基于对象的视频图象分割技术

毛燕芬 施鹏飞
(上海交通大学图象处理与模式识别研究所,上海 200030)

摘 要 随着"流媒体"技术应用的发展和 MPEG-4 基于内容的功能的提出,视频图象处理领域中,基于对象的分割技术已成为该领域的研究热点.如今视频分割研究已由基于镜头的分割发展到了通过提取视频对象面,来分割出视频对象的阶段,但目前基于对象的分割研究仍处于起步阶段,技术还很不成熟.为了推动该技术进一步发展,在深入分析分割问题本质的基础上,首先提出从分割所利用的信息角度出发来进行分割的技术;然后针对分割技术的发展趋势,深入介绍了该研究领域国内外的最新研究算法,并分析了各方法技术的贡献和不足;最后提出了一些分割技术值得进一步深入探讨的问题和研究方向.

关键词 计算机图象处理(520·6040) 基于对象的视频分割 视频对象面 MPEG-4 流媒体
中图法分类号: TP391.41 TN941.1 **文献标识码**: A **文章编号**: 1006-8961(2003)07-0726-06

Object-based Video Segmentation Technology

MAO Yan-fen, SHI Peng-fei
(*Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai* 200030)

Abstract Due to the development of streaming media and emerging of content-based functionalities in MPEG-4, object-based segmentation technology becomes a popular research in video field. Video segmentation provides an easy and efficient way for video retrieval and coding, whereas it is a difficulty issue in computer vision. From shot-based to extraction of video object planes (VOPs), video segmentation develops to extract video object(VO). In fact, shot-based segmentation is one of the primary steps in the true object-based video segmentation. Currently, segmentation technology for content-based representation is still premature. This article gives an overview of existing techniques for object-based segmentation. The performance, relative merits and limitations of each of the approaches are comprehensively discussed and contrasted. Different from the traditional classification approaches, this paper presents a new view of analyzing the recent research in this area. Some important and advanced video segmentation algorithms are analyzed and compared both in theory and experimental results. According to different information used in those algorithms, they are classified into several classes. This classification criterion is more revelatory and useful to developing new segmentation algorithm. Finally, some existing problems worth discussing and directions need to further research are proposed.

Keywords Object-based video segmentation, Video object planes (VOPs), MPEG-4, Streaming media

0 引 言

随着网络带宽的增加和多媒体技术的发展,互联网"流媒体"的应用日益广泛.国际压缩编码标准 MPEG-4 的制定,不仅使得与流媒体应用有关的技术得到了标准化,而 MPEG-4 中,基于内容的交互功能的提出,同时也对现有视频处理技术提出了挑战.通过定义视频对象面(Video object planes, VOPs),MPEG-4 提出了基于内容的交互功能.这使得视频分割研究也逐渐集中于基于对象的分割.所谓视频分割就是通过将视频序列中每一帧图象分割成任意形状的图象区域,也就是所谓的视频对象面 VOPs,并使每一个 VOP 能描述一个具有语义意义的对象或感兴趣的视频内容[1].

视频分割的传统做法是基于帧的分割,其任务主

基金项目:国家"973"项目(G1998030408)
收稿日期:2002-07-15;改回日期:2003-03-03

要集中于镜头(Shot)边缘检测,以便将视频在时间轴上分成镜头的集合,从某种程度上讲,基于镜头的视频序列分割只是真正视频分割的一个前期步骤;而基于对象的视频分割与传统的分割方法不同,它的最终的目的不仅仅是要将视频序列在时间轴上作切分,更重要的是要得到具有实际意义和使用价值的信息.因而,如何提取或分割出视频对象(Video object,VO)就成为基于对象的视频分割中一个重要的问题,然而,完全自动的提取语义上有意义的视频对象是非常困难的,相关的研究工作仍处在初期阶段.

通过提取视频对象可以很大程度地提高压缩效率,这不仅可为传输和存储视频图象提供便利,还可对静止或动态场景进行查询和交互.由于利用提取得到的"关键帧",不仅可以更好地进行检索,也能为互联网浏览和查询提供一个简便而有效的方式,因此,基于对象的视频分割技术已成为基于内容的检索、编码和视频数据库操作的关键技术,其不仅是计算机视觉研究的难点之一,同时也是新一代多媒体交互、流媒体应用等新兴领域的研究热点[2].

