

基于多特征融合的运动人体行为识别^{*}

赵海勇^a, 李成友^b

(聊城大学 a. 计算机学院; b. 网络信息中心, 山东 聊城 252000)

摘要: 提出了一种基于特征级融合的运动人体行为识别方法。应用背景差分法和阴影消除技术获得运动人体区域和人体轮廓;采用 $\mathcal{R}$ 变换提取人体区域特征,采用小波描述子提取人体轮廓特征;然后将这两种具有一定互补性的特征采用K-L变换进行融合,得到一个分类能力更强的特征;最后,在传统支持向量机的基础上,结合模糊聚类技术和决策树构建多级二叉树分类器,从而实现行为多类分类。该方法在Weizmann行为数据库上进行了实验,实验结果表明所提出的识别方法具有较高的识别性能。

关键词: 行为识别; 特征融合; 小波描述子; 支持向量机

中图分类号: TP319 **文献标志码:** A **文章编号:** 1001-3695(2012)08-3169-04

doi:10.3969/j.issn.1001-3695.2012.08.098

Human action recognition based on multi-features fusion

ZHAO Hai-yong^a, LI Cheng-you^b

(a. School of Computer Science & Technology, b. Network Information Center, Liaocheng University, Liaocheng Shandong 252000, China)

Abstract: This paper presented a recognition algorithm of human action based on multi-features fusion. First, through the background subtraction and shadow elimination, obtained human motion silhouettes and contours. It described human silhouette features by $\mathcal{R}$ transform. Then fused these features with human contour features which were extracted using wavelet descriptor by K-L transform to get a new feature which had higher discriminating power. At last, through integrating fuzzy clustering and decision tree, constructed a multilevel binary tree classifier. It evaluated the method on the Weizmann action dataset. The experimental results show that the approach has efficient recognition performance.

Key words: human action recognition; multi-features fusion; wavelet descriptor; support vector machine

0 引言

近年来,人体行为识别已经成为计算机视觉中的热点研究领域^[1]。作为人体运动分析的高级阶段,它在智能视频监控、视频注解、人机交互和虚拟现实等领域具有广泛的应用价值。然而,人体行为复杂且多样,对分类的定义模糊,不同的观察者可能将同一段行为运动识别为不同的结果:一方面,相似的行为在时间尺度上往往不同,并且同一种行为在不同的环境中通常具有不同的含义^[2];另一方面,同一段行为可能同时被判别为属于几个类别,因为行为是一些基本运动单元的自然组合。另外,数据采集的真实环境复杂多样,行为序列分割技术还不十分完善,以上种种因素使得人体行为识别成为一个非常具有挑战性的研究领域。

行为识别是对人的行为模式进行分析与识别^[3]。通常,它可以分为两个层次的任务^[4]:a)底层的特征提取和表示;b)高层的行为识别和建模。从图像序列中提取出能够合理表示人体运动的特征,对行为识别和理解至关重要。基于视觉的动作表示按特征的性质大致可以分为两类^[5]:a)二维底层图像信息特征;b)三维模型特征。可利用的二维底层图像信息包括:前景目标、前景目标的运动速度、运动轨迹信息^[6]、光流信息、前景目标的轮廓和侧影等。不同的行为特征从不同的角度反映人的运动状态特性,选择一种特征时不仅要

考虑可区分性,而且要考虑提取时的难易程度。两类基于形状的特征——人体侧影和轮廓能够鲁棒地获得,因此经常被用于行为特征提取。基于人体侧影的方法应用范围广泛,但是由于将整个身体形状所包含的像素考虑在内,计算复杂度较高。基于人体轮廓的方法只考虑人体形状边界所包含的像素,因而计算复杂度较低,但是由于需要准确提取人体轮廓,这就限制了其应用范围。

人体行为识别的识别方法中广泛采用了基于图模型的推理方法,这种方法大体上可以分为基于产生式模型(generative models, GM)的方法和基于判别式模型(discriminative models, DM)的方法。隐马尔可夫模型(hidden Markov models, HMM)是常用的产生式模型,它是一种有效的时变信号处理方法,隐含了对时间的校正,并提供了学习机制和识别能力。但产生式模型为了降低计算复杂度,往往假设观察序列是基于条件独立的,这种不确切的假设导致其不能表示时间序列中的上下文依赖关系以及序列间的特征多重重叠。为了改进产生式模型给时间序列建模所带来的问题,人们开始研究基于判别模型的方法。文献[7]将条件随机场模型(CRF)用于行为识别,CRF无须对观测进行建模,因此避免了独立性假设,并且可以满足状态与观测之间在长时间尺度上的交互。支持向量机^[8]作为一种判别式分类方法,在解决小样本、非线性及高维模式识别问题时具有特有的优势。

