视频摘要技术研究综述

**摘要：**近年来，随着互联网和多媒体技术的迅速发展，多媒体信息数据骤增。数字视频，作为主要的多媒体信息载体，广泛地应用于生活的方方面面。大量的视频一方面给生活带来了便利，以更丰富的形式获取信息，而另一方面，却给视频存储、传输、归档和检索带来巨大压力。尤其是监控视频，每天都会有数百万的监控摄像头录制视频，产生巨大的视频档案。然而即使视频档案中包括关键活动，但是找到它们就如大海捞针，这使得存档一文不值。在这样的背景下，视频摘要技术应运而生。类似于文本摘要，视频摘要是对原始视频内容的总结，通过分析原始视频数据流，从中选取有意义的视频内容来组成紧凑的摘要。视频摘要不仅可以结合视频标注技术服务于视频检索; 同时，它还能作为独立的产品应用，如电影预告片，方便生活。

**关键字：**互联网、视频摘要、监控视频

# 1研究背景及意义

随着现代社会电子化和信息化的发展，视频音频的多媒体数据已经成为信息处理领域的主要表现方式，而视频数据以其丰富的信息量和直观的视听感受，在社会生活的各个领域应用越来越广泛，人们同益面临大量迅速增加的视频数据的处理需要。在互联网服务、电子图书馆的检索、电视台的节目监控和视频点播、部门单位视频监控管理系统以及军事领域的视频情报分析等等各个应用方面，都需要我们对视频数据进行有效的组织和管理，从而提高视频的使用效率。然而，视频数据不同于字符信息，同时包含了动态的图像和声音，是一种流式的非结构化的数据。它的最大粒度是视频流，最小粒度是图像帧，中间没有自然过渡的概念层次，直接提取视频数据的内容或高层语义信息是很困难的，因此一般的数据管理技术不能从内容上对视频数据进行有效的分析和管理，导致传统计算机视频处理是基于像素的处理，难以抽象出人类可以感知的内容。通常计算机对视频图像的操作仅仅停留在对视频的播放、编辑和存储，视频信息处理的效率不高，难以满足现代信息社会的发展要求。

由于对视频数据广泛需求，推动了对视频处理和分析技术的迅速发展，视频处理一直是人们高度关注的技术领域。基于内容的视频分析技术的快速发展产生了很多新的应用和对新技术的研究和发展，面对海量的多媒体信息，如何快速浏览大容量的视频数据，以及如何获取和表现视频的内容，成为人们关注的焦点问题，二十世纪90年代以来，对非文字信息如图像、声音、视频的基于内容的检索成为一个异常活跃的研究领域。我们知道，对于一篇文章来说，摘要就是对文章的一个简单概述，它可以使人们快速了解文章的主要内容，文章的摘要己经被广泛地应用于对文献的检索中，用户完全可以只通过阅读文章的摘要，就能判断出该篇文章是否是自己感兴趣的。将这一思想应用于对视频的浏览和检索当中，就是产生了视频摘要技术。视频摘要(Video synopsis)：既以自动或半自动的方式对视频的结构和内容进行分析，从视频文件中提取出有意义的部分，并将它们以某种方式进行组合，形成简洁的、能够充分表现视频语义内容的概要。它是对长视频内容的简短总结，通常用一段静态或者动态的图像序列来表示，并对原始信息予以白柳。视频摘要技术具有非常实用的价值，可以减小视频的数据量，提高视频检索和视频浏览的速度，大大提高视频数据的利用效率。以新闻视频信息为例：通常我们可以通过多种渠道获得新闻视频信息，如数字电视、互联网等等。不同地域的电视台播放的新闻节目的内容不尽相同，有的新闻节目仅仅在某一电视台播放一次，而有的新闻节目则在同一电视台和不同电视台的不同时间多次播放。面对如此众多存在大量冗余的新闻视频数据流，无论是对于观众还是新闻视频制作和管理者，观看所以的新闻视频是不可能的。不管是对于电视观众还是视频处理专业人员，他们往往关注的是与自己相关的一些信息，如股民需要股市的情况，球迷关注球赛的报道，而政府新闻监管部门需要注意一些特定敏感领域的新闻内容等等。但是为了找到自己关注内容，不得不浪费人力和物力去浏览或回放所以的原始视频节目。如果能建立一个视频摘要系统，将大大缩短这一过程。由此可见视频摘要是提高视频处理系统效率的关键问题，视频摘要是基于内容的视频检索的重要工具，对视频应用领域的各种运用都具有重要意义，近年来已经成为视频研究的关注热点之一。

