基于 pLSA 模型的人体动作识别*

谭论正[1],夏利民[1],黄金霞[1],夏胜平[2]
(1. 中南大学 信息科学与工程学院,湖南 长沙 410075;
2. 国防科技大学 ATR 重点实验室,湖南 长沙 410073)

摘 要:提出一种基于主题模型的人体动作识别方法。该方法首先提取时空兴趣点(STIP,Space-Time Interest Point)来描述人体运动,然后提出使用慢特征分析(SFA, Slow Feature Analysis)算法计算兴趣点梯度信息不变量最优解,最后使用概率潜在语义分析(pLSA, probabilistic Latent Semantic Analysis)模型识别人体动作。SFA 计算的梯度不变量最优解可以表示时空兴趣点固有特征,能够无歧义反映时空兴趣点在空间及时间方向上的信息。同时,针对 pLSA 隐性主题正确性无法保证的缺点,算法将主题与动作标签"一对一"相关,通过监督方式得到主题,保证了训练中主题的正确性。该算法在 KTH 人体运动数据库和 Weizmann 人体动作数据库进行了训练与测试,动作识别结果正确率分别在 91.50% 和 97% 以上。

关键词:动作识别;主题模型;慢特征分析;时空兴趣点;梯度直方图
中图分类号:TP391 **文献标志码**:A **文章编号**:1001-2486(2013)05-0102-07

Human action recognition based on pLSA model

TAN Lunzheng[1], *XIA Limin*[1], *HUANG Jinxia*[1], *XIA ShengPing*[2]
(1. College of Information Science and Engineering, Central South University, Changsha 410075, China;
2. ATR State Key Lab, National University of Defense Technology, Changsha 410073, China.)

Abstract: A human action recognition method based on a probabilistic topic model is proposed. Firstly, the method extracts space-time interest points to describe human motion. Then the slow feature analysis algorithm was proposed to calculate the invariant optimal solution of the gradient information of space-time points. Lastly human actions were recognized with the probabilistic latent semantic analysis (pLSA). The invariant optimal solution of the gradient information can express the inherent characteristics of STIP, and it can also reflect the space and time information of STIP discriminatively. For solving the problem of latent topics that are not guaranteed in pLSA, the topics obtained in supervised fashion correspond to action labels one by one. Action recognition results were presented on KTH human motion data set and Weizmann human action data set. Our results show that the action recognition rates of the tow dataset are respectively more than 91.50% and 97%.

Key words: action recognition; topic model; slow feature analysis; space-time interest points; histogram of gradient

近年来,高层视觉研究的发展为人体动作识别提供技术动力,人体动作识别已经成为计算机领域中备受关注的前沿方向之一。其应用范围包括:机场监控、保安系统、病护监控、人机交互、运动与娱乐分析等。但由于存在背景杂乱、摄像机运动、遮挡、物体几何和光学差异、缩放变化、低空间和时间分辨率等问题,人体动作识别技术仍然是计算机视觉领域的难点。

目前,人体动作识别都大致分为两个步骤:一是底层视频特征提取与表示;二是高层人体动作建模与识别。底层视频特征提取方面,整体运动特征和局部运动特征被广泛应用于动作识别,如人体形状和外貌特征、关节点轨迹、局部兴趣点信息及光学流等。如 Bobick 和 Davis[1] 提出 MHI (Motion History Image)学习和识别不同的人体动作,Guo 和 Qian[2] 通过检测和跟踪人体躯干、腿和手臂等特定部位描述学习人体动作。整体特征能够在语义水平上较好地分析人体动作。但这些整体特征的致命缺陷是高度依赖人体部位的跟踪,如果出现遮挡或环境变化复杂等因素,将无法得到完整的运动信息。在高层动作建模与识别方面,动作模型主要有隐马尔可夫模型(Hidden Markov Models, HMM)[3]、动态贝叶斯网络(Dynamic Bayesian Networks, DBN)[4] 等复杂概率模型,如

* **收稿日期**:2013-05-28
基金项目:国家 863 高技术资助项目(2009AA11Z205);国家自然科学基金资助项目(50808025);教育部博士点基金资助项目(20090162110057)
作者简介:谭论正(1981—),女,湖南株洲人,博士研究生,讲师,E-mail:tanlunzheng@126.com;
夏利民(通信作者),男,教授,博士,博士生导师,E-mail:xlm@mail.csu.edu.cn

Yamato 等[3]使用 HMM 识别停车场泊车行为和 Aggarwal[4]应用 DBN 识别两个人交互姿势。然而,HMM 和 DBN 等状态模型的不足是需要引入大量假设、约束条件,同时需要设置许多参数。

