

基于张量子空间学习的人行为识别方法

凌志刚 梁彦 潘泉 程咏梅 赵春晖

(西北工业大学自动化学院, 西安 710072)

摘要 提出了一种基于张量子空间学习降维人体高维侧影数据的人行为识别方法。给定一个动作的人侧影图像序列,首先用张量子空间学习方法将目标高维侧影图像投影到低维子空间来描述人运动的时空特性,并同时尽可能地保持目标侧影图像中像素之间的空间几何信息,然后用 Hausdorff 距离度量动作之间的相似性,并在最近邻距离框架下对动作进行分类识别。为验证本文算法的有效性,设计了动作识别和鲁棒性测试 2 个实验。实验结果表明提出的算法不仅能够有效地对人行行为进行识别,且具有较强的鲁棒性。

关键词 人行为识别 张量子空间学习 主成分分析 局部保持投影

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2009)03-0394-07

Human Action Recognition Based on Tensor Subspace Learning

LING Zhi-gang, LIANG Yan, PAN Quan, CHENG Yong-mei, ZHAO Chun-hui

(College of Automation, Northwestern Polytechnical University, Xi'an 710072)

Abstract In this paper, a simple but efficient algorithm based on tensor subspace learning is proposed to reduce the dimensionality of high-dimensional silhouette data for human action recognition. For image sequences of each action, they are projected into a low dimensional subspace so that both spatial and temporal properties of the action are preserved. Further, a nearest-neighbor action recognition is carried out basing on Hausdorff distance. Two experiments for action recognition and robust test have been carried out to testify the effectiveness of introduced tensor subspace learning.

Keywords human action recognition, tensor subspace learning, principal components analysis (PCA), locality preserving projections (LPP)

1 引言

人行为识别由于在智能监控、人机交互、虚拟现实以及基于内容的视频检索等方面有着潜在的应用背景和经济价值而受到广泛关注,并已成为计算机领域最活跃的研究方向之一。然而,人行为是一种高自由度的非刚体运动,同时现实环境中存在遮挡、光照变化、阴影、动态场景、宽松的衣着以及运动的模糊性等因素的影响,使得人行为识别成为一个复

杂而又极具挑战性的任务。

随着人行为识别研究的不断深入,大量的行为识别方法被提出^[1-2],这些行为识别方法基本上可分成基于运动特征的识别方法^[3-5]和基于形状特征的识别方法^[6-8]两类。基于运动的方法是提取人体的运动特征信息,如运动方向、速度和光流信息等来表征和识别人的行为;而基于形状的方法则是通过恢复人姿态的空间几何信息如侧影、轮廓等来表征和识别行为。心理学研究表明,运动特征具有形状特征无法比拟的优势,即便是在远距离或恶劣的能见度情况下以及最小的运动系统——MLD (moving

基金项目: 国家自然科学基金项目(60634030,60702066);高等学校博士学科点专项科研基金项目(20060699032);航空科学基金项目(2007ZC53037)

收稿日期:2007-10-16; **改回日期:**2008-01-16

第一作者简介:凌志刚(1976~),男。西北工业大学自动化学院模式识别与智能化系统专业博士研究生。主要研究方向为图像处理、信息融合与机器学习等。E-mail: zgling_hunan@126.com

light display)^[9],仍然能够根据运动模式轻松地识别目标的运动类型,而形状特征在此情况下往往变得模糊不清。但目前的计算机视觉方法很难准确地从视频中提取出目标的运动特征,而形状特征却容易获得且对纹理变化不敏感,所以形状特征在行为识别中得到了广泛的应用^[10]。本文采用的方法属于基于形状特征的识别方法。

借助形状特征,人行为可以看作是由一组按时间顺序串连起来的静止姿态图像序列之间的形状时变过程,形状的时空变化反映了人身体全局和局部运动的动态信息。然而这些静止姿态图像都位于嵌入在高维图像空间的低维流形上^[11],所以许多运动识别方法借助各种降维方法如主成分分析(PCA)^[4-6],局部保持投影(LPP)^[11]等将表征目标运动的高维形状信息投影到低维空间,并用少量的特征向量来表征目标的形状信息。如Masoud等人^[5]利用无限脉冲滤波器(IIR)构造特征图像,并通过PCA方法将其投影到低维空间,最后在低维特征空间中进行运动相似性的度量与识别;Wang^[11]则直接用LPP方法将每帧中的目标侧影图像投影到低维空间,然后在低维特征空间采用最近邻框架来识别人的动作。所有这些降维方法都是将 $m \times n$ 大小的侧影图像转换成高维 $\mathbf{R}^{m \otimes n}$ 空间中的一个 $1 \times L$ 维向量($L = m \times n$),然后再对这些高维向量进行降维处理,但这种将2维图像转换到向量的排列方法是将图像中像素看作是相互独立的点,从而丢失了图像中像素之间的空间几何信息。

