

基于最优 DAGSVM 的服务机器人 交互手势识别

钱 堃 马旭东 戴先中 胡春华

(东南大学自动化学院, 南京 210096)

摘 要 针对目前服务机器人手势交互方法在输入方式自然性和识别方法可靠性方面的不足, 提出采用结合人脸和人手的姿态作为输入方式, 实现了一个基于最优有向无环图支持向量机(DAGSVM)的手势识别系统。系统采用分步细化特征检测过程, 即先粗检肤色, 然后分别利用人脸 Gabor 特征和人手边缘小波矩特征检测脸和手部, 可克服背景中的肤色干扰, 并显著提高特征提取的可靠性; 综合利用脸手区域不变矩和手的位置信息组成混合特征向量, 采用优化拓扑排序策略组织多个两分类支持向量机(SVM), 构成最优 DAGSVM 多分类器, 达到比普通 DAGSVM 更高的多分类准确率。实验验证了该方法的有效性和可靠性, 并用于实现一种自然友好的人机交互方式。

关键词 手势识别 小波矩 有向无环图支持向量机 人机交互

中图法分类号: TP242.6 **文献标识码**: A **文章编号**: 1006-8961(2009)01-0118-07

Optimal DAGSVM Based Posture Recognition for Human-robot Interaction

QIAN Kun, MA Xu-dong, DAI Xian-zhong, HU Chun-hua

(School of Automation, Southeast University, Nanjing 210096)

Abstract A vision-based posture recognition system is proposed utilizing Optimal DAGSVM (Directed Acyclic Graph Support Vector Machine) classifier to achieve natural and reliable human-robot interactions. Coarse-to-fine feature detection scheme extracts skin-colored candidate regions, followed by face and hand verifications with Gabor filtered eye features and wavelet-moments of hand edge respectively. Statistical invariant moments and relative coordinates of face and hand regions are calculated as pattern feature vectors. A set of binary SVM classifiers are combined using Decision Directed Acyclic Graph with optimal structure to construct a more accurate multi-class DAGSVM classifier. Experimental result validates the reliable performance of the approach, where a natural and friendly interaction is achieved with a service robot.

Keywords posture recognition, wavelet-moments, directed acyclic graph support vector machine(DAGSVM), human-robot interaction

1 引言

在机器人监护服务研究课题中, 利用机器人视觉系统识别服务对象的指令手势是实现智能人机交互(HRI)的一种自然友好途径^[1], 其重要研究价值

在于手势以一种自然非接触的方式为机器人提供了一种冗余的人机间沟通渠道, 尤其在语音信息存在模糊含义或声音嘈杂的环境中更加重要; 结合语音信息时, 手势可以实现精确的动词+补语或动词+宾语结构语义表达, 例如“go there”, “stop here”或“grasp it”等。

基金项目: 国家重点基础研究发展计划(973)项目(2002CB312200); 国家高技术研究发展计划(863)项目(2004AA420110)

收稿日期: 2007-04-13; **改回日期**: 2007-06-26

第一作者简介: 钱堃(1982~), 男。现为东南大学自动化学院博士研究生。主要研究方向为机器人视觉与人机交互。E-mail: martinqk@163.com

由于手势的多义性和应用目的不同,目前国内手势识别的研究主要有 3 种类型:利用衣着颜色特征识别手臂姿态^[2]控制机器人运动,但其固定衣着颜色的限制不符合人机交互的自然性原则;利用形状特征识别手形^[3],主要依赖于手形边缘的精确提取,但对于运动中的机器人视觉系统较难保证识别精度;综合利用脸和手等部位的姿态或动作作为手势特征^[4-6],人脸和手的自然肤色特征基本不受场地环境条件限制,对手势发出者的体态衣着没有额外要求,与特定人无关,因此适合室内自然环境下与机器人实时交互。

手势一般可定义为“手和手臂结合,参照人脸的相对位置所产生的各种姿态”,其识别方法包括手势特征提取和特征分类两个步骤。文献[4]采用级联的神经网络识别人脸注视方向和手臂指向,从而识别人体的指向性手势;文献[5]采用所有肤色区域作为手势特征,以径向基神经网络为分类器识别手势,该方法对背景环境中的类似肤色物体干扰

