

文章编号:1001-9014(2007)06-0437-06

红外序列图像中基于形状的人体检测

王江涛, 杨静宇

(南京理工大学 计算机科学与技术学院, 江苏 南京 210094)

摘要:对红外序列图像中的人体检测问题进行了研究,提出一种新的人体检测方法.首先采用自适应高斯混合模型对序列图像中背景进行建模,在准确分割出前景运动目标的基础上,提出了一种新的人体形状表达模型,充分考虑了多个人体发生粘连或互相遮挡的情况,并用亮度投影的方法对其进行分离;以人体表达模型作为输入向量,构建支持向量机(SVM, Support Vector Machine)对人体进行分类判别.不同红外视频序列的检测结果表明了所提出算法在单个人体和多人体情况下均具有较好的鲁棒性和可行性.

关键词:人体检测;红外序列图像;支持向量机;高斯混合模型

中图分类号:TP391.41 **文献标识码:**A

SHAPE-BASED HUMAN DETECTION IN INFRARED IMAGE SEQUENCES

WANG Jiang-Tao, YANG Jing-Yu

(Institute of Computer Science and Technology, Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract: The human detection problem in infrared image sequences was studied, and a novel detecting approach was presented. GMM (Gauss mixture model) was first adopted to construct a background model. And then on the basis of accurately segmenting the forward objects, a shape-based human representing model was designed. By taking account of occlusions and merging among multi-body, the intensity projection curve was applied to separate single ones. By using human shape models as input vectors, a SVM (support vector machine) was constructed to classify and identify the human bodies. Experimental results on different infrared video sequences show that the proposed method is robust and feasible in single body and multi-body cases.

Key words: human detection; infrared image sequences; support vector machine; gauss mixture model

引言

人体检测作为对人体进行视觉分析的基础,近年来得到了越来越多的研究人员关注,其作为智能监控、高级人机接口、人体运动分析、人体行为理解等领域一个先决条件,已成为计算机视觉领域中的一个热点.当前,人体检测的研究主要集中于可见光领域,经过十年多的研究出现了众多的算法,大量实验表明了这些算法的有效性,这方面人体检测的算法已臻成熟.由于红外成像系统中物体的图像取决于物体本身的温度和其辐射出的热量,与可见光图像相比红外图像具有其独有的优点,随着技术的不断发展,红外图像技术已经开始得到越来越多的应用^[1,2],最近几年,红外热图像领域中的人体检测问

题已开始得到国际上学者们的关注^[3,4].

红外图像中的人体的检测一般可以分为两步来完成.第一步,选择可能存在人体的兴趣区域(Regions of Interest, ROI),第二步,采用某种手段对所有的ROI进行分析,确认其是否为人体.文献[3]中,Bertozzi等人以正面人体的对称性作为特征,在一预先确定好的尺度范围内对人体进行检测,通过将候选区域与形态学人体模型进行匹配来完成人体检测,该算法在确定人体存在的位置时均需要对图像进行多尺度搜索,这影响了人体检测的实时性能.文献[4]中,Nanda等人则在采用了一种基于像素亮度的人体表达模型的基础上,通过进行样本训练得出人体亮度和非人体亮度的概率分布,并以此构造亮度概率模板来执行人体检测,然而由于图像中人

收稿日期:2006-09-21,修回日期:2007-04-27

Received date: 2006-09-21, revised date: 2007-04-27

基金项目:国家自然科学基金重点(60632050)和国家自然科学基金(60472060)资助项目

作者简介:王江涛(1979-),山东昌乐人,博士研究生,主要研究方向为模式识别、目标的检测跟踪及行为识别等.

