中山大学硕士学位论文

基于内容查询的视频图文缩略图自动生成方法与应用

Approach and Application on Automatic Generation of Video Visual-Text Thumbnail Based on Content Query

学位申请人：　戚　鑫

指导教师：　罗笑南 教授 林淑金 副教授

专业名称：　软件工程

答辩委员会主席(签名)：_____

答辩委员会委员(签名)：_____

_____

_____

_____

二零一七年五月二十三日

论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期：

学位论文使用授权声明

本人完全了解中山大学有关保留、使用学位论文的规定，即：学校有权保留学位论文并向国家主管部门或其指定机构送交论文的电子版和纸质版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆、院系资料室被查阅，有权将学位论文的内容编入有关数据库进行检索，可以采用复印、缩印或其他方法保存学位论文。

学位论文作者签名：　　导师签名：

日期：　年　月　日　　日期：　年　月　日

论文题目： 基于内容查询的视频图文缩略图自动生成方法与应用
专 业： 软件工程
博 士 生： 戚 鑫
指导教师： 罗笑南 教授 林淑金 副教授

摘要

摘要概括论文的主要信息，包括研究目的、方法、成果及最终结论。字数控制在满一页但不超过两页,硕士论文摘要一般不超过1200字。博士论文摘要一般不超过2000字。关键词是供检索用的主题词条，应采用能覆盖论文主要内容的通用词。关键词本科3-5个，硕士5-7个，博士7-9个，关键词之间用”,”分割，关键词不能是英文简写。

本科、硕士论文摘要一般采用2段，第一段研究背景（包括理论背景、应用背景）、研究环境、方法手段、影响和前景，第二段研究的内容、成果、价值、意义和不足之处。

博士论文摘要可分段介绍创新点。

关键词：学位论文，格式，模板

Title: Approach and Application on Automatic Generation of Video Visual-Text Thumbnail Based on Content Query

Major: Software Engineering

Name: Name

Supervisor: Prof. Xiaonan Luo Shujin Lin

Abstract

The American students are part of one of the most ambitious undertakings in the history of education: the American effort to educate an entire national population. The goal is-and has been since the early decades of the republic-to achieve universal literacy and to provide individuals with the knowledge and skills necessary to promote both their own individual welfare as well as that of the general public. Though this goal has not yet been fully achieved, it remains an ideal toward which has been made is notable both for its scope and for the educational methods which have been developed in the process of achieving it.

About 85% of American students attend public schools. The other 15% attend private schools, for which their families choose to pay special attendance fees. Four out of five primate schools in the United States are run by churches, synagogues or other religious groups. In such schools, religious teachings are a part of the curriculum, which also includes the traditional academic courses of reading, mathematics, history, geography and science.

Keywords: Mesh Editing, Mesh Deformation, Sketch-Based Interface, Linear Constrained Deformation, Least-Squared Editing

目 录

引言

从引言起为论文正文主体部分。页码从1开始编排。引言(前言)部分内容主要包括5个方面:为本研究课题的学术背景与环境、存在的问题、意义（不需要详细解释、只需几种说明解决了哪几个问题）、突破点、比较结果及优缺点。

注意不要与摘要内容雷同。

此部分是第一章（综述）和第五章（总结）的两章内容的总结

如果引言部分省略，可以合并到第一章综述中去。

1.正文书写格式说明：

每段落首行缩进2字；或者手动设置成每段落首行缩进2字，宋体，小四，：多倍行距（1.5倍），前段、后段均为0行，取消网格对齐选项。可以采用样式和格式里面的”正文格式”来格式化正文文本。注意：每两级标题之间一定要有过渡性的文字，避免两级标题直接相连。一般而言,硕士论文正文页数在50页以上,而博士论文页数在100页以上.

脚注书写格式说明：

①一般而言,网页地址、国家\ 地方标准都只能作为脚注,而并非出现在参考文献中.

① 网页标题,http://www.sysu.edu.cn

第 1 章 综述

近年来，随着网络带宽的不断提高，社会化网络和网络流媒体技术的发展，多媒体信息特别是视频已经成为当今信息时代主要的数据来源形式。视频缩略图在呈现视频内容上扮演着非常主要的作用，它直接决定了用户是否要点击视频以观看视频具体内容。好的视频缩略图能让用户第一时间了解视频内容，极大提高了用户检索效率。

本章首先介绍视频缩略图自动生成方法的研究背景与意义，接着从视频内容的提取，视频语义分割，视频缩略图的自动生成三个角度出发分析国内外研究现状，然后介绍本文的主要工作和创新点，最后简述本文的组织安排。

1.1 研究背景与意义

自上世纪九十年代以来，随着互联网技术，多媒体技术的发展和人们对信息需求的不断增长，越来越多的信息通过多媒体的形式展现在用户面前，例如视频，音频，动画，图像等。相比于其他多媒体数据形式，视频以其生动性、直观性、信息的丰富性备受用户的喜爱，特别是网络视频用户规模不断扩大，截至2016年12月，中国网络视频用户规模达5.45亿，较2015年底增加4064万人，增长率为8.1%①。网络视频已成为一种人们分享信息，想法，趣事的主要媒介形式。

随着视频数据的爆发式增长，以提供视频分享为主要业务的视频门户网站也在蓬勃的发展，国内有优酷②、爱奇艺③、腾讯、搜狐等数百家视频门户网站，国外视频门户网站有youTube④、Yahoo Video⑤、AOL Video等等。数量众多的视频门户网站为用户提供了海量的视频信息，极大的满足了用户对视频数据的需求，但同时也增加了用户检索视频的难度。这些视频分享网站中的视频的来源大多是由众多用户上传，上传的同时也伴随着一些视频的元信息以便于视频搜索引擎检索或者吸引其他用户点击，例如视频标题，描述，标签，缩略图等等，其中相比于标题，描述这些视频文本元信息，视频缩略图更生动，并能达到预览视频内容的效果，用户能通过缩略图直观地了解视频内容，因此视频

