

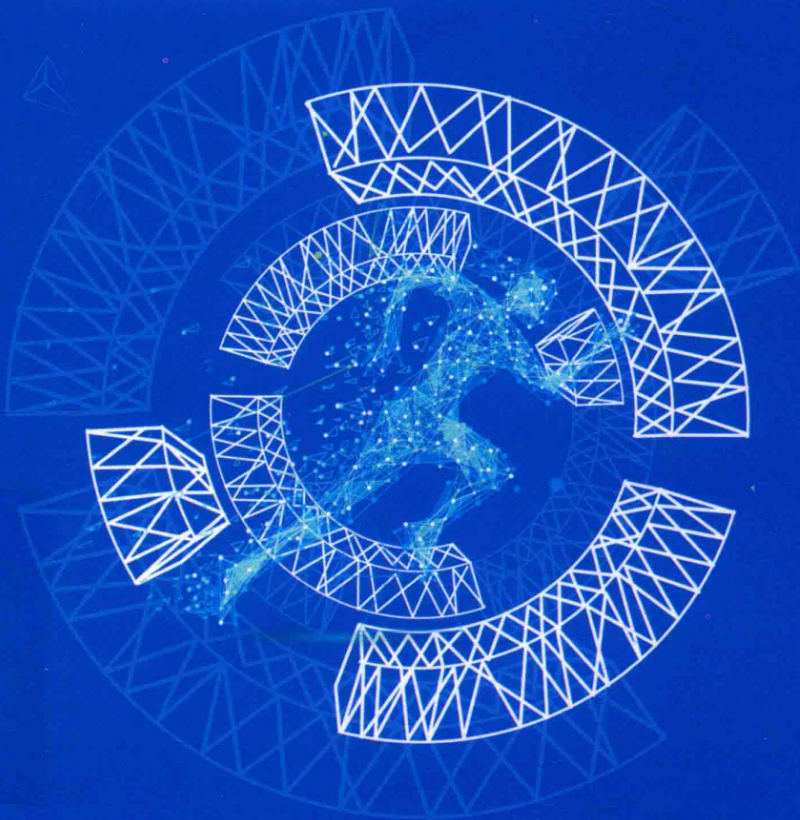
·四川省教育厅资助项目(14ZA0257)

·绵阳师范学院资助项目(QD2012A02,QD2017A006,Mnu-JY1665)

YUNDONG RENTI XINGWEI YUYI JISUAN JISHU

运动人体行为语义 计算技术

李敏 著



四川大学出版社

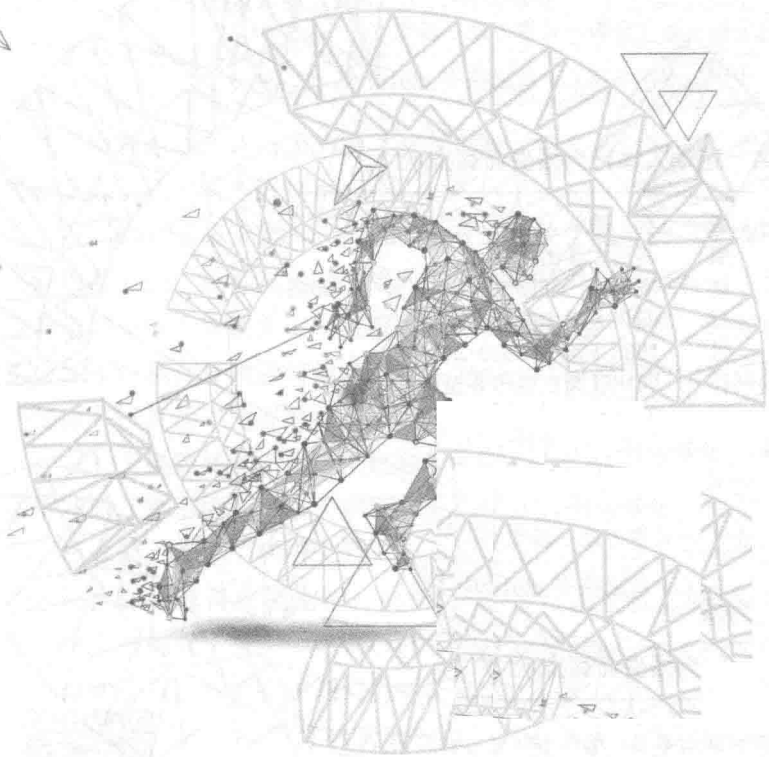
· 四
· 绵阳师范学院资

257)
006, Mnu-JY1665)

