



基于内容的视频信号与图像库 检索中的图像技术¹⁾

卢汉清 孔维新 廖 明 马颂德

(中国科学院自动化研究所模式识别国家重点实验室 北京 100080)

(hqlu @nlpr. ia. ac. cn)

摘 要 基于内容的视频信号(Video)与图像库检索是当前计算机视觉,图像数据库与知识挖掘(Knowledge Discovery)等领域研究的热点之一.较系统地介绍了该研究方向的现状.对于静态图像,介绍了基于颜色、纹理、形状、区域、目标特征的检索和交互式检索方法,对于视频序列图像,介绍了镜头检测,镜头内容表示,场景建立,运动目标分割等技术.最后分析了该领域研究存在的难点,并提出了对今后工作的展望.

关键词 内容检索,视频信号,图像数据库,计算机视觉,图像处理.

A REVIEW OF CONTENT-BASED PARSING AND RETRIEVING FOR IMAGE AND VIDEO

LU Han-Qing KONG Wei-Xin LIAO Ming MA Song-De

(National Laboratory of Pattern Recognition, Institute of Automation

Chinese Academy of Sciences, Beijing 100080)

(hqlu@nlpr.ia.ac.cn)

Abstract Content-based image/video parsing and retrieving is one of most active researches in computer vision, image database, knowledge discovery and so on. In this paper, we systemically introduce the technical development situation in this area. For image, we introduce the techniques based on color, texture, shape, region, object feature and interactive retrieval method. For video, we discuss the techniques for shot detection, shot content representation, scene construction and moving object segmentation. Finally, we analysis the difficulties in this area and make some remarks about future work.

Key words Content-based retrieval, video, image database, computer vision,

1) 国家“九七三”重点基础研究项目(G1998030500)资助课题.

image processing.

1 引言

随着计算机性能价格比的不断提高、计算机网络发展和普及、以及计算机存储容量的快速增长和存储介质价格的下降,计算机信息技术得到了快速的发展.人们常说,当今社会是一个信息时代的社会.随着信息、计算机和网络技术的快速发展,大量的图像和视频信号的存储与传输已经成为可能,我们每天都会获得以 G 字节计的数字图像、数字化的视频信号、医学图像、遥感图像、新闻图片等等…….对这些巨大量的数据如何组织、表达、存储、管理、查询、检索是对传统数据库技术提出的重大挑战,以文字为对象的传统数据库无法满足图像数据库的要求,没有对图像及视频数据的自动和有效的描述,大量信息将淹没在数据库中,无法在需要时被检索出来.因此,如何将数字图像处理、计算机视觉技术与传统数据库技术相结合,建立基于对图像及视频内容自动或半自动描述的新一代图像视频数据库就成为迫切的需求.近年来,基于内容的视频信号与图像数据库检索技术成为国内外研究热点,也成为未来信息高速公路、数字图书馆等重大项目中的关键技术.

所谓基于内容的检索(CBR,Content-Based Retrieval),是指直接根据描述媒体对象内容的各种特征进行检索,它能从数据库中查找到具有指定特征或含有特定内容的图像(包括视频片段),它区别于传统的基于关键字的检索手段,融合了图像理解、模式识别等技术,具有如下特点:

1)直接从媒体内容中提取信息线索;

2)基于内容的检索是一种近似匹配,这一点与常规数据库检索的精确匹配方法有明显不同;

3)特征提取和索引建立可由计算机自动实现,避免了人工描述的主观性,也大大减少了工作量.

近十年来,基于内容的图像、视频数据的检索是当前计算机视觉,图像数据库与知识挖掘(Knowledge Discovery)等领域最活跃的研究热点之一.每年都有相关的国际会议召开,如 SPIE 的 Storage and Retrieval for Image and Video Databases(到1998年已经开了四届),ACM Multimedia 等,还有一些新创刊的杂志,如:MultiMedia Tools and Applications,这方面的研究是这些杂志发表的论文的重要组成部分.近几年,许多国际上重要的杂志,都以此内容发表了专刊.如:IEEE Computer 28(9),IEEE Trans. On PAMI 18(8),Pattern Recognition 30(4),Image and Vision Computing 17(7).目前,国内外很多机构都在进行相关的研究工作,并取得了很多人瞩目的成就.如:IBM 的 QBIC(Query by Image Content)^[1],MIT 的 Photobook^[2],CMU 的 Informedia^[3]等,国内的一些研究单位,如中科院自动化所国家模式识别实验室、清华大学、上海交通大学、复旦大学等单位也进行了研究并开发了一些实验系统.图1是实验系统的框图.本文主要介绍系统中涉及的图像技术,数据库、查询语言等方面的内容可参见其它综述^[1~6].

