一种基于多关键字的新闻视频自动检索方法

周 生[1,2] 胡晓峰[1] 罗 批[1]
(国防大学信息作战与指挥训练教研部 北京 100091)[1] (解放军炮兵学院基础部计算中心 合肥 230031)[2]

摘 要 针对TBVR技术中人工标注存在的问题和CBVR技术的不成熟,以及对虚拟新闻系统中视频检索需求和特点的深入分析,在TBVR的基础上提出了一种基于多关键字的新闻视频自动检索方法。详细讨论了标注字典库、树形标注结构、关键字自动获取、相似度计算模型和自动检索算法等问题,并进行了实验验证。结果表明,提出的方法在对新闻视频进行检索时取得了很高的查全率,同时取得了较高的查准率,能够解决虚拟新闻系统中视频自动检索的问题。

关键词 虚拟新闻,视频检索,视频相似度,战略对抗演习
中图法分类号 TP391.9 **文献标识码** A

News Video Automatic Retrieval Method Based on Multiple Keywords

ZHOU Sheng[1,2] HU Xiao-feng[1] LUO Pi[1]
(The Department of Information Operation & Command Training, NDU of PLA, Beijing 100091, China)[1]
(Computer Centre of Basic Department of Artillery Academy of PLA, Hefei 230031, China)[2]

Abstract Aiming at the problem of manual annotation in TBVR and the immatureness of CBVR and in-depth analysis on the requirements and features of video retrieval in virtual television news, a news video retrieval method based on multiple keywords was proposed. Annotation dictionary, tree annotation structure, automatic keywords extraction, similarity computation model and automatic retrieval algorithm were discussed in detail. An experiment was made on this method. The result shows that it has attained a very high recall rate with a comparative high accuracy rate and it can solve the problem of automatic video retrieval in virtual television news.

Keywords Virtual television news, Video retrieval, Video similarity, Strategic seminar gaming

1 引言

随着网络技术和存储技术的发展,视频信息逐渐成为信息传输和存储的主体。如何在大规模的视频数据库中进行有效的检索,是多媒体领域研究的一个热点问题。目前视频检索有两种方法:基于注释的视频检索(TBVR)和基于内容的视频检索(CBVR)。TBVR利用文字对视频进行标注,然后通过数据库检索技术来检索视频,属于传统的检索方法在视频数据库中的应用。CBVR则是利用视频自身的视觉特征,如颜色、纹理、形状、运动等来检索视频[1,2]。

虽然CBVR是当前视频检索中研究的热点,但是其技术还比较初级,只是利用一些相对简单的特征来检索,技术尚没有真正成熟。并且CBVR系统都有一个人机交互的界面,需要用户干预和调整。目前大多数成熟的检索系统仍是基于传统的方法,例如著名的互联网搜索引擎Google、百度等对图片和视频的检索大都采用TBVR技术。TBVR技术存在的主要问题是对大量的视频数据进行人工标注时,工作量大和不同的人对同一视频内容理解的不同而导致的标注差异。

虚拟新闻系统[3,4]是计算机利用过去的电视新闻片断生成的、模拟未来危机事件的电视新闻。该系统采用叙事方式[5,6]展现战略对抗演习[7]中的危机事件。虚拟新闻系统需要对视频素材库进行检索,它的检索与一般的视频检索相比有3个突出的特点:第一,由计算机自动提供检索用的关键字;第二,计算机自动从候选视频中选优;第三,视频检索要服务于叙事性的需要。

针对这3个特点以及CBVR技术和TBVR技术各自的优缺点,本文在TBVR技术的基础上,提出一种综合利用多关键字的视频标注、检索和相似度计算的方法,以实现虚拟新闻系统视频检索的自动化。

2 新闻视频素材的标注

TBVR技术不可避免要对视频进行文字标注。标注主要存在两个问题:(1)人工对视频进行标注,工作量非常大;(2)不同的人对同一段视频的理解存在着差异,在标注时可能存在差别。对于第一个问题,由于自动标注技术不够成熟,没有什么好的方法。对于第二个问题,在本文中设计一个专用的

到稿日期:2009-02-20 返修日期:2009-05-05 本文受国家863项目(编号:2006AA01Z337)和国家973项目(编号:2006CB303106)资助。
周 生(1978-),男,博士研究生,研究方向为战争模拟、可视化表现,E-mail:unbend@126.com;胡晓峰(1957-),男,教授,博士生导师,研究方向为战争模拟系统与环境、军事运筹、军事信息系统工程等;罗 批(1974-),男,博士后,副教授,研究方向为战争复杂性、战争模拟、遗传算法等。