① 数据来源前瞻产业研究院:https://bg.qianzhan.com/report/detail/459/170313-223982fb.html
② http://www.youku.com/
③ http://www.iqiyi.com/
④ https://www.youtube.com/
⑤ http://video.search.yahoo.com/

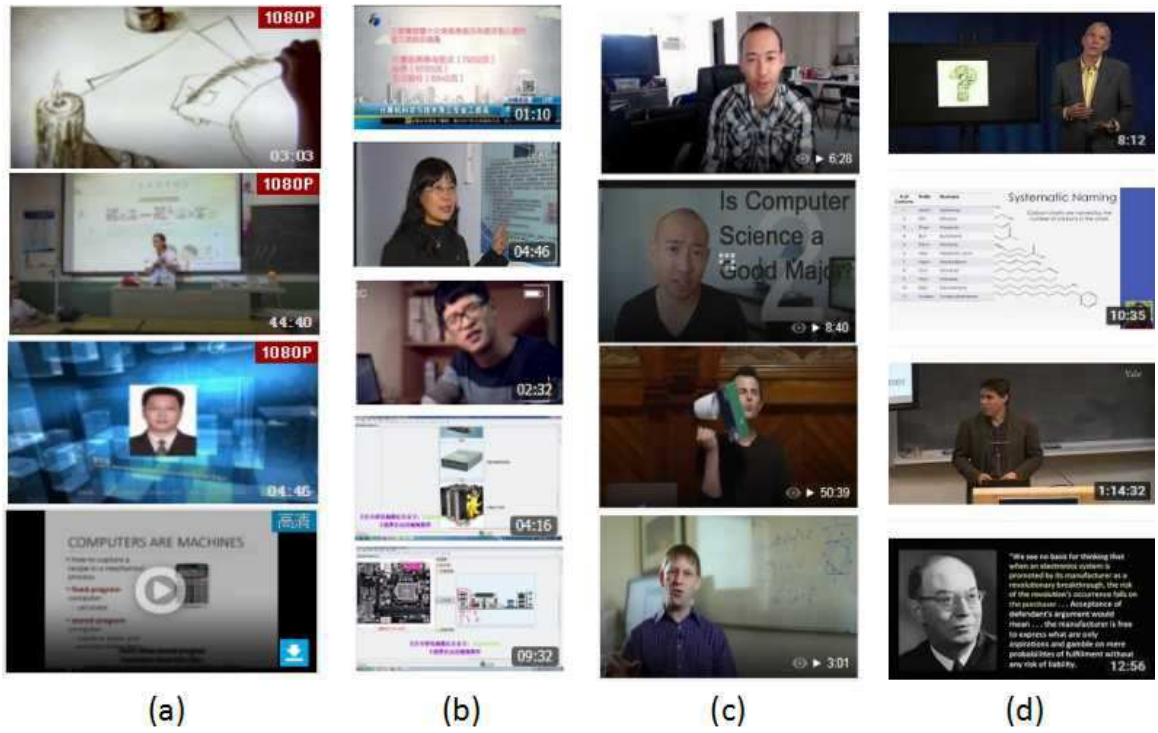


图 1-1 国内外知名视频网站以“计算机科学”/”computer science”为关键词检索的部分结果：(a)优酷视频.(b)爱奇艺视频.(c)YouTube Video.(d)Yahoo Video.

缩略图已成为一种常用的视频和用户交互的技术。Yuli[1]等人指出视频缩略图对用户浏览行为上有很强的影响力。Michael[2]等人的研究也表明高质量的缩略图大大提高了视频检索效率和用户满意度。

视频上传用户一般选择视频中的一帧作为视频缩略图，以youtube视频网站为例，当用户上传视频后，系统随机生成视频三帧画面供用户选择其中一帧作为视频缩略图，然而这样的缩略图往往无法反映视频内容，质量较差。如图1-1所示，用户以“计算机科学”或者”computer science“为关键词在国内外知名视频网站检索到的部分结果，可以看出仅从现在视频网站提供的视频缩略图用户几乎得不到任何任何有效的信息。如果用人工挑选精心生成缩略图，固然可以提高缩略图的质量，但是这样做太过费时，为每一个视频都精心打造一个缩略图是不现实的[3]。而且，视频缩略图和用户查询意图存在鸿沟，无法满足用户的要求。如图1-3所示，图中的3张子图是同一视频的3帧图片，都可以成为视频的缩略图，但它们包含有截然不同的信息：从第一张子图得出这个视频是一个TED演讲类视频；第二张图片表明视频中有讲到蒙娜丽莎的微笑；第三张子图则表明视频有讲到关于健身减肥的内容。不同的用户或者同一用户不同时间在检索视频时显然带有不同的意图，所以当两次检索结果有同一视频时，他