不稳定。文献[6]仅以脸与手的相对位置关系为手势特征,以最小距离模板匹配法实现各种手势的分类。

针对上述方法在特征和分类方法方面的不足,基于室内服务机器人提出并实现了一个可靠的手势识别系统。由粗到精(coarse-to-fine)的特征检测过程,综合利用肤色、眼手区域纹理等特征检测脸和手,克服了仅用肤色检测提取手势区域的不足;结合脸手区域像素的统计矩和手的位置信息构成混合特征向量,采用有向无环图支持向量机(DAGSVM)多分类器对特征向量进行分类。为解决其中决策有向无环图(DDAG)的拓扑排序问题,提出一种最优节点拓扑排序策略,提高了手势分类的准确率。

2 系统构成与基本原理

如图 1 所示,服务对象手势的识别需要经过 3 个主要模块的处理:

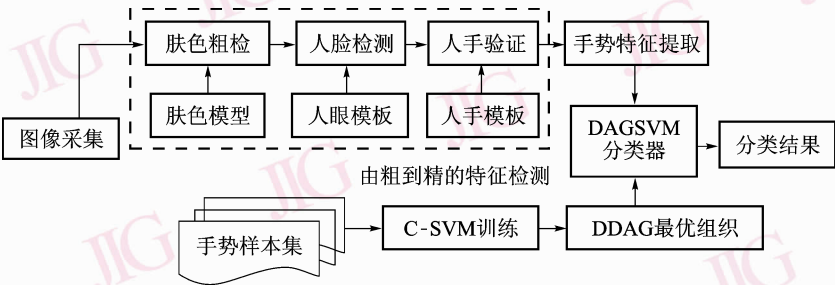


图 1 最优有向无环图支持向量机手势识别基本方法框图

Fig. 1 Framework of posture recognition with optimal DAGSVM classifier

(1)由粗到精的手势区域检测 在肤色粗检的基础上,分别引入人眼 Gabor 矩和人手边缘小波矩特征对肤色候选区域进行验证,精确检测出脸和手,并排除室内背景中常见的门窗家具等类似肤色物体的干扰。

(2)手势区域特征提取 提取脸手区域的不变矩特征,结合手的位置特征,构成混合特征向量。将测量空间的高维数据转换为特征空间表示的低维模式。

(3)手势特征分类 采用最优有向无环图支持向量机(DAGSVM)算法对手势特征进行分类。

DAGSVM 是 Platt 提出的一种解决硬间隔支持向量机(SVM)多类别分类问题的新方法^[7],基于图论中的决策有向无环图思想组织多个两类 SVM 分

类器。 k 分类 DAGSVM 只需进行 $k-1$ 次两类判别,同时 DDAG 拓扑结构具有冗余性,因而相对 OVO(一对一)、OVR(一对多)等其他 SVM 多分类法,具有分类速度快和分类精度高的特性;相对于最小距离、贝叶斯分类等其他方法,DAGSVM 又能继承支持向量机基于结构风险最小化原理、适用于小样本学习的特性。当两分类器为软间隔 C-SVM 时,DDAG 中的两分类器节点拓扑排序与 DAGSVM 最终分类结果相关,然而目前研究和应用中都是任意安排^[7-8],具有一定的盲目性。为此,针对手势特征分类问题提出了一种最优 DAGSVM 的设计策略。在训练两类分类 C-SVM 的基础上,针对各个 C-SVM 的分类能力的不同,优化 DDAG 的节点拓扑排序,从而构造最优 DAGSVM 多分类器。

3 特征检测与提取

3.1 肤色检测

肤色在 YC_bC_r 色彩空间与人眼对色彩的感知一致,并具有良好的聚类性,但是其色差分量仍非线性地依赖于亮度分量。通过对 YC_bC_r 空间进行分段非线性变换^[9],使肤色的聚类性独立于亮度,并在此空间中建立肤色标准混合高斯模型,从而获得对光照变化鲁棒性较好的肤色检测效果。

