

基于支持向量机的复杂背景下的人体检测

潘 锋 王宣银

(浙江大学流体传动及控制国家重点实验室, 杭州 310027)

摘 要 常用的人体检测方法多是基于经验风险最小化原理的传统统计理论,其性能只有在样本趋于无穷大时才有理论上的保证,而在实际应用中,学习样本通常是有限的。针对传统统计理论在人体检测中存在的不足,提出了一种基于统计学习理论——支持向量机(SVM)的人体检测方法,利用彩色空间对背景进行自适应建模提取运动目标,然后使用训练好的 SVM 进行验证是否是人体。为了简化 SVM 分类器的设计及提高机器学习的效率,提出了一种星形向量表示法用于抽取目标的特征向量,并且用实验方法得到了这种表示法的最优表示。将 SVM 与 ANN 进行比较,并且对不同内积函数的 SVM 的性能也进行了比较。实验结果表明,SVM 的性能要优于 ANN,并且采用径向基函数的 SVM 性能最好。该方法鲁棒性强,正确率高,解决了复杂背景下运动人体实时检测的一些关键问题。

关键词 支持向量机 背景相减 运动检测 特征抽取 人体检测

中图分类号: TP391 **文献标识码**: A **文章编号**: 1006-8961(2005)02-0181-06

Support Vector Machine-Based Human Detection under Complex Background

PAN Feng, WANG Xuan-yin

(The State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027)

Abstract In the field of computer vision, the research on human detection has a wide application prospect. Prevalent human detection methods usually use traditional statistical theory, which is based on empirical risk minimization. But the minimization of empirical risk over limited training data does not imply good generalization to novel test data. Aiming at the shortcomings of traditional statistical theory used in human detection, a new method based on SVM is presented in this paper. An adaptive background subtraction method combined with color is used to segment the motion objects. Then the trained SVM classifier distinguishes the motion object whether it is a human or not. In order to simplify the design of SVM classifier and improve efficiency of machine learning, a center radiating vector representation is proposed to abstract features of the object. And the optimal representation is obtained by experiments. During the machine learning, a bootstrap method is adopted to reduce the complexity of training SVM. Experiments show that the performance of SVM is better than ANN, and the radial basis function SVM has better performance than other SVMs on human distinguish. This method has strong robustness and high accuracy.

Keywords support vector machine, background subtraction, motion detection, feature extraction, human detection

1 引 言

近年来人体的视觉分析已成为计算机视觉领域中的一个重要研究方向^[1],它主要研究从包含人的序列图像中检测、识别、跟踪人体并对其行为进行描

述与理解。人体的视觉分析在智能监控、高级人机接口、人体运动分析、虚拟现实及基于内容的图像检索与储存等方面有着广泛的应用前景和潜在的经济价值。如在可视化远程监控系统中,传统的监控模式需要工作人员监视各个监控画面,这样不仅工作繁重、效率低,而且容易出错,如果能使系统具有目

基金项目: 航天支撑技术基金项目(2001-HT-ZJDX); 杭州市科技发展计划(2001121C42)

收稿日期: 2003-09-22; **改回日期**: 2004-07-12

第一作者简介: 潘锋(1976~),男,2000年于浙江大学机械与能源学院获机械电子工程专业学士学位,现为浙江大学机械与能源学院流体传动及控制国家重点实验室博士研究生。主要研究方向为智能机器与图像信息技术、机电一体化等。E-mail: pf@zju.edu.cn