本文提出了一种基于多特征融合的人体行为识别算法。

收稿日期: 2011-12-14; **修回日期:** 2012-01-16 **基金项目:** 国家自然科学基金资助项目(61104179)

作者简介: 赵海勇(1981-),男,山东聊城人,博士,主要研究方向为计算机视觉、模式识别(zhaohaiyong@lcu.edu.cn);李成友(1968-),男,高级工程师,硕士,主要研究方向为人工智能、网络管理等。

考虑到单独采用基于人体侧影的特征或者单独采用基于人体轮廓的特征表征人的行为状态都存在不足,因此本文将两种特征融合以更好地表征人的行为状态。特征级融合既能保留多特征的有效鉴别信息,又能在很大程度上消除特征的冗余。传统的特征融合方法将两组特征直接合并为新的特征向量,由于特征合并后的维数是两种原始特征的维数之和,这就导致了合并后新特征的维数急剧增加,从而使得组合后识别的速度大幅度地降低。本文采用离散 K-L 变换,对提取的多种特征进行信息压缩,实现特征的有效融合,并采用支持向量机方法实现行为识别。为了将支持向量机推广到多类分类,本文结合模糊聚类技术和决策树,构造多级二叉树分类器。首先利用模糊 C-均值聚类技术求取每类行为样本模糊聚类中心,由聚类中心确定二叉树型,然后在二叉树各节点处根据聚类中心重新构造学习样本集和 SVM 子分类器,从而构造多级二叉树分类器。

1 行为特征提取

运动人体检测是行为识别的第一步,也是最重要的一步,因为识别结果很大程度取决于检测的质量,而遮挡、光照变化以及噪声等因素又严重影响检测和分割的质量。数据采集环境为:a)摄像机静止不动;b)摄像机方向与人体运动方向近似垂直;c)环境中只有一个人运动。通过背景差分法、二值化以及阴影消除得到运动人体区域。在人被检测出来以后,提取每帧图像中人体姿态的二值侧影图像和人体轮廓。具体步骤如下:

a)在检测之前,要对图像进行去噪处理。视频序列在获取和传输时常会受到各种各样的噪声干扰,如高斯白噪声、脉冲噪声和乘性噪声等,这些噪声会给后续的处理带来很大的不便。本文采用高斯滤波和邻域去噪相结合的方法,对图像进行去噪处理。

b)建立一个无参数背景模型^[9],模型中每一个像素点的概率密度函数可由样本数据通过核密度估计方法计算得到。假设 $I = \{x_1, x_2, \dots, x_N\}$ 表示一个像素点的 N 个样本,其中 x_t 表示 t 时刻的观测值。该像素点的概率密度函数可用核密度函数估计:

$$p(x_t) = \sum_{i=1}^N \alpha_i K_h(x_t - x_i) \quad (1)$$

其中: K_h 是窗口长度为 h 的核函数; α_i 是标准化系数,通常 $\alpha_i = 1/N$ 。

在文献[9]中,核函数 K_h 满足正态分布 $N(0, \Sigma)$,其中 Σ 表示核带宽。假设三种颜色分量的核带宽相互独立, σ_j^2 表示第 j 个颜色分量的带宽,那么 $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, $\text{diag}(\cdot)$ 为对角矩阵。像素点 x_t 的概率密度函数可以写为

$$p(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_{tj} - x_{ij})^2}{2\sigma_j^2}} \quad (2)$$

