南京理工大学

硕士学位论文

足球视频语义分析

姓名：张峰

申请学位级别：硕士

专业：系统工程

指导教师：王建宇;周献中

20060601

摘 要

近年来，基于内容的多媒体信息检索已经成为一个热门研究领域，体育视频检索作为其中一部分也得到了广泛的研究。足球比赛深受世界众多球迷喜爱，因此本文以从电视转播中采集到的足球视频为研究对象，分析提取视频中的高级语义信息，如射门等事件，并提出了一个足球视频语义分析框架。主要包括这样几个部分：基于切变检测的视频分段、比赛场地检测、镜头分类、慢镜头检测、禁区线识别、比赛字幕区域检测以及精彩比赛事件识别等。

视频分段是后续处理分析的前提，由于足球视频中镜头转换方式以切变为主，因此本文提出了基于切变检测的视频分段算法。在得到的视频片段中选取关键帧，并在关键帧上做比赛场地分割处理。通过对场地分割结果中场地比例、场地中球员形状和面积等特征的分析，将镜头分为长镜头、中镜头、特写镜头和场外镜头四类。足球比赛中，当出现射门事件或犯规事件时，会出现慢动作回放镜头，因此检测慢镜头能够对精彩事件定位。本文提出了一种慢镜头检测算法，对视频片段是否慢镜头进行标注。射门事件通常是在禁区附近发生的，本文通过检测禁区线实现对禁区的检测，从而辅助射门事件的检测。如果射门成功，比分就被改写，则视频中会出现比赛比分的字幕，因此字幕检测同样可以用来辅助射门事件的检测，同时还可以作为判断射门是否成功的依据。

综上所述，本文采用慢镜头触发事件检测机制，每当检测到慢镜头后向前回溯，并根据镜头类型出现的时序分析结果，实现对精彩事件的识别，并通过禁区检测和字幕检测辅助事件识别。最后，本文以 Visual C++ 6.0 为开发平台，实现了一个足球视频自动分析原型系统。实验表明，本文提出的足球视频语义分析算法具有令人满意的效果。

关键词：足球视频，切变检测，慢镜头，字幕检测，事件检测

Abstract

Content based multimedia retrieval has been a hot research field recent years. As a part, sports video retrieval has been studied widely. Soccer is a very popular game, which has a large number of audiences all over the world. Therefore, this paper used soccer video from TV program as an example, extracting high-level semantics, such as the event when goal shot occurs. This paper presents a framework for soccer video semantic analysis, consisted of the parts as follows: video segmentation based on cut detection; field segmentation; shot classification; slow-motion shots detection; forbidden zone detection; caption region extraction and exciting events detection.

Video segmentation is the first step, this paper presents an algorithm to segment video by cut shot detecting, for most of shots transitions in soccer video are cut. Select key frames in video segments, and detect field region in each frame. Shots are classified into four types: long-view shot, middle-view shot, close-up shot and out of field shot, using field rate, shape and areas of player in the field. Slow-motion replays often appears after shoot shoting or foul occurs, so we determine if a given shot consists of a slow-motion replay to detect events. This paper presents an algorithm for slow-motion replays detecting. As shoot often happens near forbidden zone, and forbidden zone lines in a long-view shot can be detected to find forbidden zone, which can help shoot event detection. If a goal occurs, the score will change, so there will be score caption in the video. Therefore, score captions are used to help detect shoot event and determine whether goal happens.

In a word, this paper presents an algorithm triggered by slow-motion replays. With the results, go back certain seconds to find temporal sequences of shots to detect exciting events, assisted by forbidden zone detection and caption region extraction. This paper implements a prototype system for automatic soccer video semantic analysis by Visual C++6.0. Experiments have demonstrate that the algorithm is effective.

Key Words: Soccer Video; Cut Detection; Slow-motion Replay; Caption Extraction; Event Detection

1.引言

1.1 研究背景

随着宽带网络、通信器材、存储设备以及数字电视等多媒体载体及处理设备的快速发展，人们已经可以通过个人电脑，掌上电脑，数字电视，甚至3G手机，随时随地访问视频。由于多媒体信息具有很大的数据量，而且人们习惯运用高层语义概念来查询和浏览多媒体数据库，因此有必要发展多媒体内容的自动语义分析技术，实现多媒体数据库的建立、管理和检索等。Google 和百度在文本搜索领域的巨大成功也反映了人们在“信息爆炸”的今天，对快速有效检索信息的强烈渴望。在多媒体信息检索领域，基于内容的信息检索近年来得到了广泛的研究。基于内容的信息检索是指通过对视频数据从低层到高层进行处理和分析以获取其内容，并根据内容进行检索。涉及到直接根据图象和视频内容的含义，对图象和视频进行有效查询、索引、浏览、搜索与提取。

视频数据具有内容繁多并且复杂的特点，因此对视频的检索十分困难，很难建立通用的检索系统。不同种类视频都有各自的一些特点，如视频生成模型，领域知识等。具体说来，新闻报道需要能为观众提供有用简洁且准确的信息，因此镜头之间的转换以简洁为原则，各个新闻条目之间也有着比较明确的边界。电影类视频又可分为艺术片、武打片、战争片、剧情片等种类，并且每种类型的电影都有自己独特的制作手段，呈现不同的特征。体育比赛更是因为比赛项目的不同，造成内容千差万别。但是尽管不同项目的体育比赛节目具有不同的特点，同一项目的比赛却常常表现出相似的拍摄、编辑模式，根据这些信息可以建立某一类体育项目的分析框架。本文选择足球比赛视频作为体育视频检索领域的研究对象。下面将主要针对体育视频检索这一研究方向，说明其研究意义、研究现状以及发展趋势。

1.2 基于内容的体育视频检索

体育运动深受大众喜爱，观看各类体育比赛节目已经成为人们业余生活中不可或缺的一种休闲方式。随着体育视频应用和传播的日益广泛，体育视频检索已经成为基于内容的视频检索领域的一个重要分支，受到了广大研究人员的重视。体育比赛一般要持续几十分钟，因此一场体育比赛视频具有较大数据量，但是有价值的片段，即令观众感兴趣的精彩镜头通常只是整场比赛的很小一部分，因此有必要通过对体育比赛

视频的处理和分析，去除视频中不精彩的部分，并创建视频摘要，以方便观众检索和浏览，从而节省观看时间，减少资源浪费。若用手工方式实现上述任务，虽能达到目的，但工作量将会非常大，因此迫切需要解决体育比赛视频自动分析的问题。

对体育视频进行索引和建立视频摘要等工作，主要是指结合画面，声音，文本多方面信息，提取事件和物体特征等有价值的语义信息。其中画面，声音，文本等多方面信息虽然有助于高级分析，但由于文本信息有时无法获得，音频和信息流特征分别由于说话人特征和压缩方式的变化而有所波动，所以画面等视觉信息是体育视频处理中最为有效的特征。体育视频的视觉特征(画面)分析是指对低层视觉特征，节目制作特征和视频对象特征等的提取。低层视觉特征包括颜色，纹理，形状，运动特征。节目制作特征是指在节目制作过程中，工作人员为了更好的表达和展现比赛过程，根据一般制作规则和自身长期积累的丰富经验，在视频中加入的一些特殊效果。镜头转换，镜头类型，镜头长度，慢镜头回放等，都是体育视频节目中常见的制作特征。视频对象特征属于高级语义特征，主要是指球员、裁判、禁区线和球等具有语义信息的物体对象的特征。

体育比赛拥有众多需求不同的观众群体，而不同的观众群体对视频检索也有着各自不同的要求。以足球比赛为例，一般观众的检索目的主要是查询并浏览比赛中射门或进球等精彩镜头；球迷则对自己所钟情的球队和球星等特别关注，他们不仅仅满足于精彩镜头的查询，同时还希望查询所喜欢球队的相关比赛视频，以及包含所钟爱球星的视频片段；教练和球员则出于“知己知彼，百战不殆”的目的，侧重于通过比赛视频更多地了解对手的比赛策略、战术组合、球员技术动作等。因此，基于内容的体育视频检索除了需要针对比赛类型的不同进行视觉特征的提取外，还需要考虑不同观众群的检索要求，并根据领域知识采取相应的处理分析方法。本文将以一般观众的检索要求为例，对足球比赛视频进行语义分析。

1.3 足球视频分析

足球是世界上开展最为广泛的体育运动之一。近年来，足球视频作为被较早研究的对象，已经成为基于内容的体育视频检索领域最热门的一个研究方向。每年世界上都会有数千场各类联赛、杯赛等足球比赛，每场比赛至少持续 90 分钟，如果将所有比赛存储下来，那么这个视频库将会十分的庞大，同时人们多数情况下只是希望欣赏某球队或有某球星参加的相关比赛的精彩片段，因此只存储足球比赛中的精彩片段及相关语义信息，可以在满足人们需求的同时，极大的节省存储成本。为了实现根据用户提交的简单查询要求，就能够快速提供相应比赛精彩片段的剪辑，需要建立一个高效的视频摘要生成和索引机制。目前对足球比赛视频的剪辑，只能够依赖于专业的电

视制作部门投入大量的人力物力，通过人工剪辑制作完成。因此有必要开发足球视频语义内容自动分析的相关技术，主要包括视频结构化分析，特征提取和精彩事件检测等部分，并为用户提供友好查询界面，方便用户在数据库中找到并可以浏览相应视频片段。

视频结构化分析是视频处理分析的基础，这是因为视频数据是非结构化数据，具有庞大的数据量，不宜于直接处理。视频结构化分析，是指将一个完整的视频拆分成若干小的视频片段，并在视频片段上提取关键帧。足球视频分析中，主要提取的特征包括场地颜色、场地区域比例以及场地中球员形状比例等。精彩事件是足球视频中最重要的部分，主要包括射门事件、点球事件等。精彩事件检测是指在视频结构化和特征提取的基础上，根据足球比赛的领域知识，通过对特征的静态分析和时序分析，从视频中检测出射门事件、点球事件等。一个完整的足球视频分析系统还应该包括视频数据库的管理和检索，实现对大量足球比赛视频的有效管理、检索和浏览。可以看出，足球视频语义分析不仅能够为各类视频的自动分析和理解提供参考方法，还会影响到体育新闻的制作、体育视频的管理和交互式电视点播等实际应用领域。因此进行足球比赛视频的研究，具有较高的学术价值和良好的应用前景。

1.3.1 足球视频分析研究现状

近年来，国内外研究人员在足球视频语义分析方面做了大量的研究工作，主要集中在低层视觉特征的提取、运动对象的探测与跟踪、多特征融合的视频内容分析、精彩事件的检测以及比赛剪辑的生成制作等方面。所采用的研究方法也多种多样，但本质上来说主要有三种：多特征融合的方法、基于运动特征的方法和基于镜头的方法。

Yow[7]和 Gong[8]采用对象颜色和纹理特征来检测足球比赛中的精彩事件。Intille[9]和 Tovinkere [10]通过分析对象运动轨迹实现对事件的检测。Xie[11]使用运动和颜色特征分别实现了美式足球中“暂停”和“比赛”两种状态的检测。Leonardi [12]通过镜头切变检测和摄像机运动参数计算，完成事件检测。Rui [13]通过击球声的音频特征，创建了棒球比赛视频的摘要。吴川[14]等人提取运动特征，用有限状态自动机实现跳水比赛中的事件识别。Zhong[17]等人通过识别视频中出现的比分，实现对体育视频中语义信息的提取。Ekin[15]等人提出了一个框架用来对足球视频进行语义分析，通过镜头分类以及慢镜头检测实现射门事件、裁判员和禁区的检测。Li[16]等提出了一个体育视频语义分析的两步处理框架，针对足球，通过融合慢镜头检测结果、球员特写镜头检测结果和主色比例、切变镜头检测结果等实现对足球事件的检测，然后通过检测比分字幕，并将文字识别结果与事件对应起来，形成对足球比赛视频内容的分析和描述。

很多研究机构在对足球视频进行深入研究的同时，也开发了相应的处理分析系统。下面简单介绍一下目前已有的一些系统。

中科院计算所开发了一套多媒体检索系统[48]，包含对象检测子系统、识别与跟踪子系统，三维重建与动画生成子系统，音频分类子系统，场景分析子系统和精彩片段提取子系统等。从精彩事件的检测到三维动画的重建，实现了一个较为完整的视频分析系统。从其主页上提供的演示视频来看，各个子系统都具有较好的处理结果。

夏普公司于 2004 年初推出了商用体育视频分析软件——HiMPACT[49]，能够为棒球、橄榄球、足球和相扑等四种比赛节目自动制作精华片段。据称其自动剪辑的片段节目质量并不比手工剪辑的差多少。

哥伦比亚大学新媒体技术中心开发的足球视频分析系统，可以实现特定领域场景分类，确定比赛进行和间歇两种状态，通过模板匹配方法定位声音事件，通过物体跟踪实现交互浏览，并通过对静止帧的检测完成了慢镜头的定位。

Rochester 大学的体育视频分析系统，能较好的对体育比赛视频进行物体目标和事件的检测，并且最终形成精彩镜头的视频摘要。该系统已用于 2004 年奥运会，将足球比赛视频处理后传送到用户手机终端上。

Amsterdam 大学开发了名为 Goalgle[50]的基于 WEB 应用的足球视频搜索引擎，具有树形结构框架。通过此引擎，用户可以很方便的找到如进球，黄牌，红牌警告，换人等视频片段，或者含有特殊球员的视频片段。

1.3.2 存在的问题

由于计算机和人脑对数据存储处理方式的不同，视频的高层语义与低层视觉特征之间存在着天然的巨大语义鸿沟。限于当前科技水平，这种鸿沟短期内仍会存在，足球视频也不例外。很多体育比赛具有清晰的比赛过程，呈现出非常固定和有限的几种典型场景，并在时间轴上呈现良好的周期性，如篮球比赛中的每次投篮得分后，都会经历暂停比赛，重新发球，继续比赛的过程；乒乓球比赛中的每个回合都以发球开始，以得分结束，且每回合都持续较短时间。足球比赛却在时间轴上没有这种明显的周期性，足球比赛中只有在半场结束、射门或犯规等发生后，才会出现暂停比赛，重新开球的情况，其余时间一直处在比赛状态中，因此不能使用比赛回合的周期性对其进行研究，采用事件为语义单位则具有一定的难度。目前主要存在问题有：

(1)视频分段是难点。渐变目前很难得到准确的检测，因此本文采用了基于切变检测的视频分段方法，但切变精度的提高也是一个问题所在。

(2)镜头分类算法受主观因素影响，如何确定各种类型镜头，及采用怎样的分类标准，还需要进一步研究。

(3)慢镜头检测是另一个难点，目前的算法要么过于复杂，要么效果不尽如人意。

(4)精彩事件的定义因人而异，如何对各类事件进行准确的定义，有待研究。

1.4 本文研究的主要内容

本文以足球视频为研究对象，探讨了视频的语义分析方法，主要基于足球比赛本身的特点和节目制作手段，通过镜头分类，慢镜头检测，实现对足球比赛视频中的典型事件的识别。本文具体工作如下：

(1)针对足球视频固有特征，提出了一种改进的镜头边界检测方法，即利用视频帧图像的草地颜色比例，相邻视频帧直方图的相似性来检测镜头切变点。

(2)提出一种镜头分类算法，即利用草地颜色比例，球场区域中球员所占面积、个数，形状等对镜头进行分类。

(3)通过对国内外研究人员已有成果的分析比较，本文提出了一种慢镜头检测算法，实现了慢镜头在视频流中的定位。

(4)结合足球视频的领域知识，通过对镜头类型、慢镜头等的时序分析，对射门、犯规和点球事件进行了检测。

(5)以 Windows XP、Visual C++6.0 为开发平台，实现了一个足球视频语义分析原型系统，对足球视频进行分段、慢镜头检测、镜头分类，事件检测，以及数据库的管理和检索。

1.5 论文的组织和结构

本文共包括六章，后续章节基本上是按照上述研究内容逐一展开的。第二章介绍足球比赛视频分段及特征提取，包括基于镜头切变检测的视频分段，比赛场地区域分割，镜头分类和禁区线检测等；第三章是视频后期制作特效的分析，包括慢镜头的检测和字幕区域的检测；第四章是基于精彩事件识别的足球视频语义分析，通过对视频特征的时序分析，实现射门等事件的识别；第五章设计与实现了足球视频分析原型系统，包括数据库子系统，分析处理子系统和事件查询子系统；第六章对研究内容进行了总结，并对未来的研究问题进行了展望。

2 足球比赛视频分段及特征提取

2.1 足球视频语义分析框架

足球视频不是结构化数据,足球比赛也不像其他很多体育比赛具有清晰的比赛过程，呈现非常固定和有限的几种典型场景，并在时间轴上呈现良好的周期性。因此对足球视频的分析需要借助于足球比赛领域知识，比如草地颜色为比赛画面的主要颜色，场地以中线和中圈为分界对称分为两个部分，禁区部分有白线标出，大禁区有个白色弧顶；攻防过程主要发生在中场附近，射门则在靠近禁区地方发生，点球时，主罚点球队员站在点球点附近，同时守门员站立球门下面，等等。这些领域知识是进行足球比赛视频语义分析的很好依据。

进行比赛转播时，足球视频的制作方法同样也可以作为语义分析的重要依据。长期从事足球比赛转播的电视制作人员，从最大限度方便电视观众舒适的欣赏比赛的角度出发，逐步掌握了一套科学合理的摄制、编辑模式，并将其应用到足球比赛的转播节目中。这种典型的编辑模式同时也提供了丰富的语义信息。比如由于足球场地相对较大，为了让观众看清楚足球的运动路线，比赛双方的攻防形势等，摄像机往往选择离球场较远的角度拍摄，这样以比赛场地为背景的镜头就占据了整个视频的大部分时间。当发生诸如犯规、拼抢、角球、任意球、点球及射门等事件时候，镜头往往会集中在一个或几个队员身上。在发生犯规或射门等事件后，通常会有一段慢镜头重放镜头，以方便观众看清楚犯规的过程或再次感受射门的激动时刻。

本文提出了一个基于领域知识和视频编辑手法的足球视频语义分析算法。下面将首先对足球视频的语义结构进行分析，主要包括基本语义单元及其之间关系的分析。然后，介绍了语义分析过程的具体框架。

2.1.1 足球视频语义结构分析

2.1.1.1 基本语义单元

基本语义单元是体育视频中围绕一个特殊主题组织的数据集合，是表征体育视频语义内容的基本单元。基本语义单元语义倾向性较强，是周期性或者半周期性发生的能够吸引观众兴趣的视频内容单元。根据应用的不同，体育视频的基本语义单元可以有很多种。本文针对足球视频将基本语义单元做如下分类：

(1)镜头

镜头是进行视频语义分析的基本语义单元，所有根据足球视频摄制手法和编辑规则对足球视频进行的处理分析都是以镜头为基本处理单元的。镜头基本语义单元定义如图 2-1 所示：