1　视频图象序列

视频序列可看作一类特殊的三维图象,它是三维场景在二维图象平面上某一时刻的投影.所谓视频对象是指视频图象序列中具有语义意义的实体[3],它可以是任意形状的区域,一般由它的边缘轮廓来界定,其内部通常具有不一致的低层次特征,例如,颜色、亮度、纹理等.

2　视频分割

通常意义下的分割是指根据给定的标准对图象和视频进行标记分离的一种操作[2],而基于对象的视频分割则是指将视频序列按一定的标准分割成区域,其目的是从视频序列中分离出语义上有一定意义的实体,也就是要提取出 MPEG-4 所定义的视频对象.目前这方面的研究一般都是将运动对象视为有意义的实体,所以其分割是指利用对象的某些特征信息来将前景(foreground)运动物体从背景(background)中分离出来.由于分割的复杂性,国内当前的研究大都只针对静止的背景,但国外已有少量的研究开始涉及运动背景的情形.

3　分割算法研究

以往对视频分割方法的分类通常是从人的参与程度、使用的数学工具等角度出发.本文则是从分割技术所利用的信息出发来研究不同的算法,因为这样更能深入问题的本质,并可对今后的研究提供一些很有价值的启示.

分析总结国内外最新的研究成果可见,基于对象的分割技术主要可分为利用对象的运动信息、密度信息、时空亮度梯度信息以及综合利用不同信息等几大类.

3.1　利用运动信息进行分割

从运动一致性的角度看,由于前景运动对象通常有不同于背景的运动,因此,可以将运动信息作为一个非常有用的特征来进行分割.这也是目前分割算法中广泛采用和最具典型的方法.

3.1.1　运动估算

利用运动信息进行分割,首先要估算密度运动场,然后才能根据运动信息对场景进行分割.所谓运动场是指三维物体的实际运动在图象平面上的投影,而光流场则是由图象亮度随时间的变化引起的,而且光流场并不对应于运动场[2].基于光流法的分割就是通过研究光流场,从视频序列中近似计算不能直接得到的运动场,然后再根据运动场的运动特性进行分割.

光流法是用于估算运动场的一个较普遍的方法.假设 $\boldsymbol{I}(x,y;t)$ 表示连续时空亮度分布,并且沿着运动轨迹,物体的亮度保持不变,则可以得到光流约束条件(OFC——Optical flow constraint)

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{I}(x,y;t)=\frac{\partial \boldsymbol{I}}{\partial x}\cdot\frac{\partial x}{\partial t}+\frac{\partial \boldsymbol{I}}{\partial y}\cdot\frac{\partial y}{\partial t}+\frac{\partial \boldsymbol{I}}{\partial t}$$

$$=\langle\nabla \boldsymbol{I},(u,v)\rangle+\frac{\partial \boldsymbol{I}}{\partial t}=0 \qquad (1)$$

其中,$\nabla \boldsymbol{I}=\left(\frac{\partial \boldsymbol{I}}{\partial x},\frac{\partial \boldsymbol{I}}{\partial y}\right)$,$u=\frac{\partial x}{\partial t}$,$v=\frac{\partial y}{\partial t}$,$\langle\cdot,\cdot\rangle$ 为向量的内积.从上式可见,只有法向流——梯度 $\nabla \boldsymbol{I}$ 的流量分量可以被估算出来,由于正交分量在点积后消失,所以其可以是任意值,而不会改变内积值.

大家知道,运动估算存在遮挡问题和孔径问题,其中,遮挡问题指不能为覆盖/显露的背景像素建立对应;孔径问题是指未知量的个数大于已知量的个数.由此可见,需要加入附加的假设模型来模拟二维运动场的结构,才能得到唯一确定的解.设 (x,y) 为

第 k 帧的像素坐标,$(\hat{x},\hat{y})$表示第 $k+1$ 帧的像素坐标,这样在正交投影下,可以得到 6 参数的仿射模型

$$\hat{x}=a_1x+a_2y+a_3,\hat{y}=a_4x+a_5y+a_6 \quad (2)$$

而在透射投影下,则可以得到 8 参数透射模型

$$\hat{x}=\frac{a_1x+a_2y+a_3}{a_7x+a_8y+1},\hat{y}=\frac{a_4x+a_5y+a_6}{a_7x+a_8y+1} \quad (3)$$

参数模型是由多个像素结合在一起估算出来的,所以受噪声的影响较小,但参数模型只适用于刚体运动.