# 2视频摘要相关概念

**视频摘要：**视频摘要又称视频浓缩，是对视频内容的一个简单概括，以自动或半自动的方式，先通过运动目标分析，提取运动目标，然后对各个目标的运动轨迹进行分析，从原始的多媒体数据中选取具有代表性的、有意义的部分，将不同的目标拼接到一个共同的背景场景中，并将它们以某种方式组合并生成紧凑的、用户可读的原始视频的缩略。视频摘要在视频分析和基于内容的视频检索中扮演着重要角色。一个好的视频摘要系统可在最少的时间使用户从原始视频序列中获得最大的信息量。

视频摘要分为静态视频摘要和动态视频摘要。



**图2-1：视频摘要分类**

**静态视频摘要：**又称为关键帧集，是由原始视频中具有代表性的图像帧组成，以直接、分层或缩放的方式进行组合。层次的视频摘要可以用于快速导航和索引视频帧，有利于视频检索。

**动态视频摘要：**是从原始视频中选取可表达语义内容的视频片段拼接编辑得到。它本身也是一段视频，但比原视频要短得多。动态视频摘要可分为精彩集锦和全局缩略视频。精彩集锦一般由原始视频中最精彩的部分组成。如，足球比赛中的进球集锦。全局缩略视频是对整个视频内容的概括，它通过对整个时间轴上的视频片段进行组合，使用户对视频内容进行全局掌握。

# 3视频摘要的研究现状

**国外：**

在国外，视频摘要的研究起步较早，最早可以追溯到 1994年 Carnegie Mellon 大学的 Informedia 系统。最近许多大学、科研院所及有实力的大公司纷纷投入了大量人力物力进行视频摘要技术的研究。

Informedia 系统是视频摘要研究领域的先驱。它主要针对的是新闻视频，摘要的主要目的是抽取视频中的重要音视信息形成缩略视频。该项目利用 TF-IDF 方法，结合了语音、图像和自然语言的理解来实现对视频的自动描述，选取一些有代表性的帧（如包含人脸或字幕的帧等）按照一定的优先级别进行排序重组形式视频摘要。该工程将视频摘要技术嵌入在视频库系统中，使用智能的自动机制，提供从一个非常大的在线数字视频库进行内容搜索和检索，应用于教育、培训、体育和娱乐等领域。经过多年研究与发展，该项目已经进入了二期工程Informedia-II，二期在一期的基础上提高了速度和精度，包含对名称、地点、日期和时间的解释，并且加入了动态故事分割、视频事件的特性和相似性匹配技术，旨在实时处理和分析视频信息，并同期自动填充分布式专有视频数据库。

QBIC 系统（Query By Image Content）是一个商用的基于内容的图像和视频查询系统，由 IBM 公司 Almaden 研究中心推出，是最早的“基于内容”的视频检索系统。它允许对图像、视频、文本和语音多种形式的多媒体信息进行检索，并对视频也提供了基于运动特征的检索。视频方面主要利用了颜色、纹理、形状、摄像机和对象运动来描述内容。其子项目 CueVi-deod 主要研究视频摘要和索引技术，它与 TJ Watson（言论和信息检索）和Haifa（音频分析）合作，旨在解决大型视频数据库在创建、索引和使用中出现的挑战。

德国 Mannheim 大学的MOCA 系统是 W.Effelsberg 教授开发的一个包含视频、场景、片段、帧的 4 层数字视频解析与摘要系统，使用者可以在各个层次上对视频流数据进行操作。该项目研究范围集中在电影分析上，采用的算法也同样适用于一般的视频和音频材料。它抽取电影片段中最有代表意义的场景来自动生成预告片，为了加快在摘要提取的速度，一般将电影分为长短相等的 6 个部分，最后一部分影片的结尾剔除，留以悬念来吸引观众，从其余 5 部分中提取 5 种类型场景。这 5 类场景为：对话场景，高动态、变化快的场景，多动作的场景，包含有标题的场景，反映电影基本色调的场景。然后将选出来的场景序列按照时间顺序进行组织来生成视频摘要。

Video Manga 系统是由美国 FX Palo Alto 实验室开发，该系统能自动生成漫画书（ComicBook）形式的视频摘要。该算法是基于每一帧的颜色特征来进行视频分割的，根据视频片段之间的相似性，视频片段进行分析处理。如果某个片段较长或是罕见，则被认为是重要的。关键帧从重要度高的片段中提取，其在漫画书中的大小也与重要度成正比。目前该系统已经实现了一个交互式版本的图案摘要作为Mbase系统中抽象级的视频浏览工具。当鼠标移动到某一关键帧时，便能在时间轴突显该帧和相应的片段。它允许用户探索视频的间特性，点击某一关键帧，则可以从该帧所在片段的第一帧开始播放该视频。