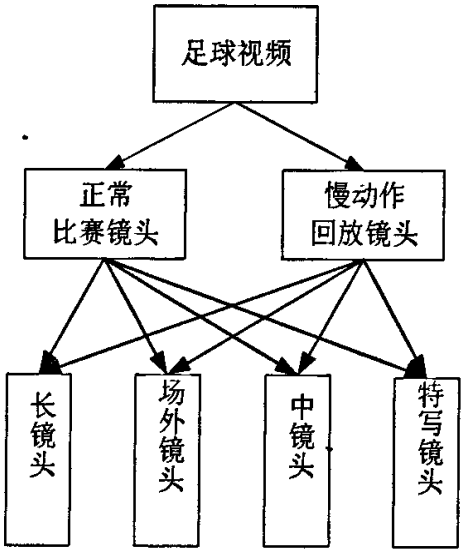


图 2-1 镜头基本语义单元

足球视频由正常速度播放的正常比赛镜头和慢速回放的慢镜头两部分组成。比赛中，当出现精彩场面或观众感兴趣的片段之后，通常会同时出现从多个不同的角度对精彩片段进行回放和慢放的镜头，称之为慢镜头。因此慢镜头本身包含着很强烈的语义信息，本文将其作为非常重要的基本语义单元。由于无论是否慢镜头都是由现场的摄像机在特定位置和角度捕捉采集的，因此根据拍摄角度与距离的不同，又可将镜头分为长镜头、中镜头、特写镜头和场外镜头四类。每一类镜头都是一个基本语义单元。

长镜头是指摄像机在远离赛场的位置拍摄到的，所展现的是整个或者半个球场的态势，给观众以全局的感觉，也是足球比赛中最常见的镜头类型。

中镜头是指赛场某一特定部分的聚焦镜头，通常会有一个人或几个人，并能够清晰看到球员的整个身体部分。

特写镜头通常只显示球员上半身，一般用来展现比赛中发挥出色的球员或者犯规球员。

场外镜头是指比赛区域以外镜头，主要表现的是观众、教练或其它情况。

(2)视频对象

足球视频中的足球、球员、裁判、边线、禁区线、球门、字幕等一些对象也具有丰富的语义，可以为视频语义分析提供有用的信息。例如射门事件和禁区线、球门以及字幕息息相关的；犯规事件发生后也可能出现裁判镜头和字幕。

一些对象如禁区线、边线、球门、字幕等，在一定的场景中是固定不变的，本文主要研究字幕和禁区线两类视频对象，并提出了禁区线和字幕的检测识别算法。

足球视频中的字幕包括台标、时间的显示、比分牌、红黄牌、队员介绍、球队技术统计和阵形等。我们主要关注与事件有关的字幕，如比分牌和红黄牌。

禁区线是禁区的边界，所以镜头中出现禁区线说明比赛的焦点在禁区附近。这时候很可能发生的就是一次进攻。射门事件和点球事件发生时，同样会有包含禁区线的镜头出现。因此，将禁区线识别出来能够很好的辅助精彩事件检测。

(3)精彩事件

足球比赛是非结构化数据，具有很大的数据量。视频数据虽然是对足球比赛很好的描述，但是离人们所熟悉的理解方式仍有很大的距离，而事件既可以实现对足球比赛的描述也较为符合人们的理解方式。足球比赛中有很多事件，如射门、犯规、点球、越位、前场任意球、换人等。本文以射门事件、犯规事件和点球事件为例对足球视频进行语义分析。

事件基本语义单元通常是镜头基本语义单元和对象基本语义单元按照一定的融合关系组成的复合基本语义单元。如射门事件和点球事件是由禁区线、特写镜头、场外镜头、慢镜头和字幕等按照一定的融合关系组成的基本语义单元。犯规事件是由裁判、中镜头、慢镜头和字幕等组成的基本语义单元。类似的可以定义其他事件。

以上将足球视频按照摄制手法、编辑方式、领域知识等作的基本语义单元分类。从上面的分类中可以看出各个类别之间存在各种关系，接下来就讨论以下它们之间的关系。

2.1.1.2 基本语义单元之间关系

由于视频数据含有丰富信息，各个基本语义单元之间的关系也错综复杂，没有严格的定义。本文从视频数据的特点出发，主要考虑以下两种基本语义单元之间的关系：

(1)融合关系

融合关系可以是低层次的，如象素级、特征级的融合；也可以是高层次的，如事件之间的融合。由于视频数据由图像序列、伴随音频和文字等构成，因此视频本身就是多种信息融合在一起的数据，基本语义单元之间也就自然存在一种融合关系，如前所述，进球事件就融合了许多低层的基本语义单元。

(2)时序关系

视频数据的另一个关键特征是时间属性。各类媒体数据在时间轴上的相互关系称之为时序关系。因此基本语义单元之间的时序关系对于视频的语义分析同样具有重要作用。如射门事件发生前会出现球门区域的长镜头，在射门发生之后通常会有一个完

成射门动作的球员的特写镜头，紧跟着会有慢动作回放整个射门过程的慢镜头，如果射门成功，还会出现观众欢呼的镜头或教练欢呼的镜头。犯规事件发生后会有一个较短的慢镜头，方便观众看清楚犯规事件是如何发生的，责任在谁，如果犯规程度比较严重，发生红黄牌，那么还会出现裁判出示红黄牌的中镜头，并在屏幕底部出现字幕，告诉观众是哪位球员犯规及得到红牌还是黄牌。其他事件各个基本语义单元之间的时序关系通过分析也可以得到。

2.1.2 语义分析框架

现有足球视频语义信息提取算法，都是以事件检测为基础的，同时遵循这样的分析步骤：首先提取低层视频特征如颜色、纹理、对象运动轨迹、摄像机运动参数等；然后利用这些特征生成语义片段；最后通过时序序列分析、有限状态自动机等方法分析语义片段，从而实现事件的检测和视频内容语义的提取。

由于摄像机运动参数和对象运动轨迹算法过于复杂，且准确率较低，本文选取视频帧主色、球场区域分割结果等作为低层视频特征。通过主色的分析和球场区域中非主色块的几何形态分析，将视频帧分为三类：以足球场地为背景的远视角镜头、运动员特写镜头和场外镜头。结合对视频中慢镜头检测的结果，通过时序序列分析，实现精彩事件的定位。另外，视频中的叠加文本包含了运动员信息、比分信息等，检测这些文本信息有助于对其内容的理解和索引，本文提出了一种基于小波变换的K均值聚类字幕区域分割算法，实现视频中字幕区域的定位。本文将采用图2-2所示框架来进行足球视频语义分析。

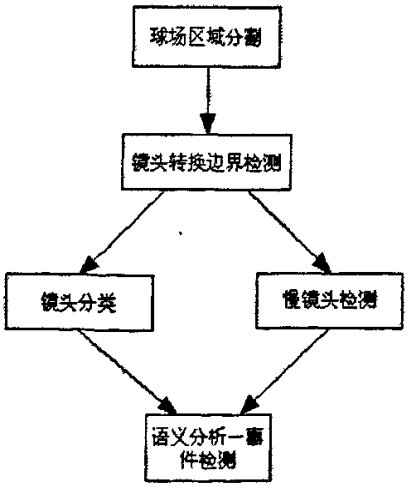


图2-2 足球视频分析框架

2.2 基于镜头切变检测的视频分段

视频中相邻的图像帧往往具有很高的相似性。通常将由若干时间连续、内容相似的图像帧构成的序列，作为一个基本物理单元，称为镜头。面对数据量巨大的足球视频，如果逐帧处理将会耗费很多时间，因此一般以镜头为基本单元对视频进行切分，之后在每个镜头中选取关键帧，作为镜头的描述。特征提取、运动分析以及检索等工作都是在视频分段后得到的关键帧上进行的。这不但提高了处理速度，降低了存储空间，还不会造成信息的大量丢失，只要提高镜头边界检测的准确度，就不至于对分析结果造成可察觉的影响。由此可见，准确的镜头边界的检测至关重要。镜头边界检测问题的研究已经持续了很多年，并出现了很多切变镜头检测算法，但由于镜头内还可能包含其他类型的变化，例如摄像机或视频对象的运动、光照的变化以及其他的噪声等，这些都会对镜头边界检测的准确率造成一定影响，因此在具体视频分析中还需要针对其特点决定采用哪一种检测算法。

2.2.1 镜头转换类型及传统检测算法

镜头转换方式可以分为两类：切变和渐变。切变是指一个镜头直接切换到下一个镜头，中间没有时间上的延迟。在发生切变的两个镜头边界处，视频帧的图像特征往往存在突变，因此，通过比较这些图像特征（如直方图比较、象素点差分、边缘比较、运动信息等）的差异，可以比较容易地检测出切变。渐变通常也称之为特效剪辑，与切变不同，它是指在视频编辑过程中加入一些空间或时间上的编辑效果，使前一个镜头渐渐转换为下一个镜头，是一种镜头间的平滑过渡，不存在明显的镜头边界，因而难以检测。足球视频中，镜头转换的方式以切变为主，渐变则往往出现于慢镜头与正常比赛切换之间，因此通过切变检测足以实现视频分段工作，本文将只考虑镜头切变检测。下面我们简单介绍一下切变镜头检测的传统算法。

(1) 基于直方图相似性比较的方法

相邻帧图像间的差异定义为：

$$D(I_1,I_2)=\frac{\sum_{i=0}^{K-1}\min(H[I(x,y,t),i],H[I(x,y,t+1),i])}{\sum_{i=0}^{K-1}H[I(x,y,t+1),i]} \tag{2-1}$$

式中：$H(\bullet)$是图像的直方图；K是图像灰度级数。

当帧差超过阈值T_2时，则认为发生一个切变，该阈值T_2可以自动确定为

$$T_2=m+s\times\sigma$$

式中：m和σ分别是帧差值的均值和标准差；s是一个较小的数。

该类算法基于这样一个前提：相同背景下相同目标的两帧图像在直方图上的差异应很小。由于直方图体现的是图像总体的灰度分布，因此对一般的运动和噪声不敏感，实际中有比较广泛的应用。同样地，不同目标的场景可能存在近似的灰度或颜色分布，所以该方法存在漏检问题。

(2) 基于灰度/颜色模板匹配的方法

这类方法直接通过两帧图像之间灰度或颜色的差值来检测镜头分割。I_1，I_2 两帧图像的相似度定义为：

$$D(I_1,I_2,i,j)=\begin{cases}|P(I_1,i,j)-P(I_2,i,j)| & \text{灰度图像}\\ \sum_{n=1}^{3}|P(I_1,C_n,i,j)-P(I_2,C_n,i,j)| & \text{彩色图像}\end{cases} \tag{2-2}$$

其中 $P(I,i,j)$ 是 (i,j) 点象素的灰度值，$P(I,C_n i,j)$ 是点 (i,j) 的颜色分量（如 RGB 等），然后我们再在局部或整幅图像上用公式(2-3)求和。

$$S(I_1,I_2)=\sum_{i=1}^{X}\sum_{j=1}^{Y}D(I_1,I_2,i,j) \tag{2-3}$$

当 $S(I_1,I_2)$ 大于某一门限时，就认为检测到一个镜头切换。该方法的最大问题是对摄像机和物体的运动比较敏感，因此当运动较剧烈时，相邻两帧的差异往往会超过预定的阈值，从而造成误检。

(3) 基于块匹配似然比的方法

这种方法将帧图像分割为若干子图像，通过如下方法计算图像差异。

$$D(I_1,I_2)=\sum_{i}\left\{\frac{\sigma(I_1,i)+\sigma(I_2,i)}{2}+\left[\frac{m(I_1,i)+m(I_2,i)}{2}\right]^2\right\}^2 \tag{2-4}$$

其中 $m(I,i)$，$\sigma(I,i)$ 分别是图像第 i 个子图像灰度的均值和方差。

该方法缺点是计算量较大，且查全率和精确率并没有明显提高。

2.2.2 一种适于足球视频的切变检测算法

足球比赛视频较一般视频有其许多特殊性，如何有效检测足球视频镜头边界仍然是一个充满挑战的课题，究其原因主要有以下几点：

(1) 与一般视频不同，在足球视频中，由于摄像机大部分时候以比赛场地为拍摄中心，而比赛场地往往有着单一的颜色(足球场地一般会呈现绿色)。这样就造成了镜头间具有强烈的颜色相关性，结果是在相继两镜头难以出现明显的颜色差异，其颜色直方图也不会有明显的变化。传统的基于帧间直方图差的镜头边界检测方法此时就难以奏效。

(2) 足球视频中存在大量摄像机运动和对象运动，如摄像机扫视和变焦广泛用于

跟踪和聚焦运动对象，结果是原有依赖检测统计变化的算法就得不到很好的效果。

为有效可靠地检测到足球视频的镜头边界，本文采用文献[15]提出的镜头边界检测算法。该算法用帧间主色比例差和颜色直方图相似性差来刻画镜头的转换：

帧间主色象素比例差G_d：第 i 帧和第 $i-k$ 帧之间帧间主色象素比例差定义为

$$G_d(i,k)=|G_i-G_{i-k}| \tag{2-5}$$

其中G_i表示第 i 帧主色象素占总象素的比例。

颜色直方图相似性差H_d：第 i 帧和第 $i-1$ 帧的颜色直方图相似性差定义为

$$H_d(i,1)=|HI(i,1)-HI(i-1,1)| \tag{2-6}$$

其中，两直方图的相似性由直方图相交法来度量。

$$HI(i,1)=\frac{1}{3}\sum_{m=1}^{3}\sum_{j=0}^{B_m-1}\min\left(H_i^m(j),H_{i-1}^m(j)\right) \tag{2-7}$$

其中B_m表示颜色分量 m 的颜色区间数，H_i^m为第 i 帧归一化的颜色分量 m 的颜色直方图。

镜头边界由H_d和G_d与一组阈值相比较来确定。当主色象素比例很低，镜头特性与普通镜头相似，在这种情况下问题就是普通镜头检测，因此可以只用一个较高的H_d阈值。对于主色象素达到一定比例的情况，同时用H_d和G_d，但要用较低的H_d。为此定义四个阈值以检测镜头边界：$T_{H_d}^{high}$、$T_{H_d}^{low}$、T_{G_d}、T_G，其中$T_{H_d}^{high}$，$T_{H_d}^{low}$为直方图相似性H_d阈值，分别用于非场地镜头和场地镜头边界的检测；T_{G_d}为帧间主色象素比例差G_d阈值，用于场地镜头的检测；T_{G_d}为主色象素比例 G 阈值，用于判别场地镜头和非场地镜头，即若G_j小于T_{G_d}则用$T_{H_d}^{high}$进行镜头检测，否则用$T_{H_d}^{low}$，T_{G_d}进行镜头检测。上述阈值的取值通过实验确定。镜头边界检测算法流程如图 2-3 所示。

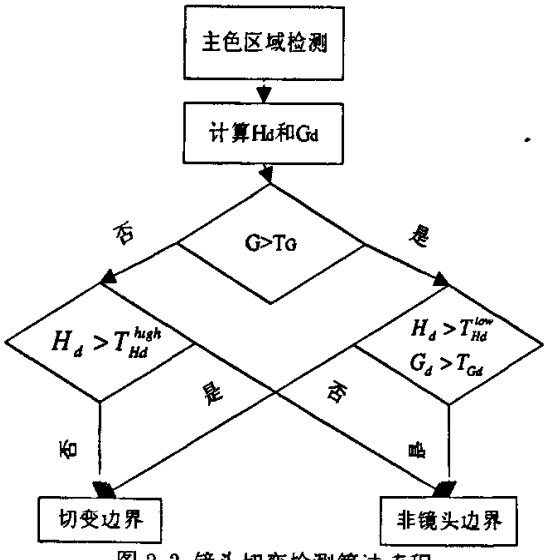


图 2-3 镜头切变检测算法流程

2.2.3 实验结果分析

本文选取的实验素材是 4 个半场足球视频，每个长约 45 分钟，分辨率为 704×576。采用查全率和准确率作为算法性能评价标准，查全率、准确率分别定义为：

$$查全率 = 正确检测数/(正确检测数 + 漏检数) \quad (2\text{-}8)$$

$$准确率 = 正确检测数/(正确检测数 + 误检数) \quad (2\text{-}9)$$

为了检验本文提出算法的检测效果，选取了传统切变检测算法中的基于直方图相似性比较的方法同时进行切变检测。检测结果如表 2-1 所示，其中应检数为人工目测结果。原型系统处理切变镜头检测效果如图 2-4 所示，左窗口显示正在处理的足球视频，右窗口是镜头切变点的图象帧，切变镜头检测结果在下方输出的同时写入数据库中。从表 2-1 中可以看出，传统切变检测算法检测足球视频中的切变镜头存在较多的漏检情况，而本文提出的算法具有较高的查全率，漏检情况减少很多，同时准确率和传统切变检测算法接近，总的说来，效果令人满意。

表 2-1 镜头切变检测实验结果

	应检数	实检数	误检数	漏检数	查全率	准确率
直方图相似性算法	1018	947	105	176	82.7%	88.9%
本文算法	1018	1092	139	65	93.6%	87.3%

图 2-4 原型系统中镜头切变检测效果图

2.3 比赛场地区域分割

2.3.1 HSI 颜色空间

一般的颜色空间如 RGB 颜色空间等，都不能模仿人眼的特点对颜色进行比较准确的描述。人们通常不会用象素各个颜色分量的值来描述一个物体的颜色，都倾向于通过其色调，饱和度和亮度加以描述。色调描述颜色所在的波段，如黄、橙、红；饱和度描述纯颜色被白色稀释的程度；亮度是一个主观描述子，实际很难度量，具体为亮度的无色感知，是描述颜色感觉的关键描述子，在描述单色图象时亮度(灰度级)非常有效。HSI 颜色空间(色调 Hue，饱和度 Saturation 和亮度 Intensity)将亮度分量和颜色信息(色调和饱和度)从彩色图像中分离出来，用亮度、色调和饱和度三个分量对物体颜色的进行描述，这种描述比较符合人们的习惯，对于人眼来说，HSI 颜色模型很自然，很直观。因此，HSI 模型是开发彩色图象处理算法的很理想的模型，基于 HSI 模型开发的算法，能够很好的模拟人眼处理彩色图象的效果。

HSI 模型的坐标系统接近圆柱坐标系统，如图 2-5 所示。对其中任一个色点 P，其 H 的值对应指向该点的向量与 R 轴的夹角。这个点的 S 与指向该点的向量长成正比，越长越饱和。在这个模型中，I 的值与该点所在平面与最下对应黑色点的距离成正比。如果色点在 I 轴上，则其 S 值为 0 而 H 没有定义，这些点也称为奇异点。奇异点的存在是 HSI 模型的一个缺点，而且在奇异点附件，R，G，B 值的微小变化会引起 H，S，I 值的明显变化。在 RGB 空间和 HSI 空间转换时要特别注意奇异点的存在。

