自然环境下日常动作的在线识别

曹媛媛[1],黄飞跃[2],陶霖密[1],徐光祐[1]
(1. 普适计算教育部重点实验室,清华大学计算机系,北京 100084; 2. 腾讯研究院,北京 100080)

摘 要: 自然环境下的日常动作识别有着广泛的应用前景和重要的研究价值. 不同于以往在结构化和孤立条件下进行的动作识别,自然环境下的日常动作是连续的,视角多变并常发生遮挡. 本文提出了分布式视觉系统下日常动作的在线识别方法. 时间轴上的滑动窗口每个时刻取一段视频帧,采用基于"包容形状"的视角无关的体态表示方法提取体态特征向量,并用隐马尔科夫模型进行识别. 动作类型的搜索空间由环境知识推理得到. 遮挡检测和部分遮挡下的体态表示也在文中进行了讨论. 实验表明本文提出的日常动作的在线识别方法能够克服日常场景给动作识别带来的困难,结果证实了方法的有效性.

关键词: 动作识别; 视角不变; 遮挡检测
中图分类号: TP391 **文献标识码**: A **文章编号**: 0372-2112 (2009) 4A-016-06

On-Line Recognition of Actions in Daily Living

CAO Yuan-yuan[1], HUANG Fei-yue[2], TAO Lin-mi[1], XU Guang-you[1]
(1. *Key Laboratory of Pervasive Computing of Ministry of Education, Department of Computer Science and Technology, Tsinghua University, Beijing* 100084; 2. *Tencent Research, Beijing* 100080)

Abstract: Recognition of actions in daily living is challenging because: (1) actions are continuous; (2) human location is changeful; (3) human body is partially occluded sometimes. This paper proposes a multi-view framework for on-line recognition of actions in daily living. Action representation based on "Envelope shape" enables view-invariant action recognition. A sliding window concatenates the feature vectors for action representation into a stream as the input to a bank of HMMs. A maximum likelihood based classifier detects action. The HMMs are chosen by an ontology based environment knowledge model. Besides, occlusion detection and action representation in occlusion are also discussed. Implementations demonstrate the efficacy of our approach.

Key words: action recognition; view invariant; occlusion detection

1 引言

动作识别一直是计算机视觉和人机交互领域重要的研究方向. 目前在这方面的研究大多数假设是在结构化和孤立(constructed and isolated)条件下进行,如,手势命令的识别. 而随着视觉监控、老年人看护系统、智能厨房、智能会议室等新应用的不断产生,对自然环境和现实生活中的动作识别已成为新的研究热点,这类动作是连续的,位置视角多变的,而且活动范围广,与环境有频繁交互. 如,老人看护中所需要的是对自然和连续的大范围场景中的动作的识别. 与传统的动作识别相比,自然环境下的动作识别面临着如下难题:如何克服由于视角,距离,遮挡等多种因素带来的干扰和不确定性;如何实现对连续动作的在线识别;如何有效利用环境信息. 本文针对以上问题提出了能够适应视角变化、距离变化、并能处理部分遮挡的连续动作的在线识别方法.

对动作的识别其实是对一段时间上的体态特征序列的分类和识别,因此体态表示成为动作识别中最基本也最关键的问题. 在自然环境中,我们不能限制人体的移动和转动,因此体态表示需要能够容忍视角、距离的变化和部分遮挡的情况. 由于距离和部分遮挡的问题可以通过摄像机规划来解决,所以重点讨论视角无关的动作识别. 视角不变的方法有这样的两难问题:为了容忍视角的变化,通常会存在信息的损耗,这将造成动作区分度的降低. 如何使得体态的表示可以在容忍视角的变化同时又保留足够的可区分信息是一个关键的问题.