标自动检测、识别及跟踪的能力,就能大大提高系统的可靠性,实现监控的自动化和智能化。

在各种以人体为研究对象的系统中,人体检测是首要步骤。由于人体的非刚性运动以及目标之间的互遮挡与自遮挡等因素的影响,使得人体检测成一项极具挑战性的课题。人体检测通常是先找出所有的候选目标然后对运动目标进行分类识别^[2-7],如利用标准矩和人工神经网络识别人形目标^[2];实时视觉监控系统 W4^[3]通过对目标形状的分析区分与定位人体,能在像户外这样的环境下检测和跟踪无遮挡的人体;利用运动检测结合傅里叶描述子抽取目标的特征向量,然后通过神经网络对目标进行分类^[4]。近年来利用立体视觉^[8]、形变模板^[9]、多信息融合^[10,11]等方法也受到了广大研究人员的重视。国内对这方面的研究还比较少^[6,7]。目前大多数人体检测方法都是基于传统统计模式识别,而传统统计模式识别多是基于经验风险最小化原理,只有在样本数趋于无穷大时其性能才有理论上的保证。而在实际应用中,样本通常是有限的,这时很多方法难以取得理想的效果。人体是个复杂的形变体,样本的采集比较困难,而且在像人体检测这样的“是”与“非”两类模式识别问题中,非人体样本的采集显得更加困难。为此提出了一种基于小样本统计学习理论——支持向量机(support vector

machine, SVM)的人体检测方法,能够较好地解决小样本学习问题。

2 基于支持向量机的人体检测原理

支持向量机是基于结构风险最小化原理的统计学习理论的一种新的模式识别方法^[12],在解决小样本非线性及高维模式识别问题中具有许多特有的优势,对小样本问题具有很好的学习与泛化能力,成为当前机器学习领域的一个研究热点。将运动目标提取与 SVM 相结合的人体检测方法,是利用运动检测提取运动目标,再通过训练好的 SVM 分类器判断是否为人体的。其原理结构框图如图 1 所示,运动目标提取采用改进的自适应背景高斯建模方法,在保持目标提取快速性的同时,增加了目标提取的准确性。为了简化 SVM 分类器的设计及提高机器学习的效率,提出了一种星形向量表示法用于目标的特征向量抽取,并且用实验方法得到了该特征向量的最佳表示。在机器学习过程中的样本收集,特别是非人体样本的收集,采用了样本自举(bootstrap)的方法,加快了样本收集的进程。实验结果表明,该方法能在复杂背景中检测单个或多个人体,并且容许目标有适度的遮挡,对人体的姿态没有特别的要求,能适应背景的变化,能在全天候条件下工作。

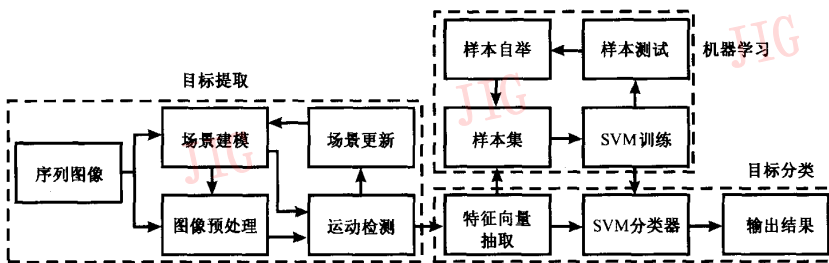


图 1 基于支持向量机的人体检测原理框图

Fig. 1 Framework of human detection based on support vector machine

3 运动目标提取

在运动检测前,假设摄像机相对背景是静止的,而且这里所说的复杂背景是指背景不单一,场景及光照有变化,但相对目标的运动变化要慢。在此采用改进的自适应背景高斯建模方法提取运动目标。

在视频采集过程中,不可避免地会带来一些噪声,通常认为这些噪声具有高斯分布。设 RGB 3 通

道为 s_r, s_g, s_b , 共有 6 个参数估计,期望值为 μ_r, μ_g, μ_b , 标准差为 $\sigma_r, \sigma_g, \sigma_b$, 对背景每一点进行高斯建模: $[\mu_r, \mu_g, \mu_b, \sigma_r, \sigma_g, \sigma_b]$ 。

背景相减按以下的方式进行:当前像素值为 s_r, s_g, s_b , 与背景模型相比,如果 3 个通道只要有一个通道满足 $|s - \mu| > 3\sigma$, 就认为该像素点为前景点,否则为背景点。