标注字典库来解决。

2.1 标注字典库

由于虚拟新闻系统是应用于战略对抗演习的，因此该系统中用于展现演习中的新闻视频主要涉及政治、经济、军事和外交4个方面，而不涉及一般新闻视频中的体育、科技、文化、猎奇等其他方面。也就是说，虚拟新闻系统中使用到的新闻视频是限于特定领域的，这就使得建立一个字典库来标注新闻视频成为可能。

虚拟新闻采用叙事性方式来展现演习中的危机事件。叙事性方式一般包括时间、地点、人物、故事情节等要素。所以，对视频片断的标注也要符合叙事的需要，对每个视频片断按照叙事方式中的时间、地点、人物、故事等要素进行标注。有些视频中可能表现的不是具体的人物，而是一些其他的实体，比如军队、民众或者股市等，所以将人物和其他实体统一抽象为对象。由于演习是对未来危机事件的推演，未来的时间不可能和过去的时间相吻合，因此时间要素不标注。地点和对象标注较为简单。比较难以处理的是新闻视频的故事情节。如果完整地描述整个故事情节并标注，会引起3个问题：(1)完整的故事情节不好标注，工作量大，不同的人有不同的理解；(2)对整个情节的标注不利于后续的检索；(3)演习中要赋予现有视频新的故事内容，所以对情节进行完整标注也没有意义。基于以上3点理由，本文对故事情节进行简化处理，用视频中人物的动作来代替。根据对象、地点和动作来设计的标注字典库的内容如表1所列。

表1 标注字典库(部分)

词条名称	类型	词性	所属国家(地区)	说明
外交部发言人	对象	名词	A	可通用
国防部部长	对象	名词	B	可通用
商业部长	对象	名词	C	可通用
海军	对象	名词	D	可通用
民众	对象	名词	E	可通用
股市	对象	名词	F	可通用
发表讲话	动作	动词	A	可通用
举行会谈	动作	动词	B	可通用
进行访问	动作	动词	C	可通用
北京	地点	名词	D	不可通用
华盛顿	地点	名词	E	不可通用

因为战略对抗演习是由参演人员扮演国家决策部门的角色，是一种角色扮演，而不是人物模仿，所以标注字典库中不收录具体的人物姓名。目前整个字典库收录词条789个，按照对象、动作、地点分类管理。所属国家(地区)用演习中参演各方来代替。说明中的可通用则表明，该标注词条除了适用于本方，也可用于其他方的标注。

2.2 三叉树树形标注结构

建立了对象、动作和地点标注字典库后，视频素材库中的视频片断就可以用ENBF描述为：

〈Video Clip〉:=〈Place〉〈Objects〉〈Actions〉

针对视频中多个对象的共现问题，比如中国国家主席和美国总统同时出现，对象又可以表示为：

〈Objects〉:=〈object+〉

用动作来描述新闻视频的故事情节，有时可能不太准确，所以对动作再进行补充描述，动作可表示为：

〈Actions〉:=〈detail+〉

根据ENBF范式，用三叉树对视频片断进行标注，如图1所示。与三叉树对应的数据结构描述如下：

```
typedef struct TripleNode {
    string tagInfo;
    string extendedInfo;
    struct TripleNode * lChild, * mChild, * rChild;
}TriNode, * TriTree;
```

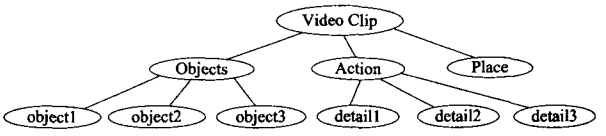

图1 新闻视频标注三叉树

对三叉树及其数据结构说明如下。

(1)三叉树的数据结构可以满足标注的需要。如果采用更复杂的树形结构，会增加标注的工作量和检索的复杂度。

(2)三叉树的数据结构可以解决一般视频中多个重要人物同时出现的标注问题。对于像六方会谈、APEC会议等这种多个国家领导人同时出现的视频标注，采用的方法是在extendedInfo中进行附加说明，而不是一一标注每个人物。

(3)三叉树的数据结构可以解决单一节点描述动作不够精确的问题，同时又不至于描述过细，增加检索的困难。

用以上的三叉树对国务院副总理王岐山与美国财长保尔森会谈的视频标注如图2所示。

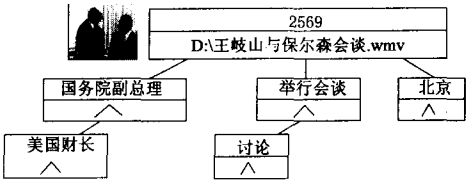

图2 标注实例

3 视频自动检索方法

要想实现视频的自动检索，需要解决3个根本的问题：(1)计算机怎样自动获取检索用的关键字；(2)视频相似度计算模型；(3)自动检索算法。下面对这3个问题分别进行论述。