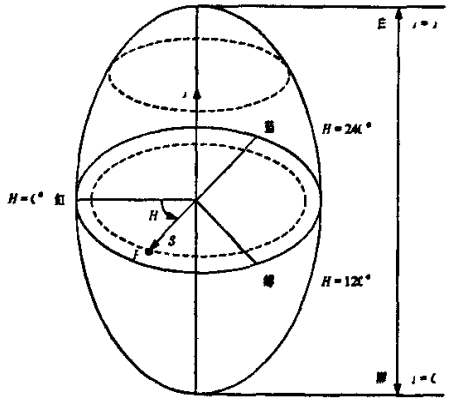


图 2-5 HIS 颜色模型

在 RGB 空间的彩色图像可以方便的转换到 HSI 空间。对任何 3 个归一化到[0，1]范围内的 R，G，B 值，其对应的 HSI 模型中的 H，S，I 分量可由下面的公式来计算：

$$H=\begin{cases}\arccos\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & R\neq G\text{或}R\neq B\\ 2\pi-\arccos\dfrac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}} & B>G\end{cases} \tag{2-10}$$

$$S=1-\frac{3}{(R+G+B)}\min(R,G,B) \tag{2-11}$$

$$I=(R+G+B)/3 \tag{2-12}$$

注意由式(2-10)直接计算出的H在[0°, 360°]之间，为使H在[0, 1]之间，再令$H=H/360°$。另外当$S=0$时，对应无色，这时H没有意义，此时定义H为0。另外当$I=0$或$I=1$时，讨论S也没有意义。

另一方面，如果已知HSI空间色点的H，S，I分量，也可将其转换到RGB空间。

2.3.2 主色提取

主色是指在一幅图像中占主要地位的颜色。足球场地以绿色为主，找到了帧图像中所有的绿色象素点也就实现了足球场地的分割。由于每场比赛场地颜色都不相同，而且比赛过程中随着光线、气候等外部条件的变化，场地颜色也会随之改变，因此进行场地分割前需要提取主色，并且在视频处理中要不断更新主色。

主色提取是场地分割第一步。由于事先未知哪幅图象帧中场地颜色占主要地位，因此需要随机抽取视频中的帧，对其进行边缘复杂度分析。本文选取Canny算子提取图像边缘，然后把帧图像分成N×N的小块，统计小块中的边缘点。由于场地区域相对较光滑，取适当阈值T_1，统计帧图像中边缘象素小于阈值的子块个数，计算其比例系数R，共抽取系数R大于阈值T_2的帧共计50帧，如此可以保证所抽取的帧可能含有一定比例的场地，以提高主色提取的可靠性。计算所有50帧的HSI颜色直方图，找出峰值颜色i_{peak}。由于场地颜色会随着场馆、天气、灯光等的不同而有所改变，直接用峰值来表示主色不是很准确，为此提出用包含峰值颜色的一定区间的颜色的均值来刻画场地主色，以保证可靠性和准确性。主色提取的计算公式如下：

$$Hist(i_{min})\geq K*Hist(i_{peak}) \tag{2-13}$$

$$Hist(i_{min}-1)<K*Hist(i_{peak}) \tag{2-14}$$

$$Hist(i_{max})\geq K*Hist(i_{peak}) \tag{2-15}$$

$$Hist(i_{min}+1)<K*Hist(i_{peak}))\tag{2-16}$$

$$mean=\frac{\sum_{i=i_{min}}^{i=i_{max}}Hist(i)*i}{\sum_{i=i_{min}}^{i=i_{max}}Hist(i)} \tag{2-17}$$

其中$Hist(\bullet)$为颜色直方图，颜色区间的上下界$[i_{\min}, i_{\max}]$由式(2-13)到(2-16)确定，K取 0.2，主色由式(2-17)计算，即主色定义为峰值i_{peak}左右颜色直方图下降到K倍$Hist(i_{peak})$的区间范围内所有颜色的均值，如图 2-6 所示。

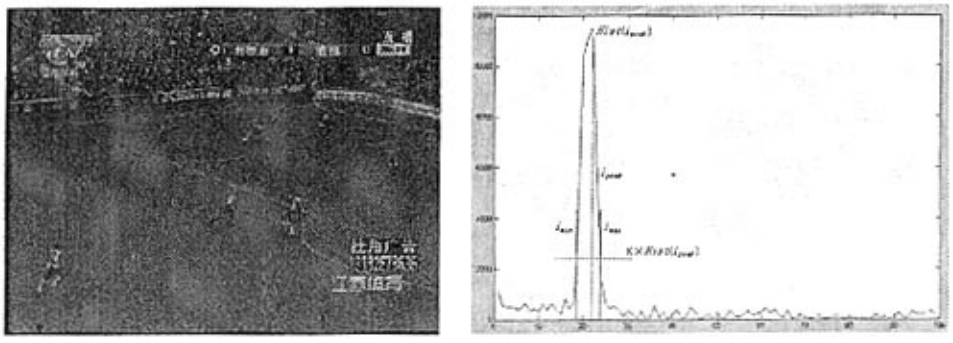


图 2-6 图像及其主色计算过程的直方图

2.3.3 基于圆柱距离的场地区域分割

确定了主色后，为提取场地区域需要对帧内象素进行是否为主色象素的判断，在此采用文献[15]提出的圆柱距离准则。具体计算公式如下：

$$d_{\text{intensity}}(j) = \left| I_j - I_{mean} \right| \tag{2-18}$$

$$d_{chroma}(j) = \sqrt{(s_j)^2 + (s_{mean})^2 - 2 s_j s_{mean} \cos(\theta(j))} \tag{2-19}$$

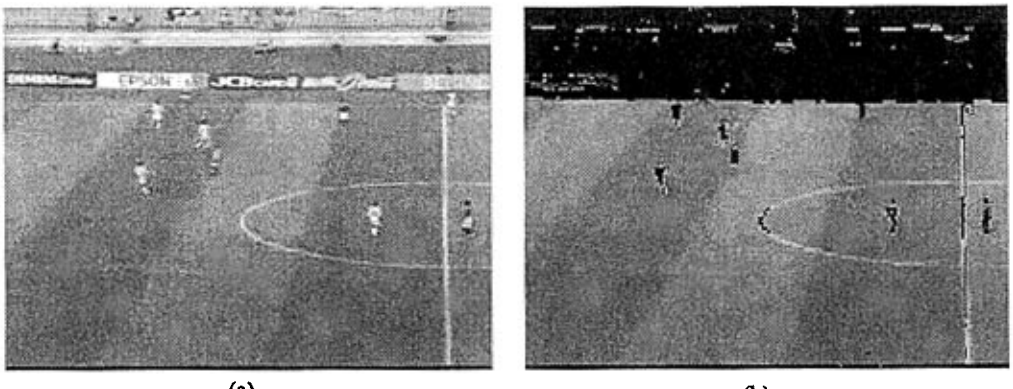
$$d_{cylindrical}(j) = \sqrt{(d_{\text{intensity}}(j))^2 + (d_{chroma}(j))^2} \tag{2-20}$$

$$\theta(j) = \begin{cases} \Delta(j) \ , if\ \Delta(j) \le 180^o \\ 360^o - \Delta(j) \ , otherwise \end{cases} \tag{2-21}$$

$$\Delta(j) = \left| Hue_{mean} - Hue_j \right| \tag{2-22}$$

$$d_{cylindrical} < T_{color} \tag{2-23}$$

对于无色象素直接用公式(2-18)计算象素到主色的距离，对于有色象素，用式



(a)　　(b)

图 2-7 原始图象及其圆柱距离分割结果

(2-19)到(2-20)进行圆柱距离计算，其中 Hue 为色调、S 为色饱和度、I 为强度。如果某一象素的圆柱距离小于阈值T_{color}，则认为其为主色象素。场地区域分割结果如图 2-7 所示。

2.3.4 场地区域分割后处理

由于足球场地外部区域里面可能包含具有主色象素的点，同样足球场地内部可能包含不具有主色象素的点，因此经过基于圆柱距离的场地区域分割之后得到的结果，在场地外部区域会出现很多孤立噪声点，在场地内部区域会出现很多空洞。为了得到更为准确的场地区域，需要对场地区域分割结果进行处理。处理步骤如下：

Step1. 对二值图象进一步做连通性分析，找到最大的一块连通区域，作为球场区域，并去掉其余连通区域。

Step2. 对得到的分割结果二值图象进行膨胀腐蚀的形态学处理，以消除毛刺、填补空洞洞，从而场地区域完整的掩膜图像。

Step3. 将其与原图像进行掩模处理，可以得到最终的场地分割结果。

2.3.4.1 形态学处理

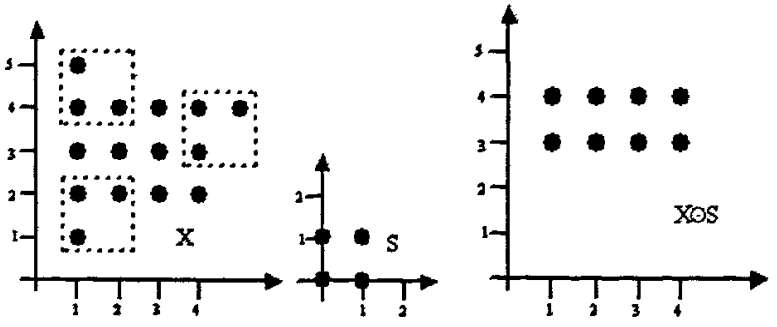
(1)图像腐蚀(Erosion)

腐蚀在数学形态学中的作用是消除物体边界的毛刺(边界点)。如果结构元素采用 3×3 的黑点块，腐蚀将使物体边界沿周长减少一个象素点。

对于一个给定的目标图像X和一个结构元素S，腐蚀运算定义为：

$$X \ominus S = \{x | S[x] \subseteq X\} \tag{2-24}$$

即当结构S在图像上移动时，结构覆盖下的相应元素点都在图像X中。腐蚀运算的具体进行过程如下图 2-8 所示。



(a) 图像 X 和结构元素 S　　(b) X⊖S

图 2-8　X 被 S 腐蚀的几何解析

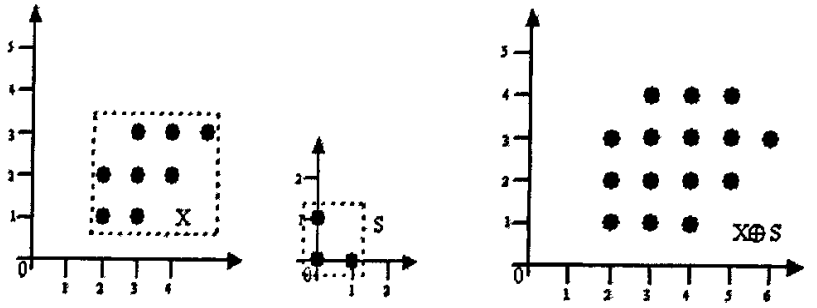
用图中的S对X进行腐蚀，由于在尖角处都只有三个点，不能与S重合，因此经腐蚀后的图像消去了这些突出部分的点。同样地，对同一图像X，结构元素S不同时，腐蚀结果也将不同。

(2)图像膨胀(Delation)

腐蚀可以看作是将图像X中的每一个与结构元素S全等的子集$S[x]$收缩为点x。那么反之，也可以将X中的每一个点x扩大为$S[x]$。这就是膨胀运算，记为$X \oplus S$，定义为：

$$X \oplus S = \{x | S[x] \cap x \neq \phi\} \tag{2-25}$$

膨胀运算的具体过程如下图2-9所示。



(a) 图像X和结构元素S　　(b) X⊕S

图2-9　图像X被结构元素S膨胀几何解析

由图可见，X被图中的结构元素S膨胀相当于在原有的X图像基础上向右方和上方各扩充了一个单位。同样地，对同一幅图像X，如果改变结构元素S的形状，则X被S膨胀就会得到不同的结果。

(3)形态学处理

选取合适的结构算子对于膨胀腐蚀的结果又一定影响。经过实验，本文选择结构元素为：

$$SE = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

首先，考虑到有的噪声点比较大，对场地区域分割结果二值图象连续进行三次腐蚀处理。为了让场地区域仍能近似保持原有尺寸，还需要进行同等次数的膨胀处理。这样就可以去掉球场区域外部的噪声点。

然后，在此结果上，对其连续进行三次膨胀处理。同样为了让场地区域仍能近似保持原有尺寸，还需要进行同等次数的腐蚀处理。这样就可以填补球场区域内部的空洞。

2.3.4.2 连通性分析

形态学处理只能去掉面积较小的噪声块，因此经过形态学处理后，在球场区域外部还可能存在较大的噪声块。这时就需要对其进行连通性分析，找到最大的一块连通区域作为球场最终区域。下面介绍基于象素标记的连通区域标记算法。

假设对一幅二值图象从左向右、从上向下进行扫描（起点在图象的左上方）。要标记当前正被扫描的象素需要检查它与在它之前扫描到的若干个近邻象素的连通性。例如当前正被扫描象素的灰度值为 1，则被它标记为与之相连通的目标象素。如果它与 2 个或多个目标相连接，则可以认为这些目标实际是同一个目标，并把它们连接起来。如果发现了从为 0 的象素到 1 个孤立的为 1 的象素的过渡，就赋 1 个新的目标标记。

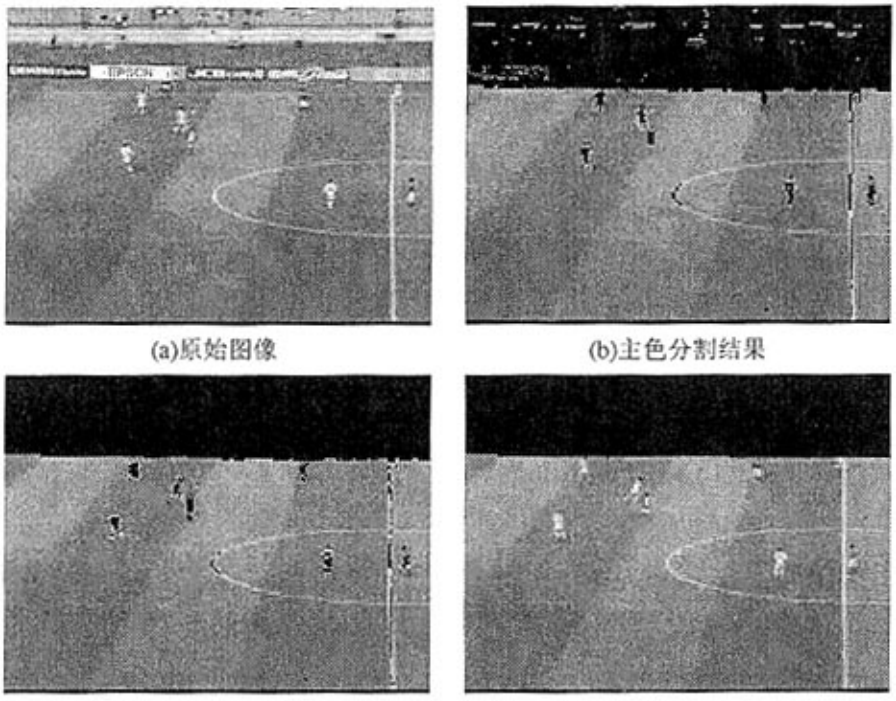
4-连通的情况，根据以上建立的概念我们可如下进行标记。假如当前象素的值是 0，就移到下一个扫描位置。加入当前象素的值是 1，检查它左边和上边的两个近邻象素（根据所用的扫描次序，当我们到达当前象素时这两个近邻象素已被处理过了）。如果它们都是 0，就给当前象素一个新的标记（根据已有信息，直到目前这是该连通区域第 1 次被扫描到）。如果上述两个近邻元素只有一个值为 1，就把该象素的标记赋给当前象素。如果它们的值都为 1 且具有相同的标记，就将该标记赋给当前象素。如果它们的值都为 1 但具有不同的标记，就将其中的一个标记赋给当前象素并做个记号表明这两个标记等价（两个近邻象素通过当前象素而连通）。在扫描终结时所有值为 1 的点都已标记但有些标记是等价的。我们所需要做的就是将所有等价的标记对归入等价组，对各个组赋一个不同的标记。然后第二次扫描图像，将每个标记用它所在等价组的标记代替。

为了给 8-连通的区域标记，我们可采用相同的方式，只是不仅对当前象素左边和上边的两个近邻象素，而且对两个上对角的近邻象素也要检查(同样，所用的扫描次序保证当我们到达当前象素时，这四个象素已被处理过了)。假如当前象素的值是 0，就移到下一个扫描位置。假如当前象素的值是 1 并且上述四个相邻象素都是 0，给当前象素赋一个新的标记。如果只有一个相邻象素为 1，就把该象素的标记赋给当前象素。如果两个或多个相邻象素为 1，就将其中一个标记赋给当前象素并做个记号表明它们等价。在扫描结束后将所有等价的标记归入等价组，对每个组赋一个唯一的标记。然后第二次扫描图像，将每个标记用它所在等价组的标记代替。

2.3.5 场地分割结果

场地分割结果如图 2-10 所示。图 2-10(a)是在一段足球视频中随机选取的具有

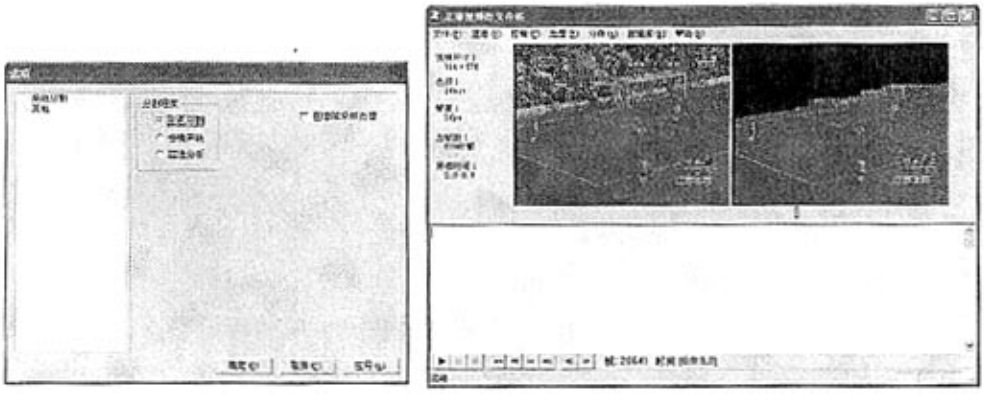
含有场地的图象帧；将图像从 RGB 空间映射到 HSI 空间后，计算所有象素点与事先计算好的主色点的圆柱距离，得到场地粗分结果如图 2-10(b)所示，可以看出非场地区域存在较多噪声块，场地内部存在较多球员造成的空洞；通过连通性分析，去除最大连通区域以外的区域，得到没有噪声块的处理结果，如图 2-10(c)所示；通过多次膨胀和腐蚀运算，可以填补场地内的空洞，与原始图像掩膜后得到最终场地区域分割结果，如图 2-10(d)所示。原型系统中场地分割参数设置和运行效果如图 2-11 所示。