针对以上问题,本文提出使用时空兴趣点描述人体运动,使用 pLSA 隐性主题模型识别人体动作。STIP 能检测空间和时间方向上人体动作状态的变化。相对于全局运动描述,时空兴趣点具有较好的旋转、平移和缩放等不变性,可有效降低复杂背景、人体形状和相机等带来的影响。甚至,在出现部分遮挡和在杂乱的背景条件下,STIP 仍能比较稳定和有效地描述识别人体动作。提取的 STIP 区域立方体,本文使用梯度立方图量化表示。空间梯度直方图(Histogram of Gradient, HOG)[5]只能对 STIP 区域立方体进行空间描述,而不能对 STIP 区域立方体时间方向变化信息进行描述。因此,本文提出使用 SFA 分析计算兴趣点立方体梯度不变量最优解。SFA 不变量最优解表示时空兴趣点固有特征,能够无歧义地反映时空兴趣点信息。在高层动作建模和识别方面,pLSA 主题模型无须引入大量假设和约束条件,它以主题作为隐性变量,通过边缘化结构得到主题,并使用 EM 算法估计参数。但是,通过边缘化结构得到的主题个数与正确性都无法保证。而主题与动作类别相关,导致动作类别个数与正确性无法保证。针对以上 pLSA 的缺点,本文将主题与动作标签"一对一"相关。在训练中,主题通过有监督方法得到并将主题与动作标签一一对应。得到的主题与词的概率结构,使用 EM 算法得到动作视频的动作向量。

本文的动作识别框架如图 1 所示,首先提取 STIP 检测视频动作的局部时空特征,提出 SFA 算法计算兴趣点梯度信息不变量最优解,然后对数据库视频训练估计 pLSA 模型参数,并对未知视频进行动作识别。

本文提出使用 SFA 算法计算兴趣点梯度信息不变量最优解,梯度不变量最优解表示时空兴趣点固有特征,能够无歧义全面地反映时空兴趣点在空间及时间方向上的信息。同时使用一个特征向量表示量化 STIP 空间块有效降低特征空间维。模型训练中,将主题与动作标签"一对一"相关并通过监督得到主题,克服了隐性主题模型主题无法保证造成的错误。通过与支持向量机(Support Vector Machine, SVM)、无监督的 pLSA、LDA[6](Latent Dirichlet Allocation)等对比试验,证明本算法能够有效地识别人体动作。

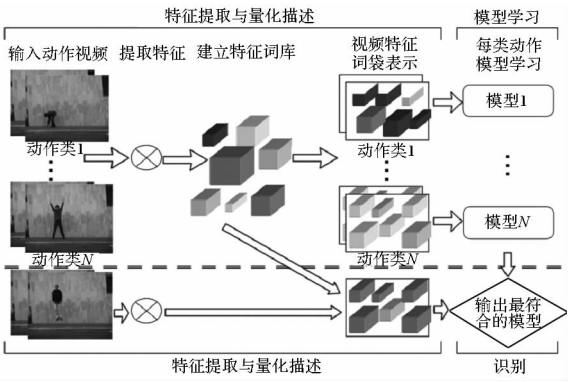


图 1　本文算法流程图

Fig. 1　Flow chart of the algorithm

1　视频局部特征表示

本文使用 Laptev[7]提出的时空兴趣点检测算子提取时间和空间上梯度变化大的像素点。在检测得到的 STIP 邻近区域,提取 3D 空间立方体,并使用该立方体表示兴趣点处的特征。Thi[8]使用 HOG 和 HOF(Histograms of Optical Flow),描述及量化时空兴趣点立方体。然而,空间 HOG 仅能描述兴趣点立方体的空间信息,无法表示立方体在时间上变化。HOF 能描述人体运动时间变化,但光学流需要匹配多帧图像中的像素点来计算视频的像素变化矢量,该方法计算复杂,进行精确与实时检测时需要专门的硬件设备。本文使用 HOG 描述量化时空兴趣点立方体,然后提出使用 SFA 算法计算兴趣点梯度信息不变量最优解。

1.1　时空兴趣点(STIP)检测

STIP 是视频中空间和时间上梯度变化显著的像素点,本文通过搜索形状及运动梯度变化大的像素进行检测。STIP[7]是 Harris 角点检测算子[9]在空间 - 时间上的扩展,能够检测出视频在空间和时间方向上的变化,STIP 检测过程如下。

