

基于知识表示与有监督学习的动态红外图像分析 *

张 蔚¹, 张天序², 沈 俊³

(1. 武汉理工大学 自动化学院 模式识别与智能系统教研室, 湖北 武汉 430070; 2. 华中科技大学 图像识别与人工智能研究所 图像信息处理与智能控制教育部重点实验室, 湖北 武汉 430074; 3. 法国波尔多第三大学, 计算机图像处理研究室, 吉伦特 波尔多)

摘要: 动态图像分析是一个令人感兴趣的研究课题, 广泛地应用于交通监测、场景监控和预警等方面。由于红外图像的噪声相对较大, 对比度不高, 因此红外序列图像的动态分析有着自己的特点。提出了一种基于知识表示与有监督学习的动态红外图像分析方法, 能有效地利用序列图像全局建模进行红外序列图像中的运动物体检测、特性判别行为分析。

关键词: 动态图像分析; 运动物体检测; 有监督学习; 知识表达; 场景建模
中图分类号: TP391 **文献标识码:** A **文章编号:** 1007-2276(2005)02-0216-05

Dynamic infrared imagery analysis method based on knowledge representation and supervised learning*

ZHANG Wei¹, ZHANG Tian-xu², SHEN Jun³

(1. Laboratory for Pattern Recognition and Intelligent System, School of Automation Wuhan University of Science and Technology, Wuhan 430070, China; 2. Key Laboratory of Education Ministry for Image Processing and Intelligent Control, Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan 430074, China; 3. Bordeaux 3 University, Laboratory for Computer Image Processing, Bordeaux, France)

Abstract: Dynamic image analysis is an interesting research field, and it is widely used in traffic monitoring, scene surveillance and warning forecast. Dynamic analysis of infrared imagery sequence has its character because of its high noise levels and low contrast. A dynamic infrared imagery analysis method based on knowledge representation and supervised learning is proposed. This algorithm can efficiently apply global modeling of sequential images to moving objects detection, pattern discrimination and behavior analysis in infrared imagery sequences.

keywords: Dynamic image analysis; Moving object detection; Supervised learning; Knowledge representation; Scene modeling

收稿日期: 2004-04-01; 修订日期: 2004-06-20

* 基金项目: 国家自然科学基金重点资助项目(F60135020); 中法先进研究计划(PRA SI 01-03)

作者简介: 张蔚(1980-), 女, 湖北武汉人, 硕士, 主要研究方向为动目标检测、目标跟踪、图像理解和图像分析。

0 引言

对于序列图像中运动目标的检测与分析,近几十年来国内外学者已经提出了多种方法^[1],如基于光流场、神经网络、差分、自适应滤波器以及特征匹配等方法。文中试图在序列图像全局建模的基础上,从宏观的角度对运动目标进行检测和分析,从而摆脱纯微观图像处理方法的局限性。

对红外序列图像进行场景建模,包括多个层次。不同层次的面向主体不同,如面向像素层次、面向区域层次和面向对象层次等,不同层次之间互相关联。从低层次到高层次的转变过程为:

低层次 → 高层次

图像处理、分析 → 图像理解

基于统计的较低层次(面向像素层次、面向区域层次)建模:主要包括平均亮度、区域纹理(基于共生矩阵的各种统计量)、灰度统计直方图、区域边缘率分布、光流矢量统计值、形状统计值和位置信息。

基于规则的较高层次(面向对象层次)建模:场景特性,如在一个“停车场”序列图像中有建筑物、小车和行人等场景对象,建筑物上有多个平行的深色条带,轮廓有直角且整齐;小车的车窗是一个亮度非常低的四边形区域,它与车身的亮度差异极大;人的亮度很高,且相对成像面积很小等。

这些特征一般称为“语义级特征”。我们在观察一幅图像时会将其分成一些视觉上有意义的区域,每个场景区域都有着相应的语义内容。基于语义的场景建模从人的视觉理解出发,着眼于提取图像中符合人类视觉概念的部分,所以可称为语义级建模。

在语义级建模中关键是语义提取,也就是规则表示。规则表示要解决两个问题^[2]:

(1) 语义特征具有模糊与不确定性,要提取图像的语义特征就必须解决特征的“抽象化”问题,将定性的语义特征转变为计算机能够理解与接受的定量的规则表述,这里,有必要将语义分解为相应的语义基元,通过分析语义基元的特点,寻找适当的图像处理与分析方法来实现规则表示。例如“建筑物上有多个平行的深色条带”这句话里面就有多个较抽象的概

念,如“多个”、“深色”、“条带”,这些词都是不严格的定性用语,必须用计算机语言对这些抽象的概念进行描述与表达;

(2) 语义特征离不开人的视觉理解,所以必须在“理解”图像的基础上进行全局建模。

1 序列图像场景建模

采用人工智能中的框架结构来描述和表示红外序列图像中的各种场景对象:

<框架名>

<公共槽部分>……

<私有槽部分>……

<约束部分>……

<公共槽部分>是由各种场景对象都具备的一些特性组成,各个场景的<公共槽部分>构造都是相同的,在实现的时候可以只用定义一个结构体。<私有槽部分>根据建模对象的不同而有着不同的内容,如<小车私有槽>、<行人私有槽>等的内容都是各自场景规则的表述。各部分又可详细表示为:

(1) <公共槽部分>

槽 1 局部灰度统计直方图。

槽 2 位置信息值。

槽 3 光流矢量统计量:

侧面 1 光流大小局部统计直方图;

侧面 2 光流方向局部统计直方图。

槽 4 形状统计量;

侧面 多个不变矩统计值。

(2) <小车私有槽>

小车的车窗所在位置是一个亮度非常低的四边形区域,与车身的亮度差异极大。另外,移动的小车所在处有着大小较为一致、位置集中的光流矢量等。

(3) <行人私有槽>

行人所在位置的像素亮度较高,纹理较为均匀,且相对成像面积很小。同样,运动的人所在处有着大小较为一致、位置集中的光流矢量等。

对红外序列图<公共槽部分>进行填槽时,<公共槽部分>的场景信息可以直接提取,填槽的过程如图 1 所示。从原始序列中挑出连续的若干帧图像作为训

练样本,通过人工参与的方法从中提取出各类静止场景与运动场景样本。这些场景样本通过预处理可以得到各种预处理图,以供后续统计工作所用。从像素级和块级等层次对场景样本进行公共槽信息数据统计,统计得到的数据填入场景框架中。

各类场景的私有槽部分实际上是由一些规则组成,在表述这些规则的时候要用到它们的公共槽中的信息。从某种意义上来说,在实际应用的时候,根据需要挑选其中的部分规则即可。

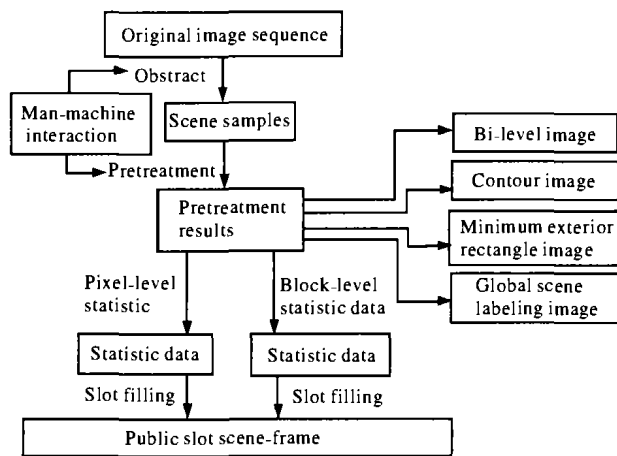


图 1 序列图像公共槽部分填槽过程示意图

Fig.1 Public slot filling process of serial images

2 运动区域检测

基于由粗到精的运动目标检测思想,先进行序列图像中运动区域的检测,然后再在这些感兴趣区域中提取真正的运动目标。这里,我们采用一种有监督学习的基于帧间图像差异显著性检验和最小代价贝叶斯判别准则的运动区域预检测算法^[3]。