在基于形状特征的识别方法中,形状的时空变化反映了运动期间人姿态的空间信息,即目标形状特征图像中像素点之间的空间关系会影响行为识别的结果。所以如果能够在降低数据维数的同时,还能有效地保持姿态的空间信息,将会提高人行为识别的结果。基于这种考虑,本文借助张量子空间学习理论,将动作序列中人的二值侧影图像看作是二阶张量并通过张量子空间学习方法将其投影到低维子空间,从而保持了像素之间的固有几何信息,然后在最近邻距离框架下识别人的行为。

2 张量子空间学习理论

假设 $\mathbf{X} \in \mathbf{R}^{m_1 \times m_2}$ 是一个 $m_1 \times m_2$ 大小的矩阵,数学上,根据张量子空间理论, \mathbf{X} 可以看作是张量空间 $\mathbf{R}^{m_1} \otimes \mathbf{R}^{m_2}$ 的二阶张量,并设 $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{m_1}), (\mathbf{v}_1,$

$\mathbf{v}_2, \dots, \mathbf{v}_{m_2})$ 分别是空间 $\mathbf{R}^{m_1}, \mathbf{R}^{m_2}$ 上的垂直正交基。则二阶张量 \mathbf{X} 可以唯一地写成:

$$\mathbf{X} = \sum_{ij} (\mathbf{u}_i^T \mathbf{X} \mathbf{v}_j) \mathbf{u}_i \mathbf{v}_j^T \quad (1)$$

式中,向量 $\{\mathbf{u}_i \mathbf{v}_j^T\}$ 构成张量空间 $\mathbf{R}^{m_1} \otimes \mathbf{R}^{m_2}$ 的基。定义矩阵 $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{m_1}] \in \mathbf{R}^{m_1 \times l_1}$ 和 $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_{m_2}] \in \mathbf{R}^{m_2 \times l_2}$,假设 $\boldsymbol{\mu}, \boldsymbol{\nu}$ 分别是空间 $\mathbf{R}^{m_1}, \mathbf{R}^{m_2}$ 中由基向量 $\{\mathbf{u}_i\}_{i=1}^{l_1}, \{\mathbf{v}_j\}_{j=1}^{l_2}$ 张成的子空间,则张量积 $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$ 是张量空间 $\mathbf{R}^{m_1} \otimes \mathbf{R}^{m_2}$ 的子空间,那么二阶张量 $\mathbf{X} \in \mathbf{R}^{m_1 \times m_2}$ 在子空间 $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$ 上的投影可以表示为

$$\mathbf{Y} = \mathbf{U}^T \mathbf{X} \mathbf{V} \in \mathbf{R}^{l_1 \times l_2} \quad (2)$$

张量子空间学习的任务就是基于特定的目标函数^[12-13],寻找变换矩阵 \mathbf{U}, \mathbf{V} 并将 \mathbf{X} 投影到 $l_1 \times l_2$ 低维子空间 $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$,使得重构误差最小。基于不同的目标函数,张量子空间学习方法有TensorPCA(TPCA)和TensorLPP(TLPP)等方法^[13]。

假设有 N 个矩阵 $\mathbf{X}_1, \dots, \mathbf{X}_N \in \mathbf{R}^{m_1 \times m_2}$,则这些矩阵均值 $\mathbf{M}^{(X)}$ 可以表示为

$$\mathbf{M}^{(X)} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \quad (3)$$

则根据式(2),矩阵在子空间 $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$ 的投影均值 $\mathbf{M}^{(Y)}$ 可以表示为

$$\mathbf{M}^{(Y)} = \frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i = \mathbf{U}^T \mathbf{M}^{(X)} \mathbf{V} \quad (4)$$

式中, \mathbf{Y}_i 是 \mathbf{X}_i 在子空间 $\boldsymbol{\mu} \otimes \boldsymbol{\nu}$ 的投影。