YUNDONG RENTI XINGWEI YUYI JISUAN JISHU

运动人体行为语义 计算技术

李敏 著



四川大学出版社

项目策划：梁 平
责任编辑：梁 平
责任校对：傅 奕
封面设计：璞信文化
责任印制：王 炜

图书在版编目 (CIP) 数据

运动人体行为语义计算技术 / 李敏著. — 成都 :
四川大学出版社, 2018. 10
ISBN 978-7-5690-2454-8

I. ①运… II. ①李… III. ①人体运动—人体科学—
行为分析—语义分析—计算方法—研究 IV. ①G804.62

中国版本图书馆 CIP 数据核字 (2018) 第 235934 号

书名 运动人体行为语义计算技术

著 者	李 敏
出 版	四川大学出版社
地 址	成都市一环路南一段 24 号 (610065)
发 行	四川大学出版社
书 号	ISBN 978-7-5690-2454-8
印前制作	四川胜翔数码印务设计有限公司
印 刷	四川五洲彩印有限责任公司
成品尺寸	148mm×210mm
印 张	6
字 数	162 千字
版 次	2019 年 7 月第 1 版
印 次	2019 年 7 月第 1 次印刷
定 价	38.00 元

版权所有 ◆ 侵权必究

- ◆ 读者邮购本书, 请与本社发行科联系。
电话: (028) 85408408 / (028) 85401670 /
(028) 86408023 邮政编码: 610065
- ◆ 本社图书如有印装质量问题, 请寄回出版社调换。
- ◆ 网址: <http://press.scu.edu.cn>



四川大学出版社
微信公众号

前 言

基于视觉的运动人体的行为识别有着广泛的应用前景，但大量的成果都集中在模版匹配法、状态空间法中。这两种方法的识别结果与人类的理解存在语义鸿沟，基于语义方法可以克服这个缺陷，但成果却很少。本书的主要目的是研究融入场景敏感度、人物身份的运动人体行为的语义计算方法。为实现此目的，本书进行了以下几方面的研究：

(1) 研究了基于 Beowulf 机群中改进的粒子滤波的三维人体运动跟踪算法。该算法通过三维人体模型参数的自动初始化、粒子数目和模板的调整来实现跟踪失效的自动恢复；并基于任务动态分配策略和低开销通信策略设计了 Beowulf 机群中的迁移式粒子滤波并行算法，用以解决粒子退化问题和提高计算速度。

(2) 为提高场景识别率和提取场景高层语义，提出了基于四层树状语义模型的场景语义识别新方法。四层语义模型包括视觉层、概念层、关系层和语义层。提取训练样本场景实物的颜色、颜色层次和轮廓得到视觉层；同类场景中实物的名称（概念单词）的交集构成了概念层；统计概念单词的频率并对概念单词的空间位置关系进行关联规则的数据挖掘，得到关系层；计算关键概念单词与 PSB 标准模型语义属性分类树的语义相似度，得到场景高层语义。计算测试样本的底层特征后，通过检索视觉层得到概念单词，通过检索概念单词的频率和空间位置关系关联规则得到场景分类。场景分类、场景高层语义、场景概念单词构成场

景语义的识别结果。

(3) 为提高身份识别率和减少识别时间,提出了一种人体身份识别的认知物理学方法。该方法将人脸特征和步态特征用数据场进行表征,利用数据间的相互作用和运动实现数据的自组织聚类,以这种非线性变换的方式实现身份特征数据的降维。对降维后的样本库以最大势函数值进行排序,实现了离散点快速检测和样本检测的二分法查找。基于改进后的 D-S 证据论对人脸和步态识别结果进行融合。

(4) 现有的语义方法未融入场景语义和人物身份,缺乏对复杂行为的有效识别和描述,因此本书提出了一种基于层次化概念空间的视频图像人体行为语义计算新方法。该方法引入认知学的概念空间,建立了层次化的运动概念空间,将复杂的运动分解为原子运动概念层、简单行为概念层和事件行为概念层。原子运动层细分为身体移动、肢体和姿态的原子运动,从大尺度、中尺度和小尺度提取了人体运动的特征。概念激活函数实现了原子运动的检测,时空逻辑规则实现了简单行为语义的识别。本书提出的七元组语义模型融入了场景语义和人物身份,实现了对事件行为的建模和识别。