3.1.2 运动分割

Murray 和 Buxton 最先提出用最大后验概率(Maximum a posterior,MAP)的方法来分割视频序列[4],所以也称 M-B 法,这种最大后验概率法属于贝叶斯法.众所周知,贝叶斯法是进行运动分割效果最好的方法之一,其主要思想是在给定观察图象 $\boldsymbol{O}$ 的条件下,寻找分割结果 $\boldsymbol{X}$ 的最大后验概率 $P_{\max}(\boldsymbol{X}|\boldsymbol{O})$.利用 Bayes 规则,$P_{\max}(\boldsymbol{X}|\boldsymbol{O})$可由 $P(\boldsymbol{O}|\boldsymbol{X})P(\boldsymbol{X})$得到.分割的先验概率模型假设为马尔可夫随机场(Markov random field,MRF)的一个样本,该先验概率 $P(X)$为 Gibbs 分布.$P(\boldsymbol{X}|\boldsymbol{O})P(\boldsymbol{X})$可通过模拟退火(SA)方法来搜寻得到使后验概率最大的标记.最大后验概率法使用的 MRF 能量函数包括空域平滑度、时域连续性等参数.贝叶斯法的最大缺点是计算复杂,不适用于实时处理.

Bouthemy 和 Francois 提出了类似的方法[5],但是它们采用的能量函数只包含空间的平滑度参数.其最后采用迭代条件模型(Iterated conditional mode,ICM)代替 M-B 法的模拟退火(Simulation anneal,SA)来进行优化,虽可使得运算速度加快,但在局部最小值的求解时仍会遇到困难.

Wang 和 Adelson 提出了一个用层次表示图象序列的方法[6],这可与 MPEG-4 的 VOP 技术联系起来.该算法首先估算光流场;然后将帧图象分成子块,再对每一个子块通过线性回归计算仿射运动参数来得到运动假设的初始集合;最后将像素通过运用迭代自适应 K 均值聚类方法来进行分组,并且使用时域中值滤波器来得到每一个对象的单一表示.

这种算法存在的不足是:(1)仿射变换不能充分描述非刚体运动;(2)没有考虑颜色、亮度等其他信息,精度完全依赖于光流估算;(3)层次结构的处理需要很长的帧序列,所以不具有实时性.

由以上分析可见,基于光流法的分割技术,因为只考虑利用光流数据来进行决策,所以受到被估算的光流场精度的限制.这也使得这些方法不可避免地受到噪声的影响,而且,分割所得的运动对象的边缘精度不够,在运动不完全情况下,则会产生分割结果不完整等问题.另外,由于运动场并不是很可靠,因此通常在重叠物体边界或纹理不突出区域产生错误,而这种错误的运动矢量会对分割结果产生明显的影响.

考虑到上面方法存在的不足,很多算法用变换检测模板(Change detection mask,CDM)$\boldsymbol{C}$ 来代替运动场,如 Mech 和 Wollborn 提出了从估算变换检测模板中提取视频对象面或对象模板的方法[7].该方法首先通过使用全局阈值来计算连续两帧图象的差值残余变换检测模板;然后通过局部自适应阈值来在迭代松弛中精细 CDM,以增强空间连续性,同时,采用时域存储器结构来记录每一个像素变化的标记,再根据对象的运动来决定存储器的时间深度 L.其对于快速运动的物体,L 很小;而对于运动较慢的物体,L 则很大,这样就可以得到好而稳定的时域一致性.

为了调节对象的边缘,以适应亮度边缘的变化,亮度边缘由 Sobel 算子计算得到.当背景边界与对象模板边界靠得很近时,这种边界自适应方法会失效,而且,这种方法没有充分利用空域信息,以得到更精确的边缘定位.但值得一提的是,文献[7]提出的二维形状估算方法由于考虑了摄像机的运动,因此对噪声具有鲁棒性.它包括 4 个步骤,图 1 为其对象模板计算的流程方块图.图中 MEM(变换检测模板的存储器)表示变换检测模板的存储器,$\boldsymbol{M}$ 表示对象模板,$\boldsymbol{D}$ 表示位移向量场(Displacement vector field,DVF).在摄像机运动估计和补偿中,采用了式(3)8 参数透射模型.