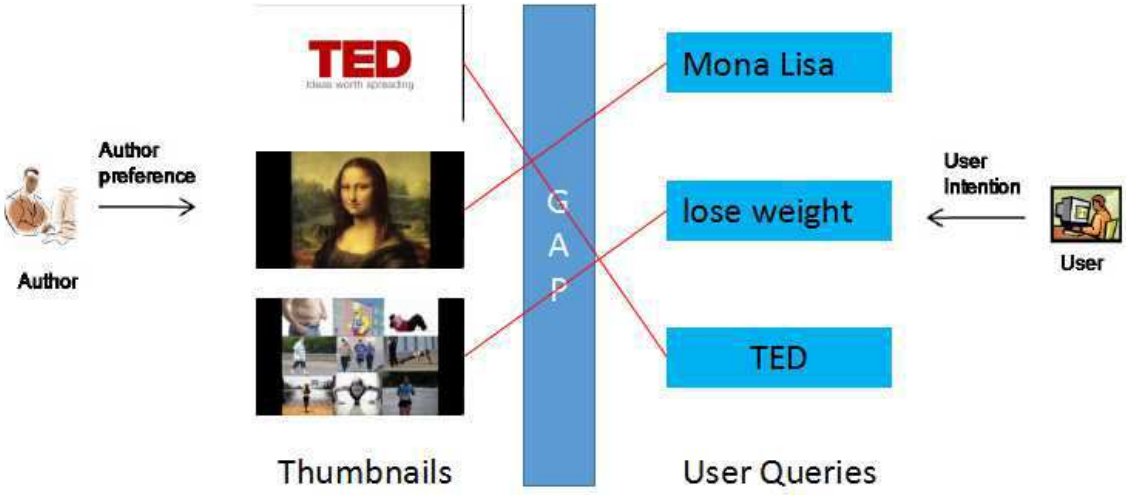


图 1-2 用户查询意图与视频缩略图之间存在的鸿沟

们关注的内容也不尽相同。但是传统的视频缩略图具有静态性，一旦生成就不可改变，且生成缩略图的时候并不知道将来用户会以什么查询词检索视频，我们当然希望以”Mona Lisa”为关键词是检索到的视频缩略图为图1-2中的第二个子图，遗憾的是传统的视频缩略图生成方法显然满足不了需求，很可能会出现用”Mona Lisa”为关键词检索到的视频的缩略图为图1-3中第一个或者第三个子图，这时视频缩略图的内容显然不合适的，与用户的检索意图背道而驰。因此用户查询意图与视频缩略图之间存在巨大鸿沟，用户在缩略图上找不到自己感兴趣的视频内容，导致用户检索效率降低，用户满意度下滑。

本文的目标就是基于视频多通道内容信息，跨越用户查询意图与视频内容的鸿沟，生成动态的并且自适应用户查询的视频缩略图。通过该目标，帮助用户快速了解自己视频中感兴趣的内容，提高检索视频的效率。

1.2 国内外研究现状

视频是一种非结构化的流媒体数据，其内容信息复杂而丰富。从视觉上看，视频是由一系列视频帧组成；从听觉上看，视频也含有大量嵌入视频的语音信号。好的视频缩略图首先要能很好的反映视频内容，所以生成视频缩略图之前首当其冲要解决的问题是提取视频内容，并理解视频有效信息，即要清楚地知道视频由几部分组成，每部分蕴含了什么信息，这就涉及到视频主题边界检测，以及视频语义内容概括。近年来，国内外学者对以上几个方面做了大量的研究，本小节下面的内容主要从视频内容提取与分析研究现状，视频主题分割研究现状，视频缩略图生成方法研究现状三个角度出发，对前人的研究加以归纳总

结。

1.2.1 视频内容提取与分析研究现状

视频内容提取与分析的目标是将复杂抽象的数据转换成计算机容易处理的格式，并提取、整合、挖掘有效的信息，其中涉及的领域非常广泛，诸如图像处理，语音识别，模式识别等。通过归纳关于视频内容提取与分析的相关文献，将提取分析的关键过程归纳为：镜头分割与关键帧提取方法，基于OCR①的视频文字提取方法，视频语音识别方法。

（1）镜头分割与关键帧提取方法

镜头分割的方法主要包含基于颜色直方图的镜头分割方法[4, 5]，基于像素颜色差异的镜头分割方法[6]，基于运动的镜头分割方法[7, 8]，基于边缘特征的镜头分割方法[9]。基于颜色直方图的镜头分割方法统计视频图像的像素灰度分布或者颜色分布，核心原理是不同镜头之前图像的灰度和色彩会发生剧烈变化。该方法简单明了，计算复杂性低，缺点对光照强度和镜头运动速度太过敏感，基于运动的镜头分割方法解决了这一缺点，主要原理是基于点或块的运动矢量的估计，可以很好地检测镜头的变化，但是该方法计算复杂性太高。基于边缘特征的镜头分割方法考虑的是边缘在局部照明变化下大部分是不变的，该方法减少了由于运动和照明导致的不变性问题，但是对于图像内容复杂的视频镜头分割不甚理想。关键帧提取方法包括：特定帧法[10]、直方图平均法[11, 12]、基于运动特征的关键帧提取方法[13, 14]、基于内容的分析方法[15]。

（2）基于OCR的视频文字提取方法

光学字符识别（Optical Character Recognition, OCR）通过光学机制识别字符，可以将数字图像中的文字信息转换为可编辑文本的过程。OCR[16]是一个计算机视觉/图像中活跃的研究领域，它通常包含两个步骤，首先是定位图像中的包含文本的区域，然后是识别区域中的文本。Neumann[17]等人提出了用于定位图像中文本的算法。OCR技术也已经发展的比较成熟，市面上已经有非常成熟的OCR识别引擎，例如Tesseract,OmniPage,Readiris等等。

（3）视频语音识别方法

视频语音识别是提取语音特征，将视频中复杂的语音信息转换为可编辑易处理的文本数据。语音识别的方法主要包括：基于统计模型的方法，基于学习的方法，基于规则的方法[18, 19, 20]。现在有一些高级的语音识别工具，英文的工具

① 维基百科OCR介绍：https://en.wikipedia.org/wiki/OCR

有MSR①,CMU Sphinx[21],Nuance Dragon②等，中文的工具主要有科大讯飞的讯飞开放平台③。

1.2.2 视频主题分割研究现状

视频主题分割不同于视频镜头分割和场景分割，它的目标是检测视频语义主题边界，把视频按语义分割成一个个语义独立，内容连贯的主题片段。通过视频主题分割，用户能清楚地了解视频结构，掌握视频内容。视频语义分割问题可以转化为文本主题分割问题，两者的本质都是把一段内容丰富，结构复杂的信息分割成内容独立的片段，研究学者对这方面进行了深入的研究。