体形状的多变性,同样需要多个模板在不同尺度的情况进行匹配。

本文中,提出了一种快速高效的红外序列图像中人体检测算法.该方法首先通过自适应背景抽取,对前景目标中的人体候选区域进行定位,克服了单幅图像中人体检测多尺度搜索引起的计算复杂度问题,在此基础上设计了一种新颖的人体形状表达模型对人体特征进行建模,进而通过训练构建支持向量机分类器对前景目标候选区域进行分类判别,同时给出了不同红外图像测试集中人体检测的结果。

1 基于自适应高斯混合模型的背景建模

1.1 红外图像中人体目标的特点

红外图像与可见光图像在图像信息方面有着很大的区别,在可见光图像中物体的颜色分布取决于物体表面的入射光线和表面的反射程度,而在红外领域中物体的图像则取决于物体本身的温度和其辐射出的热量。

与采用可见光图像进行人体检测相比,基于红外图像的人体检测有其自身的优点,同时也存在缺点.在红外成像系统中,由于人体温度总是高于其所处的环境的温度,同时人体辐射出的热量也远远高于静态背景环境,比如树木、道路等等,因此,人体属于红外图像中像素的亮度值较高的一部分,从而可以明显地与背景相对比,此外红外成像系统在很大程度上消除了光照、颜色、阴影、纹理等因素的影响对人体成像的影响,这使得红外图像特别适合于人体的检测和定位;然而,红外图像也有其不利于人体检测的一些特性,首先,非人物体,比如动物、汽车、电线盒、建筑物等都会具有较高的亮度,这使仅依靠目标的亮度来进行人体检测变得不可能,其次,红外图像中物体的亮度不只依赖于物体的温度,还受到物体的表面特征影响(辐射性能、反射性能、转递性能等等)、表面朝向、波长等因素的影响,这增加了人体检测的难度。

对于单幅红外图像中的人体检测问题,只能从上述红外图像中人体成像的特点出发,采用多尺度搜索的手段来寻找目标.而对于红外序列图像中的人体检测问题,则可以通过背景抽取来获取前景目标,避免了单幅图像中的多尺度搜索问题。

1.2 构建自适应高斯背景模型

基于自适应高斯混合模型的在线背景建模方法最先由 Stauffer 和 Crimson 在文献[7]中提出,该方法可以鲁棒地克服光照变化、重复运动、缓慢运动物

体和混乱等对背景的影响,其出发点是认为单峰分布的背景模型不足以对夹杂在图像中的噪声、光照变化等因素进行同时处理,因此,他们采用了一个高斯混合模型对图像中像素的特征值进行表达.在红外图像中该特征值即为像素的亮度.将连续序列红外图像中同一位置的像素在时域中的亮度看作为一时间序列 $\{X_1, X_2, \dots, X_t\}$,则当前像素概率的观测值采用混合高斯模型可以表达为

$$P(X_t) = \sum_{i=1}^k w_{i,t} * \eta(X_t, u_{i,t}, \sigma_{i,t}^2) \quad (1)$$

这里 $w_{i,t}$ 为第 i 个高斯分布在时间 t 的权值, $u_{i,t}, \sigma_{i,t}$ 分别为第 i 个高斯函数的均值和标准差. k 为高斯混合模型中所包含的高斯分布的个数,其大小取决于系统可用内存以及系统的计算能力, η 为高斯概率密度函数

$$\eta(X_t, u_{i,t}, \sigma_{i,t}^2) = \frac{1}{\sqrt{2\pi}\sigma_{i,t}} e^{-\frac{(X_t - u_{i,t})^2}{2\sigma_{i,t}^2}} \quad (2)$$

采用上述模型决定一个像素是否为前景运动物体的过程如下:初始时,将高斯混合模型初始化为均值预先确定、高方差、小权值的 k 个高斯分布.对于一个新的像素,如果这个像素的观测值位于高斯混合模型中某一高斯分布的 2.5 个标准偏差范围内,则认为其与该高斯分布配对.然后根据配对的情况通过下式对所有的高斯分布权值进行更新

$$w_{i,t} = (1 - \alpha)w_{i,t-1} + \alpha(M_{i,t}) \quad (3)$$

其中 α 为学习参数,一般取 $\alpha = 0.05$, $M_{i,t}$ 则为当前像素的高斯分布配对标志,如果存在一个高斯分布与当前像素配对则 $M_{i,t} = 1$, 否则 $M_{i,t} = 0$. 对权值进行更新后,对新生成一系列权值进行归一化,所配对的高斯分布的参数更新为

$$\mu_i = (1 - \rho)\mu_{i-1} + \rho(X_t) \quad (4)$$

$$\sigma_i^2 = (1 - \rho)\sigma_{i-1}^2 + \rho(x_t - \mu_i)^T(x_t - \mu_i) \quad (5)$$