上述的背景相减提取目标方法不但将目标提取出来,而且还会将目标影子也认为是目标的一部分。

为此,还需要对前面通过背景相减获得的前景点做进一步的处理,以消除影子带来的影响。由于影子对背景的颜色影响不大,所以可以利用这一点来判断是否为影子区域。为了简化处理,使用两个规一化的参数:

$$s_{rc} = s_r / (s_r + s_g + s_b) \quad (1)$$

$$s_{gc} = s_g / (s_r + s_g + s_b) \quad (2)$$

同样得到4个参数估计:期望值 μ_{rc}, μ_{gc} , 标准差 σ_{rc}, σ_{gc} 。

用类似的方法对第1阶段得到的前景点再进行处理,排除可能的影子区域。

这样在实际使用过程中共用了10个参数对背景进行高斯建模:

$$[\mu_r, \mu_g, \mu_b, \mu_{rc}, \mu_{gc}, \sigma_r, \sigma_g, \sigma_b, \sigma_{rc}, \sigma_{gc}]$$

考虑场景光照有变化,背景模型用一个简单的自适应算法进行实时更新:

$$\mu_i = \beta \mu_{i-1} + (1 - \beta) s_i \quad (3)$$

$$\sigma_i^2 = \beta [\sigma_{i-1}^2 + (\mu_i - \mu_{i-1})^2] + (1 - \beta) (s_i - \mu_i)^2 \quad (4)$$

其中, β 为更新率, $0 < \beta < 1$,依赖于帧率与场景的变化率,这里采用 $\beta = 0.9$ 。

为了减少前景目标对背景更新带来的影响,在背景更新时只考虑更新背景点,对于前景点则保持当前的模型参数。

最后对得到的二值图像进行形态滤波处理,以消除一些杂散点,光滑目标轮廓,统计各块的大小,剔除小于一定阈值的块,最终得到运动目标的形状。

人体样本的采集可以用同样的方法得到,部分人体图像样本如图2所示。



图2 部分人体图像样本
Fig. 2 Samples of human image

4 人体特征向量的抽取

尽管在理论上可以将整幅图像作为输入,但是一幅中等大小的图像就是高维的输入,这将给学习和分类带来困难,为了简化分类器的设计和提高机器学习的效率,通常进行特征提取,保留图像空间中的主要分类信息,降低特征空间的维数。目前常用的各种主成分提取方法不同程度地存在计算量大等

特点,为此提出了一种星形向量表示法用于提取目标的特征向量,如图3所示。

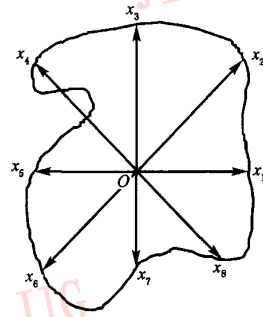


图3 星形向量表示法示意图

Fig. 3 Representation of center radiating vector

各个向量从目标轮廓的中心 O 点出发,按一定间隔呈放射状分布,指向目标的轮廓线。如图3为 $i=8$ 的星形向量表示,其中, x_1 指向为水平 0° 方向,其他向量按逆时针方向,间隔 45° 角分布。间隔的角度可以根据定义向量数的多少,可以等角度,也可以是变角度。如果某个向量不止一次穿过目标轮廓的边界,则选择距离最远的,如图3中向量 x_4 。将各向量的长度按顺序表示为目标形状的特征向量,并做规一化处理,使特征向量具有伸缩不变性。从特征向量的抽取过程中可以看出,特征向量只与目标的轮廓有关,在目标提取过程中有可能目标内部空洞不会对目标的特征向量造成影响。