目前已经提出了一些视角不变的动作识别方法. 三维模型通常是视角无关的,Campbell[1]等人提出的基于立体视觉数据的三维手势识别系统. Jin[2]建立了基于三维模型的动作识别系统. 三维模型通常参数多,训练

收稿日期:2008-11-06;修回日期:2008-12-12
基金项目:国家自然科学基金(No. 60673189)

复杂,计算量大,自由度大,一般比较难以精确获得. 二维表观模型相比三维模型使用的较多,Ogale[3]通过为每个动作建立各种不同视角下的多个表观模型来进行视角不变的动作分析. 这在训练阶段带来巨大的计算量. 另一种想法是试图合成与训练一致的视角,因为许多视角相关的研究都证明了从与运动方向正交的视角可以获得最佳的体态特征. Rao[4]使用人手质心的轨迹来描述由一只手完成的动作并利用轨迹的仿射不变量开发了可以自动运作的识别系统. 这种方法在手部动作以外的通用性还有待更深入的研究. Parameswaran[5]重点研究了视角不变的人体动作识别方法. 他选择了人体的六个连接点,并且计算每一个体态中它们的三维不变量. 这种方法依赖鲁棒的语义特征点检测或者是点对应,而这些在自然环境下是比较难实现的.

以上研究都是在已经分割好的动作视频数据库上测试,这些数据往往背景单一,动作者位置固定,无遮挡,动作过程中不存在平移和转动. 对自然环境下的动作识别的研究也多是利用简单的体态特征对分割后的视频段进行推理[6]. 只有不多的工作研究了自然场景下的连续动作的识别. Wang[7]是在单摄像机,且视角处于区分动作的最理想的方向上,提取简单的体态轮廓特征进行识别,该体态特征只有在特定视角下才对动作有区分度,因此限制了人的运动方向和动作朝向;对每帧进行孤立的分析忽略了动作前后连贯的重要特征. 还有一些工作通过分析人的轨迹来分析识别人的一系列动作[8]或者通过检测被使用物体来分析识别人的动作[9].

目前还没有适用于自然环境下的连续、位置视角多变的日常动作的在线识别方法. 目前的研究中,视角无关动作识别的研究仅仅针对固定场景下已分割无遮挡动作进行识别,而自然场景下连续动作的在线识别往往避开了对多视角下动作的识别.

本文提出了在分布式视觉系统下对视角多变的连续人体动作进行识别的方法. 方法不受人体自由移动和转动的影响,且能够解决部分遮挡和移出视场的问题. 在动作识别的体态表示阶段,我们选择视角无关的"包容形状"来描述体态,从两个近似正交的摄像机拍摄到的观测目标轮廓信息中可以方便的获得"包容形状"体态表示. 它不需要依靠任何较难实现且对误差很敏感的语义点检测和点对应过程. 对小部分遮挡或移出视场的情况,扩展的"包容形状"仍然可以有效的识别动作. 在线的识别是通过滑动窗口来实现. 一段时间上的体态特征向量被输入到一组根据当前环境知识选择的隐马尔科夫模型中进行识别,最匹配的模型表示了动作识别的结果. 本文还讨论了摄像机标定和遮挡检测等内容.

2　遮挡检测和摄像机选择

我们搭建了一个分布式的智能家居实验平台,实验场景布置如图 1 所示. 房间安装了四个固定的摄像机,为了满足提取包容形状特征提取算法对摄像机的要求,实验中每相邻两个摄像机光轴之间的夹角为π/3～2π/3,并且使摄像机的像平面与竖直轴平行,第 3 小节会具体分析.

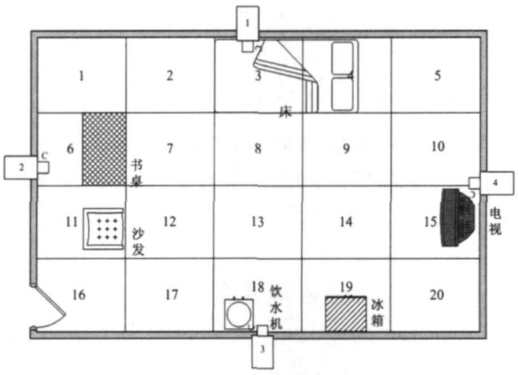

图1　实验场景布置

在分布式的视觉系统下,为避免或减少遮挡的发生,摄像机规划是首要解决的问题. 本小节先讨论遮挡检测方法,然后给出摄像机的规划方法. 这里主要讨论最常见的站立人体上/下半身被遮挡的情况. 如果可以计算出人体在当前位置时在像平面的高度,那么就可以知道是否发生了遮挡,遮挡比例是多少.