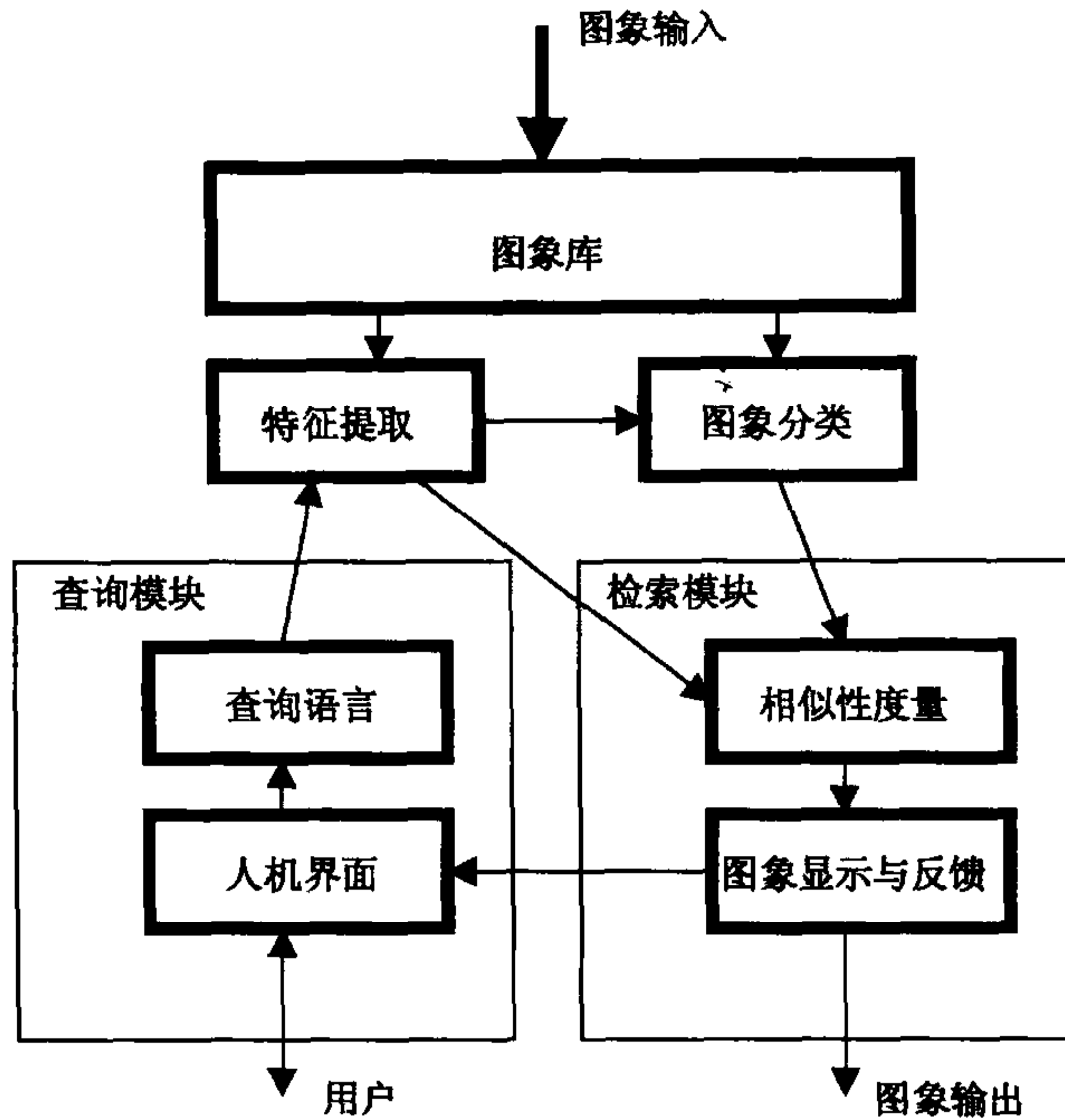


图1 基于内容的图像检索系统

2 静态图像的特征索引与检索

以图像的特征作为索引,对静态图像进行检索是目前使用最多的方法.目前比较成熟的特征索引是颜色、纹理和一些低层的、简单的形状特征和物体间方位关系^[4~6].这些特征具有计算简单,性能稳定的特点,但这些特征都有一定的局限性,因此,近几年的研究逐渐转向基于区域、目标的图像表示、语义描述等,检索则引进用户反馈等方法.

2.1 颜色

颜色是一种重要的视觉信息属性,在图像查询与检索中是一种很有用的特征.相对于其他特征,颜色特征非常稳定,对于旋转、平移、尺度变化,甚至各种形变都不敏感,表现出相当强的鲁棒性,而且颜色特征计算简单,因此成为现有检索系统中应用最广泛的特征.

颜色检索的基本思想是将图像间的相似度归结为图像颜色直方图之间的距离.这方面奠基性的工作是 Swain 和 Ballard^[7]提出的直方图交集算法,它计算两幅图像的三维颜色直方图的每一个颜色单位,并进行细致的比较. Mehtre^[8]等提出距离算法和参考颜色表方法,实验表明,每幅图像只需保留很少几种主要颜色,就能得到很好的结果.

为了得到与人的感觉类似的相似性度量,一些研究者还提出采用其他的颜色空间模型和距离测度. Zhang^[9]采用了符合人眼感觉的 H-S-I 模型. Harfner^[10]在进行直方图匹配时引入了二次型距离.另外,在提高检索对于光照的稳定性^[11]、空间分布信息的引入^[12]等方面,也出现了很多算法.

图2是一个用颜色直方图检索的例子,来自中科院自动化所国家模式识别实验室的实验系统.左上角为查询模板图像,后面则为检索到的相似图像集.



图2 基于颜色直方图的检索

2.2 纹理

纹理是与物体表面材质有关的图像特征,目前也是基于内容检索系统中的一个重要手段。

纹理分析一直是计算机视觉的一个重要研究方向,其方法主要分为两类:结构方法和统计方法.结构方法假定图像由较小的纹理基元排列而成.它采用句法分析方法,只适用于规则的结构纹理.统计方法又可进一步分为传统的统计方法、基于模型的方法,以及基于频谱分析的方法。

传统的统计分析方法始于20世纪70年代早期 Haralick^[13]等提出的基于二阶灰度统计特征的共生矩阵方法. Tamura^[14]以人的主观心理度量为标准,提出了六个基本的纹理特征,这些特征与人的感觉较一致,在许多图像检索系统中得到应用。