这种方法虽计算简单,但仍存在如下两个缺点:一是除非运动对象包含足够的纹理,否则只有遮挡的地方被检测出有变化,而对象的内部将被视作没有变化;二是,如果对象或对象的一部分在某一段时间停止运动,则会被漏检,这是不适合基于内容的分割应用要求的.为了解决这个问题,该算法使用了存储器,但检测到的对象会比实际值大.对于运动量较小,如对头肩运动的视频序列来讲,CDM 比运动场有效,但对非刚体运动,CDM 则不是很有效.

为此,Meier 等人提出用二维二值化模板来提取感兴趣的对象,并在视频序列中跟踪它[8].由于该算法基于模式识别和对象跟踪技术,从而避免了许多与运动估算有关的问题.其技术的核心是采用一

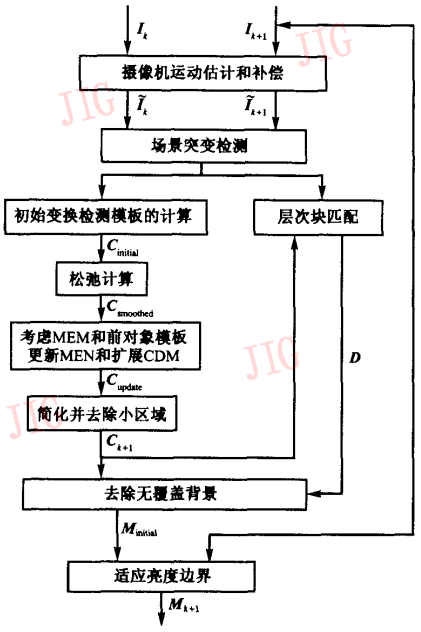
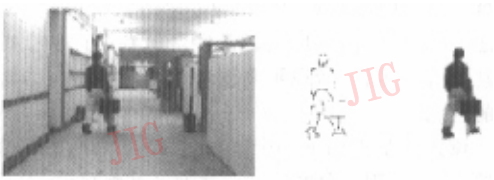


图 1　对象模板计算流程方块图

个对象跟踪器，这样在整个视频序列中就建立了对象的时域相关性. 这一点对于基于内容的检测功能是非常重要的，因为这样可以时刻跟踪对象，即使它在任意一段时间内停止运动也能跟踪.

该算法的总体思路是，首先定义运动连接分量，在颜色或亮度变化区域中自动检测出运动对象；然后用 Hausdorff 距离检测来得到二值掩模图，并在后续帧中查找相匹配的模板，由于对象有旋转和形状变化，因此需要在每一帧中对模板进行更新，这也使得与运动估算相关的遮挡问题不存在；最后跟踪器输出的是一系列二值掩模图，从中就可以提取出 VOPs. 图 2 显示了 VOP 的提取结果[8].



(a) 原图　　(b) 二值掩模图　　(c) VOP 提取结果

图 2　Hall monitor 视频序列分割结果

图 2(a)的背景较复杂，且含有很多噪声，但从分割结果来看，该算法对噪声的鲁棒性很好. 图 3 是该算法的流程图，其中主要包括运动对象检测、模型初始化、模型更新和 VOP 提取 4 个功能块. 其中运动对象检测模块首先采用最有效的 Canny 算子来得到轮廓图象；然后用 Hough 变换来检测任意形状的二值化对象；再用 Hausdorff 距离来检测查找的模板，当 Hausdorff 距离小于预定的阈值时，即得到匹配. 这种检测方法不仅计算效率高，且对噪声和形状变化具有鲁棒性，从图 2 也可见，它的实际检测效果是令人满意的.