Hearst[22, 23]提出的Textiling算法先把文本数据分成连续的，非重叠的文本块，提取文本块的”bag of word”特征，用余弦相似度衡量文本块的相似度，然后比较文本块间的”bag of word”特征间的差异，Hearst认为差异较大的文本块就是主题的边界。Satanjeev[24]等人将Textiling算法应用在会议主题分割，取得了比较好的结果。Textiling算法使用词袋向量来描述文本块的特征，但是词袋模型往往非常稀疏，而且词袋模型不能反映近义词语义的相似，例如”计算机“和”电脑“两个词在语义上很相似，但表示成词袋向量后则截然不同，所以词袋向量不利于文本语义的表示。Gally[25]等人提出了一个基于TextTiling的算法，不同于TextTiling算法，它用词的tf-idf④权重代替了单一的词频权重从而取得了更好的效果。Martin[26]等人提出了TopicTiling算法，改进了Textiling算法。TopicTiling算法基于LDA[27]主题模型，LDA是一个三层贝叶斯模型，可以认为一篇文章的每个词都按一定概率属于某个主题。如图1-3所示，TopicTiling算法首先对语料库进行LDA训练，挖掘出语料库潜在的主题，并得出每一篇文章每一词属于某个主题的概率，然后为每一个主题分配一个TopicID,这些ID被用来计算相邻文本块的余弦相似度，即把”bag of word”特征替换为”bag of TopicID”，然后再进行分割流程。TopicTiling算法利用”bag of Topic”既将冗长的”bag of word”特征降了维，又因为近义词往往属于同一主题，解决了近义词特征不同的问题。

基于TextTiling算法是具有线性复杂度，且取得了较好的效果，但是它会出现过度分割或者分割不足的问题，原因是在确定主题边界的阈值难以确定，不能自适应文档的长度。而且把Textiling算法应用在视频主题分割时，不但可以利用由音频转换而来的文本信息，还可以结合视频的图像信息加以优化，本文下

① Microsoft Speech Recognition(MSR) API: https://msdn.microsoft.com/library/ee125663.aspx
② Nuance Dragon Speech Recognition Software(Nuance) url: http://www.nuance.com
③ 讯飞开放平台网址：http://ai.xfyun.cn/
④ tf-idf百度百科https://baike.baidu.com/item/tf-idf/8816134?fr=aladdin

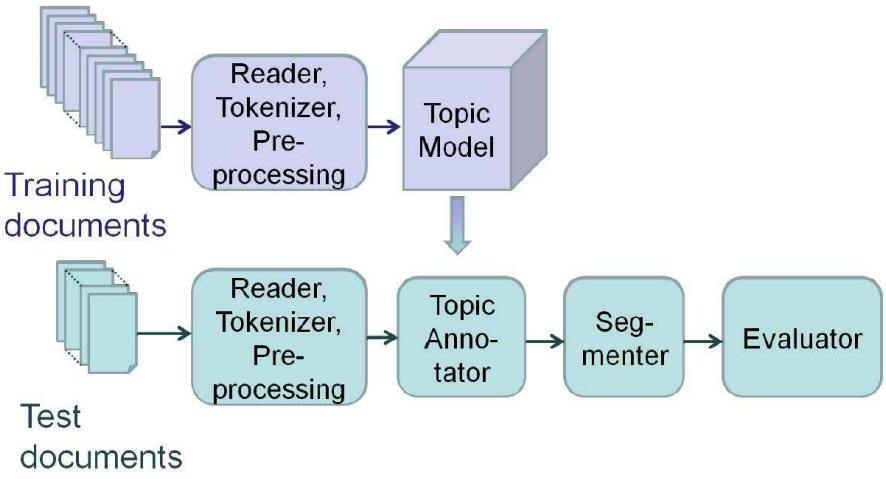


图 1-3 TopicTiling算法主题分割流程

面的章节会详细介绍。

除了无监督的方法，Nguyen[28]等人训练了一个无参数的层次化的模型解决了多人谈话的主题分割主题分割问题，认为每个文本块都与一个说话人有关，通过训练出的模型判断文本块的说话人，从而判断主题的转移。Chen[29]等人提出了一种基于自我验证的声学分割方法（SACuts）来把口语文档分割成主题片段。相比于其他方法，他们的方法仅用了声学级别的信息确定主题数量，在没有任何的额外的计算负担下解决了过度分割和分割不足的问题。Fragkou[30]利用NER（Ebtity Annotation Recognition）对语料库进行标注，算法首先对语料库每一个词、短语都被归类为提前定义后的实体类型，随后词、短语被替换成替换成实体标识符，然后用已有的分割算法，例如Choi[31]提出的C99b算法，Utiyama[32]等提出的基于统计学模型的分割算法，Kehagias[33]等人最小化全局分割代价优化算法等，进行主题分割。以上算法各有优劣，且取得了较好的效果，但应用到视频主题分割问题上会出现不适用的问题。视频数据复杂丰富，往往包含很多说话人角色，若利用Nguyen、Chen的方法进行声学特征或者说话人的识别进行主题分割，会出现把一个主题过度分割成很多小碎片的问题。且当今视频数据量增长迅猛，很难训练出鲁棒性很好的模型适用所有的视频数据，而且标注，训练语料库往往非常耗时，不能大规模推广。

1.2.3 视频缩略图生成方法研究现状

视频缩略图是视频分享网站给用户提供的友好接口，质量高的缩略图能帮助用户快速了解视频内容，吸引用户点击，从而极大提升视频检索系统的检索性能，改善用户检索视频体验。目前现有的视频分享网站往往选取第一帧或随