其中: d 表示颜色分量的个数。核带宽 σ 可以通过样本中位数绝对离差得到。如果 $p(x_t) < th$,则该像素点被看做运动目标像素点, th 为全局阈值。

c)运动目标检测的目的是从背景图像中分割提取变化区域,因此在检测运动目标的同时,运动目标投射的阴影也会被检测为运动前景的一部分。本文采用文献[10]提出的方法对前景与阴影进行分割,该方法适用于室内、户外等多种环境。

d)对提取的运动区域进行形态学处理及连通性分析,最后采用基于连通性的边界跟踪算法获取整个图像序列中行人

的侧影。图1给出了其中一帧的人体侧影和轮廓提取结果。



图1 人体侧影及轮廓提取

1.1 基于运动区域的特征提取

本文使用 $\mathcal{R}$ 变换形状描述子^[11]提取行为区域特征, $\mathcal{R}$ 变换是对 Radon 变换的一种改进,通过改进使其具有了几何不变性(平移不变性、旋转不变性和尺度不变性)。Radon 变换是计算图像在某一特定角度射线方向上投影的变换方法,对一幅离散二值图像 $f(x, y)$, Radon 变换定义为

$$T_R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy = R\{f(x, y)\} \quad (3)$$

其中: $\theta \in [0, \pi]$; $\rho \in [-\infty, \infty]$; $\delta(\cdot)$ 为 dirac-delta 函数,

$$\delta(x) = \begin{cases} 1 & \text{如果 } x=0 \\ 0 & \text{其他} \end{cases} \quad (4)$$

通过对式(3)计算分析可知, Radon 变换不满足几何不变性。为了解决上述问题,文献[11]提出了 $\mathcal{R}$ 变换,其定义为

$$\mathcal{R}_f(\theta) = \int_{-\infty}^{\infty} T_R^2(\rho, \theta) d\rho \quad (5)$$

$\mathcal{R}$ 变换本身所具有的一些属性使其适合于行为特征表示。

属性1 当图像 $f(x, y)$ 平移向量 $\mu = (x_0, y_0)$ 时:

$$\int_{-\infty}^{\infty} T_R^2((\rho - x_0 \cos \theta - y_0 \sin \theta), \theta) d\rho = \int_{-\infty}^{\infty} T_R^2(\nu, \theta) d\nu = \mathcal{R}_f(\theta) \quad (6)$$

由式(6)可以看出,平移图像并不会改变 $\mathcal{R}$ 变换的结果。

属性2 改变图像 $f(x, y)$ 的尺度,因子为 α :

$$\frac{1}{\alpha^2} \int_{-\infty}^{\infty} T_R^2(\alpha \rho, \theta) d\rho = \frac{1}{\alpha^3} \int_{-\infty}^{\infty} T_R^2(\nu, \theta) d\nu = \frac{1}{\alpha^3} \mathcal{R}_f(\theta) \quad (7)$$

改变图像的尺度,会引起 $\mathcal{R}$ 变换结果振幅的改变,因此,需要对图像进行标准化。

属性3 当图像旋转角度 π 时:

$$\int_{-\infty}^{\infty} T_R^2(-\rho, \theta \pm \pi) d\rho = - \int_{-\infty}^{\infty} T_R^2(\nu, \theta \pm \pi) d\nu = \int_{-\infty}^{\infty} T_R^2(\nu, \theta \pm \pi) d\nu = \mathcal{R}_f(\theta \pm \pi) \quad (8)$$

由式(8)可以看出, $\mathcal{R}$ 变换的周期为 π , 因此,一个 180 度的向量就能充分表示人体侧影。图2中第2列给出了六种行为姿态的 $\mathcal{R}$ 变换结果。由于环境的多样性以及不同人执行动作所造成的差异性, $\mathcal{R}$ 变换结果可能不会完全相同,但是曲线的形状大致相同(如图2第3列)。为了使 $\mathcal{R}$ 变换更具有代表性,对多次 $\mathcal{R}$ 变换结果取平均值,图2给出了六种行为姿态的 $\mathcal{R}$ 变换结果。

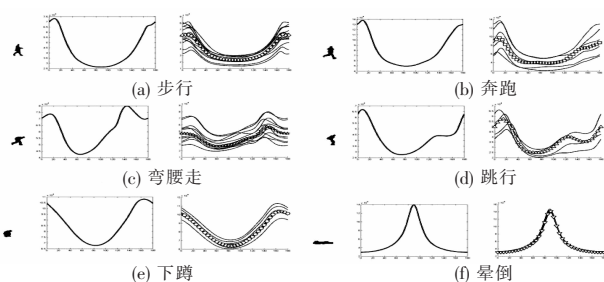


图2 六种行为姿态的

1.2 基于人体轮廓的特征提取

人体轮廓形状是人的重要特征之一,它随时间呈周期性变化,在人体运动状态描述方面,具有一定的代表性。小波描述子可以很好地对二维形状轮廓进行描述,最大的优点是对形

状的旋转、尺度和平移具有很高的鲁棒性,本文采用离散小波变换对人体轮廓进行描述。为了更加显著地表示人体特征、降低计算复杂度,将人体二维轮廓特征展开转换为 l 维距离特征^[12]。