5 支持向量机人体检测的实现

5.1 支持向量机原理

支持向量机(SVM)是在统计学习理论基础上发展起来的一种新的模式识别方法,与传统的机器学习方法普遍采用经验风险最小化原则相比,SVM采用了基于结构风险最小化(structural risk minimization, SRM)原则,使得在有限样本条件下,SVM方法比基于经验风险最小化的传统方法如人工神经网络(artificial neural network, ANN)方法具有更好的泛化能力。

SVM方法是从线性可分情况下的最优分类面提出的,如图4所示的2维两类线性可分情况。

设样本集为 $\{x_i, y_i\}, i=1, \dots, N, x_i \in \mathbf{R}^d, y \in \{+1, -1\}$ 是类别标识。 d 维空间线性判别函数的一般形式为

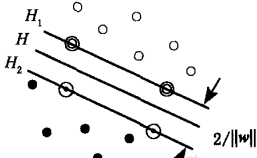


图 4 最优分类面示意图

Fig. 4 Optimal classification surface

$$g(x) = w \cdot x + b$$

分类面方程为

$$w \cdot x + b = 0$$

将判别函数进行归一化,使两类样本满足

$$|g(x)| \geq 1$$

这样分类间隔为 $2/\|w\|$,因此要使分类间隔最大要求 $\|w\|$ 最小,而要求所有样本正确分类,则要求样本满足:

$$y_i [(w \cdot x_i) + b] - 1 \geq 0 \quad (5)$$

对于线性不可分的情况,可以通过构造新的特征向量,将问题转换到一个新的空间,也就是利用内积函数的非线性映射将输入向量映射到一高维空间,SVM 能在此高维空间中给出最佳分类超平面。

构造内积函数为 $K(x_i, x_j)$ 的 SVM 可以归结为求解条件约束二次函数的极值问题:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (6)$$

$$\text{约束条件为 } 0 \leq \alpha_i \leq C, \sum_{i=1}^N \alpha_i y_i = 0$$

其中, α_i 为 Lagrange 系数,问题(6)解中非零的 α_i^* 及对应的 x_i 构成支持向量集,由此构成的支持向量机为

$$f(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i^* y_i K(x, x_i) + b^* \right) \quad (7)$$

其中, b^* 是分类的阈值,可以由任意一个支持向量按式(5)求得(因为支持向量满足其中的等式)。

5.2 支持向量机分类器的构造和训练

样本集是构造支持向量机分类器的基础,而且样本集的性能直接影响 SVM 分类器的性能。样本主要包含人体样本与非人体样本两类。对于人体样本用以下两种方式进行采集:一是采用拍摄人体行走的录像,利用前述的运动目标提取方法,获得人体样本,这里的人体行走考虑不同的人(胖瘦、高矮等),不同的行走方向(左右前后等);二是选取各种含有人体的图像,用手工方式分割出人体。由于非

人体样本的多样性,往往需要大量的非人体样本,庞大的非人体样本空间使得训练过程中迭代次数增加,影响收敛速度。为了减小非人体样本集的规模,采用自举的方式可以使样本更具代表性,从而提高分类器的训练速度及分类的正确率。非人体样本的收集从各类非人体的图像中获得,包括各类动物、车辆等。从非人体样本中抽出部分与人体样本组成训练样本集,送入 SVM 进行训练,得到一个初步的 SVM 分类器,将其余的非人体样本作为测试集,从错误区分的样本中随机挑选一部分与前面的非人体样本组成新的非人体样本集。重复此过程直至收集到足够的非人体样本。

在构造 SVM 分类器之前,非常有必要验证前面提出的目标形状的星形向量表示法的有效性,以及究竟多少维向量为最优的问题。为此从样本集中随机抽取 300 个人体样本,400 个非人体样本。将两类样本按随机方式分成两组样本集:样本集 A 及样本集 B,每一样本集中包含人体样本 150 个,非人体样本 200 个。将样本集 A 作为训练样本集,将样本集 B 作为测试样本集。分别对两组样本进行特征向量提取,特征向量维数从 2 维到 24 维,间隔为 2,这样共得到 12 组训练集数据,12 组测试集数据。将训练数据送入 SVM 进行训练,训练好后去测试相应的测试集数据。测试结果如图 5 示。从图中可以看出,随着特征向量维数的增加,分类的正确率也相应地增加,但在增加到一定维数后,分类的正确率反而有所下降,在维数 16 附近分类的正确率达到最大值(97%)。这说明维数过低,目标形状的星形向量表示不足以表示目标的形状,维数过高,对样本及样本的数量的要求增加,反而使得分类器的泛化能力降低。因此在后面的分类器设计中就采用 16 维的