**国内：**

在国内，视频摘要的研究起步较晚。国内的研究单位如微软亚洲研究院、中国科学院计算所、清华大学、浙江大学、复旦大学以及国防科技大学等也研究开发出了一些视频摘要系统,在视频镜头的分割、视频检索、镜头探测、场景聚类、关键帧提取方面取得了一些技术进展,有了比较深厚的技术积累。

国防科技大学的多媒体研究开发中心提出了一个统一、可扩充的视频摘要模型“实体一描述一效用”模型。该模型从视频实体出发,将原始视频流进行切分,生成不同层次的实体,然后采用不同的描述子如人脸描述子、关键人物描述子、直方图描述子等对实体进行描述,生成不同的效用值,并根据效用值来生成最终的视频摘要。不同的实体描述子可能完全不同,比如镜头描述采用人脸描述子,而关键帧描述采用直方图描述子,等等。该模型一方面具有可扩充性,用户可以根据自身需要添加不同的描述子;另一方面能够较好地反映用户的兴趣,不同描述子的权值可由用户来指定。

北京交通大学计算机与信息技术学院计算机研究所提出了一种概要序列形式的视频摘要生成技术。这种技术是在视频结构化的基础上,从减少冗余信息和保持本质内容两个角度出发,定义了镜头(场景)的冗余度和语义重要度,从而推导出视频内容缩减和保持的原则,然后综合这两个方面,并考虑了可理解性的因素,自上而下地从原视频片段中提取出能准确概括原主要内容的视频摘要。

# 4视频摘要的相关技术

视频摘要的关键技术主要包括：背景提取，运动目标检测。视频摘要生成的过程如图4-1所示。



**图4-1：视频摘要生成过程**

视频摘要技术一般都遵循先分后合的原则，首先要对视频进行类容理解和分析，把视频划分为合理的基本单元。现在人们多以帧作为视频的基本结构单元，所以先在通过读取视频流并分割成帧。在视频分割的基础上，对每帧图像进行处理，提取出背景。然后再提取出前景，跟踪运动目标。最后，把处理过的每帧图片重组为最终的视频摘要。

## 4.1背景建模

视频摘要中背景的提取是运动目标检测的基础，也是视频处理、理解、编码的关键。背景建模是视频图像中运动目标检测的一种方法，其思想是对图像的背景进行建模，一旦背景建模建立，将当前的帧与背景模型进行比较，根据比较结构确定前景目标及运动目标。

常用的背景建模的方法有：直方图背景建模、均值背景模型、帧间差分背景模型、混合高斯背景模型。

**直方图背景建模：**直方图统计背景模型是提取背景的一个常用方法，该算法基于运动目标在场景中运动不会长时间停留在同一个位置上的假设，在某一段时间内，视频序列特定像素位置处出现频率最高的像素值就是该点背景像素值。

直方图统计方法，就是通过存储视频中每一像素位置在序列的时间轴方向上的直方图信息（即在某一时间段内统计该像素位置处每一种可能的像素值所出现的次数），取该像素位置的直方图峰值所对应的像素值作为背景像素值。然后，用背景减除法，来获取运动目标。

该算法抗噪声干扰性好，在通常情况下提取的背景较好，但运算量大、提取速度慢。并且当运动目标速度过快的时候，对于背景中的近景，因为运动目标快速通过，背景显示的时间比较长，得到的背景图像较好；而对于图像远景，一些运动目标相连，会误把运动目标当成背景，是得到的背景图像失真。

**均值背景模型：**平均背景模型是讲运动目标看作为噪声，用累加平均的方法消除之。利用运动目标运行一段时间的序列图像进行平均得到的背景图像。



均值背景模型是一种简单、计算速度快但对环境光照额背景的多模态性比较敏感的背景建模算法。在实际应用中，往往是图像中的某些区域有亮度高的目标，而某些区域有亮度低的目标经过，通过平均算法得到的背景图像就会出现亮度分布不均的区块。另外，该算法得到的背景图像受运动目标数量的影响比较大。

**帧间差分背景模型：**考虑到摄像机固定，背景像素点的灰度值变化比较慢，而运动目标对应的像素点灰度值变化比较快。实际上每一幅的图像都可划分为两类区域：一类由灰度值变化较小的像素点构成，成为背景区域；另一类由灰度值变化比较快大的像素点构成，包括运动目标在当前帧中所在的区域和它在上一帧所在的区域，称之为运动变化区域。