20世纪80年代开始,基于随机场模型的技术被用于纹理分析. Kashyap 等^[15]提出用 CSAR(Circular simultaneous autoregressive)模型提取旋转不变的纹理特征. Cohen^[16]则采用了 GMRF(Gaussian Markov Random Field)模型. MIT 的 Photobook^[17]中采用了随机场的2-D 分解技术,得到 Periodicity, Directionality and Randomness 三种特征,用于纹理图像检索。

多尺度的思想促进了基于频谱分析方法的发展,特别是基于 Gabor 滤波器的技术在纹理分析中被广泛采用^[18,19],Gabor 滤波器可以通过调整获得图像不同方向和尺度的纹理信息. Majunath^[20]等把它用于图像检索,并把它与 MR-SAR (Multiresolution simultaneous autoregressive)和小波变换等方法做了比较。

2.3 形状

基于图像内物体形状的检索是基于内容检索当中一个最具挑战性的问题之一,因为寻找符合人眼感知特性的形状特征不是一件简单的工作^[21].首要的困难是要将不同物体从图像中分割出来,这是计算机视觉的困难问题之一,至今没有很好解决(见本文3.4节).

形状的描述也是困难的问题,常用的方法有傅立叶描述子^[22]、矩不变量^[23]、各种简单的形状因子^[24](如面积、圆度、偏心度、主轴方向)等.除了这些全局特征以外,有时也用一些局部特征(如直线段、圆弧、角点、高曲率点等^[25]),以解决遮挡问题.除这些常规方法外,人们还提出了许多基于不变性和变换的方法^[26~29].

由于基于简单特征的方法无法对形状作细致的匹配,也不能解决广泛存在的变形问题.因此实际系统中常常只用它们来作初步的过滤,最终的匹配结果由一些更复杂的方法给出,如变形模板^[30]、弹性匹配^[31]等.

2.4 区域与目标

由于颜色、纹理的检索仅适合部分图像检索的情况,且检索的正确率不高,而且,在很多情况下,人们感兴趣的并不是整幅图像,而是图像中的某些区域或目标,因此,近几年来,人们提出了基于区域或目标的图像表示和检索方法^[32~37].在这类方法中,通常需要完成图像分割,然后再利用感兴趣部分的信息进行检索.

2.5 用户反馈(Relevance feedback)

由于现有的技术实现自动语义分析还存在很大的困难,因此很多系统中将人包括在检索环路中,通过交互式的反馈来决定采用何种特征组合及各种特征的权值^[38~40].交互式的反馈过程在模式识别技术中实际上是一种“有教师的学习分类”过程.如图2中是第一次检索得到的图像,通过人机交互,使用者可指出那些图是符合他的检索期望的,机器据此调整各种特征的权值,从而使数据库内的图像分类更接近使用者的主观愿望,使检索符合使用者的个性化要求,该研究方向是当前的研究热点.

3 视频数据

对于视频图像数据,目前公认比较合适的检索单元是一段中间没有摄像机信号中断、连续拍摄得到的视频镜头(shot).因此,视频图像数据分析的第一步就是镜头检测,然后抽取有效代表镜头内容的关键帧,这时可以用静态图像检索的技术来检索关键帧,从而达到检索视频的目的.当然,视频图像数据毕竟不同于静态图像,它包含有更多的信息,如:摄像机运动信息,视频流中的目标运动等,这些都是视频序列图像分析中可以利用的重要信息.图3是中科院自动化所国家模式识别实验室视频信号检索实验系统的界面,上面为分割得到的部分关键帧图像,左下角为时空流图像,右下角为视频播放窗口.

3.1 镜头检测

镜头检测算法的研究是视频分析与检索中最活跃的研究领域之一,在该领域已发表了大量的论文,提出了许多算法.其中一些主要的方法有:1)基于像素差的方法^[41,42],这种方法简单,但对噪声和运动很敏感;2)基于统计量(如窗口内灰度均值和方差)的方法^[43];3)基于图像特征的方法;如文[44]中基于图像边缘的方法;4)基于灰度和彩色直方图的方法^[45~47],其中彩色直方图应用最为广泛,并在突变检测中得到90%以上的正确率.

由于视频数据量很大,视频的分析 and 检索将花费大量的时间,因此,为了加快速度,人们提出了许多加快运算速度的方法,如:5)压缩域中的方法^[48,49].压缩域(主要指 MPEG)方法的优点是速度快,并且可直接利用某些压缩过程中已提取的信息,如块运动矢量^[50].6)区域块法^[51],用图像中几个区域块的比较代替整幅图像的比较.7)时空流法^[52],用平面沿时间轴与视频数据相切,从而得到时空流图像,并在时空流图像上寻找视频镜头切割点,因此,大大节省了时间,提高了镜头切割点检测的速度.图3的左下角为时空流图像的一个例子.



图3 视频序列图像分析界面

镜头检测的主要困难在于渐变的检测.渐变中相邻帧图像内容没有明显的突变,难以检测到明显的切换点.目前,对渐变的镜头检测方法主要有双重比较法^[53,54]和基于各种渐变数学模型的方法^[55].

以上方法中的阈值是人工预先选取的,这对内容多变的视频序列并不合适. Gunsel^[56]和 Naphade^[57]用非监督的聚类方法来自动选取阈值,而 Vasconcelos^[58]则从统计的观点出发,假定镜头长度分布符合 Weibull 分布,以贝叶斯估计来获得自适应的阈值.

3.2 视频镜头内容表示

一个镜头由多幅图像组成,构成一个序列.目前,镜头的表示主要有两类方法:关键帧表示和基于拼接技术的背景、目标表示.