随机选取实验环境中各种光照下 24 670 个肤色样本点,在非线性变换后的 $C'_bC'_r$ 空间内分别建立人脸和人手区域肤色高斯模型 $M_f = (\Sigma_f, \mu_f)$ 和 $M_h = (\Sigma_h, \mu_h)$, μ_f 和 μ_h 分别为人脸和手的肤色均值, Σ_f 和 Σ_h 分别为人脸和手的肤色标准差。建立颜色向量 $X = [C'_b, C'_r]^T$ 的标准混合高斯肤色概率密度函数 $p(X) = w_f p_f(X) + w_h p_h(X)$, 其中,

$$p_f(X) = (2\pi)^{-1} |\Sigma_f|^{-1/2} e^{(-0.5(X-\mu_f)^T \Sigma_f^{-1} (X-\mu_f))}$$

$$p_h(X) = (2\pi)^{-1} |\Sigma_h|^{-1/2} e^{(-0.5(X-\mu_h)^T \Sigma_h^{-1} (X-\mu_h))} \quad (1)$$

w_f 和 w_h 分别是样本中人脸和人手样本肤色点所占比重。以 $p(X) \geq T_{\text{skin}}$ 作为肤色判决准则。肤色检测之后,采用形态学滤波对二值化图像进行处理,用区域搜索法找到 J 个候选区域。

3.2 人脸区域检测

采用文献[10]的人脸检测方法,首先根据人眼灰度特征,在肤色候选区域内进行人眼候选区域搜索;然后在人眼候选区域内,利用人眼模板 Gabor 不变矩匹配法计算该区域与人眼模板的相关性,找出相似人眼的矩形区域;再综合利用人眼与人脸的几何位置约束以及人脸候选区域的投影关系标记出人脸区域。

3.3 人手模板边缘小波矩验证

室内环境中的类似肤色物体如家具、门等,在颜色空间中较难与肤色区分。采用人手模板(如图 2 所示)边缘的 2D 小波矩匹配法验证处于交互姿态下的人手区域(不同尺度和旋转方向的正反面舒展手),以排除背景中相似肤色的干扰目标。2D 小波矩能同时提取全局特征和局部特征,并具有旋转不变性^[11],相比 Zernike 矩具有更好的图像局部特征描述能力和更少的信息冗余。同时,小波矩的旋转不变性能够适应手的位姿旋转变化。



图 2 人手模板与其边缘图

Fig. 2 Hand template and its edge

3.3.1 小波矩构造

$f(x, y)$ 为二值图像的分布函数,标准矩 M_{pq} 定义为 $M_{pq} = \iint x^p y^q f(x, y) dx dy$, 根据 $x = r \cos \theta, y = r \sin \theta$, 将 M_{pq} 变换到极坐标系下为

$$F_{pq} = \iint f(r, \theta) g_p(r) e^{iq\theta} r dr d\theta$$

式中, $g_p(r)$ 和 $e^{iq\theta}$ 分别是变换核的径向分量和角度分量。将 $\iint f(r, \theta) e^{iq\theta} d\theta$ 记为 $S_q(r)$, 则有 $F_{pq} = \iint S_q(r) g_p(r) r dr$ 。

径向分量 $g_p(r)$ 取为小波基函数 $\psi^{a,b}(r) = \frac{1}{\sqrt{a}} \psi\left(\frac{r-b}{a}\right)$, 即得到小波矩。采用具有最优空间频率局部特性的三次 B 样条函数作为母小波函数^[11]

$$\psi(r) = \frac{4\alpha^{n+1}}{\sqrt{2\pi(n+1)}} \sigma_w \cos(2\pi f_0(2r-1)) \times e^{-\frac{(2r-1)^2}{2\sigma_w^2(n+1)}} \quad (2)$$

式中, $n = 3, \alpha = 0.697\ 066, f_0 = 0.409\ 177, \sigma_w^2 = 0.561\ 145$ 。

小波基函数 $\psi^{a,b}(r)$ 的尺度参数 a 和平移参数 b 分别取离散值 $a = 0.5^m, b = 0.5n0.5^m, m = 0, 1, 2, \dots; n = 0, 1, \dots, 2^{m+1}$, 则小波基函数定义为 $\psi_{m,n}(r) = 2^{m/2} \psi(2^m r - 0.5n)$ 。数字图像的小波矩为

$$\|F_{m,n,q}^{\text{wavelet}}\| = \left\| \sum_{r=0}^1 \sum_{\theta=0}^{2\pi} f(r, \theta) e^{iq\theta} \psi_{m,n}(r) r \Delta r \Delta \theta \right\| \quad (3)$$

