频具有良好的压缩率和内容涵盖率。

关键词：向量空间模型；主题关键帧；视频摘要

Semantic Content Mining Approach in Video Based on Vector Space Model

XIE Xiaoneng[1,2], WU Fei[1]
(1. Institute of Artificial Intelligence, Zhejiang University, Hangzhou 310027;
2. Hangzhou Radio & TV University, Hangzhou 310009)

【Abstract】Video summarization is receiving increasing attention to mining semantic contents in huge video databases. This paper proposes a novel emantic content mining approach that mines subject keyframes by an algorithm based on vector space model. After pre-processing, video is transformed into a relational dataset of keyframe classes. Using subject keyframe detection algorithm, it keeps the pertinent keyframes that distinguish one scene from others and remove the visual-content redundancy from video content. The corresponding summary is obtained by assembling them by their original temporal order. Experiments are conducted to evaluate the effectiveness of the proposed approach with summary compression ratio and content coverage. The results demonstrate that meaningful news video summaries is generated.

【Key words】 Vector space model; Subject keyframe; Video summarization

视频摘要技术提供了对视频的高效浏览以及重要内容准确定位的方法，挖掘涵盖最大内容的视频模式是视频摘要技术研究的重点，特别是对于大量的、具有语义相关的新闻类视频数据，有效去除和发现冗余的重复信息，可大大增强视频摘要表达视频独特内容信息的能力[1]。

在海量视频集、视频新闻或记录片数据库中，反映同一主题事件的重要镜头或场景往往反复出现，虽然每次报道的表现形式不同（不同记者、主持人，不同剪辑方式等），但这些镜头或场景是语义密切相关的[1,2]。因此，通过分析这些视频中的各个场景的相似性和差异性，挖掘其关联信息，去除冗余信息，使得摘要真正由最精要的要素组成，又能表达各个场景独特的内容信息，是目前视频摘要生成面临的难题。

1 基于向量空间模型的视频数据挖掘

基于向量空间模型(Vector Space Model, VSM)在文本挖掘中被广泛应用，可以基于这种模型去发现文本潜在的概念以及概念间的相互关系，发现隐含的知识[3~5]。通常由TF/IDF公式来计算，TF和IDF分别指文档词频(Term Frequency, TF)和逆文档词频(Inverse Document Frequency, IDF)。根据TF/IDF公式，文档集中包含某一词条的文档越多，说明它区分文档类别属性的能力越低，其权值越小；另一方面，某一文档中某一词条出现的频率越高，说明它区分文档内容属性的能力越强，其权值越大。这种方法综合地考虑了一个词的出现频率和这个词对不同文档的分辨能力，因此在文档的主题提取和自动摘要生成方面是非常有效的。

在此，不妨把视频的层次化内容结构与文档进行如下类比：每个场景描述一个事件或话题，类似文档的每一个段落；组成场景的相关镜头就是句子；聚类的镜头关键帧正如文档的词条。基于视频的向量空间模型假设，本文提出新的关键帧权值算法，称之为主题关键帧提取算法。其指导思想是，能作为场景主题的关键帧不仅是在一个场景中重复出现的，而且是跟同一个场景中其它的关键帧结合起来对场景内容的区分能力比较强。如新闻节目的同一个主持人镜头虽然在多段新闻中重复出现，但与其他的关键帧的关联性显然是非常弱的，因此是冗余成分。而那些仅在某一个场景中重复出现的重要镜头显然对于理解该场景内容很有帮助，可以作为摘要内容。其基本过程如图 1 所示。

先对视频流进行镜头的切分，然后支持向量聚类，再用本文提出的主题关键帧提取算法计算主题度，挖掘出视频序列的时间关联和空间关联，提取关键帧，生成摘要。

基金项目：国家自然科学基金资助项目(60272031)；浙江省自然科学基金资助项目(M603202)
作者简介：谢晓能(1977－)，女，讲师、硕士生，主研方向：多媒体技术；吴 飞，博士、副教授
收稿日期：2006-06-01 **E-mail：**zyx_xxn@126.com

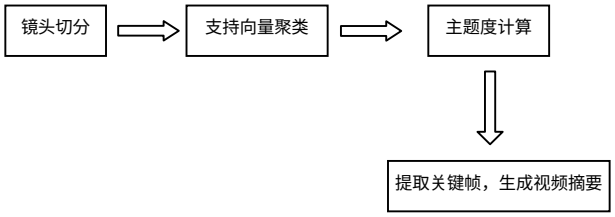

图 1 摘要生成方法的简要流程

2 视频的预处理及语义相关内容挖掘

为了挖掘非结构化的视频流数据，生成视频摘要，首要的工作是视频的预处理，其目的是将视频流转化为可采用数据挖掘技术进行分析的数据集结构。主要包括镜头切分、聚类和新闻条目切分[7,8]。

经过视频预处理之后，每个视频镜头可以归属到一个聚类子集。假设Ω为一个新闻条目，它包含了m个镜头，本文中把每个镜头的第一个关键帧作为代表帧，即：$k_1,k_2,...,k_m$，通过SVC聚类，m个镜头聚类成q个组，则

$$\Omega=\left\{G_1\left(S_{11}...S_{1l}\right),...,G_q\left(S_{q1}...S_{qr}\right)\right\}$$