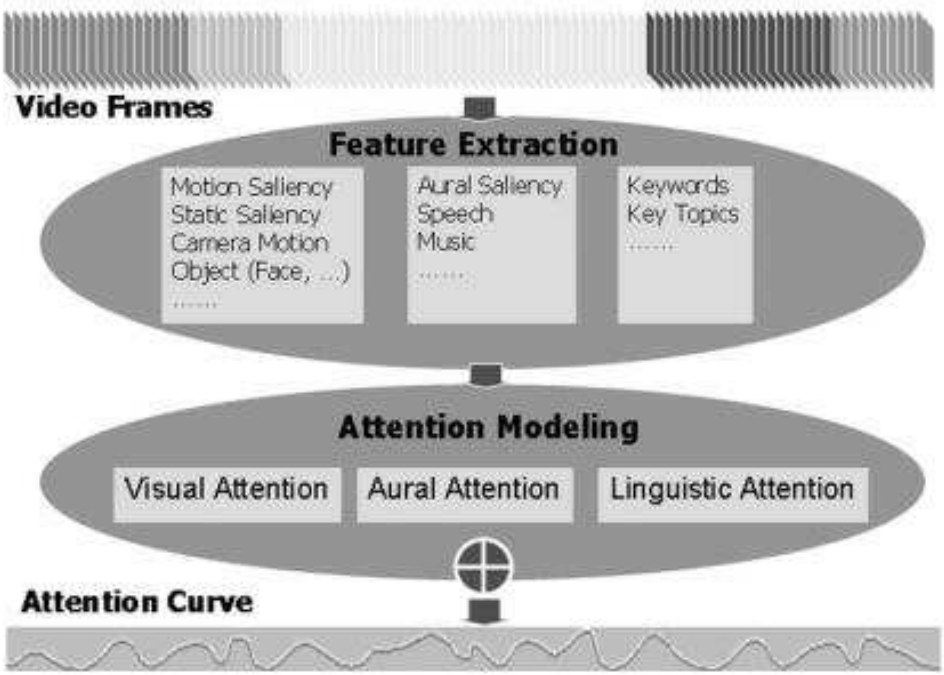


图 1-4 用户关注度模型框架

机选取一帧作为视频缩略图，或者像youtube的做法随机抽取多帧让用户选择一帧作为视频缩略图。这类方法虽然简单易行，但很难达到好的效果。视频缩略图生成研究涉及领域很广，例如计算机视觉、机器学习、数据挖掘、美学设计等等，近年来关于这方面的研究也越来越多，下面详细分析视频缩略图的研究现状。

Wolf[34]等人认为运动变化最小的视频帧最具有代表性的帧，算法首先基于光学流动分析衡量每一个视频帧的运动量，然后选取运动量最小的帧。Wolf认为拍摄视频过程中经常会为了一个镜头使用一系列相机镜头的运动来构建复杂的信息，所以往往非常重要的帧算法中描述的运动量最小，但是这只符合一些例如电影类，纪录片类视频，对一些新闻类，演讲类，教育类等镜头变化不显著的视频不是很适用。Ma[35]等人提出了一个用户关注度模型，如图1-4所示，首先对视频视觉信息，听觉信息以及视频文本信息进行特征提取，根据这些特征和用户关注度模型计算用户视觉关注度，听觉关注度，语言关注度，然后将多通道关注度融合成用户关注度曲线，取曲线最大值点的关键帧作为视频缩略图。Cong[36]等人从稀疏编码和数据重建的角度出发，核心思想是将视频帧编码成字典，如果一个视频帧的字典能够最好的重建原视频，那么就认为这一帧就是最有代表性的视频帧。相似地，Guan[37]等人提出了一个稀疏重建框架选取最有代表性的视频帧。Kang[38]等人具体度量了视频帧的代表性，认为帧的质量，图像

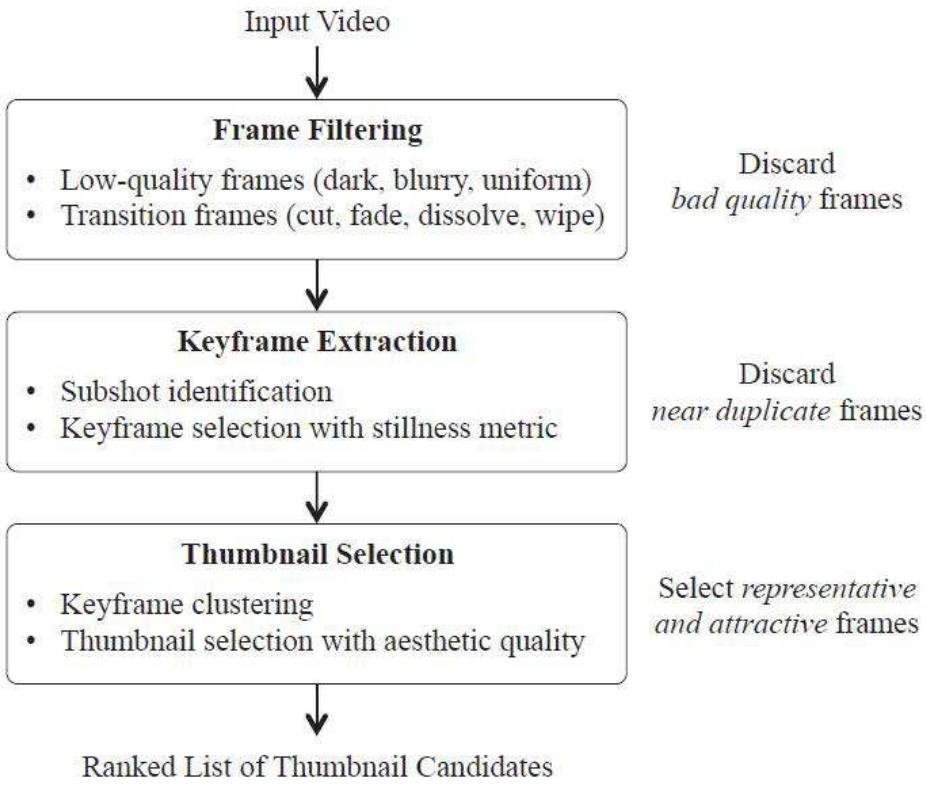


图 1-5 用户关注度模型框架

细节，内容相关性，用户关注度是衡量视频帧是否具有代表性的关键特征，然后基于高斯混合模型对视频帧代表性训练，建模，从而找出最具代表性的视频帧作为缩略图。如图1-5所示，Zhang[39]等人从信息丰富度，用户关注度，和美学特征3个方面定义了12个特征来衡量视频缩略图的质量，选取特征得分最高的视频帧作为视频缩略图。Song[40]等人的视频缩略图系统既考虑了视频内容相关性，又考虑了视觉美学质量。具体过程图1-6所示，首先过滤掉模糊，灰暗等低质量的帧；然后进行视频关键帧检测去掉重复的帧；最后进行关键帧聚类并结合美学质量评估，选取最有代表性又能吸引用户的视频帧作为视频缩略图。