(a)原始图像　(b)主色分割结果

(c)去除噪声块　(d)最终结果

图 2-10 场地分割结果



（a）参数设置　（b）场地分割效果图

图 2-11 原型系统中场地分割参数设置及效果图

2.4 镜头分类

镜头类型是拍摄特征的一种，在视频语义分析中通常可以传递重要的语义线索，结合其它特征和相关知识可以用来进行高层语义分析。比如进球事件通常伴随着特定的拍摄特征模式，如特定的镜头类型，所以对镜头类型的分析有助于进球事件的识别。

镜头的类型可以通过空间或者时空特征判断出来，比如使用颜色、形状和纹理等特征。文献[20]用两个简单的草地比例阈值把足球视频镜头分为三类：长镜头、中镜头和特写镜头，方法直观，但由于算法过于简单，所以分类效果一般；文献[42]提出了一种基于颜色模板的分类算法，将每帧的颜色特征与事先建立的各种镜头颜色模板相比较，最终得到镜头的类型。由于涉及到模板的建立，则算法鲁棒性不好，针对不同的比赛都需要重新建立镜头模板。文献[43]提出将颜色，纹理，形状信息相结合的算法来检测足球比赛长镜头。文献[15]提出了一种结合黄金分割和贝叶斯信息分类准则的足球视频镜头分类算法，算法准确率为 89%，但算法相对复杂。本文根据足球视频制作的特点提出一种新的镜头分类算法。

2.4.1 镜头类型

以摄像机拍摄时取景的内容和范围把足球镜头分为四类，如图 2-12 所示：

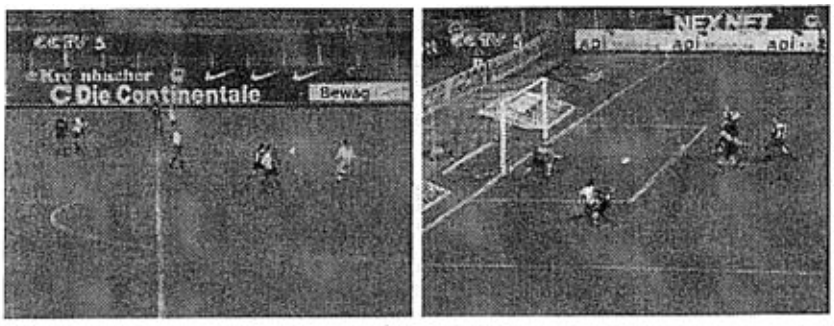
（1）长镜头：长镜头是摄像机在离场地很远的地方拍摄的，可以显示整个场地景象，用于表现比赛情况全局的镜头。长镜头可用来进一步作比赛事件的分析；

（2）中镜头：中镜头一般是比赛场景某一特定部分的聚焦镜头，通常出现一个或几个人，且可以清晰看到球员的整个身体；

（3）特写镜头：特写镜头是指通常只显示球员上半身的镜头；

（4）场外镜头：场外镜头是指非比赛场地区域的镜头，主要表现观众、教练或其它情况。

本文在每个镜头中抽取关键帧，通过对关键帧颜色、形状和空间特征的分析，将整段镜头分为上述四类。



(a) 长镜头

(b) 中镜头

(c) 特写镜头

(d) 观众镜头

图 2-12 四种镜头类型

2.4.2 特征选取

从图 2-12 中可以看出，长镜头的场景含有大块的场地，具有很高的主色比例，且由于视点较远，场地上的人员较小；中镜头中的场景含有一定比例的场地，由于是聚焦镜头，其中的人员较长镜头中的人员要大的多；特写镜头和观众镜头的最大特点是场地比例很小甚至没有，并且特写镜头的背景比较模糊、边缘不清楚，而观众镜头的背景相对较复杂，边缘信息丰富。因此，主色比例可以作为镜头类型分类的一个特征。但中镜头也可能会出现主色比例很高的情况，如果只用主色比例，只能准确的将镜头分为两类，一类是长镜头、中镜头，另一类是特写镜头和场外镜头。所以我们提出了一种两步镜头类型分类算法，首先利用主色比例将镜头分为上述两类，然后通过对场地区域内人员的形状和面积等几何特征的分析，将包含场地的镜头分为长镜头和中镜头，通过对图象的边缘复杂度的分析将含少量场地或不含场地的镜头分为特写镜头和场外镜头。

2.4.3 基于规则的镜头分类算法

据此提出一种镜头分类算法，该算法结合了一些经验知识，具体流程如图 2-13 所示。算法步骤如下：

Step1. 对足球比赛进行场地主色提取、场地分割，然后根据场地区域比例将场景分为场地场景和非场地场景。

Step2. 对于场地场景，首先用场地主色和形态学求场地区域中的人员连通区，并根据连通区外接矩形面积确定最大连通区；然后以最大连通区的长、宽、面积、长宽比例等为约束条件，把帧图像分为长镜头和中镜头。

Step3. 对于场外场景，首先对帧图像用 Canny 算子提取边缘信息；然后把它分成 16×16 的小块，统计每个小块中的边缘象素数量作为其复杂性的度量；最后统计复杂度大于一定阈值的小块的数量，作为帧图像复杂性的度量，把复杂性大于一定阈值的帧图像判为观众镜头，否则判为特写镜头。

Step4. 上面算法中用图像颜色和空间特征结合先验知识对帧图像进行了分类，由于算法计算简单，为提高分类的可靠性对镜头中的每一帧进行分类，然后根据少数服从多数原则确定最终的镜头类别。

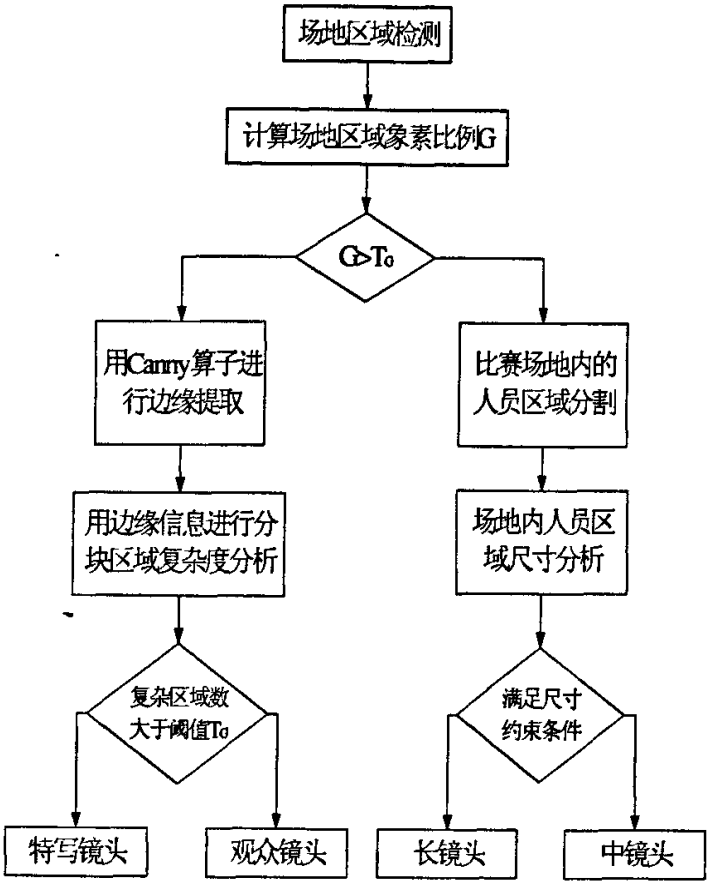


图 2-13 镜头分类流程图

2.4.4 实验结果分析

为了提高处理速度，在每个视频片段中选取关键帧，将其按照镜头类型不同分类，结果作为其所在视频片段的镜头类型。四种镜头类型的查全率和准确率分别按照式(2-8)和式(2-9)定义，分类结果如表 2-2 所示。其中特写镜头的查全率比较低，究其原因主要有两点：(1)不是所有特写镜头都是非场地镜头，结果有的特写镜头被误判为中镜头；(2)有的特写镜头其复杂度也较高，结果被误判为观众镜头。

表 2-2 镜头分类实验结果

	应检数	实检数	误检数	漏检数	查全率	准确率
长镜头	607	618	21	10	98.4%	96.6%
中镜头	236	225	20	31	86.9%	91.1%
特写镜头	102	79	3	26	74.5%	96.2%
观众镜头	73	80	11	4	94.5%	86.3%

2.5 禁区区域检测

禁区区域是足球场上的一个特殊区域，射门事件就是在这个区域发生的。禁区区域的检测可以辅助进行射门事件的分析和识别。如果禁区线出现在中镜头中，则受图象大小限制，禁区线必定十分不完整，非常不容易检测，因此本文对禁区线的检测都是在长镜头中进行的。通过对禁区中三条平行线的检测与分析，可以判断这是左右禁区中的哪一个，这种分类同样可以给足球视频的高层分析提供十分明确的语义信息，文献[44]提出了一种利用主颜色区域检测和投影法来检测球门，从而实现禁区区域检测的方法，虽然算法复杂度低，但效果一般。本文提出一种足球场球场线提取算法，并根据所提取的球场线信息对一些特定的场景如禁区进行识别。

2.5.1 图像预处理

图像在形成、传输、接收和处理的过程中，不可避免地存在着外部干扰和内部干扰。噪声恶化了图像质量，使图像模糊，甚至淹没特征，给分析带来困难。图像平滑可以有效改善图像质量，利于抽出对象特征。

本文使用高斯平滑滤波器对图像进行去噪声处理。高斯平滑滤波器是一类根据高斯函数的形状来选择权值的线性平滑滤波器。高斯平滑滤波器对去除服从正态分布的噪声是很有效的。其中，高斯分布参数 σ 决定了高斯滤波器的宽度。对图像处理来说，常用二维高斯函数作平滑滤波器。

高斯平滑滤波器使用的是高斯模板。该模板是通过采样二维高斯函数得到的。二维高斯函数具有如下形式：

$$G(x,y) = A * e^{\frac{x^2+y^2}{2r^2}} \tag{2-26}$$

其中 x 和 y 是模板中的坐标，r 是控制参数。

本文使用了较为常用的 5×5 高斯模板，如图 2-14 所示。

0	1	2	1	0
1	4	8	4	1
2	8	16	8	2
1	4	8	4	1
0	1	2	1	0

图 2-14 高斯平滑模板

2.5.2 Hough 变换

Hough 变换可以检测直线和圆等已知形状的目标，而且受噪声和曲线间断的影响小。本文采用 Hough 变换，在二值图像中检测直线。

Hough 变换的基本思想是利用点——线的对偶性。直线方程可以用 $y=kx+b$ 来表示，其中 k 和 b 是参数，分别是斜率和截距。过某一点 (x_0,y_0) 的所有直线的参数都会满足方程 $y_0=kx_0+b$。方程 $y_0=kx_0+b$ 在参数 $k-b$ 平面上是一条直线，即图像 $x-y$ 平面上的一个象素点就对应到参数平面上的一条直线。$x-y$ 平面上同一直线上的点对应在 $k-b$ 平面上的直线必定相交于一点 (k,b)。点——线的对偶性如图 2-15 所示。

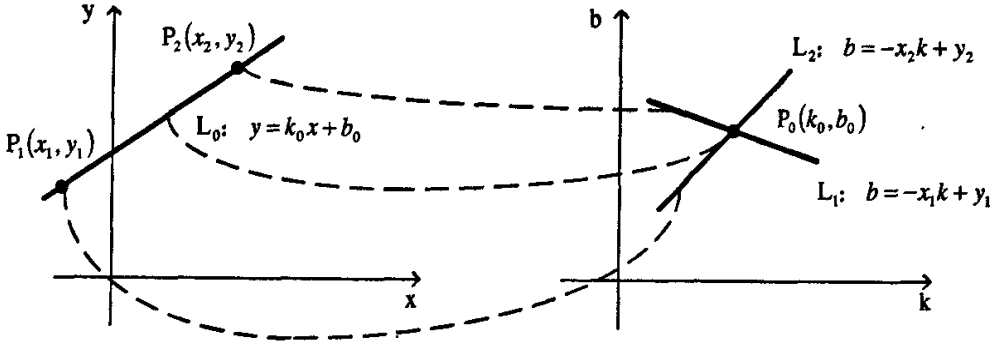


图 2-15 点——线的对偶性

$x=c$ 表示一类垂直于 x 轴的直线，无法用形如 $y=kx+b$ 的直线方程表示，因此采用参数方程(2-27)表示：

$$\rho=x\cos\theta+y\sin\theta \tag{2-27}$$

这样，图像平面上的一个点就对应到参数 $\rho-\theta$ 平面上的一条曲线上。同一条直线上的点对应在 $\rho-\theta$ 平面上的曲线将相交于一点。

2.5.3 禁区线检测

在四种镜头类型中，由于只有长镜头存在的禁区线最完整，因此在视频分段和镜头分类基础上，选取长镜头关键帧，检测其内是否存在禁区线。足球场外的直线肯定不是禁区线，因此需要通过场地分割步骤去掉。禁区线检测具体步骤如下：

Step1. 首先采用 2.3 节的场地分割算法，将长镜头中的场地区域分割出来。

Step2. 对分割好的场地区域图像做平滑处理，去除噪声影响。

Step3. 用 Canny 算子进行边缘检测，得到禁区线边缘明显的二值图像。由于场地边缘并非真正的直线，因此需要去掉场地边缘部分。然后用 Hough 边缘同时拟和所有出现的禁区线。

Step4. 对于二值图像中所有的边缘点，根据公式(2-27)在$\rho-\theta$平面就会有一条曲线，其中θ取 0 到π。这样就可以得到一曲线族。可以用图像的形式直观的表现出来，每条曲线在$\rho-\theta$平面经过的点象素值都加上 1，如图 2-16 所示。(图中红色标记都是白色曲线相交的最密集区域，白色的亮度越高，代表通过该点曲线越多。每个红色区域都是潜在的直线。)

Step5. 求出曲线族图像中亮度最高的前n个点(n取 200)，用红色标记出来，得到了若干块红色区域。每一块红色区域取m个最高亮度点(m取 5)，对其坐标(ρ,θ)求平均值，将结果代入公式(2-27)，即可得到这块红色区域所代表直线方程。

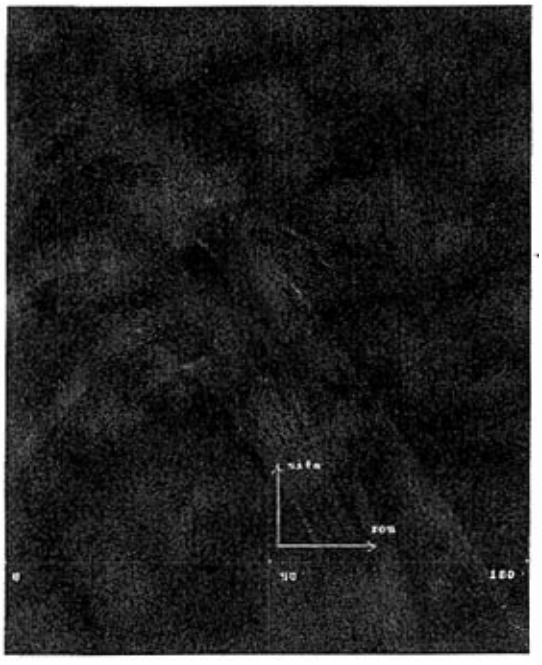


图 2-16 $\rho-\theta$平面曲线族

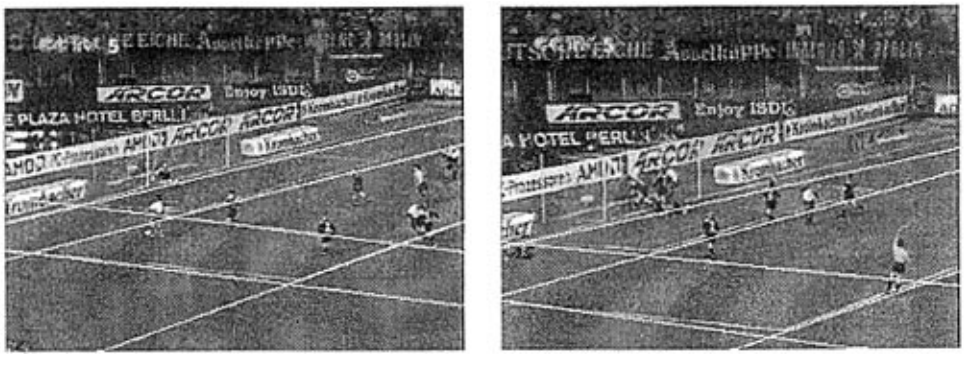


图 2-17 Hough 变换直线拟和效果图

将得到的直线显示在原图上，如图 2-17 所示。可以看出拟和得到的直线与图像中的线条基本吻合，除了小禁区左侧边线由于太短，被当作噪声去除之后，在运动员聚集的情况下，仍能将其余禁区线识别出来。

2.5.4 禁区区域判定

Hough 变换拟和结果是一组直线方程的集合，还需要确定这一组直线是否禁区线，本文采用基于模型的识别方法[46]，对禁区区域进行识别。模型如图 2-18(a)所示，由于摄像机视角的关系，该模型通常表现为如图 2-18(b)所示。

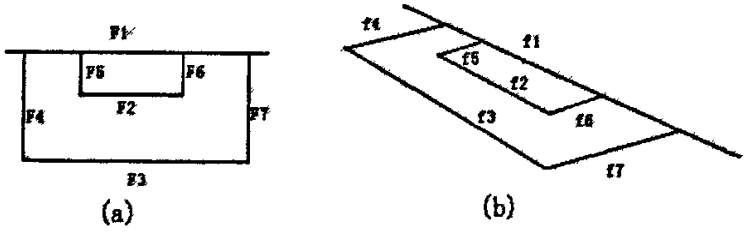


图 2-18　禁区区域线条模型

本文通过模型中直线与直线之间的相互位置约束关系来确定直线是否禁区线。具体规则如下：

(1)模型中相互平行的线段实际中也应平行或近似平行。

(2)斜率小于零的直线从右向左应该是按 f_1, f_2, f_3 顺序排列，并且斜率逐渐减小但是变化不能超过某阈值。

(3) 斜率大于零的直线从上向下应是按 f_4, f_5, f_6, f_7 排列，并且斜率逐渐增大但是变化也不能超过某阈值。

(4) 此外，在实际判断中，我们还使用如下经验公式，帮助判断和验证。设 $d_1, d_2, \cdots, d_8$ 分别是直线 $f_1, f_2, \cdots, f_8$ 的直线方程截距，则应该满足：$(d_1 - d_2)/(d_2 - d_3)$ 接近于 1/3；$(d_4 - d_5)/(d_5 - d_6)/(d_6 - d_7)$ 接近于 1/2/1.8。