由于人体本身的深度和人体到摄像机的深度相比通常较小,我们可以利用仿射摄像机模型. 这时候人体成像大小和人体到摄像机光心的距离成反比. 假设人体到摄像机光心的距离为 Z,那么可以得到

$$h = \frac{k}{Z} h_w \tag{1}$$

在式(1)中,h_w 是三维世界坐标系中人体的高度,h 则是图像中人体的高度,k 是比例系数. 对于一个特定固定放置的摄像机(相当于内外参数确定),k 是一个常数. 对于仿射摄像机,Z 也就是人体脚点到摄像机光心的距离,$Z = Z_c$.

在摄像机坐标系中,地平面可以用以下方程来表示,对于一个特定固定放置的摄像机:

$$aX_c + bY_c + cZ_c + d = 0 \tag{2}$$

其中 a, b, c, d 都是常系数.

对应地面上的点(X_c, Y_c, Z_c),它在图像坐标系的成像为(u, v). 那么由摄像机模型可以得到

$$\begin{cases} u = \dfrac{fX_c}{Z_c} + u_0 \\ v = \dfrac{fY_c}{Z_c} + v_0 \end{cases} \tag{3}$$

由式(2)、(3)可以得到

$$\frac{a}{f}(u - u_0)Z_c + \frac{b}{f}(v - v_0)Z_c + cZ_c + d = 0 \quad (4)$$

于是 Z_c 可以用如式(5)所示的形式来表示,其中对于特定固定放置的摄像机,k_a,k_b 和 k_c 也都是常系数.

$$Z_c = k_a u + k_b v + k_c \quad (5)$$

由于 $Z_c = Z$,所以最终我们可以得出

$$h = \frac{1}{k_1 u + k_2 u + k_3} h_w \quad (6)$$

其中,k_1,k_2 和 k_3 是常系数,即

$$h = S(u, v) * h_w \quad (7)$$

式(6)可以看出,人体在不同的位置,成像的高度随着人离摄像机的位置而变.成像高度的缩放倍数是人体脚点在不同的位置的图像坐标的函数 $S(u, v)$.对于不同的人,只有 h_w 不同,但是缩放比例因子 $S(u, v)$ 都是相同的.

根据式(6),对于特定摄像机,只需知道人在不同的位置时的几个高度,由于人体实际高度 h_w 已知,即已知若干组(h, u, v),就可以求出系数 k_1,k_2 和 k_3 的值.这样,当已知了人体位置和实际高度后,就可以计算出人体的图像高度,那么便可以检测到是够遮挡或者遮挡比例.

在任意时刻,当我们从 4 个同步的摄像机得到 4 帧图像时,计算每一路视频帧中人体未被遮挡的比例,平均值最大的一对相邻摄像机将被选择.包容形状特征的提取就从这两路选定的摄像机的视频帧中计算.下一小节介绍双摄像机下"包容形状"体态特征的抽取.

3 基于"包容形状"的体态表示

大多数的人体动作识别方法通常要求人体朝向相对摄像机固定,这时视角固定,从而可以基于视角相关的体态表示进行识别.然而在自然环境下的视觉应用中,通常不能限制被观测人身体的移动和转动.如老年人看护,智能会议,安全监控等.这就要求动作识别系统能够有良好的视角不变性,即在各种不同的视角下面基本保持一致的体态不变量表示.这里只讨论由人体运动引起的视角变化.理论上说,视角的变化包括围绕 x,y,z 轴三个方向的旋转分量.而在实际应用中,引起视角变化的人体的运动包括平移运动和自身的旋转,就是围绕竖直轴的旋转分量.这里介绍的算法只考虑解决这个方向上的视角变化的问题.

假设一种双摄像机的配置方式,如图 2,这两个摄像机的成像平面都和竖直轴 Y 平行,它们的光轴是正交的,同时它们像平面坐标系中的 V 轴都和 Y 轴平行.让我们来考虑人体的一个水平截面 H,在这个截面上的所有点到像平面 1 上的投影都在直线 L_1 上,而在这个截面上的所有点到像平面 2 上的投影都在直线 L_2 上.即直线 L_1 是点 p_2 的外极线,而直线 L_2 则是点 p_1 的外极线.这样人体俯仰就相当于人体所有的水平截面在自身对应的二维平面内做了一个旋转.为了发掘俯仰不变量,我们可以分析人体在某个高度的二维水平截面 H 上的切面形状在旋转时的投影变化情况.