这里 $\rho = \alpha\eta(X_t | \mu_i, \sigma_i^2)$, 其他高斯分布的参数则不变. 如果不存在任何配对,则将概率值最小的高斯分布替换为一个均值为新观测的像素值、高方差、小权值的高斯分布,对下一帧图像执行上述操作。

为了判断新观测的像素的类型(前景或背景),按照 w/σ 的值的对 k 个高斯分布降序排列,由于背景图像中的像素具有高权值和低方差,因此两者之间的比值越大,属于背景的可能性越高,在这里把权值之和大于背景阈值 T 的最前面的 b 个高斯分布被作为背景模型,即有

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b w_k > T \right) \quad (6)$$

在背景模型确定后,即可对图像中的背景像素和前景像素进行分类.如果新观测的像素与该 b 个高斯分布中的某一个配对,则认为其属于背景,否则属于前景.通过上述方法,即可自适应的对背景模型进行学习和更新.为了计算方便,本文中,令获取背景模型进行前景分割后的图像为二值图像,背景像素为 0,前景像素为 1.

2 人体目标形状的特征表达

2.1 单个人体的表达模型创建

在可见光图像中,与非人体相比人体具有明显的鉴别性特征,比如不同身体部位一般有其特定的颜色分布,因此在人体表达模型的选取上具有更大的选择空间.然而,在红外图像中人体的特点仅表现为高亮度,其他的非人体物体,如汽车等同样具备该特点,这增加了红外图像中人体检测的难度.

虽然不同的人体的形状变化较大,但是与其他非人体相比,人体还是具备自己的特征共性,本文中我们从人体的形状作为出发点,提出一种新的基于形状的人体特征表达模型.基于形状的人体目标建模过程详细论述如下.

在通过背景抽取获取图像的前景目标后,首先采用形态学算子对图像进行滤波,去除图像中的噪声,然后对所有的连通域进行连通,并求取可以将每一个连通域包含的最小矩形,将此矩形区域作为该连通域的表达窗口,在单人体的情况下,该矩形窗口即为人体表达窗口.对于每一个表达窗口,将其分为大小相等的 $m * n$ 个子矩形区域,计算每一个子区域的亮度和,然后采用该亮度和与子矩形区域面积的比值作为该子区域的特征值,这样每一个表达窗口可以用一个 $m * n$ 维特征向量来表达,令表达向量为 V ,则有

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}, \quad \text{其中}, v_{m,n} = \frac{\sum_{i=1}^w \sum_{j=1}^h im_{i,j}}{w * h} \quad (7)$$

该矩阵中每一个元素代表了处于该位置的子矩形的特征值,其中 w 和 h 分别为子矩形宽和高, $im_{ij} \in \{1, 0\}$ 为位于子矩形内 (i, j) 点的像素值.如图 1 所示,给出了一幅人体图像在划分为 5×5 个子矩形时的特征图像,其中图 1(a) 为原图,图 1(b) 给出了人体表达模型的图像,每一个方格代表一个特征值,方格的亮度则表明了特征值的大小,图 1(c) 给出了图 1(b) 所示人体表达模型的轮廓图像,从图中可以看

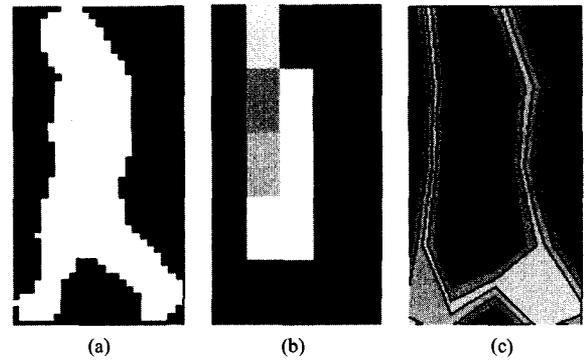


图 1 人体形状的特征表达

Fig. 1 The representing model of human shape

出很好的表达了人体的拓扑形状.