以上方法都或多或少提升了视频缩略图的质量，但都忽略了用户的查询意图。Liu[41]等人的研究改进了这一点，他们的方法首先使用Joshi[42]的方法对视频关键帧序列排序，然后计算用户查询词与图片之间的关联性，将其融入到对关键帧的排序过程中，选取排名最高的视频帧作为视频缩略图。在Liu的方法中，词与图片的关联度的计算公式1.1,

$$S(I_u, w) = P(I_u, w) \approx P(\frac{I_u}{w}) = \frac{1}{n}\sum_{i\epsilon n} s(I_u, v_i) \tag{1.1}$$

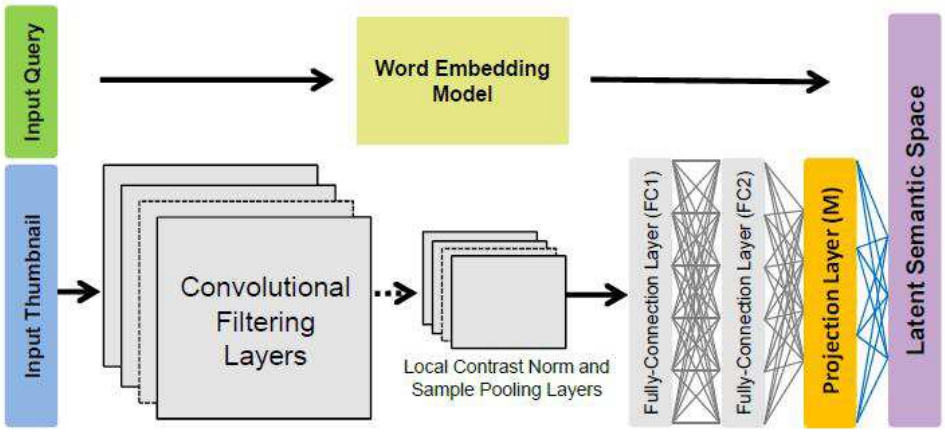


图 1-6 深度视觉-语义嵌入模型架构

其中$S(I_u,w)$表示图片I_u和查询词w的关联度，$P(I_u,w)$表示表示图片I_u和查询词w关联的可能性，$v_1,v_2,...,v_n$表示用查询词w早图片搜索引擎搜索到的topN个结果图片。由公式1.1可见，图片与查询词之间的关联度计算依赖于第三方图片搜索引擎，这种在线搜索图片的过程充满不确定性，且往往是非常耗时的。另外，关联度计算的有效性完全取决于第三方搜索引擎的有效性，所以可靠性不强。Liu[43]等人利用深度卷积神经网络（CNN）实现了一种基于多任务深度视觉-语义嵌入模型，该模型架构如图1-7所示，模型可以把用户查询词和视频缩略图都映射到潜在语义空间，通过计算查询词和缩略图在潜在语义空间中向量表示的相似度得出二者之间的关联度，跨越了用户输入的查询词语义与视频缩略图视觉的鸿沟，生成符合用户查询的视频缩略图。这个方法新颖有效，但是在计算缩略图和查询词关联度时仅仅考虑了缩略图的视觉信息，然而视频有多通道的信息，不但有视频帧等视觉信息，而且还有通过时间轴与视频帧相关联的音频信息，音频信息可通过语音识别转化为文本信息，这些与视频帧关联的文本信息能直观地反映视频帧的语义，在计算查询词和帧语义关联度时具有很大的价值。

通过对上述方法总结可以发现：（1）很多方法的本质是根据图片特征对视频的关键帧打分，然后取分数最高的视频帧作为缩略图；（2）缩略图的生成选取的标准主要考虑了视频内容相关性、图片质量、用户关注度、美学设计等等；（3）少部分方法考虑了用户的查询意图信息。但是几乎所有的方法都是只选择视频中的一帧作为缩略图，即使帧的质量很高，但由于一帧图像包含的图像信息非常有限，往往无法给用户足够有效的信息。而且，只有很少的方法考虑了用户的查询意图，他们的方法比较新颖，但仍存在一些问题，如上文所述，

文章[41]的方法具有不确定性而且非常耗时，文章[43]的方法忽略了视频的音频信息，以上方法都有较大的改进空间。

1.3 本文研究工作及创新点

视频缩略图自动生成方法是一项极具挑战的研究课题，设计数据挖掘，计算机视觉，图像处理，美学设计等多个学科，是一个综合性很强的研究领域。本文提出了一种新颖的基于内容查询的图文视频缩略图生成算法，它基于视频多通道信息，又结合用户查询内容，动态生成出既能深刻反映视频内容，又能迎合用户查询意图的视频缩略图，极大提升了用户检索视频的效率和满意度。

本文主要研究内容：

（1）视频多通道内容的分析与处理。在视觉通道上，进行镜头分割，关键帧提取，图片显著性区域检测，OCR识别；在听觉通道上，进行语音识别，LDA主题挖掘，关键词提取。

（2）视频主题分割与整合算法。对视频语音识别的结果进行主题分割，结合视频镜头分割的结果，检测视频主题边界，将视频分割成一个个内容连贯，主题鲜明的片段。分析每个片段的内容，提取每个片段最具代表性的图文，将每个片段的视频内容整合成图文二元组。

（3）视频缩略图的生成算法。基于用户查询内容和由（2）得到的每个视频片段的图文二元组，计算二者之间的关联度，选取和用户查询最相关的视频内容，再结合美学设计，生成出美观的，视频相关的，自适应用户查询的视频缩略图。

本文主要的创新点在于：

（1）提出了基于TopicTiling的视频主题分割算法，它能准确的检测出视频内容的主题边界。算法首先改进了TopicTiling深度分数阈值难以确定的问题，使其能自适应文本的长度选取合适的阈值进行主题分割；将TopicTiling这种文本主题分割算法应用到视频主题问题上来，结合视频多通道信息，对视频进行主题分割。