假设 $f(x,y,t)$ 为一图像像素,其中 (x, y) 表示像素的空间坐标,t 表示像素的时间坐标。文中使用时空参数可分离的高斯函数 $g(\cdot;\sigma_l^2,\tau_l^2)$(以下将 (x, y, t) 简化为 $(\cdot)$)与图像像素 $f(x,y,t)$ 卷积构建线性多尺度空间 L,并将 L 作为视频数据的特征模型,L 计算公式如下:

$$L(x,y,t;\sigma_l^2,\tau_l^2)=g(x,y,t;\sigma_l^2,\tau_l^2)*f(x,y,t), \tag{1}$$

其中 $*$ 表示卷积算子,σ_l^2 和 τ_l^2 分别表示独立空间尺度变量和独立时间尺度变量,高斯函数定义为

$$g(x,y,t;\sigma_l^2,\tau_l^2)=\frac{1}{\sqrt{(2\pi)^3\sigma_l^4\tau_l^2}}\cdot\exp\left(\frac{-(x^2+y^2)}{2\sigma_l^2}-\frac{t^2}{2\tau_l^2}\right) \tag{2}$$

本文使用多尺度空间 L 与高斯权重函数 $g(\cdot;s\sigma_l^2,s\tau_l^2)$($s$ 是空间尺度变量和时间尺度变量缩放系数)卷积构建时空二阶矩矩阵 μ 检测时空兴趣点,μ 计算公式为

$$\mu(\cdot;\sigma_l^2,\tau_l^2)=g(\cdot;s\sigma_l^2,s\tau_l^2)*(\nabla L(\cdot;\sigma_l^2,\tau_l^2)(\nabla L(\cdot;\sigma_l^2,\tau_l^2))^{\mathrm{T}}) \tag{3}$$

式中 $\nabla L(\cdot;\sigma_l^2,\tau_l^2)$ 为尺度空间函数 L 分别在 x、y、t 方向上的一阶导数,各分量 L_x、L_y、L_t 计算公式如下:

$$\begin{aligned}L_x(\cdot;\sigma_l^2,\tau_l^2)&=\partial_x(g*f)\\ L_y(\cdot;\sigma_l^2,\tau_l^2)&=\partial_y(g*f)\\ L_t(\cdot;\sigma_l^2,\tau_l^2)&=\partial_t(g*f)\end{aligned} \tag{4}$$

即 L 在各个方向上梯度矩阵为

$$\nabla L(\cdot;\sigma_l^2,\tau_l^2)(\nabla L(\cdot;\sigma_l^2,\tau_l^2))^{\mathrm{T}}=\begin{pmatrix}L_x^2 & L_xL_y & L_xL_t\\ L_xL_y & L_y^2 & L_yL_t\\ L_xL_t & L_yL_t & L_t^2\end{pmatrix} \tag{5}$$

要检测时空兴趣点,首先要搜索时空兴趣点所在的区域。文中通过搜索图像各像素 $f(x,y,t)$ 中二阶矩矩阵 μ 具有显著特征值区域检测时空兴趣点位置。与 Harris 角点检测算子[9]相似,文中使用阈值函数 H 检测局部极大值空间点,并将极大值点作为时空兴趣点。阈值函数定义如下:

$$H=\det(\mu)-k\ \mathrm{trace}^3(\mu) \tag{6}$$

那么,图像各像素 $f(x,y,t)$ 中的时空兴趣点在阈值函数 H 的正局部时空极大值处取得,即在阈值函数 $H>0$ 处取得,如图 2 所示,从 Weizmann 数据库提取的 STIP。

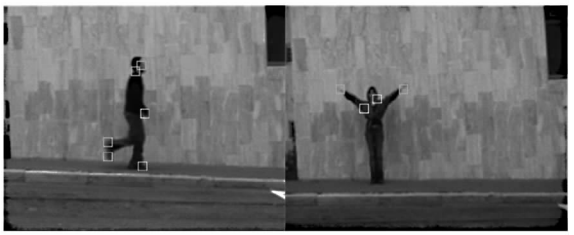


图 2 Weizmann 数据库动作时空兴趣点区域

Fig. 2 The actions' spatio-temporal interest point of Weizmann database

1.2 特征描述

将视频看作为(x, y, t)空间的立方体,利用式(1) ~(2)构建视频数据多尺度空间 L,将 L 梯度矩阵与高斯函数卷积(如式(3) ~(5))得到时空二阶矩矩阵 μ,使用阈值函数 H 检测局部极大值空间点作为视频时空兴趣点。本文使用梯度描述子 HOG 对提取的时空兴趣点区域立方体进行描述,并提取兴趣点区域梯度信号的不变量。

给定一个时空兴趣点 p,以 p 为中心提取一个 3D 立方体,立方体体积为($\Delta_x(\sigma_l)$,$\Delta_y(\sigma_l)$,$\Delta t(\tau_l)$)且 $\Delta_x(\sigma_l)=\Delta_y(\sigma_l)=18\sigma_l$,$\Delta_t(\tau_l)=8\tau_l$。假设时空兴趣点立方体起始帧时间设为 1,结束帧时间设为 T,即立方体图像块帧时间表示为 $I_t(t=1,2,\cdots,T)$。使用式(1)和(2)对立方体图像块帧 I_t 进行多尺度空间 L 计算。