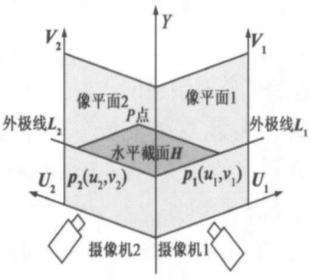

图2 双摄像机配置方案

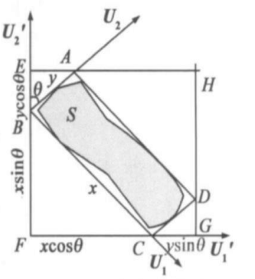

图3 旋转时水平截面在U_1轴和U_2轴上的投影

如图 3,由于两个摄像机光轴正交,所以 U_1 轴和 U_2 轴的夹角是 90 度.假设在水平截面 H 上人体外轮廓对应形状 S,它在原始的 U_1 和 U_2 轴中的投影线段是 AB 和 BC,那么 S 在矩形 $ABCD$ 里面.在另外一个旋转了某个角度 θ 的 U_1 和 U_2 轴中,它的投影在线段 EF 和 FG 中.这里我们定义原始投影线段的长度为 x 和 y,而新的投影线段的长度则是 x 和 y ′.那么,可以得到如下关系式

$$y' \le x\cos\theta + y\sin\theta \quad y' \le y\cos\theta + x\sin\theta \quad (8)$$

$$r = \sqrt{x^2 + y^2} \quad (9)$$

那么旋转后为

$$r' = \sqrt{x'^2 + y'^2} \le \sqrt{x^2 + y^2 + 2xy\sin 2\theta} \le \sqrt{2}\, r \quad (10)$$

取 r_0 是所有旋转对应的各个 r 中的最小值,那么在任何的旋转下,相应的 r 都会满足

$$r_0 \le r \le \sqrt{2}\, r_0 \quad (11)$$

和原始投影值 x 与 x 或者 y 与 y 之间比值的无限范围区间相比较,这已经是一个相当小的取值区间,也就是说我们找到了一种视角不敏感的人体表示.对于每一个高度的水平截平面,我们利用式(9)来计算一个 r 值.可以给出包容形状 r 构成的 R 向量的定义

$$\text{包容形状} \quad R = [r_1, r_2, \cdots, r_N]^T \quad (12)$$

其中 $r_1 = \sqrt{x_i^2 + y_i^2}$,下标 i 表示高度.

这样,对于每一静态帧的人体体态,我们可以得到一个包容形状 R 向量.由于这个向量构成的形状可以把人体的轮廓包围在内部,我们把这个 R 向量称作为

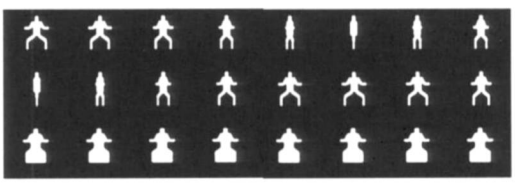
图4 不同视角下的体态

“包容形状”(“Envelop Shape”).

图 4 给出一些在不同视角下合成人体模型的包容形状图像,表示了体态“半蹲”围绕着竖直轴(Y 轴)旋转了八个不同角度时的情况. 每组前两行是两个正交摄像机拍摄的外轮廓图像,而第三行则是计算出的包容形状图像. 从图中我们可以看到在视角变化时,包容形状的变化很小.