TPCA方法的思想跟PCA类似,就是寻找由少数几个最大特征值对应的特征向量所张成的张量子空间,并使得重构误差最小,其目标函数可以表示为^[12]

$$\max_{\mathbf{U}, \mathbf{V}} \sum_i \|\mathbf{Y}_i - \mathbf{M}^{(Y)}\|^2 \quad (5)$$

$$\begin{aligned} & \text{而} \\ & \sum_{i=1}^N \|\mathbf{Y}_i - \mathbf{M}^{(Y)}\|^2 \\ &= \sum_{i=1}^N \text{tr}(\mathbf{U}^T (\mathbf{X}_i - \mathbf{M}^{(X)}) \mathbf{V} \mathbf{V}^T (\mathbf{X}_i - \mathbf{M}^{(X)})^T \mathbf{U}) \\ &= \text{tr} \left(\mathbf{U}^T \left(\sum_{i=1}^N ((\mathbf{X}_i - \mathbf{M}^{(X)}) \mathbf{V} \mathbf{V}^T (\mathbf{X}_i - \mathbf{M}^{(X)})^T) \right) \mathbf{U} \right) \\ &= \text{tr}(\mathbf{U}^T \mathbf{M}_V \mathbf{U}) \quad (6) \end{aligned}$$

式中, $\mathbf{M}_V = \sum_{i=1}^N ((\mathbf{X}_i - \mathbf{M}^{(X)}) \mathbf{V} \mathbf{V}^T (\mathbf{X}_i - \mathbf{M}^{(X)})^T)$,同样有:

$$\sum_{i=1}^N \|Y_i - M^{(j)}\|^2 = \text{tr}(V^T M_U V) \quad (7)$$

式中, $M_U = \sum_{i=1}^N \left((X_i - M^{(X)})^T U U^T (X_i - M^{(X)}) \right)$ 。可以看出, 最优投影矩阵 U, V 分别是矩阵 M_V, M_U 的特征向量, 但矩阵 U, V 不能独立计算。为了不失一般性, 取 U, V 为正交矩阵, 即 $U U^T = I, V V^T = I$, 则 M_U, M_V 可以写成

$$M_V = \sum_{i=1}^N \left((X_i - M^{(X)}) (X_i - M^{(X)})^T \right) \quad (8)$$

$$M_U = \sum_{i=1}^N \left((X_i - M^{(X)})^T (X_i - M^{(X)}) \right) \quad (9)$$

则 U, V 分别是 M_U, M_V 的特征向量并且 U, V 相互独立。当需要将矩阵降维到 $l_1 \times l_2$ 时, 只需要分别取 U, V 的前 l_1, l_2 个列向量即可。

同理, 基于 LPP^[14] 方法, TLPP 方法的目标函数可以表示为:

$$\min_{U, V} \sum_{i, j} \|U^T X_i V - U^T X_j V\|^2 S_{ij} \quad (10)$$

根据文献[13], 式(10)等价于:

$$\min_{U, V} \frac{\text{tr}(U^T (D_V - S_V) U)}{\text{tr}(U^T D_V U)} \quad (11)$$

$$\text{和} \quad \min_{U, V} \frac{\text{tr}(V^T (D_U - S_U) V)}{\text{tr}(V^T D_U V)} \quad (12)$$

式中

$$D_V = \sum_i D_{ii} X_i V V^T X_i^T, \\ S_V = \sum_{ij} S_{ij} X_i V V^T X_j^T, D_U = \sum_i D_{ii} X_i^T U U^T X_i, \\ S_U = \sum_{ij} X_i^T U U^T X_j \text{ 和 } D_{ii} = \sum_j S_{ij}$$

其中, S_{ij} 是矩阵之间最近邻图的权矩阵, 根据权值 S_{ij} 不同计算方法, TensorLPP 可以分为基于监督的方法 (S-TLPP) 和基于非监督的方法 (TLPP) 两种。由式(11)和式(12)可以看出 U, V 分别是 $(D_V - S_V, D_V)$ 和 $(D_U - S_U, D_U)$ 的广义特征向量, 并且相互依赖, 不能独立求解, 需要通过迭代方法求解 U, V 。

在人为行为识别中, 动作序列中人运动姿态的二值侧影图像可以看作是一个矩阵, 这样, 分析侧影图像的问题转化为分析矩阵数据的问题, 从而可以借助张量子空间学习理论, 将动作序列中人的二值侧影图像投影到低维子空间的同时, 还保持了人姿态图像中像素之间的固有几何信息。

3 子空间学习与动作识别

利用张量子空间学习方法, 将运动中人的二值侧影图像序列投影到低维子空间实现人动作识别, 其方法框图如图 1 所示。首先, 通过运动检测和背景减除等方法提取每帧中运动目标的二值侧影信息, 并将目标侧影图像归一化指定大小, 然后分别用 3 种不同的张量子空间学习方法将归一化后的 2 维侧影图像投影到低维子空间, 并在低维空间中采用最近邻框架识别人的动作。