HOG:图像块帧 I_t 中,各像素点梯度大小及方向计算公式如下:

$$m_t(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \tag{7}$$

$$\theta_t(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}. \tag{8}$$

通过式(7)和(8)得到 STIP 区域立方体各图像块帧各像素点的梯度大小及方向。将得到的每个图像帧各点梯度向量统计在如图 3 所示 64 维的梯度 - 方向直方图 O 上,那么每个兴趣点图像块帧 I_t 可以使用 64 维方向 - 梯度向量 $x(t)=[x_1(t),x_2(t),\cdots,x_{64}(t)]^{\mathrm{T}}$ 表示。

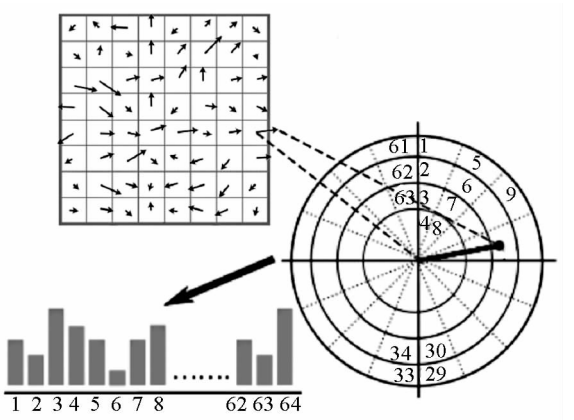


图 3 梯度 - 方向描述子

Fig. 3 Gradient direction descriptors

1.3 基于 SFA 的时空兴趣点梯度不变量提取

时空兴趣点立方体 $S=\{I_1,\cdots,I_t,\cdots,I_T\}$,图像块帧 I_t 使用 64 维方向 - 梯度向量 $x(t)$ 表示。假设输入特征块向量 $x(t)=[x_1(t),x_2(t),\cdots,x_{64}(t)]^{\mathrm{T}}$,SFA 算法的目的是找到一个非线性函数 $f(x)=[f_1(x),f_2(x),\cdots,f_L(x)]^{\mathrm{T}}$,使得该非线性函数输出 $y(t)=f(x(t))$ 的各个分量变化尽可能慢。对于非线性函数输出的变化速率,SFA 使用关于时间一阶导数的平方均值进行衡量,即

如果存在 SFA 的目标函数如下:

$$\min_{y_i}\Delta(y_i) := <(\dot{y}_i(t))^2>, \tag{9}$$

式中 $y_i(t)$ 表示第 i 个非线性函数,为了避免 $y_i(t)$ 值为常量,并使其均值为 0,将特征向量 $x(t)$ 作变换,变换后的向量为 $e(t)=x(t)-<x(t)>$。变换后的向量 $e(t)$ 是均值为 0 的特征向量。同时,为了保证输出信号方差为 1 并两两间不相关,对变换后的向量 $e(t)$ 作以下处理:

$$y_i(t)=w^{\mathrm{T}}e(t) \tag{10}$$

经过以上信号处理,目标函数(9)将变形为

$$\min J_{SFA}(w) := \min_{y_i}\Delta(y_i) := \langle(\dot{y}_i(t))^2\rangle = \langle(\dot{w}^{\mathrm{T}}e(t))(\dot{w}^{\mathrm{T}}e(t))^{\mathrm{T}}\rangle \tag{11}$$

式中 J_{SFA} 表示慢特征分析的目标函数。本文使用 D 表示数据的协方差矩阵并且 $D=\langle e(t)e(t)^{\mathrm{T}}\rangle$,那么协方差矩阵 D 的时间导数为 $\dot{D}=\langle(e(t)-e(t-1))(e(t)-e(t-1))^{\mathrm{T}}\rangle$。将协方差矩阵 D 代入式(11),得到以下目标函数:

$$\min J_{SFA}(w) := \langle(\dot{w}^{\mathrm{T}}e(t))(\dot{w}^{\mathrm{T}}e(t))^{\mathrm{T}}\rangle = w^{\mathrm{T}}\dot{D}w \tag{12}$$

由于本文仅 $y_i(t)$ 考虑线性函数,因此目标函数可以使用以下形式表示:

$$\min J_{SFA}(w) := \frac{w^{\mathrm{T}}\dot{D}w}{w^{\mathrm{T}}Dw} \tag{13}$$

上式中找到向量 w,使得目标函数 J_{SFA} 最小。求解向量 w 问题可以转化为求产生特征向量 w 并且满足 $\dot{D}w=\lambda Dw$ 的特征值 λ。输出信号 $y(t)$ 最慢的信号成分就是 $y(t)$ 投影到最小特征值 λ 的特征向量 w 的成分。