3.1 自动获取检索关键字

检索用的关键字是计算机自动从参加演习的受训者撰写的新闻稿中切分出来的。新闻稿要求与真实的新闻稿的规范相符。下面以2007年10月3日《参考消息》上的一条新闻稿加以说明。

新闻稿例子：【韩联社平壤10月3日电】韩国总统卢武铉与北韩领导人金正日于今天上午结束了第一轮首脑会晤。两位领导人就推动朝鲜半岛和平以及南北经济合作等问题展开了深入的讨论。

按照叙事性方式的要求，分别对新闻稿中的人物(或对象)、动作和地点建立字典库。字典库是按照演习的参演方(战略对抗演习由多方组成，类似于朝核问题六方会谈)独立建立的，这样可以减少检索的时间复杂度。在字典库的基础上，采用现代汉语的自动分词技术[8-10]，切分出检索用的关键字{Objects,Actions,Place}。例子中切分出来的关键字是："韩国总统,北韩领导人\会晤,讨论\平壤"(不同类型的关键字利用"\"作为分割)，将关键字同样存储到三叉树的数据结

构中，如图 3 所示。

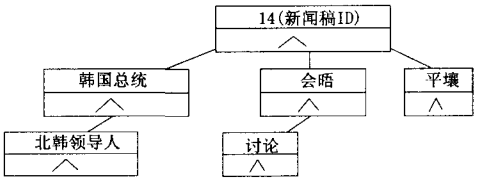

图 3　关键字自动切分实例

3.2　视频相似度计算模型

视频相似度的计算方法是解决视频自动检索和选优的根本问题。由于视频素材是按照叙事方式的地点、对象、动作 3 个要素进行标注的，而新闻稿中检索用的关键字也是按照地点、对象、动作 3 类关键字进行切分的，因此本文提出一种 OAP (Objects, Place, Action)视频相似度计算模型，如图 4 所示。

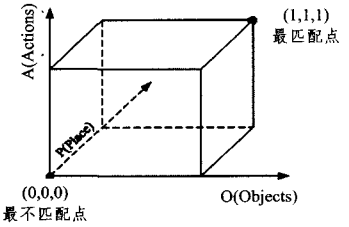

图 4　OAP 相似度计算模型

在这个三维空间模型中，O 代表对象的相似度，取值范围为[0,1]。A 代表动作的相似度，取值范围为[0,1]。P 代表地点的相似度，取值范围为[0,1]。整个视频相似度 Match 则由落在三维空间中点的欧式距离来决定。

$$Match=\sqrt{P^2+O^2+A^2} \tag{1}$$

下面分别说明 P,O,A 的计算方法。

(1)地点相似度 P 的计算

P 是 OAP 三维空间计算中较为容易的，不需要采取什么特别的处理，采用直接匹配的方法，P 的取值范围虽然为[0,1]，实际的计算结果只有两个值{0,1}。如果完全匹配，$P=1$；不完全匹配，$P=0$。

(2)动作相似度 A 的计算

A 在 OAP 三维空间模型中的计算，不能采用像 P 一样的计算方法，因为从视频标注和关键字自动切分可以得知，A 可能涉及多个，这只是一个方面。另一方面，A 在计算之前，如果不做一些转换处理，可能存在相似度计算错误。比如某段视频中动作的标注是“演讲”，而从新闻稿中提取的动作关键字是“演说”。“演讲”和“演说”在语义角度是完全匹配的，而直接比较关键字则是不匹配的。所以 A 的计算分 3 步进行。

第一步　计算切分后关键字三叉树中动作关键字的数目 n，即 Actions, Actions. lChild, Actions. mChild, Actions. rChild 非空节点的个数。

第二步　将 Actions. tagInfo, Actinos. lChild. tagInfo, Actions. mChild. tagInfo, Actions. rChild. tagInfo 进行转换处理。具体方法是为视频素材标注字典库中的每个动作词条建立一个同义词表，如表 2 所列。

表 2　动作同义词表

词条名称	字段名	类型	长度	说明
动作名称	Action	Nvarchar	20	出自标注库
同义词 1	Synonym1	Nvarchar	20	出自新闻稿
同义词 2	Synonym2	Nvarchar	20	出自新闻稿
同义词 3	Synonym3	Nvarchar	20	出自新闻稿
同义词 4	Synonym4	Nvarchar	20	出自新闻稿
同义词 5	Synonym5	Nvarchar	20	出自新闻稿