帧间差分模型是一种通过对视频图像序列中相邻两帧作差分运算来获得运动目标轮廓的方法。它可以很好的适用于存在多个运动目标的情况。当监控视频场景中出现运动物体时，帧和帧之间会出现较为明显的差别，两帧相减，得到两帧图像亮度差的绝对值，判断它是否大于阙值来分析视频或图像序列的运动特性，确定图像序列中有无物体运动。

算法步骤：

1. 把第一帧图像作为初始背景图像,迭代参数为。
2. 求当前帧与前一帧的查分二值图像：



1. 根据查分二值图像更新背景：



**混合高斯背景模型**[4]**：**混合高斯背景建模是基于像素样本统计信息的背景表示方法，利用像素在较长时间内大量样本值的概率密度等统计信息(如模式数量、每个模式的均值和标准差)表示背景，然后使用统计差分(如3σ原则)进行目标像素判断，可以对复杂动态背景进行建模，计算量较大。

在混合高斯背景模型中，认为像素之间的颜色信息互不相关，对各像素点的处理都是相互独立的。对于视频图像中的每一个像素点，其值在序列图像中的变化可看作是不断产生像素值的随机过程，即用高斯分布来描述每个像素点的颜色呈现规律【单模态(单峰)，多模态(多峰)】。

对于多峰高斯分布模型，图像的每一个像素点按不同权值的多个高斯分布的叠加来建模，每种高斯分布对应一个可能产生像素点所呈现颜色的状态，各个高斯分布的权值和分布参数随时间更新。当处理彩色图像时，假定图像像素点R、G、B三色通道相互独立并具有相同的方差。对于随机变量*X*的观测数据集，为*t*时刻像素的样本，则单个采样点其服从的混合高斯分布概率密度函数：







其中*K*为高斯分布的总数，为*t*时刻第*i*个高斯分布，为其均值，为其协方差矩阵，为方差，*I*为单位矩阵，为*t*时刻第*i*个高斯分布的权值。

算法步骤：

1. 当有新的观测值来临时，则将这个样本的像素值同当前K个模型按下式进行比较，同时计算观测点落入相应高斯分布的概率，按某以判断法则选择匹配的高斯分布。这个法则是：



C为一常数。

1. 如果所匹配的模式符合背景要求，则该像素属于背景，否则属于前景。
2. 各模式权值按如下公式更新，其中是学习速率，对于匹配的模式=1，否则=0。



1. 为匹配模式的均值与标准差不变，匹配模式的参数按照如下公式更新：







1. 如果第一步中没有任何匹配模式，则权重最小的模式被替换，即该模式的均值为当前像素值，标准差为初始较大值，权重为较小值。
2. 各模式根据按降序排列，权重大、标准差小的模式排列靠前。
3. 选前b个模式作为背景，B满足下式，参数T表示背景所占比例。



## 4.2运动目标跟踪

运动目标跟踪是视频摘要中不可缺少的环节。一般目标跟踪分为两个部分：特征提取、目标跟踪算法。其中提取的目标特征大致可以分为以下几种：

1） 以目标区域的颜色直方图作为特征，颜色特征具有旋转不变性，且不受目标物大小和形状的变化影响，在颜色空间中分布大致相同。

2） 目标的轮廓特征，算法速度较快，并且在目标有小部分遮挡的情况下同样有较好的效果。

3) 目标的纹理特征，纹理特征较轮廓特征跟踪效果会有所改善。

目标跟踪的算法大致可以分为以下四种：

**均值漂移算法**：即meanshift算法，此方法可以通过较少的迭代次数快速找到与目标最相似的位置，效果也挺好的。但是其不能解决目标的遮挡问题并且不能适应运动目标的的形状和大小变化等。对其改进的算法有camshift算法，此方法可以适应运动目标的大小形状的改变，具有较好的跟踪效果，但当背景色和目标颜色接近时，容易使目标的区域变大，最终有可能导致目标跟踪丢失。

**基于Kalman滤波的目标跟踪**：该方法是认为物体的运动模型服从高斯模型，来对目标的运动状态进行预测，然后通过与观察模型进行对比，根据误差来更新运动目标的状态，该算法的精度不是特高。

**基于粒子滤波的目标跟踪**：每次通过当前的跟踪结果重采样粒子的分布，然后根据粒子的分布对粒子进行扩散，再通过扩散的结果来重新观察目标的状态，最后归一化更新目标的状态。此算法的特点是跟踪速度特别快，而且能解决目标的部分遮挡问题，在实际工程应用过程中越来越多的被使用。

**基于对运动目标建模的方法**：该方法需要提前通过先验知识知道所跟踪的目标对象是什么，比如车辆、行人、人脸等。通过对要跟踪的目标进行建模，然后再利用该模型来进行实际的跟踪。该方法必须提前知道要跟踪的目标对象是什么，然后再去跟踪指定的目标，这是它的局限性，因而其推广性相对比较差。