$F_{m,n,q}^{\text{wavelet}}$ 实际上是 $S_q(r)$ 的小波变换,在不同的尺度因子 m 、平移因子 n 和频域因子 q 下,能提供图像 $f(r, \theta)$ 在不同尺度和频域水平上的特征。

3.3.2 人手模板小波矩匹配

针对人手模板和室内常见具有类似肤色的干扰物,以类别散度为准则,采用顺序前进(SFS)法选取 N_m 个小波矩不变量。将待验证区域归一化至人手模板的尺寸,设 F_k 为待验证区域小波矩, F_0 为人手模板小波矩,采用如下相关系数作为判别函数,验证 $J-1$ 个区域:

$$R(k, o) = \frac{\sum_{i=1}^{N_m} F_k(i) F_o(i)}{\sqrt{\sum_{i=1}^{N_m} [F_k(i)]^2} \sqrt{\sum_{i=1}^{N_m} [F_o(i)]^2}} \quad (4)$$

$k = 1, 2, \dots, J$

若 $R(k, o) > T_m$ 则认为区域 k 是人手区域。

人脸和人手区域检测结果如图3所示。图3(b)中白点代表肤色检测结果,矩形框为算法自动标记的人脸和人手区域。

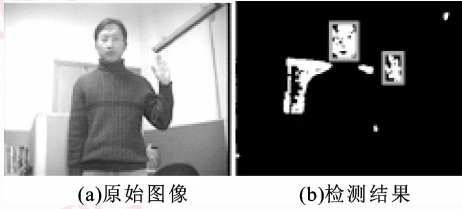


图3 肤色检测、人脸与人手区域检测结果

Fig. 3 Skin color detection and face/hand location

3.4 特征向量提取

对检测并标记出的人脸和手区域内的像素,提取基于全局图像统计特征的7阶不变Hu矩 $\{\phi_1, \dots, \phi_7\}$ 。由于Hu矩数值分布范围较大,采用如下调整方法 $\phi_i \leftarrow \log |\phi_i| (i \in [3, 7])$ 。再结合图像坐标系中手区域中心点相对于脸区域中心点的2维坐标,构成11维混合特征向量:

$$\Phi = [\phi_1, \dots, \phi_7, x_{lh}, y_{lh}, x_{rh}, y_{rh}]^T \quad (5)$$

由于Hu矩的平移、旋转和缩放不变性,该特征向量对人和机器人之间的距离变化、人在机器人视野中的平面位置变化都不敏感。同时由于在脸手检测的基础上提取区域统计特征和位置信息,大大提高了特征提取的可靠性。

4 最优 DAGSVM 分类器

4.1 DAGSVM 分类器

k 分类 DAGSVM 利用决策有向无环图 $G = (V, A)$ 组织 $k(k-1)/2$ 个两类 SVM 分类器, V, A 分别代表节点和边。如图4所示,每个两类分类器 C_{ij} 通过在 i, j 两类样本的核生成空间上最大化分类间隔,找到针对 i, j 两类的最优判决面。

对于硬间隔 SVM, 由于两类数据都远离判决面, 所以在 $1-v-1$ 判决节点 C_{ij} 做出非 i 或非 j 判决是安全的, Platt 的实验简单证明了此时 DDAG 中的

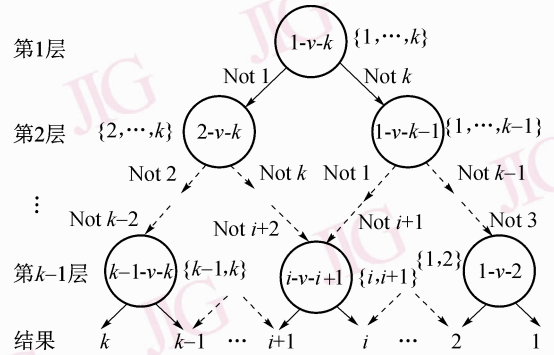


图4 决策有向无环图(DDAG)拓扑结构

Fig. 4 Topological structure of DDAG

节点顺序可以任意安排,分类结果没有显著变化^[7]。但是实际应用中存在类别之间的数据混叠和噪声,基于此,Vapnik 提出了软间隔 C-SVM。

4.2 DAGSVM 节点的优化拓扑排序策略

在 C_{ij} 为 C-SVM 分类器的情况下,各两分类器的泛化能力存在差异,此时 DDAG 中的节点拓扑排序影响 k 分类的性能。DAGSVM 的优化目标为 k 类样本 $x_i (i = 1, \dots, k)$ 的分类正确概率期望和最大:

$$\begin{aligned} \max \quad & \sum_{i=1}^k E(p(x_i | G)) \\ \text{s. t. } \quad & G = (V, A), |V| = |A| = k, V \in R \\ & R = \{C_{ij}, i, j = 1, \dots, k; i \neq j\} \end{aligned} \quad (6)$$

设分类器 C_{ij} 上本类元 i, j 的正确分类概率为 p_{ij} , 其期望值可由 $1 - T_{ij}$ 计算, T_{ij} 为 C_{ij} 的 Radium-margin Bound^[12] 泛化误差估计。

由于在分类器 C_{ij} 上本类元 i, j 的分类错误无法在低层得到纠正,同时 i, j 错误地进入下一层分类器的对象列表 **list**, 将增大该分类器分类错误的概率。因此应将泛化误差最小的分类器安排在 DDAG 的高层, 即第 a 层第 b 个节点 N_{ab} 应选取分类器 C_{st} , $(s, t) = \arg \max_{C_{ij} \in \text{list}_{ab}} \{p_{ij}\}$ 。以此策略组织所有节点得到的 DAGSVM 多分类器 G_{opt} 具有式(6)意义下的最优性。

考虑以 N_{ab} 为顶点的子图 G_n , 其分类对象列表 **list** _{ab} 中有 $n (\geq 3)$ 个待分类。不妨假设 $p_{st} = p_1$, 且 $\forall (i, j) \neq (s, t) p_{ij} = p_2, p_1 > p_2$ 。由于各样本独立分布, 非主元类被分到下一级两个节点的概率各为 0.5。某类 i 的分类正确率就是沿着其分类路径各条边的分类正确率之积。按照该计算方法, 可由归纳法求出当 N_{ab} 选取为 C_{st} 时, n 类样本分类正确概率期望和有递推关系

故

$$f(\mathbf{G}_n) = (f(\mathbf{G}_{n-1}) + 1) \cdot p_2$$

$$\begin{aligned} f(\mathbf{G}_n) &= \sum_{i=1}^n E(p(x_i | \mathbf{G}_n)) \\ &= 2p_1 p_2^{n-2} + \sum_{i=1}^{n-2} p_2^i \end{aligned} \quad (7)$$

式中, $p_1 p_2^{n-2}$ 为本类元 i, j 的分类正确概率期望, 第 2 项为其余类元的分类正确概率期望和。同理可以计算出当 N_{ab} 选为非 C_{st} 时,

$$\begin{aligned} f_1(\mathbf{G}_n) &= \sum_{i=1}^n E(p(x_i | \mathbf{G}_n)) = p_2^{n-1} + p_1 p_2^{n-2} + \\ &0.5 p_2^{n-2} + 0.5 p_1 p_2^{n-3} + \sum_{i=1}^{n-3} p_2^i \end{aligned} \quad (8)$$

两式差值为

$$\begin{aligned} &p_1 p_2^{n-2} + 0.5 p_2^{n-2} - 0.5 p_1 p_2^{n-3} - p_2^{n-1} \\ &= (p_1 - p_2)(p_2 - 0.5) p_2^{n-3} \end{aligned}$$

由于 $p_1 > p_2$, 且文中二分类器 C_{ij} 均能保证 $p_{ij} \in (0.5, 1)$, 故有式(8) > 式(9), 证实了优化拓扑排序的 DAGSVM 具有最高的多分类正确率。该策略实际上就是一种自顶向下的贪婪搜索, 每个节点都选择当前分类能力最佳的二分类器, 从而构成 k 分类的最优 DAGSVM, 步骤如下:

(1) 对第 a 层第 b 个分类节点 N_{ab} , 创建一个含有 $n(n = k - a + 1)$ 个待分类类别的有序列表 $\text{list}_{ab} = \{x_1, \dots, x_n\}$ 。

(2) 将列表元素中的 x_i 和 x_j 放至列表首尾, 其中 $(i, j) = \underset{C_{ij} \in \text{list}_{ab}}{\operatorname{argmax}} \{p_{ij}\}$ 。

(3) 取列表首尾元素构成该节点上的分类器, 即 $N_{ab} \leftarrow C_{x_i x_j}$ 。

(4) $a \leftarrow a + 1, n \leftarrow n - 1$ 。若分类器 $C_{x_i x_j}$ 做出排除 x_i 判决, 则 b 不变, 否则 $b \leftarrow b + 1$ 。排除后的其余元素生成下一层节点 N_{ab} 的列表。

(5) 返回第 2 步循环执行, 直到列表中只剩两个待分类类别。

5 实验分析

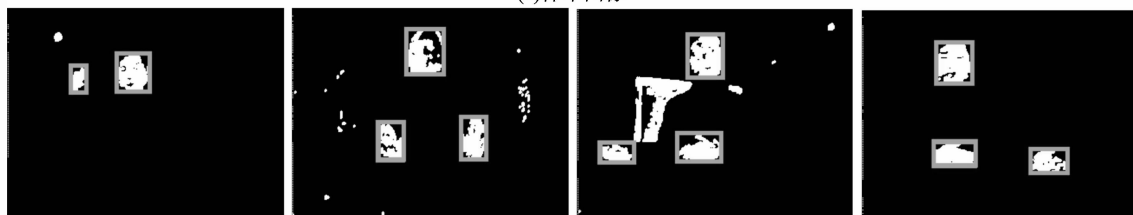
5.1 手势识别实验

在正常光照的普通实验室内, 对服务机器人在运动过程中采集到的人体手势图像进行识别。机器人为美国 ActivMedia 公司的 PeopleBot 服务机器人, 其底座为轮式差分驱动, 机载 Canon VC-C4 PTZ 彩色摄像系统, 内部装有 PXC200 图像采集卡, 采集图像分辨率为 320×240 。保持摄像头仰角 10° , 0 级光学变焦。手势发出者与机器人正面相对, 距离 $1 \text{ m} \sim 2 \text{ m}$ 。

设计识别 4 种手势的测试方案。4 个手势者每人每种手势各采集 50 个样本, 共 800 幅图像。前后 400 张分别用于训练和测试。样本图像包含室内复杂背景, 不同的人-机距离以及人体在图像中处于不同的位置。对训练图片分别采用第 3 节介绍的方法, 首先粗检肤色, 然后利用人眼 Gabor 矩特征匹配法检测人脸区域, 再对剩余的肤色候选区域提取边缘, 利用小波矩验证其是否为人手区域。人脸和人手区域检测的结果如图 5 所示。其中 Gabor 滤波器的参数采用 5 个尺度 8 个方向, 小波矩采用 $F_{1,0,0}, F_{1,0,1}, F_{1,1,1}, F_{2,1,1}$ 。



(a) 样本图像



(b) 对应手脸检测结果

图 5 指令性手势

Fig. 5 Instructional postures

按式(5)提取特征向量,训练 6 个两类 SVM 分类器 $\{C_{12}, C_{13}, C_{14}, C_{23}, C_{24}, C_{34}\}$,采用径向基 C-SVM 软间隔分类器,其高斯核参数选取为 $\gamma = 0.25, C = 10$ 。在训练两类 SVM 的基础上,使用第 4 节提出的方法构造最优 DAGSVM,如图 6 所示。

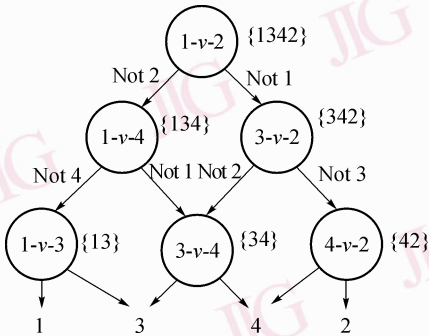


图 6 本文最优 DAGSVM 拓扑结构图
Fig. 6 Optimal DAGSVM structure

利用该最优 DAGSVM 对 4 种手势进行分类,在训练样本集上 4 种的分正确数目分别为 99,94,90,91,总正确率为 $(99 + 94 + 90 + 91)/400 = 93.5\%$,在测试样本集上 4 类测试样本的分类正确数目如表 1 所示,总正确率为 $(97 + 92 + 85 + 87)/400 = 90.25\%$ 。