(5) 如果满足上述规则的直线达到 4 条，则可以断定直线属于禁区线，当前帧包含禁区区域。

2.5.5 实验结果分析

从多个长镜头中随机抽取 1000 帧图像对识别算法进行了测试，测试结果如表 2-3 所示。查全率和准确率分别由公式(2-8)和(2-9)定义。由于有的帧图像表现为禁区一角，球场线不充分，因此查全率一般。

表 2-3 特定场景识别实验结果

	应检数	实检数	误检数	漏检数	查全率	准确率
禁区识别	592	538	31	85	85.6%	94.2%

3 视频后期制作特效的分析

负责足球足球比赛转播的电视制作人员，在长期的工作中积累了大量丰富经验，并逐步掌握了一套科学合理的摄制、编辑模式，比如何时添加字幕信息，何时制作一个慢镜头等。这些编辑制作手法不仅方便了观众欣赏比赛，还为视频处理和分析提供了丰富的语义信息。下面我们重点分析足球视频中的慢镜头的检测和字幕区域的定位问题。

3.1 慢镜头的检测

慢镜头是对比赛中的一次行为事件不同角度的回放，并让观众在视觉上产生慢动作效果。足球比赛中，当出现精彩场面或观众感兴趣的片段之后，通常会出现从多个不同的角度对精彩片段进行回放和慢放的镜头。因此慢镜头包含着很强烈的语义信息，可以在足球视频高层语义分析中作为精彩集锦的索引以及比赛事件辨识的重要元数据。

由于在不同的比赛视频中，慢镜头的产生方法可能不同，编辑手法也因人而异，并且慢镜头也没有一个很明显的边界，因此很难找出一种通用有效的慢镜头检测算法。目前研究人员通常采用寻找镜头的起始点和结束点的方法，实现慢镜头的检测，并提出了一些慢镜头检测算法。Pan[36]等人对足球视频中的慢镜头进行了探测，他们的探测基于这样一个事实——慢镜头片段通常作为一个单独的镜头出现并位于两个渐变之间，在这个基础上，首先将视频分割为单独的镜头，选取那些渐变之间的镜头作为候选慢镜头片段，然后根据慢镜头本身的诸如帧重复之类的特点作出进一步的判断。Vikrant Kobla 等人[34]对压缩域中的慢镜头进行了研究，根据慢镜头由帧重复产生这一原理，提出了一种利用 MPEG 流中的宏块、运动以及比特率等信息对慢镜头进行检测的算法，用来判断视频中是否存在慢镜头。Noboru 等人[37]针对美式足球比赛中扫换编辑效果进行探测，通过人工交互的方法找出编辑效果的模板，然后设定一个阈值，用模板在视频里进行相似匹配，相似度超过阈值的区域被认为是扫换，取两个扫换之间的片段为慢镜头。这种方法不仅可以探测扫换，还可以扩展到其他的编辑效果。H. Pan 等人[36]对空间域中的慢动作重播镜头进行了研究，提取了亮度和颜色直方图差值作为特征，提出了零交叉特征，并通过隐马尔可夫模型实现对慢镜头的建模和识别，但这种算法本身复杂，并且慢镜头边界定位的准确率有待进一步提高。

3.1.1 慢镜头的生成方式及特点

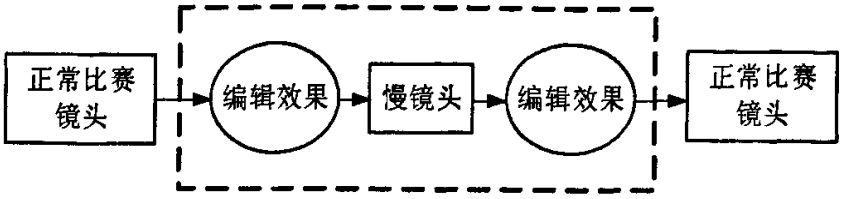


图 3-1 慢镜头结构模型

如图 3-1 所示，慢镜头过程中一般包括正常播放速率的镜头、视频编辑效果和慢镜头三种状态。编辑效果是指各种连续渐变如淡入淡出，溶解和扫换等，标志着慢镜头过程的开始或结束，通常只有几秒。此外，有的渐变还包含徽标。

慢镜头主要有两种生成方式，一种是帧重复方式，即以标准摄像机进行正常录制，然后对某些帧进行重复形成慢镜头效果，这种方法实现简单、重播速度易于控制，能够满足大多体育视频制作的需要；另一种是用高速摄像机进行拍摄，然后以正常速度进行播放形成慢镜头效果，这种方法的重播速度是固定的，而且由于制作成本较大，主要用于一些特定的科研研究领域，在足球视频转播中几乎没有使用。因此，本文仅对基于帧重复的慢镜头进行研究，并提出了一种有效的慢镜头检测算法。该算法主要利用慢镜头是由插帧来达成这一制作特点，通过对帧差序列的分析构造插帧慢镜头的模式，最后通过对慢镜头模式的检测实现对慢镜头的检测。

3.1.2 慢镜头检测算法

慢镜头检测是足球视频分析中的一个难点。本文提出了基于帧间差模式识别和帧间差分图象分析的两种慢镜头检测算法。两种算法相互独立，因此检测结果可以互补，从而提高检测精度与速度。帧间差模式识别实现一段视频中慢镜头的定位，定位结果比较精确，但是计算量较大；帧间差分图象分析则是在镜头分段的基础上对视频片段是否慢镜头进行判断，计算速度很快，但由于只是对视频片段是否存在慢镜头进行判断，所以检测结果不能达到很高的准确率。下面将分别介绍这两种方法。

3.1.2.1 基于帧间差模式识别的慢镜头检测算法

以帧重复模式生成的慢镜头，重复帧之间应该完全相同，但是考虑到噪声因素，如果两帧的帧差很小，也可认为两帧完全相同。具体步骤如下：

Step1. 计算相邻两帧间的帧差：记时刻 t 相邻两帧间帧差为 $D(t)$，由公式（3-1）计算，其中 M、N 为帧的长度和宽度；一个镜头的 $D(t)$ 如图 3-2a 所示，其中 86 到 250

为慢镜头。从图中可以看出，由于慢镜头效果是由帧重复形成的，重复帧间的差值主要是随机噪声、比较小，因此慢镜头的帧差表现为具有较大的起伏度。

$$D(t)=\frac{1}{M*N}\sum_{i=1}^{M}\sum_{j=1}^{N}(I_t(i,j)-I_{t-1}(i,j))^2 \tag{3-1}$$

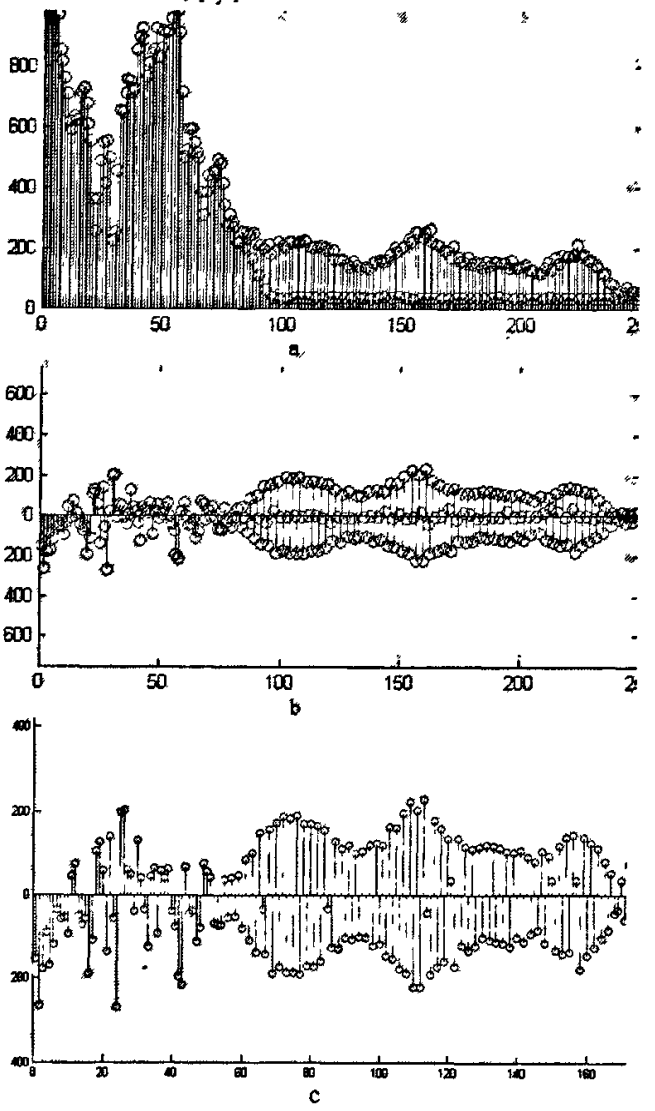


图 3-2 标准镜头和慢镜头的 $D(t)$、$DD(t)$ 和 $DD'(t)$

Step2. 对帧间差值进一步作差，记为 *DD(k)*，由公式(3-2)计算，如图 3-2b 所示；

$$DD(k)=D(k)-D(k-1) \tag{3-2}$$

Step3. 对 DD(t)进行门限处理，即如果 $-T(k)<DD(k)<T(k)$，则设 $DD(k)=0$，其中 T 为一动态门限阈值，$T(k)=\frac{\lambda}{2n+1}\sum_{k-n}^{k+n}DD(k)$，$\lambda$ 为一由实验确定的系数；

Step4. 去除 *DD(k)*序列中的零值，形成新的不包含零值的 *DD(k)*序列，记为 $DD'(k)$，如图 3-2c 所示，在 $DD'(k)$ 序列中存在明显的模式。通过对多种慢镜头

$DD'(k)$序列的观察，定义了帧重复慢镜头的四种模式，如图 3-3 所示：

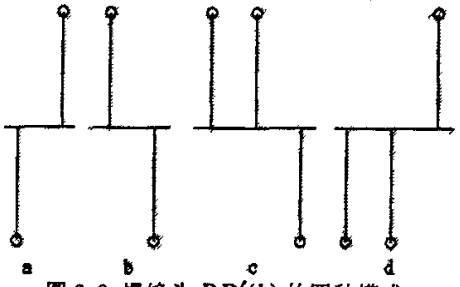


图 3-3 慢镜头$DD'(k)$的四种模式

Step5. 对镜头$DD'(k)$序列进行上述四种模式检测，只要满足其中一种即认为是慢镜头，对其进行标注，形成标签序列；

Step6. 最后通过持续时间约束，对一些孤立的标签进行前向合并，形成最终的标签序列，从中可以确定有没有慢镜头以及慢镜头的位置。

3.1.2.2 基于差分图象分析的慢镜头检测算法

在切变检测后得到视频分段结果，该算法以视频片段为一个基本分析单元，通过对相邻帧之间差分特征的分析，判断当前视频片段中是否存在慢镜头。理想情况下，慢镜头中的重复帧完全相同，它们之间的差分图象全黑（即所有象素点的颜色值为 0，并称之为 0 帧），但实际中由于噪声的存在会出现象素点不全为 0 的情况。白噪声是一种常见的噪声，如果我们只考虑是由于白噪声引起的，那么差分图象中非 0 象素点的分布也会和白噪声一样均匀，且非 0 象素点有着近似的象素值。非重复帧之间的差分图象则由于两幅图象内容的不同，非 0 象素点的象素值也会不同。对差分图象进行垂直方向的投影，即将差分图象所有象素点的象素值在垂直方向做累加，得到投影图。由于非重复帧之间的差分图象的非 0 象素点的位置和象素值分布不均匀，所以投影图会呈现较大的起伏。

根据上面的分析，我们选取差分图象的均值V_m作为一个特征，用来描述非 0 象素点的象素值偏移 0 的程度；选取差分图象垂直方向投影的标准差V_{mh}作为另一个特征，用来描述投影图的起伏程度。然后设定阈值，即可判断当前帧是否属于慢镜头的重复帧。最后根据规则完成对视频片段的判断。具体步骤如下：

Step1. 计算差分图象；

Step2. 计算差分图象均值V_m和垂直方向标准差的水平方向均值V_{mh}；

Step3. 如果V_m大于V_{mh}的 2 倍，则认为该帧属于慢镜头生成的重复帧，并计数；

Step4. 重复执行步骤 Step1- Step3，直到将整个视频片段处理完毕；

Step5. 如果认定为慢镜头生成的重复帧的数量大于 30 个，而且 0 帧的比例大于 0.3 则判断此视频片段中存在慢镜头，否则认为不存在。

3.1.3 实验结果分析

分别用两种方法实现了慢镜头的检测，都取得了较好的效果，如表 3-1 所示。

表 3-1 慢镜头检测实验结果

	应检数	实检数	误检数	漏检数	查全率	准确率
基于帧间差模式识别方法	118	125	16	9	92.4%	87.2%
基于差分图像分析的方法	118	129	21	10	91.5%	83.7%

3.2 字幕的检测

足球视频中出现的字幕包含了运动员信息、比赛比分等信息，具有丰富的语义信息。将视频字幕自动分割并识别出来有助于进行精彩事件的检测。本文通过图象帧中字幕区域的检测实现字幕帧的检测。如果图像中存在字幕区域，则当前帧是字幕帧。在字幕帧中得到的字幕区域还可以为文字识别提供较为准确的区域。

实际视频图像的背景往往比较复杂且不可预测，同时字幕的字体、大小、出现位置也不能确定，因此从视频图像中准确定位字幕区域并不是一件容易的事情。许多研究人员为此做了大量工作，概括起来主要有以下四类方法：(1)首先进行边缘检测以获得文本的边缘信息[25]，然后在水平和垂直方向投影统计，最后通过多个阈值和边缘尺寸限制来确定字幕区域。这类方法虽能快速检测文字，但需要预先设置多个阈值，因此通用性不好，检测错误率也较高。(2)针对字幕文字通常具有相似颜色这一特点，用图像分割[28]、颜色聚类[32,33]或连通区域分析[27]等方法把文字从背景中分割出来。但由于文字颜色不固定，可能存在多种颜色，因此对于背景复杂的图像和视频，检测效果不够理想。(3)根据图像纹理特征[26,27,28]判断某一个象素点或象素块是否属于文字。这类方法能适应复杂背景，但计算量大，十分耗时，且算法鲁棒性不好。(4)将图像切分成若干子块，然后用事先训练好的学习分类器(如支持向量机[30]、神经网络[29]等)对所有子块进行分类，得到字幕、非字幕两类子块。此算法检测准确率较高，但计算复杂，且分类效果受训练样本影响较大。

文字一般采用与背景有着强烈对比度的颜色，因此字幕区域表现出比其他区域更高的空间频率。小波变换具有多尺度多分辨率的特性，通过对图像进行小波变换，并重构高频细节，可以强化字幕区域的特征。本文提出了一种基于小波变换的 K 均值聚类字幕区域分割算法。相对于上述四类算法，该算法简单，不需要设置阈值，在复杂多变的背景下，对字体、大小、位置都不确定的字幕，仍能保持较高的正确率。

3.2.1 算法流程简介

本文算法流程如图 3-4 所示，首先对原图像进行两级小波分解，并分别重构两级分解得到的高频细节；然后将所有重构图像分别切分成大小为 $N\times N$ 的若干子块，计算每一子块的统计特征，形成各子块的特征向量；最后用 K 均值聚类算法对所有子块进行聚类分析，将其划分为字幕和非字幕两类，从而完成字幕区域的粗分。由于某些背景块的统计特征和字幕块的统计特征相似，某些字幕区域内子块的特征又和非字幕块的统计特征相似，因此粗分结果中可能存在噪声块，而得到的字幕区域可能存在空洞。为了去掉噪声块，并得到更完整的字幕区域，还需要对粗分结果进行后续处理。本文采用了连通性分析的方法去除噪声块，并用形态学方法填补空洞，保持字幕区域的完整性。

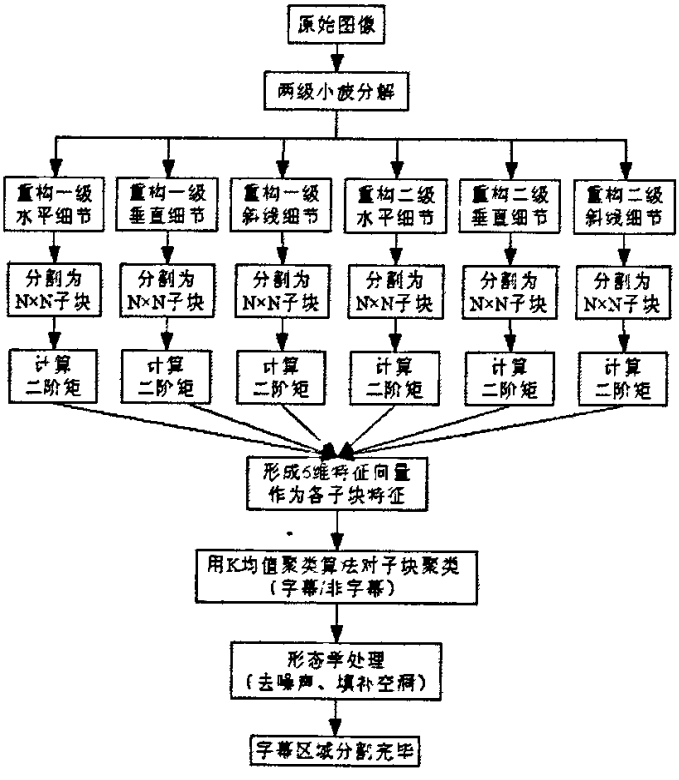


图 3-4 基于小波变换的 K 均值聚类字幕区域分割流程图

3.2.2 基于小波变换的特征提取

3.2.2.1 小波变换

小波分析(wavelet Analysis)或多分辨率分析(Multi-resolution Analysis)作为一种新兴的理论，无论对数学还是对工程应用都产生了深远的影响，广泛应用于数学、信号处理、自动控制、图像处理与分析、天体物理和分形等领域，已成为一种强有力的工具，被认为是对傅立叶分析的一个重大突破[24]。

由于小波具有良好的时频局部特性和变尺度特性，能够提供图像的逐级近似表达和边缘信息，即小波分解低通部分能够产生图像的逐级近似表达，同时带通部分所产生的细节信号提供了丰富的纹理内容。如图 3-5 所示，图 3-5(a)为原图，图 3-5(b)为一级小波分解图。从图 3-5(b)中可以发现，在三个高频子带中，字幕区域表现非常明显，由于小波的局部显微特性，小波系数大的地方总是出现在图像的边缘部分，亦即小波分解后的细节分量中有能较好地体现文本位置的信息。