**结合高斯混合模型和纹理方法**[2]**：**

结合高斯混合模型和纹理方法提取紧凑的移动物体，并尽可能消除阴影。然后，根据初始设定的轨迹，用一个粒子滤波跟踪来检查两个轨迹是否属于一个活动。然后，把这些轨迹被连接起来，产生一个更流畅的管的视频摘要。几个视频的实验结果表明我们的方法的有效性。

1. **Texture-GMM算法**

通常，阴影的纹理特征与背景相似。因此，我们整合了高斯混合模型（GMM）纹理特征的前景检测和去除阴影。

局部二进制模式（LBP）因在特征的纹理表现出较强的能力而作为这项工作的纹理特征。LBP算子的标签与相邻的一个二进制数字图像的每个像素进行比较：

 （1）

其中，表示中心像素点的灰度值，表示相等间隔P像素点的灰度值。一个特征像素的特征矢量为以像素中心，半径为R计算的直方图。

我们测量两个LBP特征值之间的相似性如下：

 （2）

其中表示LBP特征值，N表示特征值得维数。

**Texture-GMM**提取前景的方法分为3步：

1. 背景构造；
2. 通过LBP特征和GMM提取前景；
3. 前景组合。

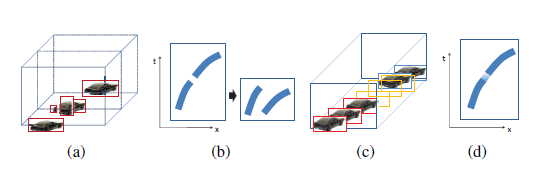
**背景构造：**在这一步，我们构造了2个背景模型。一个是高斯混合模型，每帧之前和之后剪辑的之间中位数，这个时间中位数的目的是为了构建一个背景图像，以便在背景中计算LBP特征值。当前帧前后的时间是450帧，那么总数就是900帧。

**通过LBP特征和GMM提取前景：**在这一步，每一帧提取两个前景。首先，是基于GMM的前景提取。然后，对于每一帧，计算当前帧和中央图片的每一个像素点的LBP特征。在公式2中比较了这两个特征。如果一个像素的分数小于一个阙值，它就被标记为前景。经过多次试验，阙值设置为0.6，可以达到满意的精度。

**前景组合：**一个像素如果它在基于GMM的前景和在时间中值的前景中都被标记为前景，那么它就被标记为前景。通过组合这种方案，得到最终的前景。

**2．通过跟踪连接管（tube）**

通过texture-GMM的方法，更紧凑的前景被提取，这有利于建立一个流畅管的视频摘要。然而，如果一个对象被分割成若干个部分，这个管可能断开，导致在视频摘要中的对象突然中断，如图4-2（a）和（b）所示。我们使用粒子滤波跟踪算法连接管的分段以产生流畅的管。