动作同义词表中为每一动作设计了 5 个同义词，多数少于 5 个。这也是在查阅了汉语词典中大量动词的同义词列表后做出的设计。

如果自动切分出来的动作关键字是表中 Synonym1 — Synonym5，则将其转换为对应的 Action。

第三步　依次比较 n 个关键字，每匹配一个，A 的相似度增加 $1/n$。

$$A=1/n\times\text{匹配的个数} \tag{2}$$

(3)对象相似度 O 的计算

O 在 OAP 三维空间模型中的计算，类似于 A 的计算方法，要考虑到语义匹配而字符不匹配的问题，比如“总统”和“国家领导人”的匹配计算问题。所以 O 的计算也分 3 步进行。

第一步　计算切分后关键字三叉树中对象关键字的数目 n，即 Objects, Objects. lChild, Objects. mChild, Objects. rChild 非空节点的个数。

第二步　将 Objects. tagInfo, Objects. lChild. tagInfo, Objects. mChild. tagInfo, Objects. rChild. tagInfo 进行转换处理。具体方法也是为视频素材标注字典库中的每个对象词条建立一个同义词表，如表 3 所列。

表 3　对象同义词表

词条名称	字段名	类型	长度	说明
对象名称	Object	Nvarchar	20	出自标注库
同义词 1	Synonym1	Nvarchar	20	出自新闻稿
同义词 2	Synonym2	Nvarchar	20	出自新闻稿
同义词 3	Synonym3	Nvarchar	20	出自新闻稿
同义词 4	Synonym4	Nvarchar	20	出自新闻稿
同义词 5	Synonym5	Nvarchar	20	出自新闻稿
同义词 6	Synonym6	Nvarchar	20	出自新闻稿

对象同义词表中为每一对象设计了 6 个同义词，多数少于 6 个。这是在统计了一些对象的同义词列表后做出的设计。

如果自动切分出来的动作关键字是表中 Synonym1 — Synonym6，则将其转换为对应的 Object。

第三步　依次比较 n 个关键字，每匹配一个，O 的相似度增加 $1/n$。

$$O=1/n\times\text{匹配的个数} \tag{3}$$

3.3　视频自动检索算法

视频自动检索算法按照对象、动作、地点进行匹配，其中优先对对象进行匹配比较。如果对象不匹配，则后续的动作和地点的匹配就没有任何意义。所以在 OAP 视频相似度计算模型的基础上，整个算法描述如下。

Input：一条演习中参演人员提交的新闻稿；

Output：一段与新闻稿文字内容匹配的视频片断；

声明变量：TriNode videoClip, keyWords；

第一步　按照汉语自动分词技术将新闻稿中的关键词进行切分，并存储到 keyWords 中；

第二步　对 keyWords 中的 Objects 和 Actions 节点进行

(下转第 188 页)

Z	0.0047	0.0047	0.0047	0.0047	0.0062	0.0062
Gray	0.0031	0.0047	0.0047	0.0063	0.0062	0.0078
Hilbert	0.0047	0.0047	0.0047	0.0062	0.0063	0.0078

$n=174955, d=2, M=50, m=8$

从表 4 可知，曲线的执行时间少于线性扫描和基于 R 树最短优先最近邻查询算法的执行时间。

表 5 参数 n 与执行时间之间的关系

t(s)	10 万	20 万	50 万	60 万	80 万	1 百万
Brute-force	0.539	1.073	2.7078	3.2344	4.264	5.3625
R-tree	0.0531	0.1109	0.2734	0.3219	0.4297	0.575
Z	0.0015	0.0015	0.0031	0.0031	0.0047	0.0062
Gray	0.0016	0.0015	0.0031	0.0031	0.0047	0.0078
Hilbert	0.0015	0.0015	0.0031	0.0031	0.0047	0.0047

$d=2, k=1, M=50, m=8$

从表 5 知，曲线的执行时间最短。随着参数 n 的增加，曲线的执行时间线性增长，而线性扫描和基于 R 树最短优先最近邻查询算法的执行时间成倍增加。

结束语 本文提出了基于空间填充曲线网格划分最近邻查询算法。实验结果表明，该算法的性能优于线性扫描和基于 R 树最近邻查询算法。且算法可行、有效。

参 考 文 献

[1] Roussopoulos N, Kelley S, Vincent F. Nearest Neighbor Queries [C]// Proceedings of the 1995th ACM SIGMOD International Conference on Management of Data. San Jose, CA, 1995: 71-79