目 录

第 1 章 概论	(1)
1.1 研究意义	(1)
1.2 研究现状	(3)
1.3 研究目的与研究内容	(7)
1.3.1 研究目的	(7)
1.3.2 研究内容	(8)
1.4 运动人体行为语义计算关键技术相关研究综述	(10)
1.4.1 语义及语义计算的定义	(11)
1.4.2 三维人体运动跟踪技术发展概述	(11)
1.4.3 场景语义提取与场景分类技术发展概述	(14)
1.4.4 视频中运动人体身份识别技术发展概述	(18)
1.4.5 基于语义的运动人体行为识别发展概述	(22)
1.5 运动人体行为语义计算关键技术的发展趋势	(24)
1.5.1 三维人体运动跟踪发展的趋势	(24)
1.5.2 运动人体身份识别的发展趋势	(25)
1.5.3 场景语义识别发展的趋势	(25)
1.5.4 运动人体行为语义计算方法发展的趋势	(26)
第 2 章 基于 Beowulf 机群中改进粒子滤波的三维人体运动跟踪方法	(28)
2.1 引言	(28)
2.2 方案总体设计	(30)

2.3	三维人体骨骼模型的引入	(32)
2.4	三维人体模型参数的自动初始化	(33)
2.4.1	人体关节节点的自动标注方法	(33)
2.4.2	人体关节三维坐标及旋转角度计算方法	(34)
2.5	人体视频图像特征提取的具体实现	(36)
2.5.1	运动人体轮廓提取方法	(36)
2.5.2	运动人体区域灰度和颜色提取方法	(38)
2.6	基于 Beowulf 机群中改进的粒子滤波算法的具体实现	(38)
2.6.1	经典串行粒子滤波算法运动人体跟踪的实现方法	(38)
2.6.2	跟踪模板与粒子数目的调整及跟踪失效恢复方法	(40)
2.6.3	Beowulf 机群中单目标跟踪的迁移式粒子滤波并行算法	(41)
2.6.4	Beowulf 机群中多目标跟踪并行粒子滤波算法实现方法	(43)
2.7	节点任务调度与通信策略	(46)
2.8	粒子滤波串行与并行算法时间复杂度对比理论分析	(46)
2.9	Beowulf 机群中三维人体运动粒子滤波跟踪及对比实验	(47)
2.9.1	实验环境的搭建	(47)
2.9.2	单人体运动跟踪实验及实验结果	(48)
2.9.3	多人运动跟踪实验及实验结果	(51)
2.9.4	对比实验的前提条件设计	(54)
2.9.5	跟踪失效对比实验与分析	(55)
2.9.6	跟踪计算时间对比实验	(55)

2.9.7 跟踪时间对比实验结果分析	(56)
2.10 本章结论	(57)
第3章 基于四层树状语义模型的场景语义识别方法	(58)
3.1 引言	(58)
3.2 场景语义识别方法的技术路线	(60)
3.3 基于图论改进的图像分割方法	(60)
3.3.1 基于图论阈值图像分割方法的基本原理	(61)
3.3.2 阈值图像分割权值操作的自动调整方法	(62)
3.4 四层树状语义模型的引入	(63)
3.4.1 视觉词包模型的基本概念	(63)
3.4.2 四层树状语义模型的结构	(64)
3.5 四层树状语义模型视觉层的构建方法	(65)
3.5.1 图像颜色直方图的提取方法	(65)
3.5.2 颜色空间的量化与颜色特征提取	(67)
3.5.3 颜色层次图的提取方法	(69)
3.5.4 颜色及颜色层次特征与轮廓特征融合	(71)
3.6 四层树状语义模型概念层的生成	(72)
3.7 四层树状语义模型关系层的构建	(72)
3.7.1 场景对象空间位置关系的语义描述方法	(72)
3.7.2 场景对象空间位置关系关联规则的数据挖掘	(74)
3.8 四层树状语义模型语义层的描述方法	(75)
3.8.1 高层潜在语义计算方法	(75)
3.8.2 语义层的语法生成规则	(76)
3.9 四层场景语义树分类模型的构建算法	(77)
3.10 测试图像与分类模型语义相似性计算方法	(78)
3.10.1 语义特征的获取	(78)
3.10.2 语义相似性计算方法	(78)