3.2.1 基于关键帧的表示

关键帧表示法是在构成镜头的图像序列中选择能描述镜头内容的关键图像,反映了镜头的主要内容.简单方法是选择镜头的第一帧和最后一帧两幅图像,显然关键帧的选取应与镜头内容的变化有关,变化剧烈的镜头应以较多的关键帧表示. Zhang^[59]提出了基于视频特征变化的方法,在图像的累积颜色或运动变化大于某一阈值时,就多选取一个关键

帧. Wold^[60]和 Gresles^[61]则分别采用基于运动和内容相对变化的方法,以运动或内容相对变化在时间上的极小值点作为关键帧的位置. Zhang^[62]则用聚类的方法,它的基本假设是:如果某一内容比较重要,就会有较多的帧来表现它,因此每一个大的聚类应有一个关键帧.

3.2.2 基于图像拼接技术的背景表示

关键帧是视频镜头的一种方便的表示,它可直接从原始视频数据中抽取出来,但它不可避免地会出现数据冗余,如选择不当,还有可能漏掉某些重要的内容.而融合多幅图像的背景信息通过图像拼接(MOSAIC)技术而生成一幅全景图,是一种紧凑且完全的镜头背景内容表示方法^[63~67].图4是中科院自动化所国家模式识别实验室视频信号检索实验系统做的图像拼接.通过这种方法,结合运动目标提取技术,每个镜头片段都可以用一幅全景图和一些重要的运动目标来共同表示.这种表示方法具有数据量小,包含信息多的特点,具有较好的应用前景.与此同时,还可以获得摄像机运动参数及目标运动轨迹等信息,这些都是理解视频内容的重要线索.

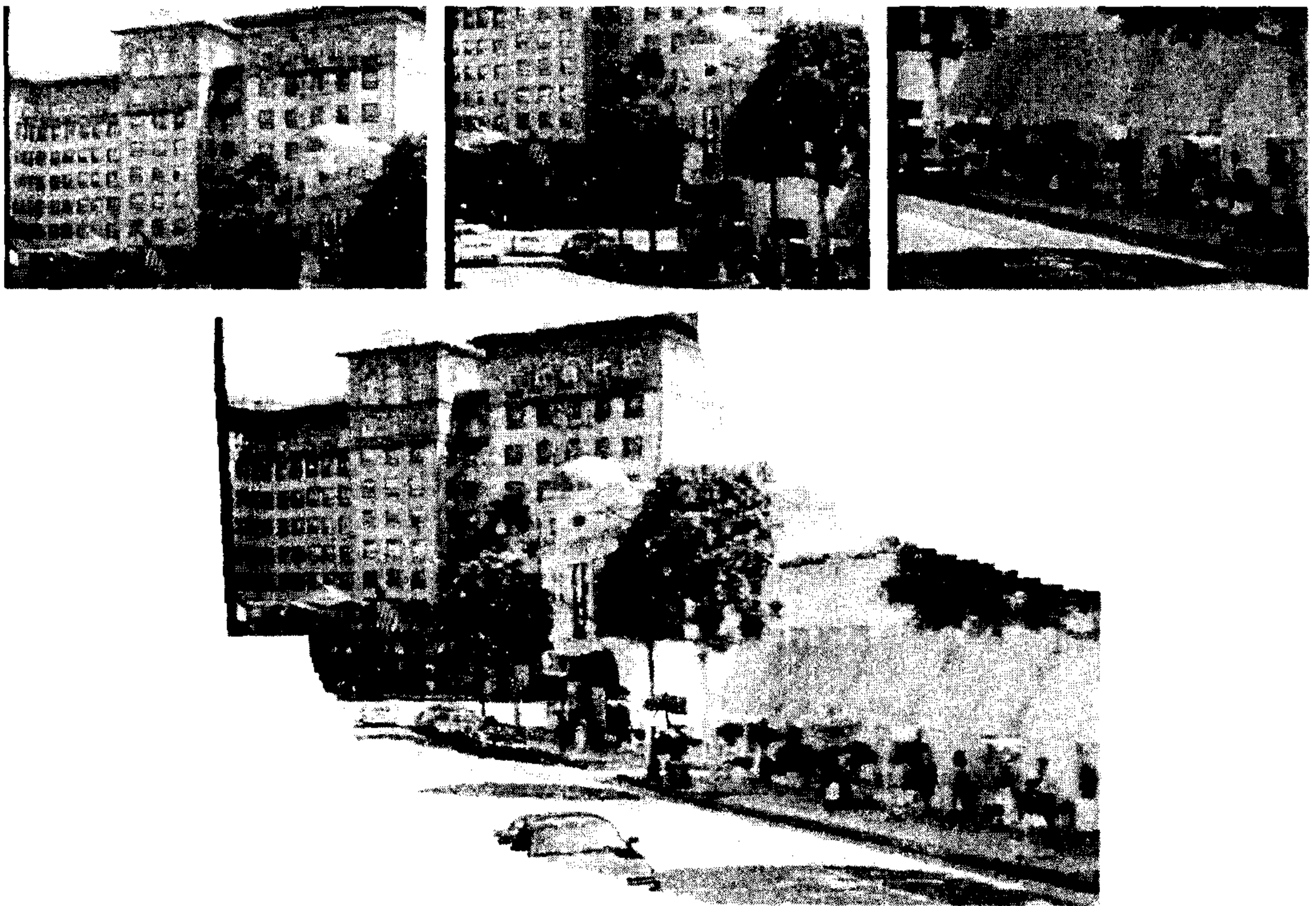


图4 图像拼接

视频图像拼接技术的关键是对摄像机运动的估计.摄像机运动参数除了可用于图像拼接,也可直接用作描述镜头的特征,或检索.由于3-D场景下任意的摄像机运动估计非常困难,因此一般算法都对场景或摄像机运动作了某些限制,以使两图像帧之间的坐标变换可用某种简单的参数模型表达.例如包含平移,旋转,及尺度变化的4参数模型^[68];最为常用的6参数仿射变换模型^[69],以及8参数2-D射影变换模型^[70];近来也出现了一些3-D场景下任意摄像机运动的参数估计算法,如平面+视差(plane plus parallax)模型^[63,64].