得到各个时空兴趣点立方体的梯度向量直方图不变量最优解 w 后,使用 $k-$mean 算法对每个时空兴趣点梯度不变量最优解进行聚类,每类作为动作特征词。

2　基于 pLSA 人体动作识别

本文以时空兴趣点作为动作特征,使用 pLSA 主题模型对人体动作进行建模识别。动作识别的第一步提取时空兴趣点。提取到的时空兴趣点区域空间使用梯度向量直方图表示,并使用 SFA 分析时空兴趣点区域梯度不变量最优解,最优解经过 $k-$mean 算法进行聚类,得到时空兴趣点特征词,最后使用 pLSA 建立人体动作模型识别。

pLSA 模型是 Thomas Hofmann[10] 提出的一种主题模型,该模型初始用于文档库结构学习。Niebles[11] 等提出使用 pLSA 模型识别和定位人体动作,并取得较好的识别效果。然而,pLSA 进行动作识别仍存在以下问题:由边缘化结构得到的隐形“主题”正确性与动作的相关性无法保证。本文将动作标记与主题一一对应相关,并在训练中主题由观察所得。在训练中,主题是有监督学习,能保证其正确性及个数,有效提高动作识别率。

2.1　pLSA 模型

假设动作视频训练集中有 N 个动作视频、M 个特征词和 K 个主题。动作视频 d_j、主题 z_k 和特征词 w_i 的联合概率结构如图 4 所示。

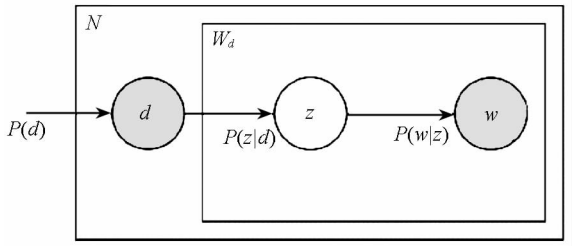


图 4　pLSA 图解模型

Fig. 4　Graph model of pLSA

假设动作视频 d_j、主题 z_k 和特征词 w_i 各量相互独立,那么给定观察对 (d_j, z_k, w_i) 的联合概率公式如下:

$$p(d_j,z_k,w_i)=p(d_j)p(z_k|d_j)p(w_i|z_k) \tag{14}$$

式中 $p(d_j)$ 是动作视频 d_j 的先验概率且 $p(d_j)=1/N$, $p(z_k|d_j)$ 表示主题 z_k 与动作视频 d_j 的概率结构,$p(w_i|z_k)$ 表示特征词 w_i 出现在主题 z_k 的概率。通过边缘化主题 z_k 得到条件概率 $p(w_i|d_j)$:

$$p(w_i|d_j)=\sum_{k=1}^{K}p(z_k|d_j)p(w_i|z_k) \tag{15}$$

式中 $p(w_i|z_k)$ 和 $p(z_k|d_j)$ 使用 EM 算法最大化目标函数计算得到,目标函数如下:

$$F=\prod_{i=1}^{M}\prod_{j=1}^{N}p(w_i|d_j)^{n(w_i,d_j)} \tag{16}$$

其中,$n(w_i,d_j)$ 表示在动作视频 d_j 中特征词 w_i 出现的次数。

pLSA 模型训练过程中,主题是由最大化目标函数得到。为克服主题无法保证问题,本文提出训练过程中,主题由观察所得。输入动作视频 d_j,将该动作视频划分为多个动作,每个动作与一个主题对应。同时,在每个动作提取特征词。动作视频 d_j 与主题 z_k 的概率结构由先验概率计算,即 $p(z_k|d_j)=\frac{n_{k,j}}{n_j}$($n_{i,j}$ 表示主题 z_k 出现在视频 d_j 的次数,n_j 表示出现视频 d_j 主题数)。视频 d_j 可表示为主题组合向量,如某视频由 1 个 walking、2 个 jogging、0 个 running 和 2 个

bending 动作组成,表示为向量(1, 2, 0, 2) 或(0.2, 0.4, 0.0, 0.4)。由得到的 $p(z_k \mid d_j)$ 代入式(15),pLSA 模型通过 EM 算法最大化目标函数,估计各特征词与主题的概率结构 $p(w_i \mid z_k)$。

2.2 动作识别

给定一个未识别动作视频 d_{test},要识别视频中人体动作,首先,利用前述方法提取视频中时空兴趣点;然后,利用梯度描述子描述量化时空兴趣点邻近立方体,并使用慢特征分析算法提取时空兴趣点不变量,得到各个时空兴趣点特征词 $w = \{w_1 \cdots w_m\}$ (m 为视频 d_{test} 包含可重复的特征词个数);最后,使用 pLSA 模型对动作视频进行识别。