为了更好的将人体与其他运动物体(比如汽车)分别出来,针对人体矩形表达窗口的尺度比保持在一定范围内的特点,对上述的特征表达向量进行加权.令人体矩形窗口的高和宽分别为 H 和 W . 则其权值可以采用 $\lambda = W/H$, 加权后的特征向量如下

$$V^* = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}, \quad \lambda = W/H \quad (8)$$

2.2 多个人体相互粘连情况下的人体特征表达

在实际场合中多人之间的互动是十分普遍的,在这种情况下,多个人体将不可避免的粘连在一起或者是互相遮挡,此时,在进行背景抽取后将得到一个包含多人的前景连通域,上述的单人表达方法不能适用与此种情形.

由于即使在发生遮挡情况下,人体的头部仍然具有比较明显的特征,因此,可以根据人体的头部特征将粘连或发生遮挡的多个人体分割为单个的人体.在如图 2(a) 所示的情况下,对获得的前景图像进行亮度垂直投影,图 2(b) 给出了投影曲线,从图中可以看出头部的位置对应于图像投影的峰值,而两个人体粘连的部位则对应于投影曲线在两个峰值间的谷底,因此可以根据此特征来确定图像分割点对单个人体进行分离,图 2(c) 给出了分割的结果.

通过判断两个峰值之间是否存在较低的谷底对发生粘连的人体进行分割的方法,同样适用于多人体的情景.在通过背景抽取获得前景目标的矩形表达窗口后,首先对窗口内的目标进行亮度垂直投影,通过上述方法将多人体图像分割为一系列的单人图像,进而重新拟合其表达窗口,即可采用 2.1 节给出的方法进行特征表达.由于人体下半部分的像素

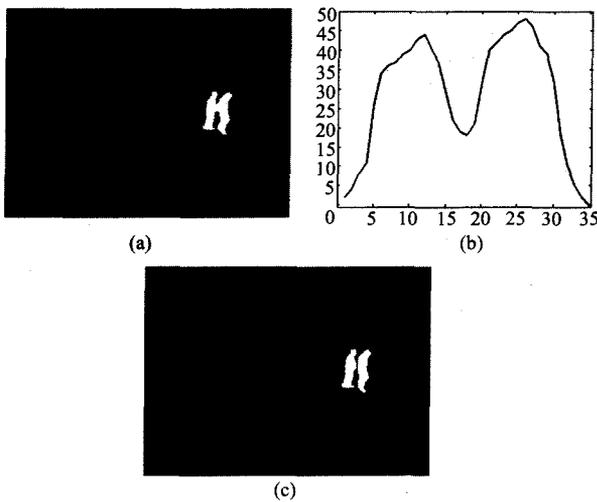


图2 多个人体的投影分割

Fig. 2 An illustration of separating multi-bodies by using projecting curve

分布比较稀疏,为了避免其对图像投影曲线的影响,在实际中我们只采用了图像上部1/2的部分来进行投影.

经过上述步骤完成人体特征的表达后,即可以量化的提取不同前景物体的特征向量,将这些特征向量作为训练样本,通过训练分类器来执行人体检测的任务.

3 基于支持向量机的人体检测

本文中我们通过监督学习方法来构建人体检测系统,在采用一系列的人体和非人体的图像训练样本对分类器进行监督训练的基础上,求取最优的分类决策函数.

支持向量机作为一种新的模式识别方法,在解决小样本、非线性及高维模式识别问题中表现出许多特有的优势,其已被广泛的应用于不同领域的模式识别问题.支持向量机的基本思想如下:首先通过

非线性变换将输入空间映射到一个高维空间,然后在这个新空间中寻找最优超平面作为分类决策函数,而这种非线性变换是通过定义适当的内积函数实现的.假定有一组训练样本集 $\{x_i, y_i\}, i = 1, \dots, n$, x_i 为 n 维特征向量, y_i 则属于 $\{+1, -1\}$ 是类别标签,首先通过函数 Φ 将 x 映射到高维空间,然后寻找下述形式的决策方程.

$$f(x) = \text{sgn}(w \cdot \Phi(x) + b) \quad (9)$$

当离超平面最近的点 $\Phi(x_i)$ 与超平面之间的距离最大时,就认为此时的分类决策函数 $f(x)$ 是最优的,输入向量 x 的类别可以由 $f(x)$ 的输出符号获得.