1) 质心计算 提取人体轮廓后,可以使用边界提取算法(如 Canny 算法)提取轮廓线上的坐标点,利用式(9)计算轮廓的质心坐标:

$$x_c = \frac{1}{N} \sum_{i=1}^N x_i, \quad y_c = \frac{1}{N} \sum_{i=1}^N y_i \quad (9)$$

其中, (x_c, y_c) 是质心坐标; N 是轮廓线上像素点个数; (x_i, y_i) 是轮廓线上点的坐标。

2) 轮廓线展开 选定轮廓线上纵坐标最大的点为起点(左下方为图像坐标原点),沿顺时针方向,利用式(10)将轮廓线展开为质心到轮廓点欧式距离的一维特征:

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad (10)$$

3) 数据标准化 为了消除图像尺度和特征长度不同造成的影响,在最大限度地保持特征完整性的前提下,需要将特征的幅度和长度值标准化,以达到利于分析数据规律和特征的目的。由式(10)可将轮廓表示为一个由 N 个元素组成的向量 $D = [d_1, d_2, \dots, d_N]$, 统计视频序列中人体轮廓线的像素点个数,95% 以上的视频序列满足 $N \in [400, 600]$, 且呈正态分布。通过边界像素的内插处理将 N 标准化: $N = 500$ 。设 d' 为标准化的距离范围,取 $d' = 100$ 。轮廓特征降维并标准化的结果如图3所示。

离散小波函数族定义如下:

$$\varphi_{j,n}(t) = 2^{-\frac{j}{2}} \varphi(2^{-\frac{j}{2}} t - n) \quad j, n \in \mathbb{Z} \quad (11)$$

如果上述小波函数族构成 $L^2(\mathbb{R})$ 的正交规范基,那么对于任意 $f(t) \in L^2(\mathbb{R})$ 都有离散小波变换:

$$C_{j,n} = [f(t), \varphi_{j,n}(t)] = \int_{-\infty}^{+\infty} f(t) \varphi_{j,n}^*(t) dt \quad (12)$$

其中: $\varphi_{j,n}^*(t)$ 为 $\varphi_{j,n}(t)$ 的共轭; $C_{j,n}$ 称为 $f(t)$ 的离散小波系数。

选择 Coiflet 小波基并用 h 表示,根据式(13)对 D 进行两层小波变换,得到轮廓的小波描述子 W :

$$W = \langle \langle D, h \rangle, h \rangle \quad (13)$$

选择 16 个低频段小波描述子系数作为人体轮廓特征。

1.3 基于 K-L 变换的特征融合

在轮廓特征和区域特征融合之前,需要对两种特征进行归一化的处理。采用 Z-score 归一化方法^[13]: 假设 X_s 表示区域特征, X_c 表示轮廓特征,经过 Z-score 归一化后的特征向量为

$$\begin{cases} X'_s = (X_s - \mu_s) / \sigma_s \\ X'_c = (X_c - \mu_c) / \sigma_c \end{cases} \quad (14)$$

其中: μ_s 和 μ_c 分别表示区域特征向量均值和轮廓特征向量均值; σ_s 和 σ_c 分别表示区域特征向量和轮廓特征向量标准差。

采用离散 K-L 变换对归一化后的特征数据进行压缩融合。设 Y 为 N 维随机向量,则 Y 的 K-L 展开可表示为

$$Y' = TY \quad (15)$$

其中, T 为正交变换矩阵,其元素由 Y 的协方差矩阵的特征向量组成,即 $T = (\varphi_1, \varphi_2, \dots, \varphi_N)$ 。 Y 的协方差矩阵记作:

$$\Sigma_y = E\{(Y - \bar{Y})(Y - \bar{Y})^T\} \quad (16)$$

其中, $\bar{Y} = E(Y) = (\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_N)$ 代表均值向量,存在