动作识别的目的是使用已经学习的模型对新输入视频序列进行动作分类,也就是使用动作向量表示输入视频。从不同的训练序列集通过学习得到与动作类相关的主题 z 和特征词 w 概率结构 $p(w \mid z)$。当给定一个新动作视频,将这个未知视频"投影" 到动作直方图上,即找到主题的混合系数 $p(z_k \mid d_{test})$。从实验中可以得到主题混合系数的先验概率 $\tilde{p}(w \mid d_{test})$,计算 $\tilde{p}(w \mid d_{test})$ 和 $p(w \mid d_{test}) = \sum_{k=1}^{K} p(z_k \mid d_{test}) p(w \mid z_k)$ 的KL距离,使用EM算法最小化KL距离,得到动作混合系数 $p(z_k \mid d_{test})$,即得到待识别视频动作组合。

3 实验结果

为验证文中算法有效性,采用两个数据库对算法进行了训练、测试,并与无监督 pLSA 主题模型及 SVM 等目前先进方法进行对比实验,两个数据库分别是:KTH 人体运动数据库和 Weizmann 人体动作数据库。两数据库的每个视频序列仅包含一个动作,对于每个数据库,使用"留一法"交叉实验。文中算法提取时空兴趣点区域,每个区域使用32bin 梯度直方图描述。实验时,用VC ++6.0实现了文中算法,测试的环境是 AMD2GHz、1G 内存的普通 PC,使用 Windows XP 操作系统。

KTH 数据库包含六种人体动作,即走、拳击、拍手、挥手、慢跑和跑,这六种动作都由 25 个人在四个不同场景(户外、缩放变化户外、不同衣着户外和户内)下执行,同时带有摄像机运动。本数据库的关键帧如图 5 所示。

为证明本文使用有监督 pLSA 模型识别 KTH 数据库视频动作的有效性,首先使用文中提出有监督的 pLSA 模型进行动作识别实验,再与使用 SVM 识别动作实验结果进行对比。文中算法识别结果混淆矩阵如图 6(a)所示,本文算法整体动作识别精度为 91.50%,图中跑和拍手两个动作识别精度比较低,原因是跑与慢跑动作相似、拍手与挥手动作相似导致特征相似。图 6(b)[12] 为使用慢特征分析算法提取数据库动作视频时空兴趣点不变量后,利用 SVM 进行动作识别的结果混淆矩阵。图 6(b)中的实验同样提取动作视频时空兴趣点并使用慢特征分析抽取兴趣点不变量表示视频动作,该实验使用 SVM 的整体识别精度为 84.67%,低于本文的整体识别精度 91.50%。同时,由图 6(a)与图 6(b)比较可知,本文使用有监督 pLSA 模型除动作拍手和跑外,其他动作识别精度均高于 SVM 识别精度。本文算法代码大小对识别精度影响如图 6(c)所示,代码词大小为 600 时识别结果最好。

图 5 KTH 数据库关键帧

Fig. 5 Key frame of KTH database

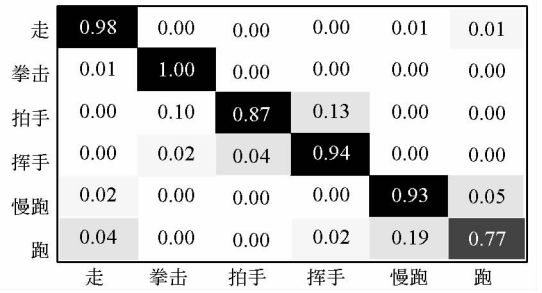
为进一步证明本文算法的有效性,表 1 将本文算法与其他方法实验结果进行对比。Niebles 等[11] 使用无监督 pLSA 模型识别率 81.5%,该方法在训练过程中主题与参数均通过 EM 算法估计。Yang 等[13] 提出半监督方法,使用 S - LDA (91.20%)模型识别动作,在训练中主题和词由观察所得,识别率要高于 Niebles 等识别率。文中提出的算法,在训练过程中主题由观察所得,所以识别结果(91.50%)要好于 Niebles 等[11] 结果,并与 Yang 等[13] 识别效果相似。Nowozin 等[14] 使用 SVM 进行识别,识别率低于文中算法。