（2）提出了基于内容查询的图文视频缩略图的算法，其创新点在于：不在像传统思路那样选取一帧图像作为视频缩略图，而是生成由若干张图像的显著性区域和相应的文字拼合的缩略图；提出了将视频主题片段的整合成图文二元组的方法，并实现了视频主题片段二元组与用户查询词的关联度度量算法；缩略图不仅反映了视频多个主题的内容，而且自适应用户查询意图。

1.4 本文的组织安排

本文主要内容安排如下：

第1章，本文综述。首先介绍视频缩略图的研究背景和意义，接着从视频内容提取与分析，视频主题分割，视频缩略图生成方法三个角度分析了国内外研究现状，最后针对现有方法的局限性，提出了本文的主要研究工作和创新点。

第2章，提出了基于视频多通道内容的主题分割与内容整合算法。首先介绍视频多通道内容提取分析的方法；接着阐述基于TopicTiling的视频主题边界检测算法，根据主题边界，将视频分割成内容连贯的主题片段；最后阐述整合这些片段内容的方法。

第3章，提出基于查询的视频图文缩略图生成算法。首先介绍用户查询与视频内容相似性度量算法，查找与用户查询最相关的若干个视频主题片段图文内容；然后阐述如何利用美学设计知识将上述图文内容排版到一张视频缩略图中；最后是视频缩略图算法的实验结果和评估。

第4章，基于内容查询的视频缩略图生成算法应用。首先设计了一个视频检索系统；然后详尽的阐述该系统包括整体框架，模块布局以及交互应用设计的实现等；最后是该系统的实验评估。

第5章，本文的工作总结和今后工作的展望。

第 2 章　基于视频多通道内容的主题分割与内容整合算法

2.1　视频多通道内容分析与处理

2.2　基于TopicTiling的视频主题边界检测

2.3　视频主题内容整合方法

2.4　本章小结

$$x_i = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{2.1}$$

$$x_i = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{2.2}$$

第 3 章　基于查询的视频图文缩略图自动生成算法

内容概括。

3.1　用户查询与视频内容关联性度量算法

此处省略N个字……

3.2　视频图文缩略图排版算法

此处省略N个字……

3.3　实验结果及评估

此处省略N个字……

3.4　本章小结

此处省略N个字……

第 4 章　基于内容查询的视频缩略图生成方法应用

内容概括。

4.1　视频缩略图生成系统应用设计

此处省略N个字……

4.2　视频缩略图生成系统应用实现

此处省略N个字……

4.3　视频缩略图生成系统应用评估

此处省略N个字……

4.4　本章小结

此处省略N个字……

第 5 章 总结与展望

内容概括。

5.1 本文工作总结

此处省略N个字……

5.2 今后工作展望

此处省略N个字……

参考文献

[1] Gao Y, Zhang T, Xiao J. Thematic video thumbnail selection[C]//Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009: 4333-4336.

[2] Christel M G. Evaluation and user studies with respect to video summarization and browsing[C]. SPIE, 2006.

[3] Song Y, Redi M, Vallmitjana J, et al. To Click or Not To Click: Automatic Selection of Beautiful Thumbnails from Videos[C]//Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016: 659-668.

[4] Nagasaka A, Tanaka Y. Automatic Video Indexing and Full-Video Search for Object Appearances.[C]//Visual Database System Ii, Elsevier. Dallas, United States: ACM, 1991:113 – 127.

[5] Swain M, Ballard D. Color indexing[J]. International Journal of Computer Vision, 1991,7(1):11 – 32.

[6] Zhang H, Kankanhalli A, Smoliar S. Automatic partitioning of full-motion video[J]. Multimedia Systems, 1993, 1(1):10 – 28.

[7] Bouthemy P, Gelgon M, Ganansia F. A unified approach to shot change detection and camera motion characterization[J]. IEEE Transactions on Circuits and Systems for Video Technology,1999, 9(7):1030 – 1044.

[8] Courtney J. Automatic video indexing via object motion analysis[J]. Pattern Recognition,1997, 30(4):607 – 625.

[9] Zabih R, Miller J, Mai K. A feature-based algorithm for detecting and classifying production effects[J]. Multimedia systems, 1999, 7(2): 119-128.

[10] Divakaran A, Sun H. Descriptor for spatial distribution of motion activity for compresse dvideo[J]. 2000, 2:392 – 398.

[11] Yeung M M, Liu B. Efficient matching and clustering of video shots[C]//Image Processing, 1995. Proceedings., International Conference on. IEEE, 1995, 1: 338-341.

[12] Zhang H J, Wu J, Zhong D, et al. An integrated system for content-based video retrieval and browsing[J]. Pattern Recognition, 1997, 30(4):643 – 658.

[13] 董晨晨. 镜头边界检测与关键帧提取技术研究[D]. 南京: 东南大学, 2010.

[14] 朱曦, 林行刚. 视频镜头时域分割方法的研究[J]. 计算机学报, 2004, 27(8):1027 – 1035.

[15] Jeannin S, Jasinschi R, She A, et al. Motion descriptors for content-based video representation[J]. Signal Processing Image Communication, 2000, 16(1-2):59 – 85.

[16] Ciresan D C, Meier U, Gambardella L M, et al. Convolutional neural network committees for handwritten character classification[C]//Document Analysis and Recognition (ICDAR), 2011 International Conference on. IEEE, 2011: 1135-1139.

[17] Neumann L, Matas J. Scene text localization and recognition with oriented stroke detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Sydney, Australia:IEEE, 2013:97 – 104.

[18] 赵亚琴. 基于内容的视频片段检索技术研究[D]. 南京:南京理工大学, 2007.

[19] 冯哲. 基于内容的视频检索中的音频处理[D]. 上海:复旦大学, 2004.

[20] 闫乐林. 基于视听信息的视频语义分析与检索技术研究[D]. 北京:北京邮电大学, 2012.