表 1 测试集上 4 种手势识别结果
Tab. 1 Recognition result on test set

类别	识别结果(数目)			
	w_1	w_2	w_3	w_4
w_1	97	—	2	1
w_2	—	92	5	3
w_3	—	7	85	8
w_4	—	6	7	87

分别选取 OVO-SVM、OVR-SVM、普通 DAG-SVM (其余 23 种拓扑排序结构随机选取,结果取平均)3 种算法在测试图片集上与之比较。另外,在 People-Bot 机器人机载的 PentiumIII 850 MHz 计算机上统计各算法分类一个手势(不包括特征检测与提取)的平均耗时。结果如表 2 所示。

实验结果表明 DAGSVM 比 OVO-SVM 少进行一半的两分类,比 OVR-SVM 的各个两分类 SVM 具有更少支持向量数目,因而运算更快。DAGSVM 具有冗余纠错结构,比推广误差无界的 OVO-SVM 和 OVR-SVM 具有更高的准确率。同时,由于分类节

点拓扑排序的优化,本文最优 DAGSVM 还取得了比普通 DAGSVM 更高的多分类准确率。另外,采集未训练手势者的 4 种手势图像共 100 幅来测试本文的分类器,其手势分类正确率也能达到 86%,证实了本文方法具有良好的实用性与可靠性。

表 2 不同的多类 SVM 分类器识别结果比较
Tab. 2 Comparison of different multi-class SVM classifiers

多类分类器	4 类别总分类正确率(%)	算法耗时(ms)
OVO-SVM	$327/400 = 81.75$	49.6
OVR-SVM	$342/400 = 85.5$	42.3
普通 DAGSVM	$355/400 = 88.75$	24.2
最优 DAGSVM	$361/400 = 90.25$	23.5

5.2 在线交互实验

手势者使用手势指引服务机器人在室内导航,机器人在线识别手势并做出相应运动。图 5 所示的 4 种手势 w_1, w_2, w_3, w_4 分别代表控制机器人启动、停止、左转和右转(相对于机器人自身位姿)。一般每次手势保持时间大于 1 s,改变手势时的切换时间不超过 2 s。服务机器人限速 100 mm/s,转向时角度 θ 的改变步长为 1° 。一次连续地用手势控制机器人启动-右转-左转-停止的结果如图 7 所示。

算法中最优 DAGSVM 分类器的训练和拓扑排序优化均为离线完成,不影响实时性能,机器人能够根据视觉系统识别的结果快速稳健地响应用户的手势指令。该手势导航实验验证了本文方法较好地适用于在线人机交互,各种指令手势直观易学,且服务对象在交互过程中感受自然良好。

6 结 论

针对运动中的室内服务机器人及其视觉系统,提出了一种识别人体指令性手势的方法。结合人眼和人手模板特征匹配法验证人脸人手区域能够较好地克服仅用肤色检测提取手势特征的不足;提取脸手区域统计不变矩和手的位置信息构成模式特征向量;以各类样本分类正确概率期望和最大为目标函数,优化决策有向无环图中的两分类器节点拓扑排序,构成最优 DAGSVM 多分类器,提高了多分类的整体准确率。实验验证了该方法能够利用人脸部和手部自然特征,有效可靠地识别人体指令性手势,进而实现一种用户友好的人机交互。