将静止的背景块的集合看作是类 R_1 ,运动的块的集合是类 R_2 。令 $P(R_1)$ 和 $P(R_2)$ 分别为静止块和运动块的出现概率。同样,令 $p(x|R_1)$ 和 $p(x|R_2)$ 分别为一对图像中静止块和运动块的各自的平均灰度差分的概率分布函数。这样,运动块的检测就可以看成是一个两类的模式分类过程:静止块或运动块。可以用最小风险贝叶斯判决准则来选择最佳的分类门限 t 。

假设门限 t 将 R_1 分成 R_{11} 和 R_{12} ,其中 R_{11} 为静止区域的正确分类。令 R_{11} 的代价为 $C_{11}=0$, R_{12} 为将静止

区域分到运动物体区域的错误分类,其代价为 C_{12} 。同样,门限 t 可能将 R_2 分成 R_{21} 和 R_{22} ,其中 R_{22} 为运动区域的正确分类。令 R_{22} 的代价为 $C_{22}=0$, R_{21} 为将运动物体区域分到静止区域的错误分类,其代价为 C_{21} ,则总代价 $r(t)$ 可以表示为:

$$r(t) = C_{12}P(R_1) \int_t^{+\infty} p(x|R_1) + C_{21}P(R_2) \int_0^t p(x|R_2) \quad (1)$$

这里 $r(t)$ 可通过调整 t 而达到最小化。最小的 r 所对应的 t 值就是灰度差分值的最佳分割门限,即:

$$r(\hat{t}) = \min r(t) \quad (2)$$

在运动区域预检测的基础上,还要进行以下处理:

(1) 对检测图中的连通域进行标记,凡空间连通的子区域赋予同一整数标记值,否则赋以不同标记。整数标记值从 1 开始,加 1 递增,最大的标记值就表示连通域的个数。

(2) 去除每个连通域内部的微小孔洞。

3 基于场景建模的运动目标检测与分析

采用运动区域预检测算法得到的检测图中有的区域存在虚警,也就是说有些区域中某些位置并没有运动物体存在,却被错误检测出来。因此,我们希望从宏观上对这些区域作一个判别,从中剔除那些虚假的区域。第 1 节所做的场景建模事实上是建立了一个宏观的场景库,在此全局建模的基础上可以对运动区域预检测图做如图 2 所示的操作。

首先判断检测出来的运动区域与静止场景是否有重叠的区域,如果没有,认为不存在虚假运动区域,否则就把这些重叠的区域标记在空白模板上。然后,分析当前帧在重叠区域与静止场景相应位置的特性是否一致,如果一致则说明该区域实际上不是运动目标所在的区域,应该从运动区域检测图中去掉,否则认为该区域是运动目标所在区域,予以保留。对所有的重叠区域进行上述操作后得到的运动区域检测图就可以进行局部二值分割,从而得到运动目标检测结果图。