[21] Walker W, Lamere P, Kwok P, et al. Sphinx-4: a flexible open source framework for speech recognition[J]. Sun Microsystems, 2004:1 - 18.

[22] Hearst M A. TextTiling: Segmenting text into multi-paragraph subtopic passages[J]. Computational linguistics, 1997, 23(1): 33-64.

[23] Hearst M A. Multi-paragraph segmentation of expository text[C]//Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 1994: 9-16.

[24] Banerjee S, Rudnicky A I. A TextTiling based approach to topic boundary detection in meetings[J]. 2006.

[25] Galley M, McKeown K, Fosler-Lussier E, et al. Discourse segmentation of multi-party conversation[C]//Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003: 562-569.

[26] Riedl M, Biemann C. TopicTiling: a text segmentation algorithm based on LDA[C]//Proceedings of ACL 2012 Student Research Workshop. Association for Computational Linguistics, 2012: 37-42.

[27] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.

[28] Nguyen V A, Boyd-Graber J, Resnik P. SITS: A hierarchical nonparametric model using speaker identity for topic segmentation in multiparty conversations[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012: 78-87.

[29] Chen H, Xie L, Feng W, et al. Topic segmentation on spoken documents using self-validated acoustic cuts[J]. Soft Computing, 2015, 19(1): 47-59.

[30] Fragkou P. Text Segmentation using Named Entity Recognition and Co-reference Resolution in English and Greek Texts[J]. arXiv preprint arXiv:1610.09226, 2016.

[31] Choi F Y Y. Advances in domain independent linear text segmentation[C]//Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference. Association for Computational Linguistics, 2000: 26-33.

[32] M. Utiyama and H. Isahara. A statistical model for domain independent text segmentation.In Proceedings of the 9th EACL, pages 491 - 498, 2001.

[33] R. Kern and M. Granitzer. Efficient linear text segmentation based on information retrieval techniques. In Proceeding of the International Conference on Management of Emergent Digital EcoSystems, 2009.

[34] W. Wolf, “Key frame selection by motion analysis,” in IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2.IEEE, 1996, pp. 1228 - 1231.

[35] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang, “A generic framework of user attention model and its application in video summarization,” IEEE Trans. on Multimedia, vol. 7, no. 5, pp. 907 - 919, 2005.

[36] Y. Cong, J. Yuan, and J. Luo, “Towards scalable summarization of consumer videos via sparse dictionary selection,” IEEE Trans. on Multimedia, vol. 14, no. 1, pp. 66 - 75, Feb 2012.

[37] G. Guan, Z. Wang, S. Lu, J. D. Deng, and D. D. Feng, “Keypointbased keyframe selection,” IEEE Trans. on Circuits and Systems for Video Technology, vol. 23, no. 4, pp. 729 - 734, 2013.

[38] H.-W. Kang and X.-S. Hua, “To learn representativeness of video frames,” in ACM International Conference on Multimedia, Singapore, November 2005, pp. 423 – 426.

[39] Zhang B, Wang Z, Tao D, et al. Automatic Preview Frame Selection for Online Videos[C]//Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on. IEEE, 2015: 1-6.

[40] Song Y, Redi M, Vallmitjana J, et al. To Click or Not To Click: Automatic Selection of Beautiful Thumbnails from Videos[C]//Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016: 659-668.

[41] C. Liu, Q. Huang, and S. Jiang. Query sensitive dynamic web video thumbnail generation. In ICIP, pages 2449 – 2452, 2011.

[42] D. Joshi, J. Wang, and J. Li, “The story picturing engine: finding elite images to illustrate a story using mutual reinforcement,” ACM SIGMM workshop on MIR, 2004.

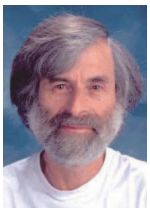
[43] Liu W, Mei T, Zhang Y, et al. Multi-task deep visual-semantic embedding for video thumbnail selection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3707-3715.

附录

作者简历

列举发表论文、专著、专利、标准及参加科研项目情况（版权所有必须归中山大学）

可以按如下模板：



1985 年，LaTeX 问世，它构筑在TeX 的基础之上，并且加进了很多新功能，使得用户可以更为方便地利用TeX 的强大功能。右图就是其开发者，美国著名计算机科学家、数学家Lamport 博士，他曾在康柏和惠普工作，目前任职于微软公司。

发表的论文：

[1] Xin Chen, Hefeng Wu, Xiang Li, Xiaonan Luo, and Taisheng Qiu, Real-time visual object tracking via CamShift-based robust framework, International Journal of Fuzzy Systems, 14(3), 2012, 262 269

[2] Xin Chen, Hefeng Wu, Xiang Li, Xiaonan Luo, and Taisheng Qiu, Real-time visual object tracking via CamShift-based robust framework, International Journal of Fuzzy Systems, 14(3), 2012, 262 269

专利/标准：

[1] 陈欣，罗力耕，陈湘萍, 一种基于数字电视中间件的视频点播方法及系统, 中国发明专利, 授权号：ZL201110130102.1.，授权日期：2012.09.05

[2] 陈欣，罗力耕，陈湘萍, 一种基于数字电视中间件的视频点播方法及系统, 中国发明专利, 授权号：ZL201110130102.1.，授权日期：2012.09.05

参与课题（注明项目编号）

[1] 课题名称：广东联合基金项目《数字几何媒体智能技术及应用研究》
项目编号：U0935004
主要工作：参与视频目标检测与跟踪技术的理论研究，撰写并发表相关国际会议论文2篇

[2] 课题名称：广东联合基金项目《数字几何媒体智能技术及应用研究》
项目编号：U0935004
主要工作：参与视频目标检测与跟踪技术的理论研究，撰写并发表相关国际会议论文2篇

获奖情况：

[1] 2012年度中山大学研究生国家奖学金

[2] 2013年度中山大学研究生国家奖学金

致谢

由衷感谢我的导师罗笑南教授，本文是在他的指导下完成的。

某某人

某年某月某日