$$\Sigma_y \varphi_i = \lambda_i \varphi_i \quad (17)$$

其中: $\lambda_1 > \lambda_2 > \dots > \lambda_N \geq 0$ 是 Σ_y 的特征值; φ_i 为对应的特征向量。

对提取的特征向量进行 K-L 变换后,得到新的特征数据,其中每个特征点是原来 N 个特征的线性组合,选择前 K 维数

据作为最后的融合特征。

2 行为识别

支持向量机^[14] (SVM) 是解决小样本模式识别问题的有效方法,其基本原理是通过用核函数将输入样本空间非线性变换到一个高维空间,然后在这个高维空间求广义最优分类超平面。SVM 是针对两类问题提出来的,为实现对多个类别的识别,需要对 SVM 进行相应的改进。目前,已经提出多种方法将 SVM 推广到多类分类问题,如用多个两类分类器来实现多类分类,或者用层次型两类分类器实现多类分类等。本文结合模糊聚类技术和决策树构造多级二叉树分类器。决策树从一组无次序、无规则的事例中推理出决策树表示形式的分类规则,用树形结构表示决策集合,是一种高效的分类器,容易产生关联规则。每个内部节点表示在一个属性上的测试,每个分枝代表一个测试输出,而每个树叶节点代表类或类分布。

为确定二叉树结构,首先利用模糊聚类技术求取每类样本的聚类中心 $C = \{C_1, \dots, C_K\}$ 作为根节点;再把它聚类成两类,将各聚类中心对应的样本数据分别标记为正类 P_{level1} (即左子树) 和负类 N_{level1} (右子树),构造第一级 SVM 子分类器;然后,将 P_{level1} 对应的聚类中心又聚类成两类,并将各聚类中心对应的样本数据也分别标记为正类 P_{level2} (即左子树) 和负类 N_{level2} (右子树),构造第二级 SVM 子分类器,对负类 N_{level1} 做同样操作;用同样的方法构造第三级子分类器,依次下去,直到每个子类只包含一个聚类中心点。这样多级二叉树构造成功, K 类行为需要构造 $K-1$ 个 SVM 子分类器。

3 实验与分析

采用 Weizmann 数据库对算法效果进行验证。该数据库共包含 90 个低分辨率的视频序列 (180 × 144 像素, 50 fps 所有动作分别由 9 名实验人员完成,每人完成 10 个自然动作,动作包括 run、walk、skip、jumping-jack、jump-forward-on-two-legs (or jump)、jump-in-place-on-two-legs、gallop sideways、wave-two-hands、wave one-hand、bend 等。由于受阴影、颜色相似性等因素的影响,实际提取的人体剪影存在孔洞、毛刺等噪声。本文对其中的 run、walk、skip、bend、jump 及 sideways 进行实验。将 9 人完成的动作序列分成 3 组,每组序列由 3 人完成,其中一组序列用于训练构造支持向量机多级二叉树分类器,其余两组用于测试。

3.1 基于 K-L 变换的特征融合

利用侧影和轮廓提取方法,提取行为序列的所有人体侧影和轮廓,图4给出了示例图像及其相应的人体侧影和轮廓。对每种行为序列,分别提取每帧图像的区域特征和轮廓特征。

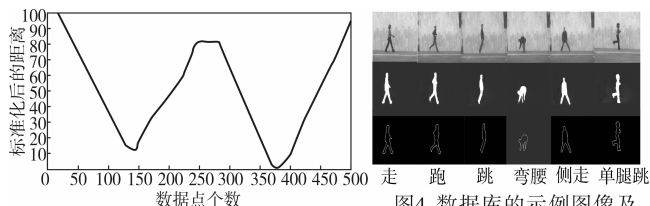


图3 距离特征归一化

图4 数据库的示例图像及对应的人体剪影和轮廓

3.2 支持向量机多类分类器构建

为了识别六种行为,采用模糊 C-均值聚类技术得到 5 个聚类中心 $\{C_1, C_2, C_3, C_4, C_5\}$, 构建支持向量机多级二叉树分

类器如图5所示,每个聚类中心对应一种行为。

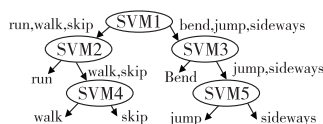


图5 支持向量机多类分类器

采用径向基核函数作为支持向量机的核函数。径向基核函数定义如下:

$$K(x_i, x_j) = \exp \left\{ -\frac{\|x_i - x_j\|^2}{\sigma^2} \right\} \quad (18)$$