3.11	实验与实验分析	(79)
3.11.1	功能验证实验与分析	(79)
3.11.2	对比实验与分析	(81)
3.12	本章结论	(82)
第4章	视频中运动人体身份识别的认知物理学方法	(84)
4.1	引言	(84)
4.2	运动人体身份识别认知物理学方法实现流程	(86)
4.3	数据场的引入	(87)
4.3.1	数据场势值和场强矢量的引入	(87)
4.3.2	数据场中影响因子的优选方法	(89)
4.4	基于数据场的人脸信息表征	(89)
4.5	基于数据场的步态信息表征	(90)
4.5.1	运动人体步态轮廓特征提取	(90)
4.5.2	运动人体下肢关节步态特征提取	(91)
4.5.3	运动人体步态特征的数据场描述	(92)
4.6	步态数据库样本中人脸区域的提取方法	(93)
4.7	运动人体身份识别的具体实现	(94)
4.7.1	基于力学和动力学原理的数据非线性降维	(94)
4.7.2	基于数据场样本库人的身份自动识别算法	(96)
4.7.3	基于数据场样本库人的身份自动识别算法的时间复杂度分析	(97)
4.7.4	D-S证据论的引入与改进	(97)
4.8	实验及对比分析	(100)
4.8.1	功能验证实验	(100)
4.8.2	对比实验与分析	(102)
4.9	本章结论	(104)

第 5 章 视频场景中运动人体时空关系的自然语言描述方法	
.....	(106)
5.1 引言	(106)
5.2 13 元组运动人体时空描述模型的引入	(107)
5.3 关键帧的提取方法	(108)
5.4 参照物与运动人体距离关系自然语言描述	(109)
5.5 场景中运动人体时空关系的自然语言描述语法规则	
.....	(110)
5.6 实验及实验分析	(110)
5.6.1 实验流程设计	(110)
5.6.2 人工图形序列实验	(110)
5.6.3 视频序列实验	(113)
5.7 本章结论	(115)
第 6 章 基于层次化概念空间的运动人体行为语义计算方法	
.....	(116)
6.1 引言	(116)
6.2 运动人体行为语义计算的处理框架的引入	(120)
6.3 基于混合云模型的概念空间建立方法	(121)
6.3.1 概念空间理论概述	(121)
6.3.2 基于混合云模型概念空间的定义	(122)
6.3.3 混合云模型概念空间的建立流程	(123)
6.4 原子运动概念空间层的建立方法	(125)
6.4.1 人体运动样本的选择	(125)
6.4.2 运动概念空间建立性质维的选择	(127)
6.4.3 身体姿态原子运动概念空间的建立及概念激活	
方法	(128)
6.4.4 肢体的原子运动概念空间的建立及概念激活	
方法	(130)

6.5	简单行为概念层的建立方法	(135)
6.5.1	时间逻辑关系的引入	(135)
6.5.2	空间逻辑关系的引入	(137)
6.5.3	简单行为的建模方法	(137)
6.6	简单人体运动行为语义计算方法	(139)
6.6.1	简单人体运动行为知识库的建立方法	(139)
6.6.2	简单行为语义计算流程	(141)
6.7	事件行为概念层的建立与语义计算方法	(141)
6.7.1	事件行为概念空间的建模	(142)
6.7.2	事件行为概念场景语义的具体描述	(143)
6.7.3	事件行为概念人物元组的具体描述	(144)
6.7.4	事件行为因果关系推理方法	(144)
6.8	运动人体行为语义解释的自然语言规则	(145)
6.9	实验及实验分析	(146)
6.9.1	功能性验证实验流程和实验环境搭建	(146)
6.9.2	简单人体运动行为知识库建立实验	(148)
6.9.3	简单人体运动行为语义计算实验	(148)
6.9.4	事件人体运动行为语义计算实验	(150)
6.9.5	对比实验的实验平台搭建	(153)
6.9.6	对比实验结果与讨论	(154)
6.10	本章结论	(156)
第7章	总结与展望	(157)
7.1	研究总结	(157)
7.2	研究展望	(158)
	参考文献	(160)

第 1 章 概论

1.1 研究意义

绝大多数研究人类起源的专家认为，通过语言进行交流是人类区别于其他生物最显著的特征，也是人类文明发展的关键一步。随着信息技术的整体发展，每天都有海量的视频信息产生，而让计算机系统能像人的思维方式一样，将视频序列信息进行抽象，转化成自然语言，一直是计算机视觉领域的研究热点。用语言进行信息的记录和信息的传承以及信息的交流是人类特有的能力，所以按照人的认知和思维过程来建立计算机系统从输入视频图像信息到输出自然语言是一个很有价值的技术研究方向。