表 1 KTH 数据库结果比较

Tab. 1 Results' comparison of KTH database

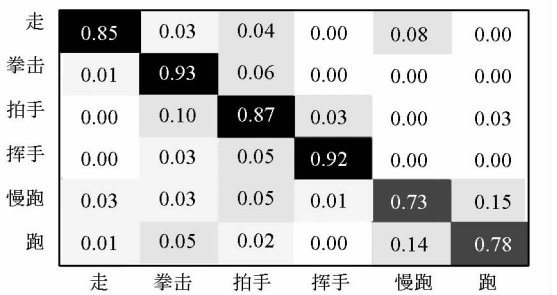
算法	识别精度(%)
本文方法	91.50
pLSA[11]	81.50
S - LDA[13]	91.20
SVM[14]	87.04

Weizmann 数据库仅有一个场景,共有 90 个视频,由 9 个人执行 10 个不同的动作,每个视频持续 2 ~ 3s。10 个动作分别为弯身、开合跳、原地



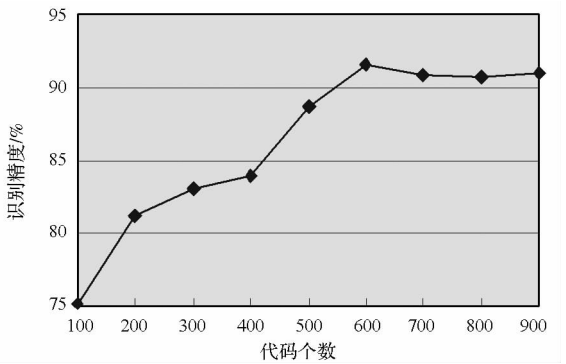
(a)有监督 pLSA 动作识别结果混淆矩阵,
检测精度为 91.50%

(a) The result confusion matrix of supervision pLSA action recognition, the overall detection accuracy is 91.50%



(b) SVM 动作识别结果混淆矩阵,检测精度为 84.67%

(b) The result confusion matrix of SVM action recognition, the overall detection accuracy is84.67%



(c)不同代码个数对本文算法识别精度影响

(c) Effects of different code number on the recognition accuracy of this paper

图 6　KTH 数据库识别结果

Fig. 6　The recognition results of KTH database

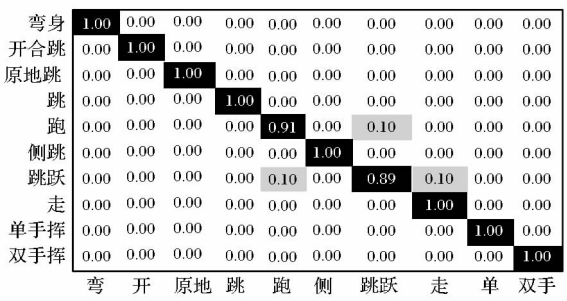
跳、跳、跑、侧跳、跳跃、走、单手挥动和双手挥动。该数据库视频无场景切换且无摄像机运动,关键帧如图 7 所示。

文中算法在 Weizmann 数据库的总体识别精度为 97.00%,识别结果混淆矩阵如图 8(a)所示。从矩阵中可以看到,除了跑、跳跃,文中算法的识别率都是 100%。跳跃动作与跑动作相似,因而识别效果低于其他动作。图 8(b)[12] 为使用慢特征分析算法提取数据库动作视频时空兴趣点不变

图 7　Weizmann 数据库关键帧

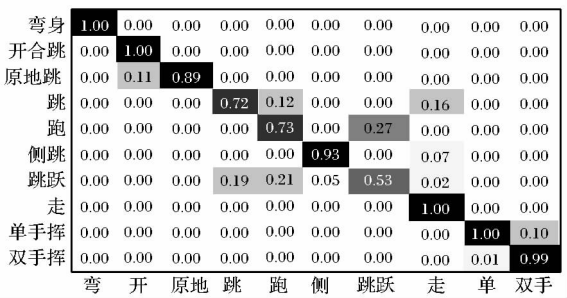
Fig. 7　Key frame of Weizmann database

量后,利用 SVM 进行动作识别的结果混淆矩阵,



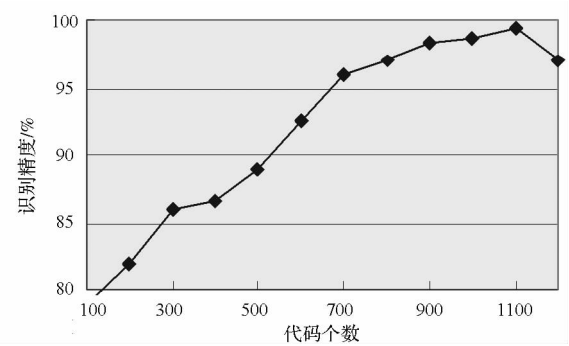
(a)有监督 pLSA 动作识别结果混淆矩阵,
检测精度为 97.00%

(a)The result confusion matrix of supervision pLSA action recognition, the overall detection accuracy is97.00%