视频图像拼接技术的困难还在于:1)由于视频图像数量很多,虽然相邻帧间的运动估

计可能很准确,最终的累积误差却很大. Irani^[65]用动态拼接方法来解决此问题,即每帧图像都与前面生成的拼接图配准,而不是与原来的前一帧图像配准, Sawhney^[66]提出效果更好的全局配准技术. 2) 视频图像中常包含目标运动,它们会对摄像机运动的估计产生影响,一般采用鲁棒性技术来消除这些影响^[71,72].

3.3 视频的高层语义表示

对于视频来讲,浏览与有明确目的的检索是同样重要的. 它已成为基于内容视频检索的新的研究方向.

Yeo 和 Yeung^[73]提出了场景转换图(STG, Scene Transition Graph)的概念,他们以一个有向图的节点表示场景,边表示时间上的转换. 通过对 STG 的化简,可以去掉不重要的镜头,并得到视频的紧凑的表示. 还有作者通过选取视频中的重要片段(high light),获得视频摘要(video summary),由于单纯从图像中获得语义信息非常困难, Smith 等^[74]提出了综合采用视频中的图像,语音,文字信息的方法.

3.4 目标分割

目标分割的结果对视频图像分析具有重要的意义. 一个视频场景通常由背景和目标组成. 而视频序列中的活动目标往往是视频图像序列的重要组成部分,常常是查询的事件主体. 目标分割是计算机视觉最困难的问题之一,目前提出的方法很多. 这里,主要就具代表性的方法作一介绍.

Colonnese^[75]利用高阶矩对一组序列图像中的相邻帧间变化区域进行检测,得出四阶矩变化区域图,然后用之估计相邻帧间的运动矢量. 用整组序列图像的运动矢量估计值作比较,以确定运动和静止区域.

Mech^[76]首先估计出相邻两帧间的变化区域掩码(CDM). 在这个掩码中,只有目标运动对应的变化点才被标注为变化区域. 然后在得到的 CDM 中用层次块匹配方法估计每一点的运动矢量,以此找到上一帧的对应点并判定其是否属于目标掩码 OM. 最后把前面得出的 OM 边界调整到空域分割得到的边界以提高分割结果的准确性.

Jae^[77]用数学形态学滤波器作图像简化和求形态梯度,得到图像的空域分割区域. 然后对相邻两帧间的变化图像进行 Neyman-Pearson 检测而得出变化区域掩码. 再把这个变化区域掩码覆盖到前面空域分割的图像中. 若一个区域中的多数点属于变化区域掩码,则这个区域属于目标,否则属于背景.

以上三种方法为 MPEG 组织所推荐的三种视频目标分割方法. 这三种方法都利用了相邻两帧间的时域差分来得到目标的运动信息,再用统计学方法消除噪音的干扰. 这对于摄像机固定的情况是适宜的,对于摄像机运动的情况,则需在摄像机运动补偿后再进行时域差分,但时域差分的可靠性严重地依赖于摄像机运动估计的正确性,因此,若出现难以正确估计摄像机的运动参数的情况(如摄像机运动过大或场景复杂运动模型难以描述),则以上方法不可行.

另一大类方法是基于视频图像的运动估计和分割,目标是要把图像中存在的运动估计出来,并根据运动一致性准则把序列图像分割成不同的区域. 这些方法基本上可以分为两类:外在方法(按一个全局方案估计多运动模型)和隐式方法(用一个局部方案确定运动的不连续性). 其中外在方法又可以分为序列方法和同步方法.

序列方法^[78]是单运动估计方法的简单拓展. 其基本思路是,先用鲁棒性运动估计方

法估计出主运动,用主运动去配准图像,得到去掉主运动后的分割区域,在分割区域中再估计相对主运动,如此循环直至收敛.序列方法的主要问题是:

- 局部运动不能太大,这是由鲁棒性估计子的崩溃点决定的.
- 相对主运动的估计严重地依赖于前一级主运动估计的精确性.
- 这里区域的分割完全依赖于运动估计,而运动估计很易受到噪音的干扰,且运动估计的运动边界往往是不精确的.

相对序列方法而言,同步方法能给出更精确的结果. Jepson^[79]和 Yuille^[80]使用鲁棒性方法和混合模型(Mixture Model),把图像片用不同运动对应的层次来表示,不同的层次有着不同的运动模型.他们假设数据点在参数空间的分布满足混合模型,用EM算法来检测和分离属不同运动的区域,同时估计出每一区域中运动模型参数.这实际上就是不断地交替进行分割和参数估计,直至得到满意的结果.

上面的方法都是试图从整幅图像中找出最优的运动估计,即属于外在方法.与此不同的是,隐式方法试图用一个局部方案确定运动的不连续性.其中最简单的是基于光流场进行聚类而得出多运动估计和分割的方法^[81,82].这种方法首先计算稠密的光流场,然后用运动模型去匹配到光流场以得到分割结果.这个方法的问题在于,准确地计算稠密的光流场本身就很困难且需要很大的计算量,此外,由于估计和分割是分开的,估计造成的误差很容易传到分割阶段.

针对上述缺点, Ju^[83]局部地引入了参数化模型,他们在图像局部的区域中应用EM算法估计多个仿射运动,同时结合正则化方法,给出所谓的“壳与骨”模型,能给出光滑的层化运动模型.这种方法在光流估计灵活性与参数化模型估计的精确性和鲁棒性中取了一个平衡.