上述问题可以转化为在约束 $y_k f(x_k) \geq 1 - \xi_k$ 下,求取下式最小值的问题

$$\varphi(w, \xi) = \frac{1}{2} \|w\|^2 + C \left(\sum_{k=1}^m \xi_k \right) \quad (10)$$

这里 w 形式为 $w = \sum_{k=1}^m \alpha_k^* y_k \Phi(x_k)$, 可以定义如下的 lagrange 函数来求解上述问题

$$Q(\alpha) = \sum_{k=1}^m \alpha_k - \frac{1}{2} \sum_{k,l} \alpha_k \alpha_l y_k y_l K(x_k, x_l) \quad (11)$$

其约束为 $\sum_{k=1}^m y_k \alpha_k = 0, 0 \leq \alpha_k \leq C$, 这里 $K(x_k, x_l) = \langle \Phi(x_k), \Phi(x_l) \rangle$ 为内积函数. 此时相应的判别函数式(9)也变为

$$f(x) = \text{sgn} \left(\sum_{k=1}^m \alpha_k^* y_k K(x_k, x) + b^* \right) \quad (12)$$

这就是支持向量机.

上述论述可以看出,在构建支持向量机时需要确定所采用的内积函数,由于在实际应用中,与其它内积函数相比径向基核函数表现出更为优异的性能,在本文中采用了径向基核函数作为支持向量机的内积函数.