将视频图像信息转化为自然语言的技术手段叫作语义计算，本书从运动人体行为语义计算入手，研究动态视频中运动人体行为的语义理解，具有重要的科学意义。该研究属于多学科交叉的研究，涉及认知科学、人工智能、计算机科学等领域。

从认知科学的角度来看，虽然人类有很多感知外界的方式，但通过视觉获取的信息占信息获取量的 80% 左右，所以探索人类视觉系统的工作原理具有重大的科学意义。目前，人类视觉系统获取信息、转化信息、加工信息和存储信息的原理，特别是将动态图像信息转化成自然语言这一“认知压缩”的机理尚未明确。

从人工智能的角度来看，计算机视觉的最终目的就是希望机

器能有和人一样的视觉系统，能够实现对运动物体的跟踪、行为语义提取，进一步用自然语言对行为进行描述和规划。视觉系统是人工智能机器的关键感知系统，因此研究动态图像序列的语义计算理论与模型是人工智能领域的重大科学问题。

从计算机科学的角度来看，计算机系统需要新一代的人机交互技术。键盘、鼠标、显示屏等传统的交互技术已经远远不能给使用者提供更好的用户体验。运动人体行为语义计算则为计算机系统的人机交互提供了全新的理论支持。

研究动态视频中运动人体行为的语义理解技术有着广泛的应用价值，主要包括：①类人机器人技术；②智能监控系统；③虚拟现实技术；④新一代人机交互接口；⑤人体运动分析；⑥基于模型的视频信息压缩编码^[1]。就智能监控系统来说，视频监控系统在全球范围内都已经是常用的安保措施，但目前的监控系统还没有真正实现智能化，应用时主要依靠人工值守。人工值守很容易出现疲倦和疏漏问题，现在监控系统迫切需要在无人值守的情况下对视频信息进行描述和理解从而实现报警和预警。让计算机系统协助或代替人来实现智能监控，就需要实现对动态图像序列语义计算。虚拟现实技术在工业、教育、科研辅助领域都有着重大的应用价值，动态图像序列的语义计算的实现，是虚拟现实技术在人体信息获取方面的技术基础。人体运动分析主要应用在医疗和运动员训练领域，在运动人体行为语义计算关键技术的研究中，人体运动跟踪是其关键技术之一，人体运动跟踪技术的发展对人体运动分析将有直接的推动作用。随着互联网技术的发展，海量的视频信息出现在网络上，如何对海量的视频信息进行压缩编码和快速检索是人们不得不面对的问题。运动人体行为语义计算将为视频信息压缩编码和快速检索提供技术支持。

综上所述，运动人体行为语义计算具有重大的理论意义和广泛的应用价值。

1.2 研究现状

运动人体行为语义计算是人体行为识别技术的实现方法之一，研究其现状首先需要研究人体行为识别技术的发展现状。人体行为识别技术广泛的应用前景和丰厚的经济价值，吸引着全世界范围内政府、大学、科研机构、商业机构对其投入大量人力、财力来推动该项技术的不断发展。1997年，美国国防高级研究计划局（DARPA）先后设立了视频监控重大项目 VSAM（Visual Surveillance And Monitoring）和 HID（Human Identification at a Distance）计划^[2]。

欧盟于1999年由 Frameworks 程序委员会设立重大项目以资助 ADVISOR 系统。该项目立足于民用功能，主要是针对公共场合的个体和群体的行为识别和分析以及人机交互方式的研究。英国雷丁大学在英国政府的支持下，先后开展了公共场合人群监视和行为识别的鲁棒性方法、视频序列理解的评估、拥挤场合公共安全的综合监控等项目。奥地利和欧盟科学基金会联合资助了大型的视频监控项目——飞行器周围环境、分类车辆和个人的跟踪以及行为的检查和解释^[3]。近几年，无论是美国还是欧盟，基于视觉的行为识别技术在军品和民品中已经有了成熟的应用。