另有一些方法在运动估计之前先进行空域分割,选择出一定的区域,每一个区域假定包含一个运动. Black^[84]首先计算出一个粗糙的光流场,然后在分割后的一致性区域中用运动模型去匹配光流场,对于运动模型不正确的情况,再用局部变形去改进之. Ayer^[78]也把空域分割和运动估计结合在一起.这类方法的问题在于一个区域只包含一个运动的假设是否成立,一般而言,只有空域分割足够细,该假设才能成立,但空域分割越细,所需计算量越大,也越易受到噪音干扰,使得参数化模型估计的优势荡然无存.若空域分割较粗,每一个区域较大,计算量会小很多,但难以保证空域边界包含着运动边界,即有可能一个区域中包含着几个运动.

针对上述问题, Liao 和 Li 等^[85,86]利用多尺度数学形态学工具,采用了自上而下的空域分裂策略,根据不同的运动复杂度使用不同尺度的空域分割,组成一个区域金字塔,使之动态地满足一个区域只包含一个运动的假设,从而在计算正确性和效率间取了一个很好的平衡.

4 总结与展望

本文简单介绍了近十年来本领域研究较普遍和常用的技术和一些主要的参考文献.九十年代初,人们的研究主要集中在图像的颜色、纹理特征和一些简单的形状特征的提取技术上,在视频序列图像中,则主要是镜头的切分与关键帧的选择等技术的研究.近年来,

本领域的研究则逐渐走向更接近人类心理学和人类视觉特点的研究技术,提出了基于区域、目标物体的分析方法^[37~42]. 由于图像的语义理解目标还存在一定的困难,因此,当前的许多研究将人加在系统中,提出了许多交互式的技术^[38~40]. 在视频镜头表示上,提出了基于图像拼接的技术^[63~67].

目前,有关语义层的研究,尽管也有一些报道,但总的来说,由于其研究还有较大的困难,还没有太多的研究成果. 对图像语义理解、视频序列图像中人的行为识别和分析、视频摘要等,将是今后研究的重点.

基于图像内容的图像视频检索的研究涉及到多个研究领域,它综合利用了计算机图像处理与计算机视觉研究各方面的技术,同时对这些技术的研究和发展也起到了推动作用. 要达到实用的水平,除了上面提到的技术以外,数据库的管理、数据的组织、查询的方式、检索的效率等,也是今后研究中需要解决的问题.

参 考 文 献

- 1 Flickner M, Sawhney H S, Niblack W, Ashley J *et al.* Query by image and video content: the QBIC system. *IEEE Computer*, 1995, **28**(9): 23~32
- 2 Pentland, Picard R W, Sclaroff S. Photobook: Tools for content-based manipulation of image database. *International Journal of Computer Vision*, 1996, **18**(3)
- 3 Wactlar H D, Kanade T, Smith M A, Stevens S M. Intelligent access to digital video: informedia project. *IEEE Computer*, 1996, **29**(5): 46~52
- 4 Marsicoi M De, Cinque L, Levialdi S. Indexing picture documents by their content: A survey of current techniques. *Image and Vision Computing*, 1997, **15**: 119~141
- 5 白雪生, 徐光佑. 基于内容的图像检索及其相关技术的研究. *机器人*, 1997, **19**(3): 231~240
- 6 Sven Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 1998, **31**(8): 983~1001
- 7 Swain M J, Ballard D H. Color indexing. *International Journal of Computer Vision*, 1991, **7**(1): 11~32
- 8 Mehtre B M, Kankanhalli M S *et al.* Color matching for image retrieval. *Pattern Recognition Letters*, 1995, **16**(3): 325~331
- 9 Zhang Y J, Liu Z W. Comparison and improvement of color-based image retrieval techniques. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases VI, 1998, **3312**: 371~382
- 10 Harfner J, Sawhney H S, Equitz W, Flickner M M, Niblack W. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(7): 729~736
- 11 Funt B V, Finlayson G D. Color constant colors indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(5): 522~529
- 12 Stricker, M A, Dimai A. Color indexing with weak spatial constraints. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases IV, 1996, **2670**: 29~40
- 13 Haralick R M, Shanmugam K, Dinstein I. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, **3**(6)
- 14 Tamura H, Mori S, Yamawaki T. Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, 1978, **8**(6)
- 15 Kashyap R L, Khotanzad A. A Model-based method for rotation invariant texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(4): 472~481
- 16 Cohen F S, Fan Z, Patel M. Classification of rotated and scaled textured images using Gaussian Markov random