我们对摄像机光轴非正交情况下的包容形状进行了扩展研究. 当光轴的夹角为 $\pi-\alpha$,由立体视觉的知识,可以推导出两个像平面夹角为 α,按照图 5 可以分析如下:

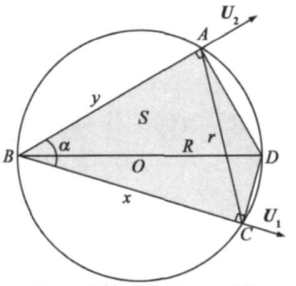

图5　光轴夹角是$\pi-\alpha$时的水平截面图

我们令 β 为:

$$\begin{cases}\beta=\alpha, & \text{当 } \alpha<\pi/2\\ \beta=\pi-\alpha, & \text{当 } \alpha>\pi/2\end{cases} \tag{13}$$

则可以定义扩展的包容形状为:

$$r=\sqrt{x^2+y^2-2xy\cos\beta} \tag{14}$$

由几何关系我们可以得到:

$$R\sin\beta=r \tag{15}$$

取 r_0 是所有旋转对应的各个 r 中的最小值. 那么在任意旋转下,r/r_0 的最大取值是 α 的一个函数,记做 $f(\alpha)$,它的取值对应了包容形状的视角不变性. 我们画出 $f(\alpha)$ 的对应函数曲线如图 6 所示,$f(\alpha)$ 越小表示包容形状 r 的视角不变性越强. 由函数曲线来看,当 $\alpha=\pi/3$ 或 $\alpha=2\pi/3$ 时,$f(\alpha)$ 取值最小,也就是说此时视角不变特性最好.

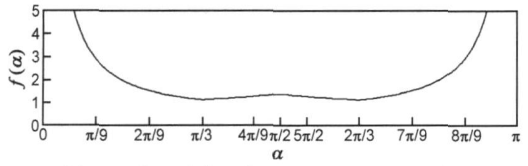

图6　包容形状的视角不变性:$f(\alpha)$函数曲线

4　动作识别系统和实验

经过人体检测、特征提取和体态表示,人体动作就可以转化为时序特征向量序列. 这时候的动作识别可以看成是对时序特征向量序列的识别问题. 这里我们采用隐马尔可夫模型 HMM(Hidden Markov Model)来进行动作的训练和识别.

我们用 $\lambda=(\pi, A, B, M, N)$ 来表示某个动作的模型,观测值序列为 $O=O_1, O_2, \cdots, O_T$,其中 O_t 表示时刻 t 的观测值. 在 HMM 的公式表示中,λ 的每个状态表示一个动作;$A=\{a_{ij}\}$ 表示从状态 i 到状态 j 的转移概率;$B=\{b_i(O_t)\}$ 是状态 i 在 t 时刻的观测值的概率;$\pi=\{\pi_i\}$ 表示状态 i 的初始状态概率向量. M 表示观测向量的维度,N 为 HMM 模型的状态数.

实验中,每一帧提取出基于包容形状的体态表示被抽象为数学形式上的特征向量,那么每一个动作序列就生成了对应的时序特征向量序列,这样人体动作就被抽象为特征向量序列,这个序列就作为 HMM 模型的观测值序列. 观测值概率密度函数我们用高斯概率密度函数来表示:

$$b_j(X)=\sum_{k=1}^{K}c_{jk}b_{jk}(X)=\sum_{k=1}^{K}c_{jk}N(X,\mu_{jk},\Sigma_{jk})$$
$$1\leq j\leq M \tag{16}$$

其中,$N(X,\mu_{jk},\Sigma_{jk})$ 为多维高斯概率密度函数,μ_{jk} 为其均值向量,Σ_{jk} 为方差矩阵,K 为组成 $b_j(X)$ 的混合概率密度个数,c_{jk} 为组合系数,且 $\sum_{k=1}^{K}c_{jk}=1$.

我们采集了相应的视频段并建立了自己的动作数据库. 这个数据库包含七个不同的动作者,每个动作者表演了生活中常见的六种动作,分别是“伸手”、“弯腰”、“坐”、“躺”、“站起”和“行走”. 这些动作涵盖了日常生活中人的四肢和躯干的基本运动,具有一定的典型性. 动作者会在三个任意视角下面执行每一个动作三遍. 图 7 显示了我们样本数据的一些示例,前两行是两个摄像机的图像,第三行,第四行是前景分割提取得到的人体轮廓,最后一行则是包容形状向量生成的图像. 动作序列都是在任意视角下面采集的.