(a)原始图像　　(b) 小波分解图

图 3-5 视频图像一级小波分解图

本文选择 Haar 小波，因为 Haar 小波在检测边缘信息方面具有良好的性能，并且 Haar 小波计算简单，可用掩模运算来实现。Haar 小波的尺度函数和小波函数可分别描述为：

$$\phi(x)=\sum_{k\in Z}p_k\phi(2x-k)=\phi(2x)+\phi(2x-1) \tag{3-3}$$

$$W_H(x)=\sum_{k\in Z}q_k\phi(2x-k)=\phi(2x)-\phi(2x-1) \tag{3-4}$$

其中 $\phi(x)=\begin{cases}1 & for\ 0\le x<1\\ 0 & otherwise\end{cases}$，对于两尺度有序列 p_k：$p_0=p_1=1$，$p_i=0, i\ge 2$，同样对于 q_k：$q_0=q_1=1$，$q_i=0, i\ge 2$。

对于一幅图像 I：

$$I(x,y)=\begin{bmatrix} i_{0,0} & i_{0,1} & \cdots & i_{0,2N-1} \\ i_{1,0} & i_{1,1} & \cdots & i_{0,2N-1} \\ \vdots & \vdots & \vdots & \vdots \\ i_{2N-1,0} & i_{2N-1,1} & \cdots & i_{2N-1,2N-1} \end{bmatrix}_{2N\times 2N}$$

其二维 Haar 小波变换可用如下 Mallat 算法实现：

$$LL_{x,y}=\frac{1}{4}\sum_{k_1,k_2=0}^{1}p_{k_1}p_{k_2}i_{k_1+2x,k_2+2y}=\frac{1}{4}(i_{2x,2y}+i_{2x,2y+1}+i_{2x+1,y}+i_{2x+1,2y+1}) \quad (3\text{-}5)$$

$$LH_{x,y}=\frac{1}{4}\sum_{k_1,k_2=0}^{1}p_{k_1}q_{k_2}i_{k_1+2x,k_2+2y}=\frac{1}{4}(i_{2x,2y}-i_{2x,2y+1}+i_{2x+1,y}-i_{2x+1,2y+1}) \quad (3\text{-}6)$$

$$HL_{x,y}=\frac{1}{4}\sum_{k_1,k_2=0}^{1}q_{k_1}p_{k_2}i_{k_1+2x,k_2+2y}=\frac{1}{4}(i_{2x,2y}+i_{2x,2y+1}-i_{2x+1,y}-i_{2x+1,2y+1}) \quad (3\text{-}7)$$

$$HH_{x,y}=\frac{1}{4}\sum_{k_1,k_2=0}^{1}q_{k_1}q_{k_2}i_{k_1+2x,k_2+2y}=\frac{1}{4}(i_{2x,2y}-i_{2x,2y+1}-i_{2x+1,y}+i_{2x+1,2y+1}) \quad (3\text{-}8)$$

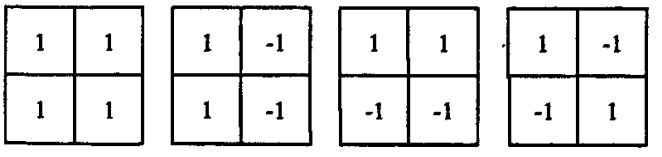
即可对图像 I 的 Haar 小波分解可用如图 3-6 所示的 Haar 模板作掩模运算实现，其计算效率显而易见是比较高的。

首先把视频帧划分成 N×N 的子窗口，对于每个子窗口，作三级子波分解，每一级对应 4 个子波分量，即近似分量 LL、水平 LH、垂直 HL 以及对角细节分量 HH，然后对每一个分量计算均方值 (m)、二阶(μ2)、三阶(μ3)中心矩作为纹理特征，共计 3×4×3=36 个特征。

$$m(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I(i,j) \quad (3\text{-}9)$$

$$\mu_2(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I(i,j)-M(I))^2 \quad (3\text{-}10)$$

$$\mu_3(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I(i,j)-M(I))^3 \quad (3\text{-}11)$$



a 低频分量　b 水平细节分量　c 垂直细节分量　d 对角细节分量

图 3-6 Haar 小波模板

3.2.2.2 特征提取

首先对图像 $I(x,y)$进行两级子波分解。由于低频分量 $A_j^d f$只是原图像的近似，因此只重构高频细节分量 $D_j^{HL}f$、$D_j^{LH}f$和 $D_j^{HH}f$，然后将所有重构的图像分别切分为大小为

$N\times N$的子块(本文中 N取 16)，并对分别计算所有子块的二阶中心矩(μ_2)作为各子块的特征量，即每个特征向量含有 6 个特征，计算公式如下：

$$E(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}I(i,j) \tag{3-12}$$

$$\mu_2(I)=\frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(I(i,j)-E(I))^2 \tag{3-13}$$

其中，N为子块的边长，$I(i,j)$为各子块图像数据，$E(I)$为各子块的均值。

3.2.3 基于 K 均值聚类的字幕区域分割

3.2.3.1K 均值聚类

K 均值(或者 ISODATA)聚类是模式识别中的经典算法，具有计算简单，能够动态聚类，自适应性强等特点，并有着广泛的应用领域，尤其解决模式分布呈现类内团聚状的问题时，该算法能取得很好的聚类结果。该算法是一种最普遍的不断迭代调整 K 个聚类质心的算法。和树聚类方法相比它有一个明显的不同，那就是在一个任意多样本集合的基础上得到一个事先顶好类别数的聚类结果。这个可以选择的类别数，可以通过以下方法得到：事先在一个较小的数据集合上进行树聚类得到聚类数，或者在经过多维比例尺度变换处理后得到的降维空间内进行树聚类得到的聚类数。

算法的中心思想是取定 K 类，并选取 K 个初始聚类中心，按最小距离原则将各模式分配到 K 类中的某一类，之后不断地计算类心，同时调整各模式的类别，最终使各模式到其判属类别中心的距离平方之和最小。算法步骤如下：

Step1. 任选 K 个模式特征矢量作为初始聚类中心：$z_1^{(0)}$，$z_2^{(0)}$，…，$z_K^{(0)}$，令 n=0。

Step2. 将待分类的模式特征矢量集{x_i}中的模式逐个按最小距离原则分划给 K 类中的某一类，即如果$d_{ij}^{(n)}=\min_j\left[d_{ij}^{(n)}\right]$，$i=1,2,\cdots,N$，则判$x_i\in\omega_j^{(n+1)}$，式中$d_{ij}^{(n)}$表示$x_i$和$\omega_j^n$的中心$z_j^n$的距离，$n$表示迭代次数。于是产生新的聚类$\omega_j^{(n+1)}(j=1,2,\cdots,c)$。

Step3. 计算重新分类后的各类心

$z_j^{(n+1)}=\frac{1}{n_j^{(n+1)}}\sum_{x_i\in\omega_j^{(n+1)}}x_i$，$j=1,2,\cdots,K$，式中$n_j^{(n+1)}$为$\omega_j^{(n+1)}$类中所含模式的个数。

Step4. 如果$z_j^{(n+1)}=z_j^{(n)}(j=1,2,\cdots,K)$，则结束；否则$n=n+1$，转至 Step2。

3.2.3.2 字幕区域分割

从本质上讲，字幕区域检测就是两类模式分类问题，即将图像中所有子块分为字幕与非字幕两类。此前有研究人员使用支持向量机或神经网络的方法对其分类，但是这两种方法都比较复杂，且分类效果受训练样本影响较大。因此，本文提出使用 K

均值聚类算法对子块进行聚类分析，所有的子块都被划分为两类——字幕或非字幕，然后将这两类的聚类中心与事先从多个样本得到的聚类中心相比较，确定哪一个聚类中心代表字幕，哪一个代表非字幕，这时得到一个字幕区域粗分结果，如图 4b 所示。由于两个聚类中心的相差比较大，因此样本的选择对于类别的判定影响是很小的。如果得到的两个聚类中心和已知非字幕聚类中心的距离都小于已知字幕聚类中心，那么就说明当前图像中没有字幕出现，从而可以检测一幅图像中是否字幕帧。

在背景图像十分复杂的情况下，一些表现和字幕块类似特性的背景图像被错判为字幕块是难以避免的。所幸的是这种情况通常表现为一些小的孤立区域，而视频、图像中的文字大多具有水平聚集排列或垂直聚集排列的特点，因此通过连通性分析就能消除绝大多数孤立噪声块。此外，字间空格或标点符号等由于在块中比例过小，很容易被聚类到非字幕块中，形成了字幕区域的空洞，如图 3-7b 所示。最后，以固定尺寸划分图像子块，有可能造成边界文字的缺失，即文字的一部分划分到字幕块，另一部分划分到非字幕块。针对区域空洞和文字缺失的问题，本文分别采用形态学的方法加以处理。其具体算法如下：

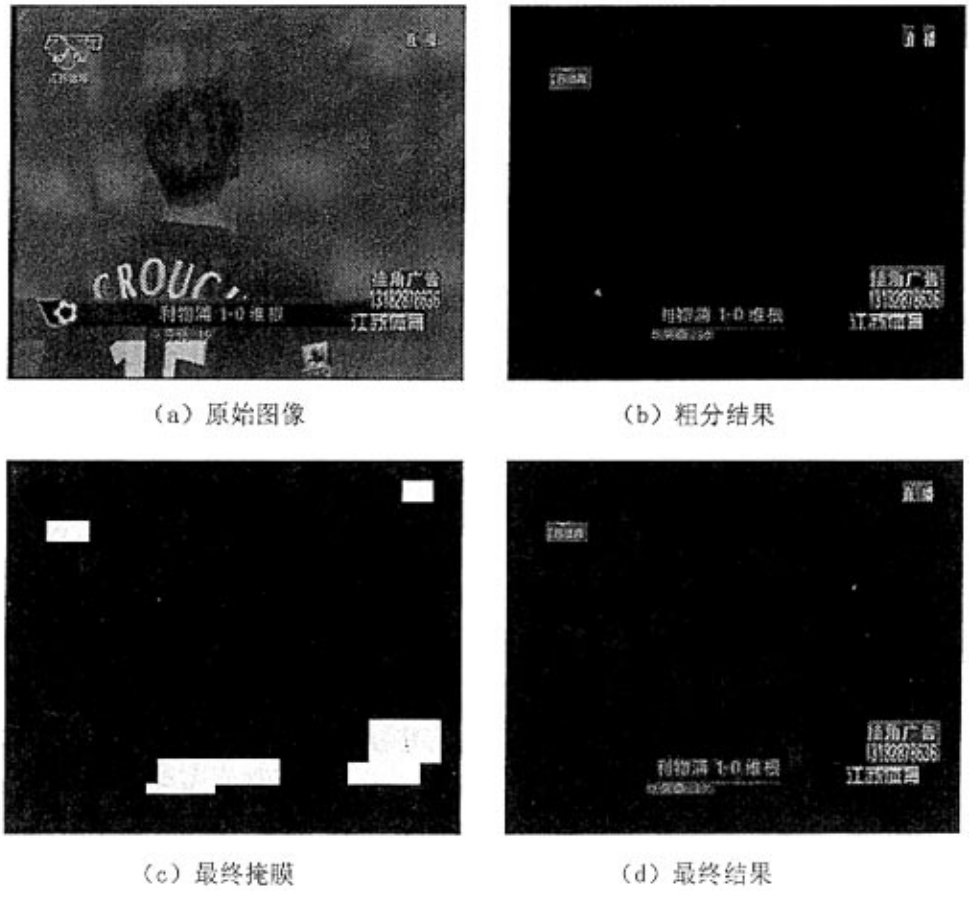
Step1. 根据聚类结果标注所有初选字幕块，1 表示字幕块， 0 表示非字幕块，每个块对应一个“象素”，于是所有子块形成一幅二值图像 $I_{块}$；接着对 $I_{块}$进行连通性分析，计算所有连通区域面积，面积小于 N(N 为限定的字幕区域至少覆盖的块数，本文选 N=3)的连通区域包含的初选字幕块被判断为噪声块，并从 $I_{块}$中去除。

Step2. 去除噪声块后，得到的字幕区域内可能还存在空洞；选择适当的结构算子(本文选取 3×3 的结构算子 E_1)对 $I_{块}$进行膨胀处理，可以填补空洞；再用同一个结构算子对 $I_{块}$进行腐蚀处理，以保证字幕区域大小在处理前后一致。

Step3. 将字幕子块对应的所有象素标记为 1，其余标记为 0，可得到字幕区域掩膜图像，如图 3-7c 所示。

Step4. 最后要做的就是保持边界文字的完整性。选择适当结构算子(本文选取 2×2 的结构算子 E_2)，对掩膜图像进行膨胀运算。将结果与原始图像进行掩膜运算，得到最终字幕区域分割结果，如图 3-7d 所示。

$$E_1=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \qquad E_2=\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$$



（a）原始图像　　（b）粗分结果

（c）最终掩膜　　（d）最终结果

图 3-7 字幕区域分割结果

3.2.4 实验结果分析

本文选取了一段长约 45 分钟的半场足球视频作为实验素材，分辨率为 704×576。实验中每秒取一帧，共计 2765 帧，检测结果如表 3-2 所示。球员衣服及其上面号码的影响，造成了误检数较高。在时效性上，由于需要计算所有子块的统计特征，计算量较大，仍不能实时处理，算法需要进一步优化。

表 3-2 基于小波变换的 K 均值聚类字幕区域分割实验结果

	应检数	实检数	误检数	漏检数	检到率	检准率
字幕帧	1982	2119	208	71	96．4%	90．2%

4 基于精彩事件识别的足球视频语义分析

足球比赛中发生的精彩事件是比赛中的亮点，也是人们最关心和有兴趣反复观看的部分。用精彩事件来描述足球视频，在不影响观众欣赏比赛的情况下， 极大的减小了数据量，不仅方便观众实现对视频的快速浏览，同时也为足球比赛视频的自动剪辑带来便利，并使得基于内容的检索成为可能。

比赛中出现的诸如射门、点球、犯规等精彩事件属于足球视频的高层语义，本章通过对这些事件的检测实现对足球视频的语义分析。

4.1 精彩事件检测的研究现状

目前，有很多研究人员在这方面做了大量的研究工作，主要有以下两种思路：一是综合利用视觉特征、音频特征或字幕信息等来实现精彩事件的检测；二是通过对视频中出现的慢镜头的检测与定位实现精彩事件的检测。

第一种思路采用的是多特征融合的方法，主要是对经过结构分析之后得到的比赛片段进行分析。这类方法中显著的难点是低层特征的选取和提取，如在射门事件检测中实现球和球门的探测与跟踪是一件很困难的事情。而且在足球比赛视频中，要实现对文本信息的提取和探测也存在一定难度。

第二种思路通过探测比赛视频中的慢镜头来获取精彩事件。这种方法基于这样一个事实：慢镜头通常出现与精彩事件发生之后，是对精彩事件的慢动作回放，且慢镜头和正常的比赛视频之间存在编辑特效，如渐变，徽标等。由于慢镜头检测算法仍然不够成熟，所以容易造成误检和漏检的情况。

实现精彩事件的检测识别，有很多研究人员提出了一些算法，主要集中在以下几种方法上：马超等提出了基于 HMM 的足球视频语义推理算法，实现了角球事件、进球事件、任意球事件和点球事件的检测。文献[34]提出了基于规则的进球事件检测。文献[40]用有限状态自动机对跳远事件进行了建模和识别。文献[45]提出了基于语义镜头时序关系分析的方法。

本文结合这两种思路，选取视觉特征、字幕特征和慢镜头作为精彩事件检测的特征，通过制定规则和对镜头类型的时序分析实现精彩事件的识别。

4.2 基于规则和时序分析的精彩事件识别

每个事件的发生都有一个从开始到结束的过程。例如，射门事件发生前会出现球

门区域的长镜头，在射门发生之后通常会有一个完成射门动作的球员的特写镜头，紧跟着会有慢动作回放整个射门过程的慢镜头，如果射门成功，还会出现观众欢呼的镜头或教练欢呼的镜头；犯规事件发生后会有一个较短的慢镜头，方便观众看清楚犯规事件是如何发生的，责任在谁，如果犯规程度比较严重，发生红黄牌，那么还会出现裁判出示红黄牌的中镜头，并在屏幕底部出现字幕，告诉观众是哪位球员犯规及得到红牌还是黄牌；点球事件发生前会有持续较长时间的包含小禁区区域的中镜头，由于罚点球也是射门，所以点球事件也会表现出普通射门所拥有的特征。本节通过不同类型的镜头的时序关系分析，辅助以慢镜头检测，字幕区域检测以及禁区检测，实现对射门、犯规和点球事件的识别。

4.2.1 射门事件

射门事件是足球比赛中的亮点，是足球比赛最重要的组成部分，构成了足球视频摘要的最重要内容，是其高层语义内容分析的重要组成部分。本小节主要对射门事件检测进行了讨论，利用前面得到的镜头分类结果，慢镜头检测结果，字幕区域检测结果及禁区检测结果，结合射门事件的视频制作特点，设定特定的射门事件判断规则，提出了一种对射门事件进行检测的新算法。

由于射门事件总是在禁区区域附近发生的，因此是否出现禁区区域可以作为判断是否发生射门事件的一个很好的依据。禁区内的白线是禁区的边界，也是描述禁区的最好特征，本文通过对禁区线的检测与分析确定是否出现了禁区区域。

在发生射门事件之后，如果得分，那么还会出现字幕信息，显示比赛双方的队伍名称以及比分，因此可以把字幕区域是否出现作为判断射门事件的另一个依据。本文通过对字幕区域的检测与定位实现字幕帧的判断。

射门事件的出现通常伴随特定模式的制作特征，一旦出现射门事件，为了进一步渲染这一激动人心的过程，以及更清楚地更细致地欣赏这一过程，视频制作者首先会传递场上的情绪给电视观众，这种情绪通常由一个或多个特写镜头来表达，如激动不已的运动员，兴奋的观众，其中这些与比赛无关的镜头成为非比赛镜头。然后用来自不同角度摄像机拍摄的多个慢速重播镜头来更详细重现这一过程，且慢镜头前后通常会有特效编辑来强化这一过程。最后以一个长镜头的开始表示这一过程的结束。以此为模式定义射门事件检测模板，如图 4-1 所示，要求：

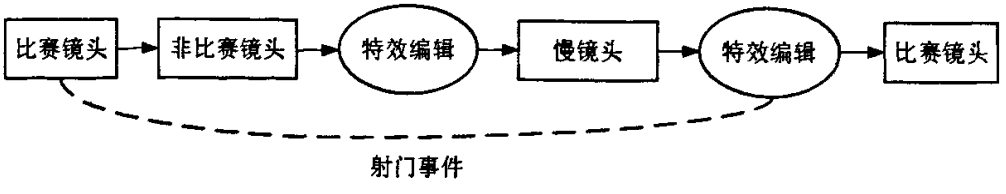


图 4-1 射门事件检测模板

①比赛镜头主要描述的是禁区场景，镜头中主色所占的比例大于40%；

②非比赛镜头持续时间大于30帧小于120帧；

③至少出现一次非场地镜头，这镜头可以是球员特写或观众镜头；

④至少存在一个慢镜头，射门事件是足球比赛中最重要的事件，因此发生之后通常会有多个角度的慢镜头对其进行慢动作回放；

⑤重播镜头的相对位置，跟在非场地镜头后面，且夹在两个特效编辑之间。

⑥慢镜头持续时间大于10秒。

在实际检测过程中，由于各种镜头类型频繁交替出现，所以不能通过镜头类型来触发射门事件检测过程。在检测到慢镜头之后，开始包括射门事件在内的各种事件检测是一个可行的方案。本文算法步骤如下：