- field model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, **13**(2):192~202
- 17 Liu F, Picard R W. Periodicity, Directionality, and randomness; Wold features for image modeling and retrieval. *IEEE Transactions on Patter Analysis and Machine Intelligence*, 1996, **18**(7):722~733
- 18 Jain A K, Farrokhnia K. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 1991, **24**(12):1167~1186
- 19 Jain A K *et al.* Object detection using Gabor filters. *Pattern Recognition*, 1997, **30**(2):295~309
- 20 Manjunath B S, Ma W Y. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(8):837~842
- 21 Scassellati, Alexopoulos S, Flickner M. Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgments. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases II, 1994, **2185**:2~14
- 22 Persoon E, Fu K. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977, **7**:170~179
- 23 Belkasim S O, Shridhar M, Ahmadi M. Pattern Recognition with moment invariant: a comparative study and new results. *Pattern Recognition*, 1991, **24**:1117~1138
- 24 Pala P, Santini S. Image retrieve by shape and texture. *Pattern Recognition*, 1999, **32**:1517~5277
- 25 Mokhtarian F. Silhouette-based isolated object recognition through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(5):539~544
- 26 丁险峰,孔维新,杨青,卢汉清,马颂德. 基于许瓦兹表示的形状检索. 计算机工程与科学, 1998, **20**(A1):121~123
- 27 Kliot M, Rivlin E. Invariant-Based Shape Retrieval in Pictorial Databases. *Computer Vision and Image Understanding*, 1998, **71**(2):182~197
- 28 Mokhtarian F, Abbasi S. Retrieval of similar shapes under affine transform. In: Lecture Notes in Computer Science **1614**:566~574
- 29 Shen D G, Wong W H, Horace H S. Affine-invariant image retrieval by correspondence matching of shapes. *Image and Vision Computing*, 1999, **17**(7):489~499
- 30 Jain A K, Vailaya A. Image retrieval using color and shape. *Pattern Recognition*, 1996, **29**(8):1233~1244
- 31 Bimbo A D. Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(2):121~132
- 32 Belongie S, Carson C, Greenspan H, Malik J. Color-base and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval. In: Proceedings of ICCV'98, 1998
- 33 Siebert, Vancouver. Segmentation-based image retrieval. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases VI, 1998, **3312**:14~24
- 34 Campbell N W, Mackeown W P J, Thomas B T, Troscianko T. Interpreting image databases by region classification. *Pattern Recognition*, 1997, **30**
- 35 Malki J *et al.* Region queries without segmentation for image retrieval by content. In: Lecture Notes in Computer Science, 1999, **1614**:115~122
- 36 Tuytelaars t, Van Gool L. Content-based image retrieval based on local affinely invariant regions. In: Lecture Notes in Computer Science, 1999, **1614**:493~500
- 37 Carson, Thomas M *et al.* Blobworld: A system for region-based image indexing and retrieval. In: Lecture Notes in Computer Science, 1999, **1614**:509~516
- 38 Rui Y, Huang T S, Mehrotra S. Relevance feedback techniques in interactive content-based image retrieval. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases VI, 1998, **3312**:25~36

- 39 Minka T P, Picard R W. Interactive learning using a society of models. *Pattern Recognition*, 1997, **30**
- 40 Cicca G, Schettini R. Using a relevance feedback mechanism to improve content-based image retrieval. In: *Lecture Notes in Computer Science*, 1999, **1614**:107~114
- 41 Otsuji K, Tonomura Y and Ohba Y. Video browsing using brightness data. In: *Proceedings of the SPIE*, 1991, **1606**:980~989
- 42 Zhang H J, Smoliar S W. Automatic partitioning of full-motion video. *Multimedia System*, 1993, **1**:10~28
- 43 Nagasaka and Tanaka Y. Automatic video indexing and full-video search for object appearances. In: *Proceedings of 2nd Working Conf. Visual Database Systems*, 1991, 119~133
- 44 Zabih R, Miller J and Mai K. Feature-based algorithms for detecting and classifying scene breaks. In: *Proceedings of ACM Multimedia'95*, 1995, 189~200
- 45 Boreczky J S and Rowe L A. Comparison of video shot boundary detection techniques. In: *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases*, 1996, **2670**:179~179
- 46 Sanchez J M, Binefa J *et al.* Local color analysis for scene break detection applied to TV commercials recognition. In: *Proceedings of Third International Conference of Visual Information and Information System*, 1999, 237~244
- 47 Bhandarkar S M, Warke Y S, Khombhadia A A. Integrated parsing of compressed video. In: *Proceedings of Third International Conference of Visual Information and Information System*, 1999, 269~276
- 48 Yeo L, Liu B. Rapid scene analysis on compressed video. *IEEE Transaction on Circuits and Systems for Video Technology*, 1995, **5**:533~544
- 49 Ngo C W, Pong T C, Chin R T. A Survey of video parsing and image indexing techniques in compressed domain. *ISSPR'98*, 1998, 231~236
- 50 Meng J H, Juan Y J and Chang S F. Scene change detection in a MPEG compressed video sequence, In: *Proceedings of SPIE: Digital Video Compression: Algorithms and Technologies*, 1995, **2419**:14~25
- 51 Xiong W, John C M Lee, Ma R H. Automatic video structuring through shot partitioning and key frame selection. *Machine Vision and Applications: Special Issue on Storage and Retrieval for Still Image and Video Database*, 1997, **10**(2):51~65
- 52 Kong Weixin, Ren Yao, and LU Hanqing. A new scene breakpoint detection algorithm using slice of video stream. In: *Proceeding of IAPR International Conference of Multimedia Information Analysis and Retrieval*, 1998
- 53 Kong W X, Ding X F, LU H Q, MA S D. Improvement of shot detection using illumination invariant metric and dynamic threshold selection. In: *Proceeding of Third International Conference of Visual Information and Information System*, Amsterdam, 1999, 277~282
- 54 Zhang H J and Smoliar S W. Developing power tools for video indexing and retrieval. In: *Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases II*, 1994, **2185**:140~149
- 55 Hampapur, Jain R and Weymouth T E. Production model based digital video segmentation. *Multimedia Tools and Applications*, 1995, **1**:38
- 56 Günsel, Ferman A M, Tekalp M. Video indexing through integration of syntactic and semantic features, In: *Proceedings of WACV'96*
- 57 Naphade M, Mehrotra R, Ferman A M, Warnick J, *et al.* A high performance algorithm for shot boundary detection using multiple cues. In: *Proceedings of ICIP'98*, 1998
- 58 Vasconcelos N and Lippman A. A Bayesian video modeling framework for shot segmentation and content characterization. In: *Proceedings of IEEE Workshop on Content-based Access to Image and Video Libraries*, 1997
- 59 Zhang H J, Low C Y, Smoliar S W, Wu J H. Video parsing, retrieval and browsing: an integrated and content-based solution. In: *Proceedings of ACM Multimedia'95*, 1995, 15~24
- 60 Wolf W. Key frame selection by motion analysis. In: *Proceedings of ICASSP'96*, 1996