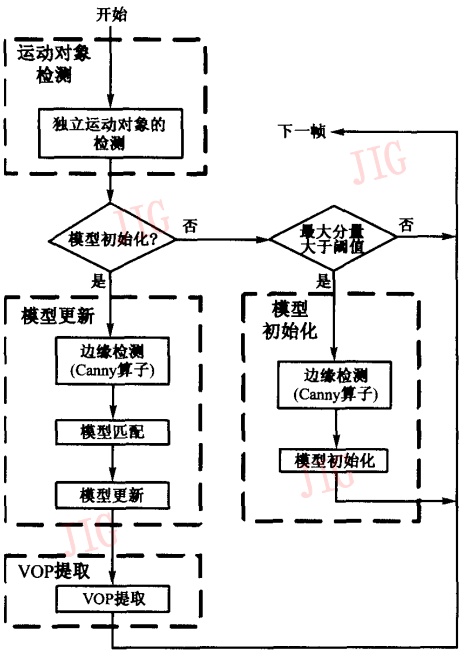


图 3　VOP 提取流程图

这里采用式(2)6 参数仿射模型，并用最小中值二乘法(Least Median Square，LMS)代替线性回归来估计参数. 由于对象的旋转或形状改变，因此需要每一帧均更新模型，对于复杂背景或摄像机运动的情况，这一步比较困难. 考虑到对象的各个部分运动速度等特性不一样，这里采用如下两种更新方法：一种适用于相对整体运动变化较慢的部分；另一个适用于变化或运动较快的部分. 最后综合两者即可得到好的更新机制. 这种分割方法在视频对象运动量较大的时候，能获得较好的分割效果.

因为复杂背景对 VOP 的提取不利，所以需要在模型匹配和更新之前先去除复杂背景，但是如果利用帧间差的方法，则可能会去除在两帧中停止运动的对象，而且对噪声敏感. 这里先采用计算一个像

素被归为边缘的频率的方法来去除复杂背景,如果这个频率超过预定的阈值,则认为这个像素是边缘的一部分,并将其去掉,否则,归为运动对象;之后,若找出每一行的第一和最后一个模板点,则位于这两者之间的像素,即为 VOP. 而对每一列也采取相同的策略来寻找 VOP.

实验结果表明,该算法对于背景较复杂,如 Coastguard 序列的 VOP 提取和对于有快速非刚体运动的 Bowing 序列都非常有效,但是由于物体的不完全运动,因此对于 Grandma 序列的分割存在不完整性,这也是基于运动信息的分割算法存在的普遍问题. 图 4 为运动不完全时的分割结果[8],这种情形下具有语义意义的某些部分(身体部分)没有被分割出来.

图 4 运动不完全时的分割结果

总之,这种方法的优点是可以很好地跟踪对象,即使物体在某一时间段停止运动也能跟踪,但是,其分割的结果需依赖于由第 1 帧图象得到的初始分割结果,而且,通过全局运动补偿在后续帧中精确匹配背景相当困难. 另外,如何得到初始模板,并及时更新也存在较大困难.

3.2 利用密度信息进行分割

Neri 等人提出了利用高阶统计方法(Higher-order statistics,HOS)将运动对象从静止背景分离出来的自动分割算法[3],该算法主要利用了密度信息,同时还考虑了运动背景的情况. 该算法采用四阶矩累积量来检测密度变化区域,累积量图象从一系列连续的图象中得到. 这里假设背景噪声具有高斯分布,因为高阶累积量对高斯过程不敏感,所以当噪声是高斯有色噪声时,高阶累积量在理论上可以完全抑制噪声的影响[9]. 通过累积量阈值化后得到的模板包括运动区域和没有运动的区域两种标记. 最后采用四级联开-闭形态算子来去除二值化图中的一些孤立的小区域,以得到更精确的边界.

该算法已在 MPEG-4 的核心实验 N2 中被反复校验,其中对典型序列 Akiyo、Mother and daughter、Container 和 Hall monitor 等进行了实验. 结果表明,该算法对室内和室外场景序列都很有效,但是当背景运动相对前景运动较快时,会产生一些分割错误,而且,分割算法的精度需依赖于高阶统计和形态滤波的结果. 但该算法的计算复杂度相对前面介绍的方法有所下降.

3.3 利用时空亮度梯度信息进行分割

Hotter 和 Thoma 等人提出了基于分层结构的自顶向下法[10],其主要利用时空亮度梯度信息. 通过对不同的运动物体计算出不同的运动参数来将不同运动对象分割出来.

该算法首先用变换检测模板检测当前帧,假设每一个连通的变化区域对应于一个对象,则从变化最大的区域开始,首先直接从时空图象的亮度和梯度图中估计运动参数;然后用最小二次方判定法将参数模型从前一帧到下一帧拟合到整个变化区域上去;最后,根据单一模型和每个区域或其子域的拟合程度,把整个区域分割成连续的小区域. 该算法不同于前面的 MAP 法,后者是依据于一些归并准则把许多子区域分组组合在一起,以形成分割区域.