(b)SVM 动作识别结果混淆矩阵,整体检测精度为 87.90%

(b)The result confusion matrix of SVM action recognition, the overall detection accuracy is 87.90%



(c)不同代码个数对识别精度影响

(c) Effects of different code number on the recognition accuracy

图 8　Weizmann 数据库识别结果

Fig. 8　The recognition results of Weizmann database

识别精度为87.90%,低于本文使用有监督 pLSA 识别动作精度 97.00%。同时,由图 8(a)与图 8(b)可知,本文使用有监督 pLSA 模型各个动作识别精度均高于 SVM 识别精度。本文算法代码大小对识别精度影响如图8(c)所示,代码大于1100后对识别效果影响不大。

为进一步证明本文算法的有效性,表 2 将本文算法与其他方法实验结果进行对比。从表2 可看到,同样由于采用无监督学习过程,Zhang 等[15]使用 pLSA(92.30%)模型及 Liu[16]使用 LDA 模型(71%)的识别效果低于半监督学习的本文算法。在本数据库中,Jhuang 等[17]使用 SVM 识别效果比本文效果好。原因是文献[17]将复杂动作跳跃去除后再进行识别,所以识别效果为98.8%,好于本文识别效果。

表 2 Weizmann 数据库结果比较

Tab. 2 Results' comparison of Weizmann database

算法	识别精度(%)
本文方法	97.00
pLSA[15]	92.30
LDA[16]	71.00
SVM[17]	98.8

4 结论

人体动作识别已经成为计算机领域的一个重要研究方向,本文提出一种基于主题模型的人体动作识别方法,主要工作和创新点:

(1)利用时空兴趣点来描述人体运动。

(2)提出采用 SFA 算法计算兴趣点梯度信息不变量最优解,并使用梯度不变量最优解聚类作为动作特征词。由于 HOG 只能描述 STIP 区域立方体空间信息,而无法描述时间方向上的变化,本文梯度不变量最优解表示时空兴趣点固有特征,能够无歧义全面地反映时空兴趣点在空间及时间方向上的信息。

(3)提出了基于概率潜在语义分析模型识别人体动作,在 pLSA 模型训练中,将主题与动作标签"一对一"相关,并通过监督得到主题,克服了隐性主题模型主题正确性无法保证造成的错误,提高识别率。

参考文献(References)

[1] Bobick A, Davis J. The recognition of human movement using temporal templates[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23 (3):257 - 267.

[2] Guo F, Qian G. Monocular 3D tracking of articulated human motion in silhouette and pose manifolds[J]. Image and Video Processing, 2008, 2008(3):1 - 18.

[3] Yamato J, Ohya J, Ishii K. Recognizing human action in time-sequential images using hidden markov model [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1992: 379 - 385.

[4] Park S, Aggarwal J K. A hierarchical bayesian network for event recognition of human actions and interactions [J]. Multimedia Systems, 2004, 10(2):164 - 179.

[5] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005:886 - 893.

[6] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Machine Learning Research, 2003(3):993 - 1022.

[7] Laptev I, Lindeberg T. Space-time interest points [C]// Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003:432 - 439.

[8] Thi T H, Cheng L, Zhang J, et al. Structured learning of local features for human action classification and localization [J]. Image and Vision Computing, 2012, 30(1):1 - 14.

[9] Harris C, Stephens M. A combined corner and edge detector [C]//Alvey vision conference, 1988: 147 - 151.

[10] Hofmann T. Probabilistic latent semantic indexing [C]// Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999:50 - 57.

[11] Niebles J, Wang H C, Li F F. Unsupervised learning of human action categories using spatial-temporal words [J]. International Journal of Computer Vision, 2008, 79(3): 299 - 318.

[12] Zhang Z, Tao D C. Slow feature analysis for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3):436 - 450.

[13] Yang W, Mori G. Human Action Recognition by Semilatent Topic Models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(10):62 - 74.

[14] Nowozin S, Bakir G, Tsuda K. Discriminative Subsequence Mining for Action Classification[C]//Proceedings of the 11th IEEE International Conference on Computer Vision,2007:1 - 8.

[15] Zhang J G, Gong S G. Action categorization by structural probabilistic latent semantic analysis[J]. Computer Vision and Image Understanding, 2010, 114(8): 857 - 64.

[16] Liu P, Wang J, She M, et al. Human action recognition based on 3D SIFT and LDA Model[C]//Proceedings of IEEE SSCI Workshop on Robotic Intelligence in Informationally Structured Space, 2011:12 - 17.

[17] Jhuang H, Serre T, Wolf L, et al. A biologically inspired system for action recognition [C]//Proceedings of the 11th IEEE International Conference on Computer Vision, 2007:1 - 8.