- 61 Gresles P O, Huang T S. Gisting of video documents: A keyframe selection algorithm using relative activity measures. In: Proceedings of the 2nd Int. Conf. on Visual Information Systems, 1997
- 62 Zhang Y T, Rui Y, Huang T S. Mehrotra S. Keyframe extraction by unsupervised clustering. In: Proceedings of ICIP'98, 1998
- 63 Sawhney H, Ayer S, and Gorkani M. Model-based 2D&3D dominant motion estimation for mosaicing and video representation. In: Proceedings of 5th ICCV, 1995, 583~590
- 64 Irani M, Anandan P. Video indexing based on mosaic representations. In: Proceedings of The IEEE, 1998, **86** (5):905~921
- 65 Irani M, Anandan P and Hsu S. Mosaic based representations of video sequences and their applications. In: Proceedings of ICCV'95, 1995, 605~611
- 66 Sawhney H, Hsu S, and Kumar R. Robust video mosaicing through topology inference and local to global alignment. In: Proceedings of ECCV'98, 1998, 103~119
- 67 Davis J. Mosaics of scenes with moving objects. In: Proceedings of International Conference of CVPR'98, 1998, 354~360
- 68 Zheng Q and Chelleppa R. A computational vision approach to image registration. *IEEE Transactions on Image Processing.*, 1993, 311~325
- 69 Teodosio L and Bender W. Salient Video Stills: Content and context preserved. In: Proceedings of ACM Multimedia'93, 1993
- 70 Mann S, Picard R W. Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Transactions on Image Processing*, 1997, **6**(9):1281~1295
- 71 Black M J, Anandan P. The robust estimation of multiple motions: affine and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 1996
- 72 Sawhney S, Ayer S. Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(8):814~830
- 73 Yeo B L and Yeung M M. Classification, simplification and dynamic visualization scene transition graph for video browsing. In: Proceedings of the SPIE: Storage and Retrieval for Image and Video Databases VI, 1998, **3312**:60~70
- 74 Smith M A, Kanade T. Video skimming and characterization through the combination of image and language understanding techniques. In: Proceedings of CVPR'97, 1997, 775~781
- 75 Colonnese S, Russo G. Automatic moving objects and background segmentation by means of higher order statistics. IS&T Electronic Imaging '97, Conference: Visual Communication and Image Processing, 1997, **3024**:8~14
- 76 Mech R, Wollborn M. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera. *Signal Processing: Special Issue on Video Sequence Segmentation for Content-based Processing and Manipulation*, 1998, **66**:219~232
- 77 Jae Gark Choi, Munchurl Kim, Myoung Ho Lee, Jisang Yoo. Automatic segmentation based on spatio-temporal information. MPEG 97/1743, 1997
- 78 Ayer S, Bigun L. Segmentation of moving objects by robust motion parameter estimation over multiple frames. In: Proceedings of the 4th Int. Workshop on Time-Varying Image Processing and Moving Object Recognition, 1993, 295~302
- 79 Jepson A D, Black M J. Mixture model for optical flow computation, In: Proceedings of CVPR, 1993, 760~761
- 80 Yuille P, Stolorz P. and Ultans J. Statistical physics, mixture of distributions and the EM algorithm. *Neural Computation*, 1994, **6**:334~340

- 81 Huang Y, Palaniappan K, Zhuang X, Cavanaugh J E. Optic flow field segmentation and motion estimation using a robust genetic partitioning algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17(12):1177~1190
- 82 Weber J, Malik J. Rigid body segmentation and shape description from optical flow under weak perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(2):139~143
- 83 Ju S X, Black M J, Jepson A D. Skin and Bone. Multilayer, locally affine, optical flow and regularization with transparency. In:Proceedings of CVPR, 1996, 307~314
- 84 Black M J, Jepson A. Estimating multiple independent motions in segmented images using parametric models with local deformation. In:The Workshop on Motion of Non-rigid and Articulated Objects, 1994, 220~227
- 85 Ming LIAO, Yi LI, Songde MA, Hangqing LU. Robust tracking of video objects through topological constraint on homogeneous motion. In:Proceedings of Third International Conference of Visual Information and Information System, Amsterdam, 1999
- 86 李毅,廖明,马颂德,卢汉清. 梯度水线的尺度空间. 中国图形图像学报,1999,4:265~269

卢汉清 研究员. 1961年4月生, 1992年1月在华中理工大学电子与信息工程系获博士学位. 在图像处理与理解领域进行了多年的研究, 共发表论文50多篇. 曾荣获中科院自然科学二等奖和文化部科技进步三等奖. 目前的研究方向为图像理解及应用、多媒体技术、医学图像处理、图像图形学技术融合.

马颂德 研究员. 1946年7月生, 1968年毕业于清华大学自动控制系. 1983年获法国巴黎大学计算机系博士学位, 1986年获法国国家博士学位. 曾获欧洲计算机图形学最佳论文奖与最佳技术奖, 中国科学院青年科学家一等奖. 在国际学术刊物和会议上发表了200余篇论文. 1999年被国家科技部聘任为国家重点基础研究发展规划项目“图像、语音、自然语言理解与知识发掘”项目首席科学家. 近年来主要研究方向为计算机视觉计算理论、机器人视觉、多媒体技术、图像图形学技术融合等.