为了在连续的视频上进行动作识别,我们在时间轴上取一个长度为 n 的滑动窗口,实验中设置 $n=30$,在 n 帧长的视频段上,计算每帧的人体体态特征向量组成特征向量序列,那么这段视频段就被识别为最匹配的 HMM 模型表示的动作. 其中最匹配的意义为:对于每个动作的 HMM 模型中,用 Viterbi 算法计算可得到最优意义上的状态序列 $Q=\{q_1, q_2, \cdots, q_T\}$,最优的条件是 $P(Q,O|\lambda)$ 最大;比较不同动作的 HMM 模型得到的 $P(Q,O|\lambda)$ 的值,概率最大的认为最匹配.

在识别和训练过程中需要注意几个问题. 当没有遮挡时,要对包容形状进行尺度正规化. 把所有包容形状进行尺度拉伸,到同一个长度,以使得特征向量的维

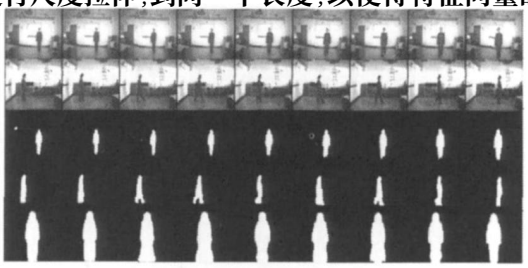
图7　一组步行的动作序列

度都一样. 当站立人体的下半部分有遮挡时,检测得到的未被遮挡的部分人体经尺度归一化后得到的包容形状特征向量为$[r_1', r_2', \cdots, r_{h1}']$,为了减少数据训练和识别的计算量,我们假设人体腿部无动作,如果直立静止人体包容性状特征向量表示为$[r_1, r_2, \cdots, r_H]$,其中H表示尺度正规化后的包容性状特征向量的维度,那么恢复后的人体包容性状特征向量表示为$[r_1, r_2, \cdots, r_{H-h1}, r_1', r_2', \cdots, r_{h1}']$.

图8是系统运行的流程图. 实验中摄像机的个数$k=4$. 程序运行的步骤如下:在已经同步的四路视频中任选相邻两路检测到人体的视频帧,采用Hu[10]提出的双摄像机下人体定位方法计算人体的脚点位置,这种定位方法的好处是在人体被部分遮挡或者部分移出视场的情况下,仍然可以确定脚点的位置. 计算遮挡比例,从四路视频中选出无遮挡或遮挡最少的相邻两路视频图像I_1和I_2生成"包容形状",用"包容形状"每个高度上的半径r作为特征向量,为了提高运算效率,我们用PCA将特征向量降到8维. 与之前$n-1$帧的特征向量序列被送入一组HMM模型进行识别.