**图4-2：（a）一个物体被分为若干个部分；（b）在原始视频（左）和管中断的概要视频（右）的管片段；（c）通过估计的矩形（黄色）跟踪从管（红）到另一管（蓝）；（d）连接的结果。**

指出和的隐藏状态，并且观察时间t内的数据并给出可能的结果，滤波估计是，符合递归：

（3）  
其中是过度模式指定对象在两帧之间如何移动，是观察模型，描述对象在特定状态的可能性。



一组M粒子的加权被用作近似过滤分布，

 （4）

粒子来自，第i个粒子的权值被选来作为可能情况的数据。

在一个图片碎片中，先验分布的跟踪过程是基于HSV的直方图，该直方图作为参考直方图。使用二阶自回归动力学模型作为过渡模型。观测模型为：

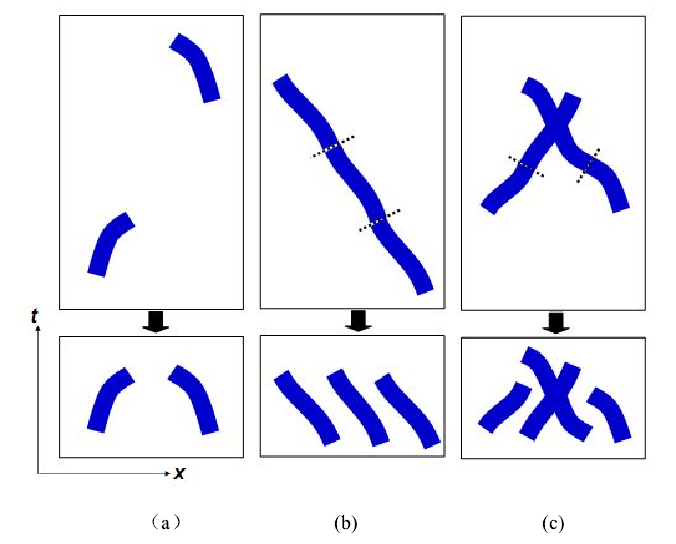
 （5）

其中定义了参考HSV直方图和当前直方图之间的距离。

对于初始的轨迹片段，跟踪器是用来连接断开的片段。假设每个轨迹（管）是由每帧中一系列包围框中的对象表示，，其中表示对象周围的包围框。在集合管中，我们从最后一个包围框开始，在之后的k帧，连续跟踪图片碎片。当存在一个管从一帧开始，其第一个包围框与跟踪的矩形有明显的重叠，那我们就计算两矩形中的直方图之间的距离。如果距离小于某一阙值，我们就把和视为同一个活动对象连接起来。（如图4-2（c））

## 4.2目标轨迹组合和优化

获得运动物体的轨迹之后，我们就需要对这些运动轨迹进行时间上的平移、组合和优化，从而实现视频的浓缩。图 3.16 展示了目标轨迹组合优化的原理[6]。



**图：活动时间在时间轴上的移动**

可以用图来阐释运动物体轨迹重组的原理：在图 3.16 中，上图表示原始视频中的运动物体序列，下图为经过时间轴平移组合后的效果图。（a）图中，原始视频中不同时出现的物体，在视频摘要中同时播放。（b）图中，原始视频中一个长时间出现的单个运动目标轨迹被分割成多个时间间隔，在摘要视频中进行同时播放。（c）通过将运动轨迹分割成段落，避免了互相遮挡的运动轨迹平移后对视频摘要效果的影响。

上面展示原始视频中活动轨迹的分布，经过在时间轴上进行转移组合后，被紧凑地分布在摘要视频中轨迹中。为了能否实现如上的视频浓缩效果，这一小节将叙述如何建立适当的数学模型解决轨迹的组合优化问题，这个也是视频摘要中的核心问题。

### 4.2.1能量目标函数

首先，我们定义不同运动轨迹之间交互作用的能量函数。通过对这个能量函数最小化，

就能在摘要的视频中获得最多的活动物体，同时避免不同运动对象之间的冲突和覆盖。假定

B 为所有运动轨迹的集合，每个运动轨迹 b 通过原始视频中一个有限的视频流来定义：。

在视频摘要生成过程中，原始视频中的运动物体轨迹通过一个时间映射函数

M 进行时间轴上的平移，成为摘要视频中的一段运动物体轨迹 。 表示将原始视频中轨迹 b 映射到了摘要视频中，M(b)=Ø 表示轨迹 b 没有被映射到摘要视频中去。我们定义能量函数如下，通过最小化如下的能量函数，即可实现运动轨迹的组合和优化。



其中，为活动事件时序一致性惩罚函数，为活动能量损失函数，为摘要视频轨迹之间的活动冲突能量代价函数。，是用户自定义系数。能量代价函数的意义在于，在众多的映射方案当中，找出其中一种方案使得代价损失最小。

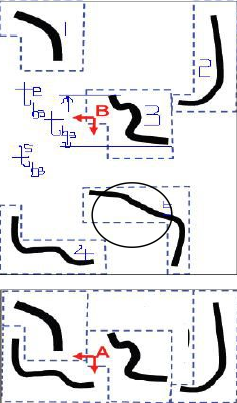
（1）活动能量代价

为了尽可能减少摘要视频中运动物体的丢失，最大化地保留原有的原有的活动事件，定义活动能量代价函数，用以惩罚视频摘要中没有出现的运动轨迹而造成的能量损失。若原始视频中运动轨迹 b 没有被映射到摘要视频中去，即 M(b)=Ø ，那么对 b 的惩罚为：





其中每个轨迹b通过能量函数来描述，为原始视频背景图像的一个像素点，为原始视频中输入图像的一个像素点，为运动轨迹持续的时间长度。若，只有部分被包含在视频摘要中，则只惩罚未被包含的部分。如下图所示，为活动能量代价损失示意图：



**图：活动能量代价损失图**

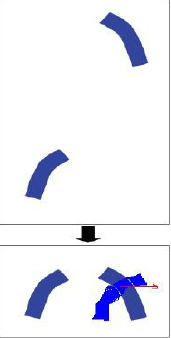
从图可以看出，上部分为原始视频，下部分为摘要视频，其中有5个运动物体轨迹，一种1-4轨迹都浓缩到了最后的摘要中，而5轨迹却出现了丢失，从而在能量函数中，产生这样一个活动损失函数。