Step1. 以切变点为边界对视频进行分段；

Step2. 对所有视频片段进行镜头分类和是否为慢镜头的判断；

Step3. 出现慢镜头之后，检测慢镜头持续时间是否满足要求。如果满足要求，继续下一步检测，否则停止检测；

Step4. 在慢镜头中查询是否出现了包含禁区的视频帧。如果存在则继续下一步检测，否则停止检测；

Step5. 检测慢镜头前面的非慢镜头片段；是否出现了球员的特写镜头。如果没出现则停止检测，否则说明发生了一次射门事件，但未知是否得分；

Step6. 检测慢镜头前后的非慢镜头片段，是否出现了观众欢呼或者教练欢呼的镜头。如果出现则继续下一步检测，否则停止检测；

Step7. 检测慢镜头后面的非慢镜头片段，是否出现了标示比赛比分的字幕帧。如果出现，则说明这次射门事件是一次成功的射门，否则说明射门未成功。

经过上述步骤的检测之后，就能够获得射门事件的大体位置，并可以获知这次射门事件是否得分。因为射门动作发生的很快，如果加上进攻组织过程，那么在慢镜头片段前面抽取30秒左右的视频片段，即可作为这次射门的描述。以此为基础，根据不同的需求，可以实现对比赛的剪辑，例如将比赛视频中的所有射门镜头组合起来，就得到了比赛的射门集锦。

4.2.2 犯规事件

足球比赛中，犯规并不是一件光彩的事情，但却是无法避免的。犯规有多种情况，有的是战术犯规，比如对方快速反击时候，己方防守人员没有回到自己的位置，这种情况下通过犯规的方式暂停比赛的进行，有助于己方人员及时回防。有的是恶意犯规，这类犯规是指某些球员通过犯规的方式向对方球员做出伤害。有的是无意犯规，比如

球员本意要将足球铲断，却因为对手速度很快，铲到了对方球员。情节严重的会被处以黄牌，甚至红牌。情节比较轻的并不会得到红黄牌处罚，比赛短暂中断后继续进行，由对方球员发任意球。在紧张激烈的比赛中，无论何种犯规，都需要裁判在最快的时间予以判罚。可是由于裁判的视觉范围有限，以及主观因素的影响，可能会出现误判的情况。这个时候需要将比赛录像调出来仔细观看。教练和球员也可以通过犯规部分录像了解对方球员风格，战术意图，并减少己方的无故犯规。犯规事件的检测也可以满足普通观众的其他不同要求。再者，本文在射门事件检测中使用了慢镜头作为判断依据，而有的犯规事件也会有个短暂的慢动作回放，以方便观众看清楚犯规是如何发生的，所以犯规事件可以通过对射门事件检测规则稍微更改后检测出来。

犯规事件并没有一个特定发生区域，因此不能像射门事件那样通过禁区区域的检测对犯规事件进行判断。但是犯规事件也有自己的特点，镜头类型在时间轴上也体现了一定的特征。犯规事件发生前，比赛一直正常进行，没有任何犯规事件发生的迹象。犯规事件发生后，往往会跟随一个慢动作回放镜头，而这个回放镜头持续时间比较短，通常只有几秒而已。慢动作回放片段的镜头类型一般是中镜头或特写镜头，但大多数情况下是中镜头，因为只有这样才能够方便观众看清楚犯规的过程。没有慢动作回放片段的犯规动作，通常不是足以影响比赛结果或情节轻微的犯规，本文将忽略这种情况，因此短暂的慢镜头是犯规事件的一个特征。如果犯规情节比较严重，裁判出示了红黄牌，那么在慢镜头之后的相邻视频片段中，会出现包含裁判的中镜头或者特写镜头，而且会持续一段时间，并且在屏幕上会出现红黄牌和犯规球员的字幕信息。

慢镜头检测和字幕定位在前面章节中已经得到了解决。裁判的检测一直是足球视频分析中的一个难点，这是因为每场比赛裁判衣服都没有统一的颜色，而且裁判本身在图象帧中所占的面积比较小。由于本文忽略不出现慢动作回放的犯规事件，而且一般较轻的犯规由于没有判罚红黄牌，也不会出现裁判镜头，所以本文采用红黄牌字幕信息的检测来判断犯规是否出现红黄牌。

经过上面的分析，本文定义犯规事件检测模板，如图 4-2 所示，要求：

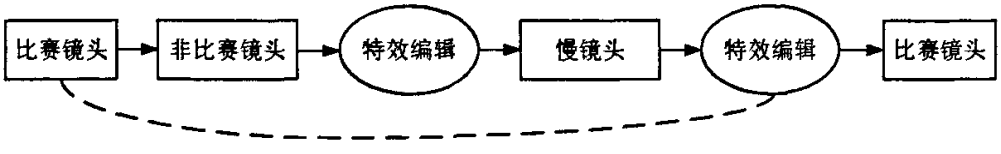


图 4-2 犯规事件检测模板

①至少存在一个慢镜头；

②慢镜头中存在球员中镜头或特写镜头；

③重播镜头的相对位置，跟在非场地镜头后面，且夹在两个特效编辑之间。

④慢镜头持续时间只有几秒，通常小于 10 秒；

由于犯规发生之前，比赛一直正常进行，所以不能通过镜头类型的改变来触发犯规事件检测过程。与射门事件检测一样，由慢镜头触发犯规事件检测。本文算法步骤如下：

Step1. 以切变点为边界对视频进行分段；

Step2. 对所有视频片段进行镜头分类和是否为慢镜头的判断；

Step3. 出现慢镜头之后，检测慢镜头持续时间是否满足要求。如果满足要求，继续下一步检测，否则停止检测；

Step4. 在慢镜头中查询是否出现了中镜头或特写镜头。如果存在则继续下一步检测，否则停止检测；

Step5. 检测慢镜头后面的非慢镜头片段，是否出现了标示红黄牌及犯规球员信息的字幕帧。如果出现，则说明这次犯规事件情节比较严重，得到了红黄牌，否则情节比较轻微。

经过上述步骤的检测之后，就能够获得犯规事件的大体位置，并可以获知这次犯规事件是否吃牌。由于犯规发生前比赛正常进行，所以犯规时刻的视频并不能很好的展现给观众犯规的过程，而慢动作回放部分是犯规时刻球员动作的慢速回放，且这时应该选取焦距比较近的摄像机采集到的视频，因此只需要提取犯规事件的慢镜头部分作为对这次犯规的描述即可。将犯规部分剪辑提取出来，可以方便有需要的观众随时浏览。

4.2.3 点球事件

点球本身属于射门事件的一种。与一般的射门不同，点球是对对方球员在禁区内犯规的惩罚，所以点球事件发生之前会有一个犯规事件发生，并且由于这次犯规事关点球能否判罚，因此犯规发生后肯定会跟随慢动作回放镜头。在罚点球之前的镜头中，禁区线和球门都非常清晰，而且会持续一段时间。摄像机也会对主罚队员和守门员特别青睐，通常会有他们的特写镜头。罚出点球之后，视频在时序上呈现和一般射门事件类似的特征。如果球进了，就会有球员、观众或教练欢呼的镜头，并在之后的视频片段中出现包含比赛比分的字幕信息；否则会出现球员懊恼的镜头，同时也不会出现字幕。

禁区区域检测和字幕区域检测的算法上文已经介绍。基于上面的分析，本文制定如下判断规则：

①点球事件发生前一定出现了犯规事件；

②在罚出点球之前的视频帧中，一定存在清晰的禁区区域；

③至少出现一次非场地镜头，这镜头可以是球员特写或观众镜头；

④至少存在一个慢镜头，而且慢镜头持续时间大于 10 秒；

⑤重播镜头的相对位置，跟在非场地镜头后面，且夹在两个特效编辑之间。

类似与射门事件和犯规事件的检测，点球事件同样由慢镜头触发。本文算法具体步骤如下：

Step1. 以切变点为边界对视频进行分段；

Step2. 对所有视频片段进行镜头分类和是否为慢镜头的判断；

Step3. 出现慢镜头之后，向前查询是否在较短时间内出现了犯规事件。如果有则继续下一步检测，否则停止检测；

Step4. 检测慢镜头持续时间是否满足要求。如果满足要求，继续下一步检测，否则停止检测；

Step5. 在慢镜头中查询是否出现了包含禁区的视频帧。如果存在则继续下一步检测，否则停止检测；

Step6. 检测慢镜头前面的非慢镜头片段，是否出现了球员或守门员的特写镜头以及清晰的禁区区域。如果没出现则停止，否则检测到点球事件，但是否进球得分仍无法判断；

Step8. 检测慢镜头后面的非慢镜头片段，是否出现了标示比赛比分的字幕帧。如果出现，则说明罚中点球，否则罚失点球。

经过上述步骤的检测之后，就能够获得点球事件的大体位置。由于点球事件是一次特殊的射门事件，所以按照射门事件抽取视频片段的方式将点球过程提取出来，并加以点球标注提示。在做视频摘要时候，就能够将点球事件独立于射门事件列出，方便观众查询浏览。

4.3 实验结果分析及小结

对于射门事件的检测，其查全率和查准率为 90.9%和 58.8%，从中可以看出查全率还是可以的，查准率比较低，表明仅以禁区作为约束还是不够的，但在足球视频中但凡出现慢镜头，一般都表示出现了重要情况，因此作这种辨识最起码可表示禁区发生了重要或精彩事件，因此还是有意义的。

表 4-1 事件检测实验结果

	应检数	实检数	误检数	漏检数	查全率	查准率
射门事件	17	23	9	3	82.4%	60.9%
犯规事件	26	29	8	5	80.8%	72.4%
点球事件	2	2	0	0	100%	100%

本章提出了一种精彩事件识别算法，并实现了射门事件、犯规事件和点球事件的检测。但是足球比赛是千变万化的，各种特征的提取和融合也是一件复杂的工作，所

以各种规则和特征提取的算法都需要进一步改进。比赛中的音频信息同样对事件的识别与定位有良好的作用，所以在以后的工作中，将考虑融合音频特征对精彩事件进行识别，如裁判的哨声意味着比赛的中断，观众的欢呼声意味着比赛的精彩。

5 原型系统设计与实现

5.1 系统总体结构

基于本文提出的足球视频语义分析算法，在PM1.5G CPU、512M内存和Windows XP平台下，采用Visual C++6.0为开发工具，实现了足球视频分析的原型系统，下面介绍该原型系统的结构、功能组成及实验结果。

本原型系统由三部分组成：数据库子系统、分析处理子系统和事件查询子系统。数据库子系统负责所有视频素材以及分析处理结果等的存储和管理，包括视频存储位置、视频信息、镜头分段结果、慢镜头检测结果及精彩事件识别等。分析处理子系统是原型系统的核心部分，完成了对足球视频的语义分析和处理，包括基于切变镜头检测的镜头分段，场地区域的分割，慢镜头的检测以及精彩事件的识别等；事件查询子系统为用户提供了友好的检索界面，方便用户对数据库进行访问，按照自己的需求查询相应的视频片段。原型系统的总体框架如图5-1所示。

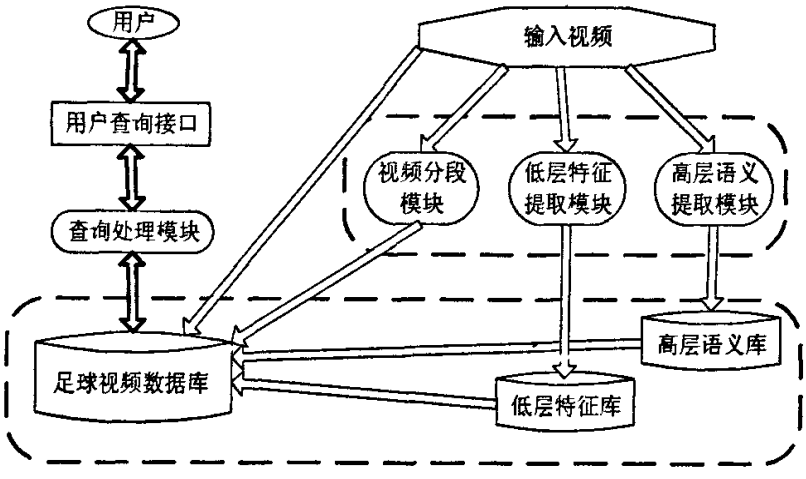


图5-1 系统总体框架

5.2 数据库子系统

足球视频数据库系统定义4个数据表：第一个是足球视频表，里面存放体育视频文件在数据库中的编号、文件完整路径、总帧数、总时间、比赛双方队伍名称、比分、黄牌数、红牌数等全局语义信息；第二个是视频片段表，存放的是通过切变镜头检测

后得到的视频片段的起止帧、是否是慢镜头及视频片段的附加描述等信息；第三个是镜头类型表，存放长镜头、中镜头、特写镜头、场外镜头等镜头类别、起止位置等信息；第四个是精彩事件表，存放事件类别、起止位置等信息。如图 5-2 所示。

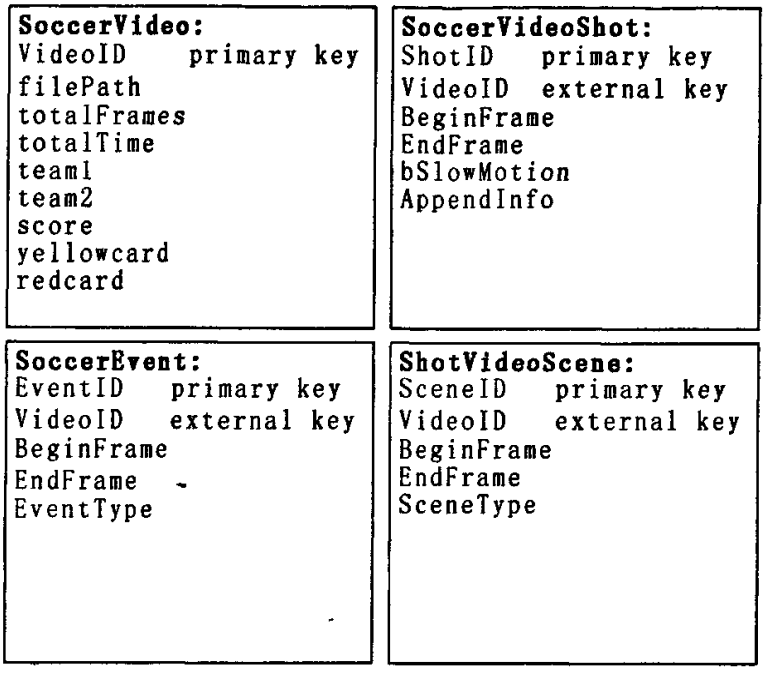


图 5-2 足球视频数据库结构

在足球视频数据库子系统中，可以直接进行数据库的修改，比如添加、删除或修改记录。系统运行界面如图 5-3 所示。

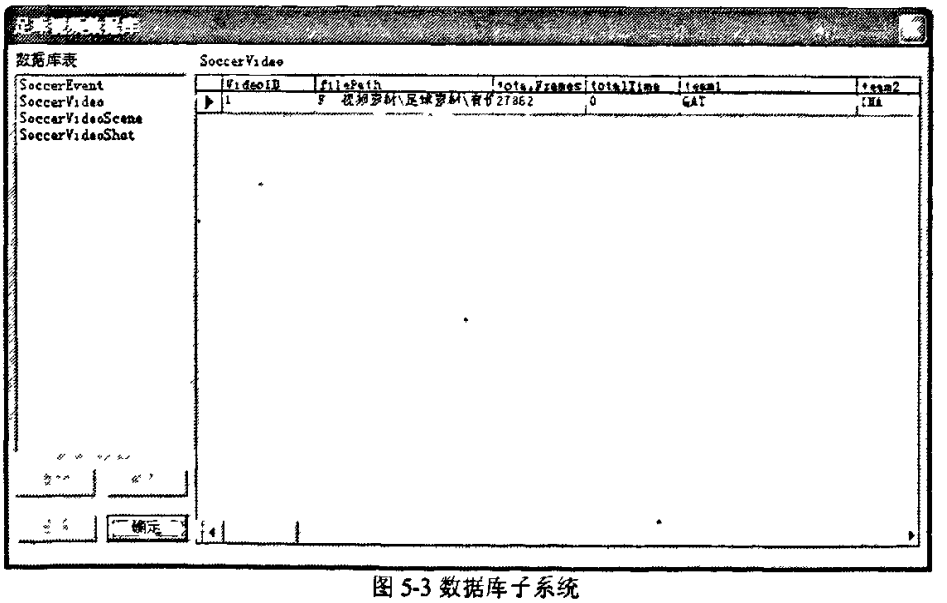


图 5-3 数据库子系统

5.3 分析处理子系统

足球视频分析处理子系统包括切变镜头检测、球场分割、镜头分类、慢镜头检测及精彩事件检测等功能模块，是原型系统的核心子系统，完成了对足球视频的语义分析。其中各个功能模块实现的功能如下：

(1)切变镜头检测

视频结构化是实现基于内容的视频检索的第一步，其基础是镜头分段。镜头分段是通过检测和识别镜头间的切变转换来实现的。切变是指一个镜头直接转换成下一个镜头的过程，中间没有时间上的延迟。在发生切变的两个镜头边界处，视频帧的图像特征往往存在突变，因此，通过比较这些图像特征（如直方图比较、象素点差分、边缘比较、运动信息等）的差异，可以比较容易地检测出切变。该模块采用第三章第二节介绍的基于帧间主色比例差和颜色直方图相似性差的多阈值镜头边界检测算法对体育视频的切变镜头边界进行分析和检测，分析和检测结果作为后续体育视频语义分析的基础。镜头分段得到的视频片段作为后续处理的基本单元，并存入对应数据库中。

(2)球场分割

该模块功能为足球视频的预处理，是进行镜头分类的基础。通过提取场地主色，通过计算每个象素点和主色的圆柱距离，并设定合适阈值，得到球场区域的粗分结果。然后对粗分得到的球场区域进行腐蚀和膨胀的形态学处理，最后进行连通区域的分析，得到最终的场地分割结果。场地分割结果蕴含着场地面积、场地内球员个数、大小等信息，这些信息都是镜头分类时候采用的特征。