在我国，虽然人体行为识别技术起步较晚，但是也引起了政府和科研机构的足够重视。中国科学院成立了生物识别与安全技术研究中心，该中心隶属于中科院自动化所模式识别国家重点实验室。北京大学、清华大学等高校也设立了各种科研课题对此领域开展研究。国家自然科学基金委、国家863项目、国家973项目都对此领域的各类项目进行了立项资助。2005年以来，我国公安部、交通部、科技部联合实施了“天网工程”“平安城市”“智能

交通系统”等一系列项目，搭建了大量的视频监控软硬件平台。这些平台为运动人体行为识别提供了广阔的应用空间。此外，军事、新形势下的反恐等领域也为运动人体的行为识别提出了迫切的发展需求。

人体行为识别技术是一个复杂、多样和充满挑战的领域，近年有大量的学者进行了研究，并且取得了丰富的研究成果。表 1.1 罗列了近年的主要研究成果，表 1.2 罗列了主要的技术手段和算法的发展情况。

表 1.1 人体行为识别技术主要发展成果

时间	作者	主要贡献
1994	Aggarwal ^[4]	普通非刚性运动的识别
1994	C. Cedras ^[5]	运动检测
1994	C. Cedras ^[6]	运动和运动轨迹跟踪
1997	Aggarwal ^[7]	人体运动分析：身体部分，多重跟踪
1997	Bobick ^[8]	关于运动的机器视觉
1999	Aggarwal ^[9]	图像序列的单镜头和多镜头跟踪
1999	Gavrila ^[10]	基于 2D 和 3D 模型的人体行为识别
2001	T. B. Moeslund ^[11]	运动捕捉和跟踪，姿势估计和识别
2002	王亮 ^[12]	人体运动检测和行为识别
2003	Buton ^[13]	场景理解
2003	Wang ^[14]	基于身体部分跟踪实现行为识别
2004	胡卫 ^[15]	自动场景中的视频监控系统
2005	Valera ^[16]	自动监控系统
2006	Yilmaz ^[17]	人体运动跟踪
2006	Moeslund ^[18]	初始化、跟踪、姿态估计、识别
2007	Poppe ^[19]	基于模型的运动自由方法

续表1.1

时间	作者	主要贡献
2007	Gandhi ^[20]	行人保护行为模型
2007	Pantic ^[21]	HCI 语境下的行为识别
2008	Ko ^[22]	视频监控、行为分析
2008	Kumar ^[23]	多模型框架下的数据融合
2008	Turaga ^[24]	人体运动检测
2008	Zhou ^[25]	运动跟踪和恢复
2009	Enzweiler ^[26]	行人检测系统
2009	Wei ^[27]	人体行为识别
2009	G. Lavee ^[28]	视频事件的理解（抽象和模型）
2010	Geronimo ^[29]	行人保护系统
2010	Gandamo ^[30]	人体行为识别在运输上的应用
2010	Poppe ^[31]	图像描述和行为分类模型

表 1.2 人体行为识别算法主要发展成果

模型框架、学习算法和技术	采用该算法的文献
改进的利他向量量子化算法 (AVQ)	[32]
基于马尔科夫模型	[33], [34]
明确的状态持续 ESD-HMM	[35]
基于改进贝叶斯的半监督 HMM	[36]
基于混合高斯模型的 HMM	[37]
持续状态的 HMM	[38]
双模型 HMM	[39]
无限隐马尔科夫模型 (iHMM)	[40]
双观察 HMM (MOHMM)	[41]

模型框架、学习算法和技术	采用该算法的文献
HDP-HMM	[42]
HMM-SVM	[43]
马尔科夫随机场 (MRF)	[44]
概率规则内容分析 (MPPCA)	[45]
贝叶斯模型分类器	[46]
动态贝叶斯网 (DBN)	[47]
簇 (分类器)	[48]
泊松模型	[49]
模糊 K 均值算法	[50]
模糊 C 均值算法	[51]
共现统计和二元共现簇	[52]
支持向量机 (SVM)	[53]
双层支持向量机 (SVM)	[54]
模糊神经网络	[55]
基于规则	[56]
基于距离函数	[57]
基于相似函数	[58]
概率潜在语义分派	[59]
自组织矩阵	[60]

从表 1.1 可以看出, 人体行为识别技术研究成果已经涵盖了其技术手段的各个环节, 包括运动检测、轨迹跟踪、场景理解、运动跟踪与恢复、数据融合等。从表 1.2 的研究成果看出, 在人的行为识别与理解技术研究领域的主要三个手段中, 模版匹配法^[57-60] (典型的模板匹配法有基于规则、基于距离函数、基于