滑动窗口每次取出的一段视频如果用所有的HMM模型进行识别,会带来巨大的计算量. 由于在日常生活

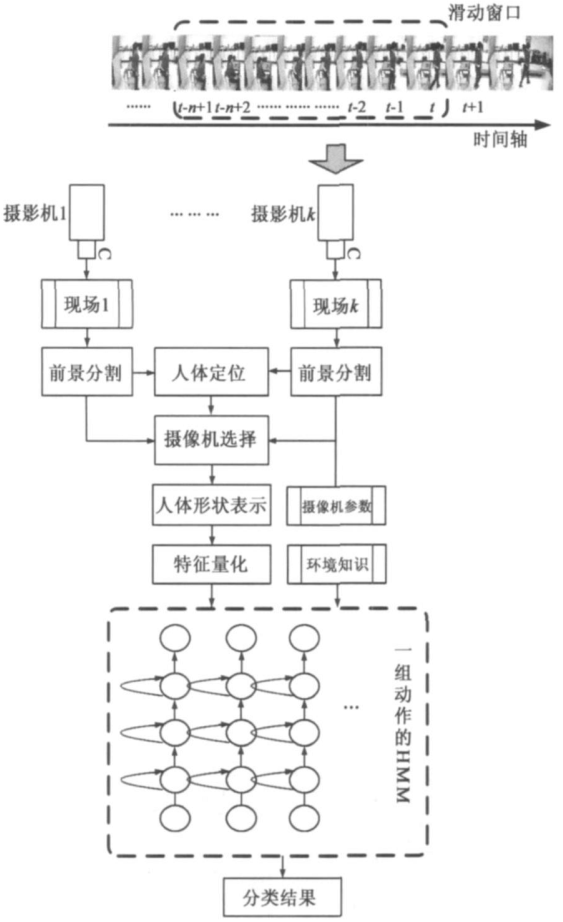

图8 系统流程图

中,人们可能的动作与当时所处的环境有着密切的关系,人所处的位置可以指导对动作的理解. 例如"坐"这个动作通常发生在椅子边,而"躺"通常发生在床上,当"躺"在地板上时,可能是发生了晕倒. 环境信息为我们进行动作识别提供了重要的线索. 因此我们建立了基于本体论(ontology)的环境知识模型. 场景中的家具、物品、摄像机和功能区域等都被看作本体,也可以称为实体(entitiy). 模型中记录了这些实体的属性和功能以及可能导致的动作. 其中,实体的属性包括坐标、尺寸、与其他实体的关系等,模型按照家居功能区域(卧室、厨房、起居室等)分类;图9展示了智能家居中环境知识模型中的一部分. 实验中,当人体位置靠近某个"实体",与这个"实体"相关联的动作就被作为对当前待识别动作的预测,因此避免了用所有的HMM模型一一识别,而只选取预测动作的HMM进行识别. 在环境知识的指导下算法效率得到很大提高,也减少了识别错误率,同时由动作识别上升到动作理解阶段.

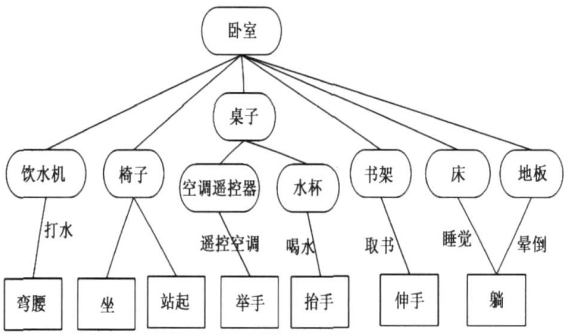

图9 部分环境知识模型:场景为卧室的例子

动作识别的实验在由四个表演者表演的4段视频上进行. 每段视频中都包括如下七个动作:"接水""睡觉""晕倒""行走""坐下""站起""伸手". 4个表演者按照自己习惯的动作顺序进行表演,因此完成动作的顺序是不尽相同的,同时,动作幅度、角度、以及姿态、顺序都没有限制,尽可能还原真实的生活状态. 实验场景如图10.

图10 不同视角下的实验场景

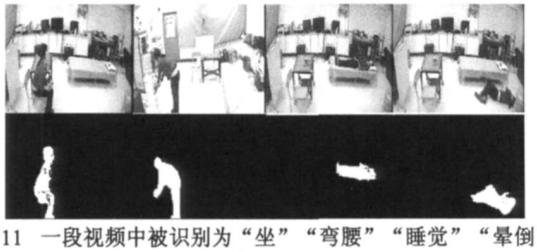

图11 一段视频中被识别为"坐""弯腰""睡觉""晕倒"(自左向右)的视频帧举例

我们在四段视频上的动作分析正确率如表 1 所示. 正确率按如下公式计算

$$\text{动作 } A \text{ 的正确率} = \frac{\text{正确识别的帧数}}{\text{动作 } A \text{ 的总帧数}} \quad (17)$$

表 1　各种动作在不同视频中的识别率

	视频 1	视频 2	视频 3	视频 4
坐	96.4	89.9	92.3	97.0
睡觉	89.1	86.7	88.5	90.0
接水	97.3	83.9	94.2	97.5
晕倒	86.6	85.4	88.7	89.0
行走	98.0	93.9	97.3	99.1
伸手	90.8	79.6	95.8	95.2
站起	93.4	83.3	92.8	96.5

从动作识别结果可以看到,视频 4(图 10 第 4 行)的结果普遍比其他视频较好,是因为视频 4 中的表演者衣着颜色与背景差异较大,能够从人体分割的结果中抽取到较为准确的体态特征,因此得到较高的识别率. 结果最差的是视频 2(图 10 第 2 行),因为视频 2 中表演者的裤子与地面颜色相近,导致体态特征抽取不够准确. 动作"晕倒"的识别率普遍比其他动作要低,是因为晕倒是作为一个异常事件检测,没有对"晕倒"这个动作本身建模,而是由动作"躺"和人体位置不在床上共同推理出. 从识别结果来看,方法对生活场景下的自然动作有较高的识别率,能够较好的处理遮挡、平移和转动等难题. 但是在视频 2 上较低的识别率也反映出了对底层处理结果的依赖性.