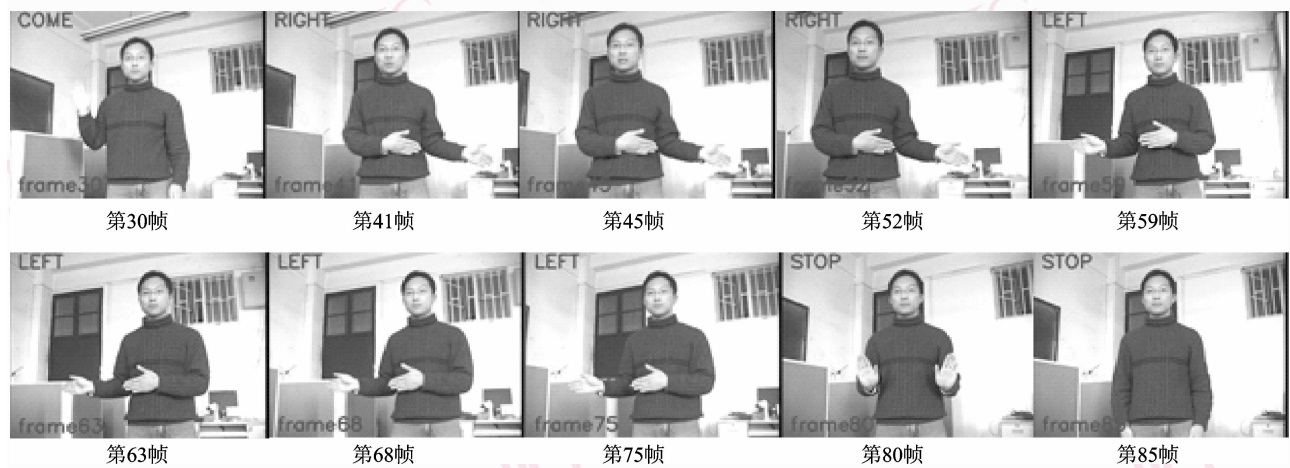


图 7 在线交互实验的结果序列图像

Fig. 7 Sequential images of on-line interaction result

该方法适用于已知基本交互意图下的静态手势实时识别。对于意图未知的情况,将进一步引入动作和场景上下文的研究,以实现对交互意图的理解和识别。

参考文献 (References)

1 Kortenkamp D, Huber E, Bonasso R P. Recognizing and interpreting gestures on a mobile robot[A]. In: Proceedings of the 13th National Conference on Artificial Intelligence[C], Portland, USA, 1996:915-921.

2 Waldherr S, Romero R, Thrun S. A gesture based interface for human-robot interaction[J]. Autonomous Robots, 2000, 9(2): 151-173.

3 Stiefelhagen R, Fugen C, Gieselmann P, et al. Natural human-robot interaction using speech, head pose and gestures[A]. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems[C], Sendai, Japan, 2004:2422-2427.

4 Richarz J, Martin C, Scheidig A, et al. Estimating pointing gestures in monocular images for mobile robot instruction[A]. In: Proceedings of IEEE International Symposium on Robot and Human Interactive Communication[C], Hatfield, UK, 2006:546-550.

5 Boehme H J, Braumann U D, Corradini A, et al. Person localization and posture recognition for human-robot interaction[A]. In: Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction [C], London, Springer-Verlag, 1999: 117-128.

6 Kuno Y, Murashima T, Shimadat N, et al. Understanding and learning of gestures through human-robot interaction[A]. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems[C], Takamatsu, Japan, 2000:2133-2138.

7 Platt J C, Cristianini N. Large margin DAGs for multiclass classification [J]. Advances in Neural Information Processing Systems, 2000, 12(3): 547-553.

8 Hsu C W, Lin C J. A comparison of methods for multi-class Support Vector Machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2):415-425.

9 Hsu R L, Mottaleb M A, Jain A K. Face detection in color images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5):696-706.

10 Hu C H, Ma X D, Dai X Z, et al. A method for fast face-detection based on normalization Gaussian mixture model [J]. Journal of Southeast University(Natural Science Edition), 2007, 37(3): 389-394. [胡春华,马旭东,戴先中等. 一种基于标准混合高斯模型的快速人脸检测方法[J]. 东南大学学报(自然科学版), 2007, 37(3): 389-394.]

11 Shen D G, Horace H S. Discriminative wavelet shape descriptors for recognition of 2-D patterns[J]. Pattern Recognition, 1999, 32(2): 151-165.

12 Vapnik V, Chapelle O. Bounds on error expectation for support vector machines[J]. Neural Computing, 2000, 12(9):2013-2036.