[2] Hjaltason G, Samet H. Incremental Distance Join Algorithms for Spatial Databases[C]// Proceedings of the 1998th ACM SIGMOD International Conference on Management of Data. Seattle, Washington, 1998: 237-248

[3] Cheung King Lum, Fu Ada Wai-chee. Enhanced nearest neighbour search on the R-tree[J]. ACM SIGMOD Record, 1998, 27(3): 16-21

[4] 刘永山，薄树奎，张强，等. 多对象的最近邻查询[J]. 计算机工程，2004，30(11)：66-68

[5] 徐红波，郝忠孝. 基于 Hilbert 曲线的近似 k-最近邻查询算法[J]. 计算机工程，2008，34(12)：47-49

[6] 徐红波. 空间填充曲线映射算法研究[J]. 科技信息，2007，24(35)：88-89

(上接第 183 页)

同义词转换处理；

第三步 将视频片断标注信息读入到 videoClip 中；

第四步 比较keyWords. Objects 和 videoClip. Objects信息，根据式(3)计算对象相似度 O，如果 $O=0$，则转到第八步；

第五步 比较keyWords. Actions 和 videoClip. Actions 信息，根据式(2)计算动作相似度 A；

第六步 比较 keyWords. Place 和 videoClip. Place 信息，计算地点相似度 P；

第七步 根据 OAP 模型，利用式(1)计算整个相似度 *Match*；

第八步 若所有视频未匹配完毕，则指向下一个视频片断，转到第三步；

第九步 按照 *Match* 取值大小，进行排序；

第十步 选择 *Match* 的最大值作为候选视频，若 *Match* 的最大值出现相等的情况，则用随机概率模型选择其中之一作为候选视频；

第十一步 算法结束。

4 实验结果分析

新闻视频自动检索方法在战略对抗演习中进行了应用，演习中的实验环境如下。

新闻稿总数：54，56，57(演习中 3 个不同回合)；

标注字典库：789 个词条；

自动切分字典库(每方单独建立)：最长 557 条记录，最短 35 条记录，平均 128 条记录；

视频素材库大小：6130 个视频片断，总计 23.5G。

实验结果如表 4 所列。在实验过程中对未被选中的视频进行跟踪记录。由于采用了基于关键字的 TBVR 技术，因此在查全率方面没有任何问题。同时，由于设计了同义词表，解决了语义匹配而字符不匹配的问题，因此取得了较高的查准率，但仍然存在少数遗漏的现象。对于这种现象需要补充和更新同义词表。

表 4 实验结果

测试集	查准率	查全率
54 条(回合 1)	94.7%	100%
56 条(回合 2)	94.5%	100%
57 条(回合 3)	95.1%	100%

结束语 本文在分析虚拟新闻系统新闻视频自动化检索的特点和现有视频检索方法不足的基础上，提出了一种基于 TBVR 的新闻视频自动检索方法，并对一般的 TBVR 中字符不匹配而语义却匹配的问题设计了解决方案，同时设计了视频相似度计算模型。实验结果表明，该方法能够很好地解决虚拟新闻系统中视频自动检索的问题，同时对于其他类似的视频检索问题有一定的借鉴和参考意义。

参 考 文 献

[1] 唐波，刘雨，孙茂印. 基于数据库的视频检索[J]. 电视技术，2005(2)：20-24

[2] 肖平，黄薇，冯刚. 基于内容的新闻视频检索技术研究[J]. 计算机与数字工程，34(10)：83-86

[3] 董献洲，胡晓峰，吴琳，等. 虚拟新闻的表达与生成及其系统设计与实现[J]. 系统仿真学报，2006，18(12)：3634-3636

[4] 陈芳莉，胡晓峰，吴琳，等. 虚拟新闻模拟系统的研究与设计[J]. 计算机仿真，2007，24(8)：5-7

[5] 罗卫光. 电视深度报道的叙事学研究[D]. 广州：暨南大学，2004

[6] 张军华，王晓勇. 电视新闻叙事的视角转换与主题建构[J]. 广西师范大学学报：哲学社会科学版，2005，41(3)：59-61

[7] 司光亚，胡晓峰，吴琳. “沉浸式”战略决策训练模拟系统研究与实现[J]. 系统仿真学报，2006，18(12)：3581-3583

[8] 卢亮，张博文. 搜索引擎原理、实践和应用[M]. 北京：电子工业出版社，2005

[9] 孙宾. 现代汉语文本的词语切分技术[R]. 北京：北京大学计算语言学研究所，2003

[10] 苗夺谦，卫志华. 中文文本信息处理的原理与应用[M]. 北京：清华大学出版社，2007