（2）活动冲突能量代价函数

将不同时间段的运动物体轨迹在时间轴上进行平移组合的同时，不可避免出现活动轨之间出现遮挡、交叉等情况。为了能够较为流畅观看视频摘要，需要最大化避免这种遮挡和碰撞的发生。对于摘要视频中的任意两个不同的活动轨迹，其出现的时间段分别为，当两个时间段从在交集时即，定义之间的活动冲突能量代价函数为：



如下图所示，为活动冲突能量代价损失示意图：



**图：活动冲突代价损失示意图**

由图可以看出，从原始视频中进行运动轨迹的平移时，出现不同运动轨迹的碰撞或者重合，从而产生了活动冲突能量代价。

（3）时序一致性代价

时序一致性代价用于保持原始视频中事件发生的先后顺序。这种时序一致性代价的设定能够最大化的保证视频中原有的故事情节和运动对象之间的交互关系。举例说明，在原始监控场景中出现两个人交谈的情形，通过定义时序一致性代价，在摘要视频中依然保持它们之间的这种交互关系。实际上这种交互关系难以检测，原始视频中两个轨迹之间交互关系通过时空距离来衡量。定义如下：

如果，那么：。其中，为在第t帧内两个轨迹所在目标内最近的两个像素之间的欧几里得距离，决定轨迹之间的交互成都。

在原始视频中两个并行的轨迹 ,假定在进行轨迹的平移后，被映射到前面去了，如果轨迹在摘要视频中不享有共同的时间段，那么他们的交互性将呈指数级地减少：

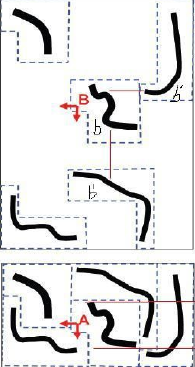


其中是一个定义事件间仍存在交互的时间程度参数。

时序一致性代价函数通过惩罚运动轨迹间时序关系被破坏的情况来保持原始视频中两个轨迹之间时间上的顺序性。时序一致性代价函数如下：



其中c为一个常数，用来作为时序性关系被破坏的惩罚。如下图所示，为时序一致性代价损失示意图：



**图：时序一致性代价损失示意图**

由图可以看出，从原始视频中进行运动轨迹的平移的时候，运动轨迹B出现的先后顺序出现了颠倒，从而导致产生了时序一致性代价。

（4）能量目标函数最小化

在义了运动物体轨迹间的能量函数之后，可以通过各种基于马尔可夫随机场的方法进行能量最小化，比如基于置信度传播 (Belief Propagation)的算法、基于图割(Graph Cuts)的算法。实验中使用贪心算法，也能取得很好的效果。

贪心算法是指，在对问题求解时，总是做出在当前看来是最好的选择。也就是说所做出的仅是在某种意义上的局部最优解，并不是整体的最优解。贪心算法不是对所有问题都能得到整体最优解，但对范围相当广泛的许多问题他能产生整体最优解或者是整体最优解的近似解。

贪心算法如下：

1、随机选定一个初始解；

2、Do while（终止条件）

1. 在某个领域函数所定义的领域范围内，按照某个（随机）扰动产生策略，得到一个新解；
2. 对新解进行评估，得到；
3. 如果（或者），即新解比老解好，则令；
4. 否则，。

3、End Do

在所有的时间映射 M 空间中，都使用了贪心算法，对轨迹数据进行组合优化。以上部分就是视频摘要的核心思想。在实现了运动物体轨迹的重新优化组合之后，就可以通过图像拼接和像素融合等技术，可以得到最后的视频摘要的结果。

# 5视频摘要的发展趋势

近5年，多媒体技术日趋成熟，网络视频数量正以无法估量的速度不断增加。视频摘要技术的研究也相应地有所偏倚，一方面市场应用对视频摘要的实时性和交互性的要求越来越高，而另一方面人们迫切地希望能依据个人喜好对视频在线提取摘要。基于此，实时视频摘要的研究越来越得到学者的青睐。

实时视频摘要通常将在线的网络视频分析作为分析对象。为了提高效率节省时间，一方面可以选取简单的特征或者改进的算法，如STIMO采用改进的Furthest-Point-First( FPF) 算法对预滤波后的视频帧提取的颜色直方图数据进行聚类，对30 帧的故事板，STIMO 相较于采用K-means 算法时间上快了20 倍。而OLAM( on-line abstraction module) 对10 min 的CCTV 新闻在线生成摘要，通过选取简单的特征，处理时间仅为1 min。另一方面，可通过直接在变换域中对视频进行处理节约解码时间，Almeida直接在变化域中对视频进行处理，提取I帧的DC 值生成DC 图像，再进行特征分析。