为了验证特征融合后对识别率的提升,表1给出了三种特征下等错误率(REE)。惩罚参数为 $C = 10$, $\varepsilon = 0.001$,核函数参数 $\sigma = 0.2$ 。通过比较可知,本文的融合方法比单一特征方法在识别性能上有了提高,等错误率降低到0.32%。

表1 等错误率比较

	区域特征	轮廓特征	融合特征
REE/%	2.4	1.4	0.32

图6给出了六种行为的识别结果,总体来说本方法对于简单的日常行为具有较高的识别率,表中对弯腰的识别率最高,而对走、跑和单腿跳识别率较低。这主要是因为本文所用识别方法是以动作序列的人体形状特征为基础的,如果两种行为中人体姿势形状相似,就容易引起错误识别。

walk	0.86	0.12				0.04
run	0.10	0.87				0.03
jump			0.94			0.06
bend				1.00		
sideways	0.07	0.02			0.91	
skip	0.05	0.07				0.88
	walk	run	jump	bend	sideways	skip

图6 六种行为的识别结果混合矩阵

4 结束语

本文提出了一种基于特征融合的人体行为识别方法。首先利用背景差分法和阴影消除技术分别提取轮廓特征和人体区域特征,将这两种具有一定互补性的特征采用K-L变换进行融合,从而得到一个分类能力更强的特征。最后,在传统支持向量机的基础上,结合模糊聚类技术和决策树构建多级二叉

树分类器,从而实现行为多类分类。实验结果表明,本文提出的识别方法具有较高的识别性能。未来的工作方向为选择更为优越的融合策略。

参考文献:

- [1] RONALD P. A survey on vision-based human action recognition[J]. *Image and Vision Computing*, 2010, 28(6): 976-990.
- [2] 黄天羽,石崇德,李凤霞,等.一种基于判别随机场模型的联机行为识别方法[J]. *计算机学报*, 2009, 32(2): 275-281.
- [3] 黎洪松,李达.人体运动分析研究的若干新进展[J]. *模式识别与人工智能*, 2009, 22(1): 70-78.
- [4] 韩磊,李君峰,贾云得.基于时空单词的两人交互行为识别方法[J]. *计算机学报*, 2010, 33(4): 1-11.
- [5] 徐光祐,曹媛媛.动作识别与行为理解综述[J]. *中国图象图形学报*, 2009, 14(2): 189-195.
- [6] BASHIR F I, KHOKHAR A A, SCHONFELD D. Object trajectory based activity classification and recognition using hidden Markov models[J]. *IEEE Trans on Image Processing*, 2007, 16(7): 1912-1919.
- [7] SMINCHISESCU C, KANSUJIA A, METAXAS D. Conditional models for contextual human motion recognition[J]. *Computer Vision and Image Understanding*, 2006, 104(2-3): 210-220.
- [8] QIAN Hui-min, MAO Yao-bin, XIANG Wen-bo, et al. Recognition of human activities using SVM multi-class classifier[J]. *Pattern Recognition Letters*, 2010, 31(2): 100-111.
- [9] ELGAMMAL A, HARWOOD D, DAVIS L. Non-parametric model for background subtraction[C]//Proc of ECCV. 2000: 751-767.
- [10] 褚一平,陈勤,黄叶珏,等.基于判别模型的视频前景/阴影自动分割算法[J]. *模式识别与人工智能*, 2008, 21(6): 849-855.
- [11] TABBONE S, WENDLING L, SALMON J P. A new shape descriptor defined on the Radon transform[J]. *Computer Vision and Image Understanding*, 2006, 102(1): 42-51.
- [12] 张浩,刘志镜.加权DTW距离的自动步态识别[J]. *中国图象图形学报*, 2009, 15(5): 830-836.
- [13] 敦文杰,穆志纯.基于特征融合的人脸人耳多生物身份鉴别[J]. *天津大学学报*, 2009, 42(7): 636-641.
- [14] 姚勇,赵辉,刘志镜.一种非线性支持向量机决策树多值分类器[J]. *西安电子科技大学学报*, 2007, 34(6): 873-876.