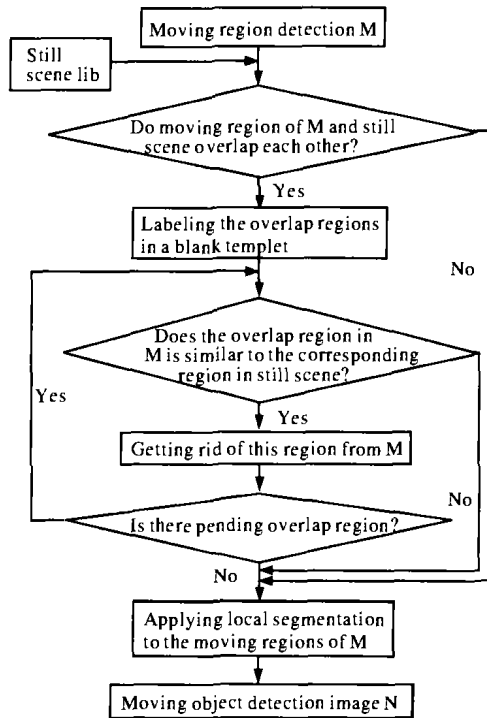


图 2 基于场景库的运动目标检测流程图

Fig.2 Flow chart of moving object detection based on scene lib

4 实验结果

实验对象为自行拍摄的“停车场”红外序列图像(大小为 320×280,256 灰度级)。从原始样本帧中人工提取各种典型场景对象,用于分类建模。从“停车场”序列图像中选取 43~46 共四帧图像作为训练样本,对第 43 帧中的静止场景人工加轮廓,如图 3 所示。其中静止场景还可以分为两类:恒静场景与暂静场景。恒静场景是指建筑物、地面障碍物等绝对静止的场景对象。暂静场景包括小汽车、小巴等暂时停止不动的场景对象。实际应用时,每隔若干帧要对暂静场景对象



图 3 第 43 帧中的静止场景人工加轮廓图

Fig.3 Manually drawn contour image of still scenes in the 43rd frame



图 4 第 43 帧中的运动对象人工加轮廓图

Fig.4 Manually drawn contour image of moving objects in the 43rd frame

进行刷新检测,若其发生改变则应从静止场景库中将其除去。对第 43 帧序列图像中的运动对象人工加轮廓,如图 4 所示。从第 43 帧序列图中提取出来的“建筑物”、“小车”、“小巴”、“行人”以及“停车泊位”最小外接矩形图如图 5 所示。

令相邻的两帧图像为一个图像对,例如,第 49 帧和第 50 帧图像构成第 49-50 帧图像对。根据上述算法对第 58-59 帧图像对进行运动区域预检测,得到的结果如图 6 所示。可以看出图中标记的 a 区域底部有一高亮度条带子区域,对照原始第 59 帧图像可知该高亮度区域属于一辆静止的小巴,希望能消除该虚假运动区域。根据已经建立的静止场景库,采用宏观判别消除运动区域预检测图中的虚警,得到的结果如图 7 所示。

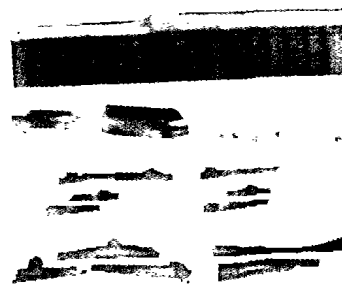


图 5 第 43 帧中“建筑物”、“小车”、“小巴”、“行人”以及“停车泊位”提取图

Fig.5 Sample images of “building”, “car”, “mini bus”, “passenger” and “berth” abstracted from the 43rd frame

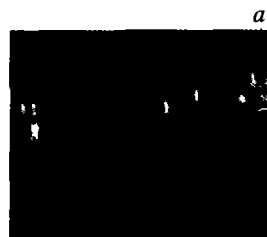


图 6 第 58-59 帧图像对运动区域预检测图

Fig.6 Moving region detection of the 58-59th frame

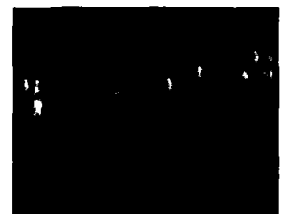


图 7 图 6 消除虚警后的结果检测图

Fig.7 Detection result of Fig.6 after eliminating false warning

前面建立的场景库包括车辆和行人的灰度分布信息,如图 8~图 11 所示(其中行人的标号如图 13 所示)。横坐标为灰度值 n ,纵坐标为频数 $f(n)$ 。可以看出车、小巴和行人的灰度分布特性以及大小尺寸是有

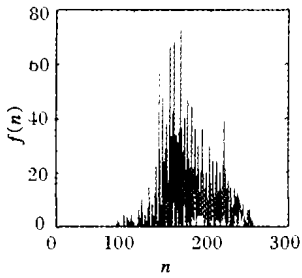


图8 小车的灰度分布直方图

Fig.8 Histogram of car

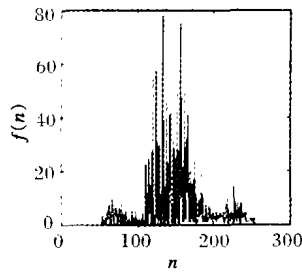


图9 小巴的灰度分布直方图

Fig.9 Histogram of mini bus

较大差异的。可以构造一个线性分类器,通过场景库中的信息进行样本训练,从而对运动目标进行分类。

根据<行人私有槽>中提到的规则“行人所在位置的像素亮度较高,纹理较为均匀”,可以对被判为行人所在的运动区域直接做局部二值分割,分割出来的高亮度区域就是行人目标。对图 7 进行局部二值分割得到运动目标预检测图,如图 12 所示。可以看出,检测的结果与实际运动目标比较吻合。