网络视频的不断增多，对实时视频摘要的要求也将越来越高。但妄图仅通过选取简单的特征来提高速度将使生成的摘要语义不完善，从而降低摘要质量。而目前压缩域的摘要技术研究较少，主要集中于I 帧DCT 系数分析、运动矢量分析和宏块类型分析3 方面。并且，由于不同视频编解码方式的差异，在具体分析时算法仍须进行调整，较为复杂。

Fu 等人在2010 年第1 次提出了多视角视频摘要，即对多视角、多摄像机拍摄的视频生成摘要。在实际生活中，对于一些重要场所，如银行、办公室，为了安全考虑通常会架设多台摄像机从不同角度进行监控摄像。对于此类视频，由于在同一时刻对同一场景有不同视角的镜头描述，因此在实际摘要时除了横向地在时域上选择具有代表性的内容，还应纵向地对同一时刻不同视角的镜头进行分析。

Fu 等人通过构建镜头的时空序列图来展示同一时刻不同视角、同一视角不同空间的视频内容信息。首先，将不同视角获得的视频分割为镜头，然后利用超图显示不同视角的镜头关系，超边代表镜头间的相关性。再将超图转化为时空序列图，每个点值显示镜头的重要度值，各边的权重记录镜头间的相似度，最后利用随机漫步进行事件聚类，并最终得到多视角的故事板摘要。Leo 等人在交通要道上间隔一定距离架设了的两台摄像机对路况进行监控，并对两个摄像机拍摄得到的视频提取摘要再结合运动分析进行异常事件监测。多视角视频摘要不仅需要对同一个视频中不同语义单元的进行识别区分，还必须实现同一语义单元在不同视角下的匹配和比较工作。研究难度较大，目前这方面的研究也较少。但是由于多视角拍摄视频广泛地存在于日常生活中，因此该研究具有较大的现实意义。

所以，视频摘要技术未来的一个重要的研究方向将会是实时的视频摘要和多视角视频摘要。

# 6.总结

综上所述，本文介绍了视频摘要技术的研究背景和意义，和视频摘要的相关概念，以及现阶段国内外的研究现状。具体依据背景建模、运动目标跟踪两个视频摘要生成的主要步骤总结了视频摘要的主要技术。并总结了视频摘要领域大展的新趋势：实时视频摘要和多视角的视频摘要。

视频摘要技术未来的一个重要研究方向将会是视频摘要的商品化。视频摘要本身具有很大的商业前景，而当将其作为成熟商业产品进行包装时，仍需考虑许多现实问题：

1. 视频摘要的实时性和交互性。不同用户拥有不同的偏好，从而对摘要长度、摘要内容的要求有所差异。允许用户依据个人偏好定制需求获得专属摘要将是视频摘要商业化必须考虑的问题。随之，为了保证用户的满意度，摘要的实时性也必须得到满足。
2. 视频摘要的便携性。生成高质量的摘要，离不开硬件设备。而即使是一台笔记本，由于缺乏移动性，也将大大影响摘要技术的传播。一旦生成摘要的硬件设备太过繁复，则视频摘要终将沦为“实验室”产品。因此，怎样让用户利用身边常见的电子设备，如手机、平板等，随时随地对自己拍摄的或者自己感兴趣的视频生成高质量摘要，也是摘要商品化需解决的问题。

## 参考文献

1. C. Wolf and J. M. Jolion, "Integrating a Discrete Motion Model into GMM Based Background Subtraction," Pattern Recognition (ICPR), 2010 20th International Conference on, Istanbul, 2010, pp. 9-12.
2. M. Lu, Y. Wang and G. Pan, "Generating fluent tubes in video synopsis," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 2292-2296.
3. T. Maeda and T. Ohtsuka, "Reliable background prediction using approximated GMM," Machine Vision Applications (MVA), 2015 14th IAPR International Conference on, Tokyo, 2015, pp. 142-145.
4. 朱碧婷,郑世宝. 基于高斯混合模型的空间域背景分离法及阴影消除法[J]. 中国图象图形学报,2008,10:1906-1909.
5. 宋杨. 基于高斯混合模型的运动目标检测算法研究[D].大连理工大学,2009.
6. 徐子豪. 海量视频数据摘要技术研究[D].南京邮电大学,2014.
7. 王田,刘伟宁,韩广良,杜超,刘恋. 基于改进MeanShift的目标跟踪算法[J]. 液晶与显示,2012,03:396-400
8. 李昱辰. 基于粒子滤波的视频目标跟踪方法研究[D].兰州理工大学,2013.