利用时空图象亮度梯度信息的方法需依赖于亮度梯度信息,由于图象的亮度梯度对观测噪声非常敏感,所以分割精度易受噪声影响. 这种方法的优点是不需要光流场的估计和运动信息.

3.4 利用运动和亮度信息进行分割

通常情况下,由于利用单一信息不能得到满意的分割效果,因此,现在越来越多的研究开始利用信息融合技术来将根据不同信息分割得到的结果进行融合,以得到较好的分割效果.

Pedersini 等人提出了同时利用运动场和亮度信息的分割方法[11],并通过聚类过程来分析运动场,以得到仿射模型,而无效区域则由基于马尔可夫场的区域估算获得. 该方法的基本思路是由对象在三维场景中的运动来产生相应二维图象序列的一致性运动场. 因为图象序列的某些区域在时间轴上的运动具有一致性,所以可将图象分割成具有一致运动的区域.

该方法基于如下一个事实,即当场景由一些数目较小,但面积较大的运动一致性区域组成时,运动模型可通过一部分无效区域的线性回归来精确获得. 这也就是说,可以在运动场不可靠的地方分离出对象区域. 除此之外,算法中还提出了运动补偿亮度帧差(Motion compensation lightness difference,MCLD)的概念.

对 CIF 格式的图象序列 Flower garden 和 Table tennis 进行测试的结果表明,在场景物体有明显的可见边界和足够的纹理特征时,该方法可以得到很好的分割效果.

4 结　论

本文从分割技术所利用的信息出发,研究分析了不同的算法的优缺点.通过分析可以看到,分割技术尤其是基于对象的分割技术还存在很多问题,需要进行继续深入的研究.

综上所述,尚有如下一些问题尚待解决:(1)以上模型研究,由于大都只考虑静止背景下,一个运动对象的情况,所以运动背景以及复杂背景中多运动对象的分割是今后需要研究的一个方面;(2)现有算法对于只有局部运动或运动不充分对象的检测问题还未能很好解决,而且大都不能分离重叠的前景物体,这是通过估算变换检测模板进行分割的一个缺陷;(3)如何有效地去除掉阴影、反射和噪声对前景对象提取产生的干扰也是一个难点;(4)现有算法的实时性问题和快速算法的研究也是非常重要的,因为这两个问题的解决会直接影响分割算法应用于实际系统的速度和程度.

参 考 文 献

1 Meier T, Ngan K N. Automatic segmentation of moving objects for video object plane generation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1998,8(5):525～538.

2 季白杨,陈纯,钱英. 视频分割技术的发展[J]. 计算机研究与发展,2000,38(1):36～42.

3 Neri A, Colonnese S, Russo G *et al*. Automatic movingobject and background separation[J]. Signal Processing, 1998,66(2):219～232.

4 Murray D W, Buxton B F. Scene segmentation from visualmotion using global optimization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1987,9(2):220～228.

5 Bouthemy P, Francois E. Motion segmentation andqualitative dynamic scene analysis from image sequence [J]. International Journal of Computer Vision, 1993,10(2):157～182.

6 Wang J Y A, Adelson E H. Representing moving imageswith layers[J]. IEEE Transactions on Image Processing, 1994,3(5):625～638.

7 Roland Mech, Michael Wollborn. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera [J]. Signal Processing, 1998, 66(2):203～217.

8 Meier M, Ngan K N. Video segmentation for contented-based coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8):1190～1203.

9 詹劲峰. 视频分割理论与应用研究[D]. 上海:上海交通大学,1998.

10 Hotter M, Thoma R. Image segmentation based on object oriented mapping parameter estimation [J]. Signal Processing, 1988,15(3):315～334 .

11 Pedersini F, Sarti A, Tubaro S. Combined motion and edge analysis for a layer-based representation of image sequences[A]. In: Proceedings of IEEE International Conference on Image Processing [C], Lausanne, Switzerland, 1996, 1:921～924.

毛燕芬　1975 年生,2001 年毕业于西北工业大学自动控制系控制理论与控制工程专业,现为上海交通大学图象处理与模式识别研究所博士生.主要研究方向为视频图象处理与智能交通系统.

施鹏飞　1940 年生,上海交通大学图象处理与模式识别研究所所长,IEEE 高级会员,教授,博士生导师.研究领域为图象分析、模式识别、智能技术与系统.发表论文 80 余篇.