5　总结

本文提出了一个分布式视觉系统下日常动作的在线识别方法. 基于"包容形状"的体态表示解决了生活中观测视角多变的难题;时间轴上的滑动窗口选择一段特征向量序列用 HMM 识别,实现了在线处理;环境知识模型对识别过程进行指导,提高了算法效率和准确性,而且给动作赋予了更高层次的语义. 本文还讨论了通过人体定位、高度校准的方法解决了平移导致人体尺寸变化的问题;通过检测遮挡可以选择最优视角的摄像机避免或减少遮挡的影响,同时扩展的"包容形状"也能够解决部分遮挡的动作识别. 实验结果证明了此方法的准确性和有效性.

参考文献:

[1] L W Campbell, D A Becker, A Azarbayejani, A F Bobick, A Pentland. Invariant features for 3D gesture recognition[A]. Proceedings of International Conference on Automatic Face and Gesture Recognition[C]. Vermont, USA: IEEE, 1996. 157 - 162.

[2] N Jin, F Mokhtarian. Image-based shape model for view-invariant human motion recognition[A]. Proceedings of Conference on Advanced Video and Signal Based Surveillance[C]. London: IEEE, 2007. 336 - 341.

[3] A S Ogale, A Karapurkar, Y Aloimonos. View-invariant modeling and recognition of human actions using grammars[A]. International Conference on Computer Vision, Workshop on Dynamical Vision[C]. Beijing, China: Springer Verlag, 2005.

[4] C Rao, A Yilmaz, M Shah. View-invariant representation and recognition of actions[J]. International Journal of Computer Vision, 2002, 50(2): 203 - 226.

[5] V Parameswaran, R Chellappa. Using 2D projective invariance for human action recognition[J]. International Journal of Computer Vision, 2006, 66(1): 83 - 101.

[6] P C Chung, C D Liu. A daily behavior enabled hidden Markov model for human behavior understanding[J]. Pattern Recognition, 2008, 41(5): 1572 - 1580.

[7] Y Wang, K Huang, T N Tan. Abnormal activity recognition in office based on R transform[A]. Proceedings of IEEE Conference on Image Processing[C]. San Antonio, Texas: IEEE, 2007. 341 - 344.

[8] N T Nguyen, D Q Phung, S Venkatesh, H Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. San Diego, CA, USA: IEEE, 2005. 955 - 960.

[9] J Wu, A Osuntogun, T Choudhury. A scalable approach to activity recognition based on object use[A]. Proceedings of IEEE Conference on Computer Vision[C]. Rio de Janeiro, Brazil: Springer Verlag, 2007. 1 - 8.

[10] W M Hu, X Zhou. Principal axis-based correspondence between multiple cameras for people tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2006, 28(4): 663 - 671.

作者简介:

曹媛媛　女,1982 年生于山东烟台,博士研究生,主要研究领域为图像视频信号处理和计算机视觉等.
E-mail: caoyy1205@gmail.thu.edu.cn

黄飞跃　男,1979 年生于江苏南通,2008 年获得清华大学计算机博士学位,现在腾讯公司任职. 主要研究领域为计算机视觉、数字图像处理、模式识别等. E-mail: huangfeiyue79@hotmail.com

陶霖密　男,1962 年生于浙江省黄岩县,清华大学计算机系副教授,主要研究领域为计算机视觉、人机交互.
E-mail: linmi@tsinghua.edu.cn

徐光祐　男,1940 年生于上海,教授,博士生导师,主要研究领域为计算机视觉、移动机器人视觉导航、多媒体技术、自然的人机交互、普适计算等. E-mail: xugy.dcs@gmail.com