其中，S_{ij}为第 i 组中第 j 个镜头的关键帧。本文用 A、B、C、D 等符号来表示某一种聚类。例如：假设某个包含 7 个镜头的视频聚类后的结果是：ABACDBA，构成如下 4 个组：A(含 3 个镜头)，B(含 2 个镜头)，C(含 1 个镜头)，D(含 1 个镜头)，则此视频可以表达为

$$\Omega=\{A(S_{11},S_{12},S_{13}),B(S_{21},S_{22}),C(S_{31}),D(S_{41})\}$$

然后，可以采用基于向量空间模型的方法来挖掘这些视频镜头所蕴涵的关系。假设视频 V 包含 n 个场景（即新闻条目），将视频转化为向量的形式表示，

$$V=\{k_1,\omega_1;k_2,\omega_2...k_i,\omega_i...k_n,\omega_n\}$$

ω_i为关键帧 k_i 聚类后得到的 S_{ij} 在视频 V 中的权值。这个权值又称“主题度”，是用来判断一个关键帧为主题关键帧的概率，计算公式为

$$\omega(k_i)=tf(k_i)+\sum_{j=1,j\neq i}^{m}\overrightarrow{Fcross}(k_i,k_j)$$

其中，m 为场景中关键帧的个数。

$$\overrightarrow{Fcross}(k_i,k_j)=df(\{k_i,k_j\})/df(\{k_i\})$$

指包含关键帧 k_i 的场景中也同时包含关键帧 k_j 的概率，其中，$df(\{k_i,k_j\})$是同时出现 k_i 和 k_j 的场景的个数，$df(\{k_i\})$是包含 k_i 的场景的个数。

在这个方法中，能反映主题的这些关键帧的“上下文关系”即语义相关被强调，即当一个关键帧出现在一个以上场景的时候，不仅要确定它在这段场景中的重要性，还必须确定它对整个摘要的影响程度（对不同场景的分辨能力），二者综合考虑构成权值，即主题度。

3 实验建立和分析

根据以上的算法，在 Windows2000 下用 C++6.0 实现了整个算法。为了评价生成摘要的有效性，本文采用了摘要压缩比（Summary Compression Ratio, SCR）和内容覆盖率（Content Coverage, CC）两个评判度量公式：SCR 为摘要的关键帧数目占视频总关键帧数目的比例，CC 为自动生成的关键帧符合人为判定的关键帧的比例，用来评价摘要完整表达原来视频的能力。对于一个同样的 SCR，如果 CC 越大，则显然生成的视频摘要越好，越能完整表达原来视频的内容。

选取中央电视台新闻频道总共 50 多个小时的视频节目进行测试。按照统一格式录制为每段 5min，再随机将数量不等的片段进行组合，形成不同长度的视频片段用于实验。实验结果如表 1 所示。

表 1 新闻视频的视频语义相关内容挖掘结果

视频长度(min)	总镜头数	聚类数目	SCR(%)	CC(%)
5	41	15	17.1	46.2
10	87	47	25.3	45.0
15	81	34	32.1	47.8
20	93	30	20.4	57.7
30	194	48	11.3	66.7

从表 1 可看出这个方法生成的新闻视频摘要有很好的内容表达能力，达到 50%左右，同时摘要的压缩比没有比传统的方法明显地增加，这主要归因于这个算法对那些重复场景敏感的识别，使得摘要压缩率没有明显升高，这正是本算法优点的体现。

4 结论

本文对视频通过镜头切分和聚类将视频流转化为基于向量空间模型的数据集结构，并且采用主题关键帧提取方法去挖掘含有语义相关内容的视频作为摘要信息，通过这种视频语义相关内容挖掘生成的视频摘要具有良好的压缩率和内容涵盖率。

参考文献

1 Hu J, Zhong J, Bagga A. Combined-media Video Tracking for Summarization[C]//Proceedings of the 9th ACM International Conference on Multimedia. 2001: 502-505.

2 Yahiaoui I, Merialdo B, Huet B. Comparison of Multiepisode Video Summarization Algorithms[J]. EURASIP Journal on Applied Processing, 2003, (1): 48-55.

3 Salton G, Wong A, Yang C S. On the Specification of Term Values in Automatic Indexing[J]. Journal of Documentation, 1973, 29(4): 351-372.

4 Salton G, McGill M. Introduction to Modern Information Retrieval[M]. New York: McGraw-Hill, 1983

5 Navarro G, Raffinot M. Fast and Flexible String Matching by Combining Bit-parallelism and Suffix Automata[J]. ACM Journal of Experimental Algorithmics, 2000, 5(4): 1-36.

6 Sebe N, Lew M S, Smeulders A W M. Video Retrieval and Summarization[J]. Computer Vision and Image Understanding, 2003, 92(2/3): 141-146.

7 Hsu W, Chang S F. A Statistical Framework for Fusing Mid-level Perceptual Features in News Story Segmentation[C]//Proc. of IEEE International Conference on Multimedia and Expo.. 2003: 413-416.

8 叶朝阳, 吴　飞, 庄越挺. 鲁棒的镜头边缘检测融合算法[J]. 计算机辅助设计与图形学学报, 2003, 15(11): 1386-1392.