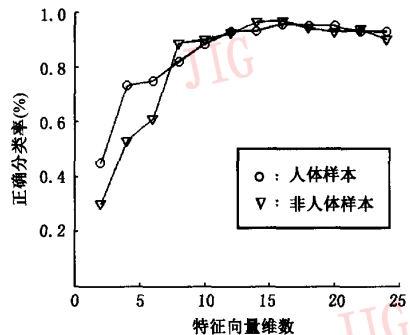


图 5 测试样本集分类结果

Fig. 5 Result of Classification on the test set

特征向量。

最终得到的样本集包含人体样本 1 536 个,非人体样本 1 860 个,将 16 维的特征向量作为输入,取优化约束条件 $C=200$,送入 SVM 进行训练,最后得到的支持向量 532 个,构成最终的 SVM 分类器。

6 实验及其结果分析

为了测试 SVM 分类器的性能,准备了两个测试集。测试集 I 是独立于训练集的图像集,采用摄像头与图像采集卡得到的含有人体的图像 1 187 张,含有非人体的其他各类运动物体(包括动物、汽车等) 1 213 张。测试集 II 采用一个人体的步态数据库^[13],解压缩后抽取了部分作为测试集,其中含有人体图像 620 张。为了便于比较,还设计了一个人工神经网络(ANN),采用 3 层结构, BP 学习算法。测试的结果如表 1 所示。从实验测试结果中可以看出,由于 SVM 对小样本问题具有较好的学习与泛化能力, SVM 比 ANN 具有较高的识别正确率, SVM 的人体检测率分别为 96.9% 和 97.6%,非人体的检测率相对较低,但也达到 89.9%。

表 1 SVM 与 ANN 比较结果

Table 1 Compare results of SVM and ANN

	测试集 I				测试集 II	
	人体(1 187 张)		非人体(1 213 张)		人体(620 张)	
	SVM	ANN	SVM	ANN	SVM	ANN
人体	1 150	1 062	123	158	605	34
非人体	37	125	1 090	1 055	15	586
正确率 (%)	96.9	89.5	89.9	87.0	97.6	94.5

不同的内积函数具有不同的分类性能,因此对 3 种常用的内积函数的 SVM 进行了比较实验,目的在于寻找哪一种内积函数的 SVM 最适合于人体的分类。结果如表 2 所示,表 2 中多项式阶数 q 取 3,径向基的 σ^2 参数取 0.5。实验结果表明,采用径向基函数的 SVM 性能最好。

表 2 不同核函数的 SVM 比较结果

Table 2 Compare results of different kernel function

	测试集 I		测试集 II
	人体(1 187 张)		非人体(1 213 张)
	人体(1 187 张)	非人体(1 213 张)	人体(620 张)
多项式	88.1	85.6	89.2
径向基函数	96.9	89.9	97.6
S 形函数	95.5	88.4	95.7

最后将算法实施在实时序列图像的处理。使用了一台 CPU 为 P III 800,内存 256M 的微机,视频采集卡为 SDK-2000。室内与室外环境的实验结果表明,系统能快速可靠地将正常行走的人体检测出来,如图 6(a)所示,系统没有将运动的汽车认为是人体目标。系统也容许目标有适当的遮挡(遮挡面积小于 1/4 目标的大小如图 6(d)所示),处理速度可达 6fps。但系统目前只能处理单个人体或互不遮挡的多个人体(如图 6(a)~(c)所示),对于多个互有遮挡的人体的处理将是下一阶段的工作。