训练样本集的选取对整个分类器的性能有着关

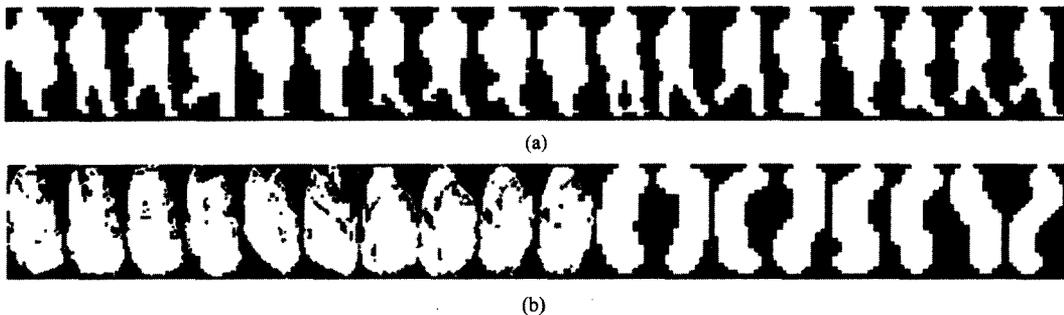


图3 训练中的部分人体样本和非人体样本

Fig. 3 Human samples and non-human samples in the training dataset

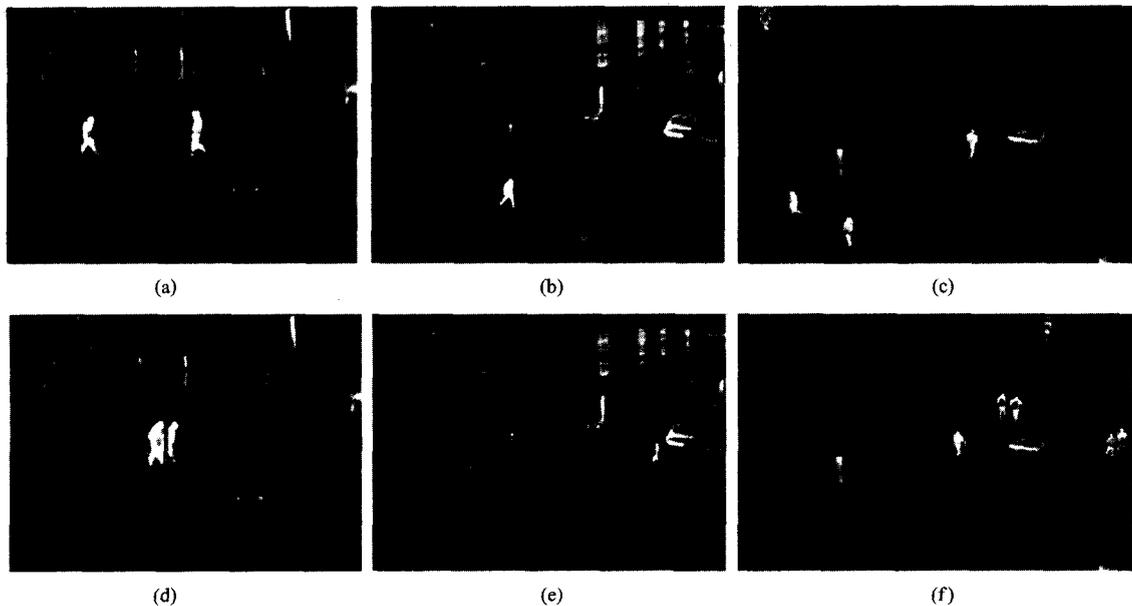


图4 部分图像的人体检测结果
Fig. 4 Human detection results of some test images

键性的影响,有效的训练样本是分类器能够正确分类的保证.在本文中,人体训练样本取自于不同的视频序列,包含了形状各异的人体.对于非人体样本,为了增加其多样性,除了包含有红外视频中的其它运动目标(如汽车)外,我们人工地加入了一部分与人体相似的物体.图3给出了样本集中的部分人体样本和非人体样本.

首先要确定训练中人体表达向量的维数,从整个样本集中取100个人体样本和150个非人体样本作为训练样本集训练支持向量机,其他样本则用来测试所训练的分类器的分类精度.为了计算方便,令人体表达向量的维数为 n^2 .表1给出了特征向量取不同维数时支持向量机对人体样本和非人体样本的正确分类率,从表中可以看出,在一开始随着样本维数的增加精度不断提高,当维数达到16附近时分类精度达到最高点,然后降低.在维数为16时正确分类率达到了97%,这表明只要选取适当的特征表达维数,人体形状表达向量即可准确的对人体进行表达.经过综合比较取人体表达向量维数为16,重新建立训练样本集.

表1 输入向量维数不同时的分类性能

Table 1 Performance of classifier with various dimensions input vectors

维数	4	9	16	25	36	49
人体样本分类率	38.5%	74.5%	97%	68.5%	65%	39%
非人体样本分类率	88.43%	85.95%	97.52%	95.87%	100%	97.52%

在确定了输入向量的维数后,采用交叉训练方法确定支持向量机的惩罚参数 $C = 150$,径向基核函数宽度为0.5,构建支持向量机进行分类判别.

4 实验与分析

为了验证所提出的人体检测算法的性能,准备了3个不同的红外视频测试集,这些测试集均与训练样本集相独立.3个测试集中,测试集1取自Teraviv Motion IR Database数据库,而测试集2和3取自于OTCBVS Benchmark Dataset数据库,这些测试集中既包括单人体红外视频序列,也含有存在多个人的红外视频序列,测试集1为多人测试集,由230幅图像组成,一共包含有460个人体;测试集2为单人体测试集,包含200幅图像,每幅图像中均含有一个人体;测试集3由23幅图像组成,该测试集共包含101个人体.所有的人体检测均在配置为AMD2.0G内存480M,运行环境为Matlab6.5的PC机上完成.

图4分别给出了3个测试集中的部分图像及其检测结果,在这些图像中检测到的人体区域均以矩形框表示;其中图4(a)和图4(d)取自测试集1,图4(b)和图4(e)取自测试集2,而图4(c)和图4(f)则取自测试集3.从检测的结果来看,即使在人体发生部分遮挡的情况下,仍然可以对其正确检测,这表明文中给出的人体表达模型和分类模式是有效的.