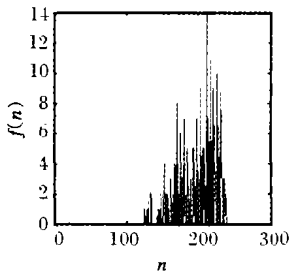


图 10 第 43 帧中行人 1 的灰度分布直方图

Fig.10 Histogram of passenger 1 in the 43rd frame

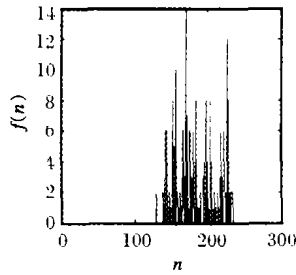


图 11 第 43 帧中行人 2 的灰度分布直方图

Fig.11 Histogram of passenger 2 in the 43rd frame



图 12 第 58-59 帧图像对运动目标检测图

Fig.12 Moving object detection of the 58-59th frame pair

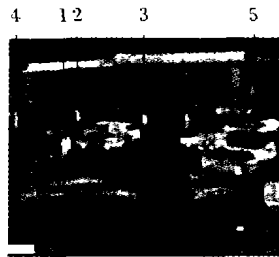


图 13 第 43 帧行人编号

Fig.13 Passenger numbers in the 43rd frame

将运动场景库中的行人对象分为单人与双人(或双人以上)两类,如图 13 所示行人 1 为双人,行人 2~

5 都是单人。统计第 43~46 这四帧图像中他们的大小,结果如图 14 所示,横坐标为图像的帧序号 m ,纵坐标为对象在图像中包含的像素数目 $g(m)$ 。通过机器学习我们完全可以找到一个最佳分类门限 \hat{n} 来对行人的类别进行判定。另外,对运动目标检测图,可以框出感兴趣的运动目标,通过分析统计目标所在位置的光流矢量大小和方向,得到该目标的运动方向和运动速率的估计值^[4]。以图 12 为例,其中感兴趣运动目标 1 和目标 2 的局部光流场如图 15 所示。最后,得到的对目标 1 的描述是:向左行走的行人,双人,运动速率估计值为 2.41 像素/帧。类似地,对目标 2 的描述是:向右行走的行人,单人,运动速率估计值为 0.98 像素/帧。

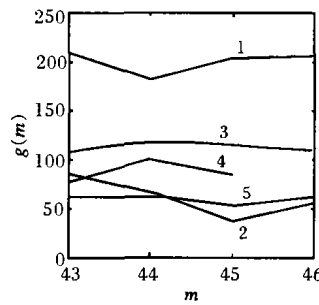


图 14 第 43~46 帧中行人 1~5 的大小

Fig.14 Sizes of passenger 1~ 5 in 43rd-46th frames



图 15 图 12 中目标 1 和 2 的局部光流场图

Fig.15 Local optical flow field of object 1 and 2 in the Fig.12

参考文献:

[1] Ricquebourg Y, Bouthemy P. Real-time tracking of moving persons by exploiting spatio-temporal image slices[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2000, 22: 797-808.

[2] 章毓晋. 图像工程下册: 图像理解与计算机视觉[M]. 北京: 清华大学出版社, 1999.

[3] Zhang T X, Zhang W, Shen J. Detection and analysis of moving objects in infrared image sequences based on supervised learning[A]. The Third Int'l Symposium on MIPPR[C]. 2003, 5286.564-571.

[4] Chang M M, Tekalp A M, Sezan M I. Simultaneous motion estimation and segmentation[J]. IEEE Trans Image Processing, 1997, 6: 1326-1333.