图 6 人体检测结果

Fig. 6 Images of human detection

7 结论

在计算机视觉领域,人体检测具有广泛的应用前景。本文基于小样本统计学习理论对运动人体的检测提出了一种新方法——基于支持向量机的人体检测方法。并提出了一种星形向量表示法抽取目标的特征向量,并且用实验方法得到了该特征向量的最佳表示,简化了 SVM 分类器的设计,提高了机器学习的效率。通过对背景的自适应高斯建模消除了噪声及场景变化对目标提取带来的影响。采用了样本自举的方法,加快了样本收集的进程。理论和实验结果表明, SVM 的性能要优于 ANN,并且采用径向基函数的 SVM 性能最好。本文提出的方法鲁棒性强、正确率高、快速性好,能在复杂背景中检测出单个或多个互不遮挡的人体,为复杂背景下运动人体实时检测提供了理论和技术基础。

参考文献 (Reference)

- 1 Gavrilu D M. The visual analysis of human movement: A survey [J]. *Computer Vision and Image Understanding*, 1999, 73(1): 82 ~ 98.
- 2 Nicolaou C A, Egbert A L, Lacher R C, *et al.* Human shape recognition using the method of moment and artificial neural networks [A]. In: *International Joint Conference on Neural Networks [C]*, Washington, DC, the USA, 1999: 3147 ~ 3151.
- 3 Haritaoglu I, Harwood D, Davis L S. W4: real-time surveillance of people and their activities [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809 ~ 830.
- 4 Toth D, Aach T. Detection and recognition of moving objects using statistical motion detection and Fourier descriptors [A]. In: *12th International Conference on Image Analysis and Processing [C]*, Mantova, Italy, 2003: 1 ~ 6.
- 5 Satoh Y, Tanahashi H, Niwa Y, *et al.* Robust human detection from complex background by radial reach filter [A]. In: *Proceeding of IEEE TENCON'02 [C]*. Beijing, China, 2002: 533 ~ 536.
- 6 Liu X D, Su G D, Zhou Q, *et al.* A visual intelligent outdoor surveillance system [J]. *Journal of Image and Graphics*, 2000, 5(12): 1024 ~ 1029. [刘晓冬, 苏光大, 周全等. 一种可视化智能户外监控系统 [J]. *中国图象图形学报 (A 版)*, 2000, 5(12): 1024 ~ 1029.]
- 7 Pan F, Wang X Y, Wang Q Q. Human detection based on head and shoulder feature in intelligent surveillance system [J]. *Journal of Zhejiang University (Engineering Science)*, 2004, 38(4): 397 ~ 401. [潘锋, 王宣银, 王全强. 智能监控中基于头肩特征的人体检测方法研究 [J]. *浙江大学学报 (工学版)*, 2004, 38(4): 397 ~ 401.]
- 8 Ran Y, Zheng Q. Multi moving people detection from binocular sequences [A]. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing [C]*, HongKong, 2003: 37 ~ 40.
- 9 Tate S, Takefuji Y. Video-based human shape detection by deformable templates and neural network [A]. In: *Sixth International Conference on Knowledge-Based Intelligent Information & Engineering Systems [C]*. Crema, Italy, University of Milan, 2002: 280 ~ 285.
- 10 Han J, Bhanu B. Detecting moving humans using color and infrared video [A]. In: *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems [C]*, Tokyo, Japan, 2003: 228 ~ 233.
- 11 Xu F, Fujimura K. Human detection using depth and gray images [A]. In: *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance [C]*, Miami, Florida, 2003: 115 ~ 121.
- 12 Vapnik V. *The Nature of Statistical Learning Theory [M]*. Berlin: Springer-Verlag, 1995.
- 13 University of California, San Diego. Gait data [EB/OL]. <http://pages.cpsc.ucalgary.ca/~boyd/gait/experiment.html>, 1996-12-12.