表 2 人体检测的实验结果

Table 2 Experimental results of human detection test

	图像大小	图像数	人体数	检测率	平均运行时间
测试集 1	240 × 320	230	460	98.5%	0.054s
测试集 2	240 × 320	200	200	100%	0.0371s
测试集 3	240 × 360	23	101	98.8%	0.120s

为近一步分析所提出的检测算法性能,在表 2 中详细的给出了各个测试集的检测结果以及每幅图像需要的平均检测时间.从表中可以看出,在 3 个测试集中均取得了等于或高于 98.0% 的检测率,这表明了算法的有效性,其中在单人体测试集 2 中取得的检测效果最好,而在多人测试集 1 和 3 中,则出现了部分人体未能检测到的问题,这主要是由于多个人体之间发生严重遮挡、相互重叠时导致无法判断人体的个数.从运行时间来看,图像的大小和存在人体的多少决定了所需要的执行时间,然而即使对人体较多的测试集 3,在 Matlab6.5 软件环境下平均处理速度达到了 8fps,基本上可以满足实时的需要.由整个检测结果来看,无论是算法的准确度还是实时性能都达到了较高的水平.

5 结语

基于红外成像系统的人体检测技术作为一个新兴的研究领域,有着广阔的应用前景.本文针对红外序列图像中人体的检测问题,提出一种新颖的人体检测算法.该算法中,首先通过高斯混合模型背景抽

取确定图像中的前景目标,然后确定前景目标的表达窗口,同时对发生粘连的多个目标进行分割,在此基础上,提出了一种新的人体形状表达模型对目标窗口区域进行特征表达,最后通过监督学习来训练支持向量机分类器,对前景中的人体目标进行分类判别.理论分析和实验结果均表明所提出的算法是有效可行的,为进一步的在红外序列图像中的进行人体分析研究提供了理论基础.

REFERENCES

- [1] CHENG Jian, ZHOU Yue, CAI Nian, *et al.* Infrared object tracking based on particle filters [J]. *J. Infrared Millim. Waves* (程建,周越,蔡念,等.基于粒子滤波的红外目标跟踪. *红外与毫米波学报*), 2006, 25(2): 113—117.
- [2] LING Jian-Guo, LIU Er-Qi, LIANG Hai-Yan, *et al.* Infrared target extraction method based on kernel density estimation [J]. *J. Infrared Millim. Waves* (凌建国,刘尔琦,梁海燕,等.基于核密度估计的红外目标提取方法. *红外与毫米波学报*), 2006, 25(6): 434—438.
- [3] M Bertozzi, A Broggi, PGrisleri *et al.* Pedestrian Detection in Infrared Images [C]. Proceedings of the IEEE International Conference on Intelligent Transportation Systems, Shanghai, 2003, 662—667.
- [4] Harsh Nanda, Larry Davis. Probabilistic template based pedestrian detection in infrared videos [C]. Proceedings of IEEE Intelligent Vehicle Symposium, Versailles, 2002, 15—20.
- [5] C Stauffer, W Grimson. Adaptive background mixture models for real-time tracking [C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Colorado, 1999, 2: 246—252.

2008 年《红外与毫米波学报》征订启事

《红外与毫米波学报》着重反映红外与毫米波领域的概念、新成果、新进展,是中国与国际红外与毫米波学术界交流最新研究成果的平台.其宗旨是努力做到发扬科学民主、提倡学术讨论、活跃学术思想、促进国内外同行间的学术交流、促进红外与毫米波科学技术的不断发展.

《红外与毫米波学报》主要报道红外与毫米波领域的最新研究成果和技术进展,刊登在红外物理、凝聚态光学性质、低能激发过程、飞秒光谱学、非线性光学、红外光电子学、红外与毫米波领域的元器件、系统及应用、智能信息和人工神经网络等方面有创新的研究论文、研究简报.读者对象为国内外红外与毫米波领域的科研人员、工程技术人员及高等院校师生、研究生等.

《红外与毫米波学报》为《SCI》、《EI》、《CA》、《SA/INSPEC》、《AJ》等国际著名检索体系收录.为“中国自然科学核心期刊”、“中国科技核心期刊”.

本刊由中国科学院上海技术物理研究所和中国光学学会共同主办,中国科学出版社出版,为中、英文混排双月刊,大 16 开本、80 页,定价:10 元/期.国内外公开发行,全国各地邮局均有订阅,邮发代号:4-335.

如中国光学学会会员向本刊编辑部订阅,可享受九折优惠.

本刊地址:上海市玉田路 500 号《红外与毫米波学报》编辑部 邮政编码:200083

电话:021-65420850 * 73206 传真:021-55393960 电子邮箱:jimw@mail.sitp.ac.cn