(3)镜头分类

该模块功能在场地分割的基础上，首先利用场地区域比例把场景分为场地场景和场外场景，在此基础上，对场地场景通过用场地区域中的球员区域形状约束的分析，把它分为全局场景和局部场景，对场外场景通过分块区域复杂度分析，把它分为观众场景和特写场景，最后通过持续性约束，对场景标签序列进行合并形成长镜头、中镜头、特写镜头和观众镜头，结果存入相应的数据库，作为检索索引以及进行进一步语义内容分析的元数据。

(4)慢镜头检测

该模块主要利用慢镜头是由插帧来达成这一制作特点，通过对帧差序列的分析构造插帧慢镜头的模式，最后通过对慢镜头模式的检测实现对慢镜头的检测，结果存入慢镜头数据表，作为精彩集锦的索引以及比赛事件辨识的元数据。

(5)精彩事件检测

该模块功能包括两个方面，一是利用灰度形态学算子和 Hough 变换，检测禁区场景；二是利用上述提取的语义内容元数据，通过制定相应规则，对镜头类型进行时序

分析，实现对射门事件、犯规事件和点球事件的检测，并将相关检测结果存入对应的数据库。

分析处理子系统的运行界面如图 5-4 所示：

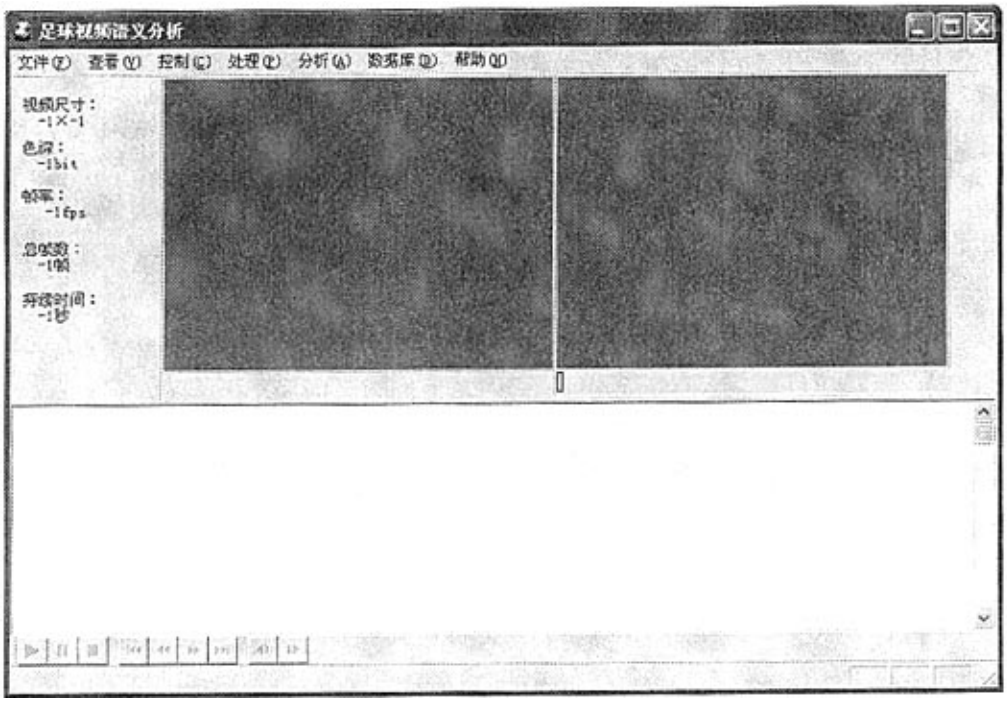


图 5-4 分析处理子系统界面

5.4 事件查询子系统

足球视频分析的目的是让用户能从大量的足球视频数据中快速高效的找到所需视频片段，并可以直接浏览感兴趣的足球视频片段或关键帧。检索时，从选择框中选定项目，如球队名称、事件、比赛等，然后点击查询。即可得到满足检索条件的结果列表，并可选择事件查看方式或镜头查看方式。点击列表中的结果会在左下角的信息框中显示片段的开始时间、持续时间、开始帧号、持续帧数、是否慢镜头、文件名称及文件路径。在右上角可以实现已选择片段的浏览播放。在右下角是视频片段的附加描述信息，事件查询界面如图 5-5 所示。

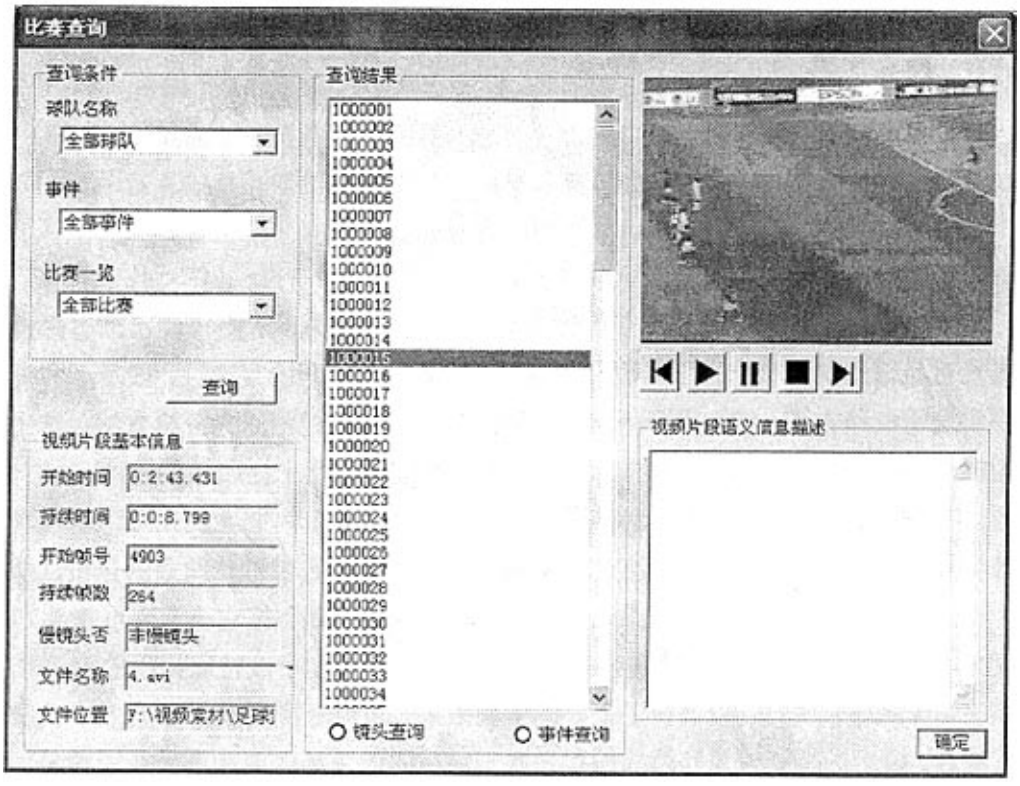


图 5-5 事件查询子系统界面

6 结论及展望

6.1 本文的主要工作

本文探讨了足球视频语义分析领域的一些问题，可以看出，虽然已经有很多研究人员对足球比赛视频进行了研究，但是足球比赛视频本身的复杂性，以及不断变化的摄制编辑制作方式，增加了分析难度。本文提出的足球视频语义分析方法，首先通过对视频中切变镜头的检测完成视频分段，之后在视频片段上提取关键帧，在关键帧上，通过足球比赛场地主色提取、场地分割，利用场地区域比例把场景分为场地场景和场外场景，进一步，对场地场景通过对场地区域中的球员区域形状约束的分析，把它分为长镜头和中镜头，对场外场景通过分块区域复杂度分析，把它分为观众镜头和特写镜头，效果基本令人满意。本文提出的基于K均值聚类和小波特征提取的字幕帧提取和字幕区域定位算法，也取得了良好的实验效果。利用慢镜头是由插帧来达成这一制作特点，提出的基于帧间差模式识别和基于差分图像分析的两种慢镜头检测算法，取得了较好的实验效果。通过场地分割、球场线检测、Hough 变换直线拟合和球场线空间排列规则等对禁区场景辨识，取得了较好的识别率。利用上述提取的语义信息，通过制定相应足球视频编辑规则，对足球比赛射门事件、犯规事件及点球事件进行了检测，取得了令人鼓舞的结果。

最后设计实现了一个足球视频语义分析的原型系统。该系统由数据库子系统、分析处理子系统和事件查询子系统三部分组成。其中，数据库子系统负责所有视频素材以及分析处理结果等的存储和管理，包括视频存储位置、视频信息、镜头分段结果、慢镜头检测结果及精彩事件识别等。分析处理子系统是原型系统的核心部分，完成了对足球视频的语义分析和处理，包括基于切变镜头检测的镜头分段，场地区域的分割，慢镜头的检测以及精彩事件的识别等；事件查询子系统为用户提供了友好的检索界面，方便用户对数据库进行访问，按照自己的需求查询相应的视频片段。

6.2 进一步工作的方向

本文设计实现了一个足球视频语义分析的原型系统，初步实现了足球视频数据库的管理，精彩事件的检测及简单的事件查询。但是由于时间的限制，以及该课题本身的难度，同时受作者水平限制，本文尚有很多待完善之处。

足球比赛本身并没有一个很好的结构，如何对足球视频建立语义模型，以更好的

对其进行描述是一个需要解决的问题。在精彩事件检测方面，虽然取得了一定的成果，但是由于只使用了视觉特征和字幕特征等，而且是基于精彩事件按照事先知道的编辑制作模式，所以面对视觉特征不明显，没有清晰的编辑制作模式的视频会出现漏检的情况。音频信息如裁判哨声和观众喝彩声都是有用的语义信息，可以更好的对精彩事件定位与检测，因此考虑融合音频特征也是将来工作的一个方向。本文虽然实现了慢镜头检测算法，但是仍然不够成熟，还存在漏检和误检的情况，如何实现对慢镜头更高效的检测仍需要进一步研究。

致 谢

谨以此文献给所有支持和关心我的朋友们，在此向他们表示深深的谢意。

首先衷心感谢我的导师王建宇教授和周献中教授，他们严谨的治学态度、广博的学识、专业方面的造诣指引我前进的方向并给我前进的动力。而王教授高尚的品格更深深的感染着我。可以说我在研究生期间的成长无一不包含着王教授的心血和栽培。从本科三年级开始，周教授就指导我在迈向科学顶峰的道路上一步一步踏踏实实前进。周教授对我的培养与教诲同样让我受益终生。

我还要对郭玲副教授等在大学期间对我无私的教诲的老师表示深深的感激。

此外教研室的各位师兄同学对我的关心和帮助也同样让我无法忘怀，尤其是吴奎博士、辛动军博士、井祥鹤博士、江金龙博士、陈志伟博士、杨建新博士、魏大宽博士、张帆硕士、袁夏硕士，还有与我同一研究小组的史迎春博士、何新博士、赵亚琴博士、骆文硕士。在大学期间我们建立了深厚的友谊，我无法忘记和他们在一起工作学习时那紧张而愉快的时光。

最后我要由衷的感谢我的家人，没有他们二十多年来对我的谆谆教诲以及生活上的无微不至的关心，就不会有我的这篇论文的诞生，在这里对他们一并表示感谢。

参考文献

1. 庄越挺, 潘云鹤, 吴飞. 网上多媒体信息分析与检索. 清华大学出版社, 2002.
2. 章毓晋, 基于内容的视频信息检索. 科学出版社, 2003.
3. A. Nurat Tekalp. 数字视频处理[M]. 电子工业出版社, 1998.
4. 章毓晋. 图像处理和分析. 清华大学出版社, 1999.
5. 章毓晋. 图像理解与计算机视觉, 清华大学出版社, 2000.
6. K. Messer, *et al.* A unified approach to the generation of semantic cues for sports video annotation. Signal Processing, 2005(85):357-383.
7. D. Yow, *et al.* Analysis and presentation of soccer highlights from digital video. In Proceedings Asian Conference on Computer Vision, 1995.
8. Y.H. Gong, L.T. Sin, *et al.* Automatic parsing of soccer programs. In Proceedings IEEE International Conference Multimedia Computer System, 1995: 167-174.
9. S. Intille and A. Bobick. Recognizing planned, multi-person action. Computer Vision Image Understand, 2001, 81(3):414-445.
10. V. Tovinkere, *et al.* Detecting semantic events in soccer games: towards a complete solution. In Proceedings IEEE Conference Multimedia Expo, 2001.
11. L. Xie, S.-F. Chang, A. Divakaran, and H. Sun. Structure analysis of soccer video with hidden Markov models. In Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing, 2002.
12. R. Leonardi and P. Migliorati. Semantic indexing of multimedia documents. IEEE Multimedia, 2002, 9(2), 44-51.
13. Y. Rui, A. Gupta, and A. Acero. Automatically extracting highlights for TV baseball programs. In Proceedings ACM Multimedia, 2000.
14. W.S. Zhou, *et al.* Online knowledge and rule-based video classification system for video indexing and dissemination. Information System, 2002(27): 559-586
15. A Ekin, A.M. Tekalp. Automatic soccer video analysis and summarization. IEEE Transactions On Image Processing, 2003, 12(7):796-807
16. B.X. Li, *et al.* Bridging the semantic gap in sports video retrieval and summarization. J. Vis. Commun. Image R., 2004(15):393-424.
17. D. Zhong, S.F. Chang. Structure analysis of sports video using domain models[A]. Proc of IEEE International Conference on Multimedia and Expo[C]. Tokyo, Japan: IEEE, 2001: 67-69.

18. 金红, 周源华. 一种基于模型的扫换检测方法[J]. 软件学报, 2001, 12(3):468-474
19. 王东辉, 朱淼良. 一种用于自动视频分段的WIPE转换检测和模式识别方法[J]. 计算机研究与发展, 2002, 39(2):247-253.
20. P. Xu, L.X. Xie, S.F. Chang. Algorithms and system for segmentation and structure analysis in soccer video. In Proceedings IEEE International Conference Multimedia and Expo, 2001:928-931.
21. Y. Luo, *et al.* Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks. Computer Vision and Image Understanding, 2003(92):196-216.
22. T.Y. Liu, *et al.* Shot reconstruction degree: a novel criterion for key frame selection. Pattern Recognition Letters, 2004(25):1451-1457.
23. 赵丕锡, 胡滨, 王秀坤, 李国辉. 足球视频的结构分析与概要. 计算机工程与应用, 2005, 41(30):166-168.
24. 陈武凡. 小波分析及其在图像处理中的应用. 科学出版社, 2004.
25. 蔡波等. 数字视频中字幕检测及提取的研究和实现. 计算机辅助设计与图形学学报, 2003, 15(7): 898-903.
26. A.K. Jain, *et al.* Automatic text location in images and video frames[J]. Pattern Recognition, 1998,31(12):2055-2076.
27. Wu. V, Nanmatha R, Risema E. Text Finder: An automatic system to detect and recognize text in images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(11): 1224-1229.
28. H.P. Li, *et al.* Automatic text detection and tracking in digital video[J]. IEEE Transaction on image processing, 2000, 9(1):147-156.
29. 李朝晖等. 小波－神经网络在视频文本自动检测中的应用[J]. 广州大学学报(综合版), 2001, 15(5):36-39.
30. 庄越挺等. 基于支持向量机的视频字幕自动定位与提取[J]. 计算机辅助设计与图形学学报, 2002, 14(8):750-753.
31. 刘骏伟等. 基于 SVM 和 ICA 的视频帧字幕自动定位与提取. 中国图象图形学报, 2003, 8(11):1334-1350.
32. 黄晓东等. 用小波变换及颜色聚类提取的视频图像内中文字幕[J]. 计算机工程, 2003, 29(1):43-44.
33. 章东平等. 自动定位彩色图像中的文本[J]. 浙江大学学报(工学版), 2005, 39(2):229-233.
34. V Kobla, D. Dementhon. Identifying sports videos using replay, text, and camera

motion features. Proceedings SPIE, 2000, 3972:332-345.

35. H. Pan, B. Li, *et al.* Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions. In Proceedings IEEE International Conference Acoustics, speech, signal processing,2002.
36. B. Li, Sezan, M.I. Event detection and summarization in sports video. In Proceedings of the IEEE Workshop on Content-Based Access to Video and Image Libraries. IEEE CS Press, Los Alamitos, CA, 2001.
37. N. Babaguchi, *et al.* Event based indexing for broadcasted sports video by intermodal collaboration. IEEE Trans. Multimedia, 2002,4(5):68-75.
38. 杜威, 廖永琚, 刘国翌, 刘彦红, 陈睿, 李华. 足球比赛场景的三维再现和自动解说. 计算机辅助设计与图形学学报, 2002, 14(9):853-859.
39. J. Assfalg, *et al.* Semantic annotation of soccer videos: automatic highlights identification. Computer Vision and Image Understanding, 2003(92):285-305.
40. 吴川, 马宇飞, 贺玉义, 钟玉琢, 张宏江. 体育视频中基于语义推理的事件检测方法. 清华大学学报(自然科学版), 2003, 43(4):507-509.
41. D. Zhong, *et al.* Real-time recognition and event detection for sports video. J. Vis. Commun. Image R.,2004(15):330-347.
42. M.R. Naphade. On Supervision and statistical learning for semantic multimedia analysis. J. Vis. Image R., 2004(15):348-369.
43. H.C. Lee, *et al.* Iterative key frame selection in the rate-constraint environment. Signal Processing: Image Communication 2003(18):1-15.
44. 胡涛, 何静, 张志刚. 一种检测足球视频中射门镜头的方法. 电视技术, 2005, 274(4):83-85.
45. 文军, 谢毓湘, 老松杨. 足球比赛视频中的精彩镜头分析方法. 计算机工程, 2004, 30(6):159-161.
46. Grmson WEL. Object Recognition by computer: The Role of Geometric Constraints. MIT press, 1990.
47. 史迎春. 基于内容的视频检索语义提取若干问题研究:[博士学位论文], 南京理工大学, 2005.
48. http://www.jdl.ac.cn/en/project/mrhomepage/index.htm
49. http://www.sharptechnologyventures.com/tech/himpact.php
50. http://www.goalgle.com/

附 攻读硕士学位期间作者发表论文情况

1. 王建宇, 张峰, 周献中, 史迎春, 骆文. 利用小波变换和 K 均值聚类实现字幕区域分割. 计算机辅助设计与图形学学报, 录用待发表.
2. Ling Guo, Ying-Chun Shi, Xian-Zhong Zhou, Feng Zhang. Location and Extraction of Broadcast in News Video Based on QGMM and BIC. The Fifth International Conference on Computer and Information Technology(CIT'05), 2005:662-667.
3. Shi YingChun，Zhou XianZhong，Luo Wen，Zhang Feng. Automatically Parsing and Labeling Video Based on Camera Motion Qualitative Analysis. Proceedings of the 2004 8th International Conference on Control, Automation, Robotics and Vision, Kunming, China December 2004:1543-1548. **(EI 收录号：05159037878)**