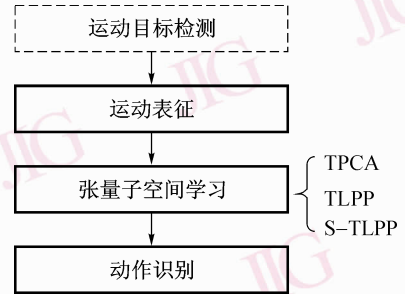


图 1 本文人为行为识别方法总框图

Fig. 1 Principal diagram of human action recognition

3.1 运动表征

运动目标检测是动作识别的第一步, 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目标检测和分割的质量。本文并不讨论人的运动检测和分割问题, 假设运动目标能够很好检测出来。在人被检测出来以后, 通过轮廓跟踪等方法提取每帧图像中人姿态的二值侧影图像。为方便后续处理, 将每帧中目标的二值侧影图像归一化为同样大小 $(m \times n)$, 并且保证归一化后的图像在尽可能多的保留人的前景侧影图像的同时也能够保存目标侧影在归一化后图像中的比例属性, 并用这些归一化后的侧影组成的图像序列作为行为张量子空间学习的输入。

3.2 张量子空间学习

为验证张量子空间学习分析人二值侧影图像的有效性, 分别用 TPCA、TLPP 和 S-TLPP 3 种不同的方法来学习子空间投影矩阵。

假设有 c 类不同的动作, 每一类动作中有 $N_i (i = 1, \dots, c)$ 组由目标二值侧影图像组成的序列, 并设第 i 类第 j 组序列的长度为 $l_{i,j} (i = 1, \dots, c; j = 1, \dots, N_i)$ 。则所有训练的侧影图像样本数为: $L =$

$\sum_{i=1}^c \sum_{j=1}^{N_i} l_{i,j}$, 同时假设 $\mathbf{X}_{i,j,k}$ 是第 i 类动作、第 j 组序列中第 k 帧的目标侧影图像。则由式(3), 样本均值为

$$\mathbf{M}^{(X)} = \frac{1}{L} \sum_{i=1}^c \sum_{j=1}^{N_i} \sum_{k=1}^{l_{i,j}} \mathbf{X}_{i,j,k} \quad (13)$$

由式(8)和式(9), 则可以很容易地计算出 TPCA 方法下的最优投影矩阵 \mathbf{U}, \mathbf{V} , 分别取 \mathbf{U}, \mathbf{V} 的前 l_1, l_2 个列向量组成的矩阵 \mathbf{U}', \mathbf{V}' 就是最终要求的子空间投影矩阵。

基于 TLPP 和 S-TLPP 的张量子空间学习相对复杂一些, 为避开子空间学习过程中额外参数选择的影响, 在子空间学习时采用简单的 0-1 策略和 K 最近邻域方法以及监督的方法来计算权值 S_{ij} , 即 S_{ij} 可以表示为

$$S_{ij} = \begin{cases} 1 & \mathbf{X}_i, \mathbf{X}_j \text{ 属于同一类动作} \\ 0 & \mathbf{X}_i, \mathbf{X}_j \text{ 不属于同一类动作} \end{cases} \quad (14)$$

$$\text{或 } S_{ij} = \begin{cases} 1 & \mathbf{X}_i \text{ 在 } \mathbf{X}_j \text{ 邻域中或 } \mathbf{X}_j \text{ 在 } \mathbf{X}_i \text{ 邻域中} \\ 0 & \text{其他} \end{cases} \quad (15)$$

采用式(14)计算权值 S_{ij} 的方法为基于监督的 TensorLPP 方法(简称为 S-TLPP); 采用式(15)计算权值 S_{ij} 的方法为基于非监督的 TensorLPP 方法(简称为 TLPP)。同时, 为度量两个数据点 \mathbf{X}_i 和 \mathbf{X}_j 之间的距离, 采用两者之间角度的 cosine 函数来度量其距离, 即

$$d_{ij} = \cos(\mathbf{X}_i, \mathbf{X}_j) = \frac{\mathbf{X}_i \cdot \mathbf{X}_j}{(|\mathbf{X}_i| \cdot |\mathbf{X}_j|)} \quad (16)$$

3.3 动作识别

行为识别问题可以简单地看作是时变特征数据的分类问题, 即将测试序列与预先标定的代表典型行为的参考序列进行匹配, 那么行为识别的关键问题之一是如何度量参考序列与测试序列之间的相似性。由于人动作执行的速度往往不一样, 而且不同的人其动作时空特征不同, 即便是同一人的同一种动作, 在不同的场合、时候也会不同, 所以在匹配行为序列时必须能够处理相似运动模式在空间和时间尺度上的变化。Hausdorff 距离为 2 个序列之间的相似性提供了一个很好的解决办法。用文献[5]描述的 Hausdorff 距离变形方法作为动作序列之间的相似性测量。假设 2 个动作序列分别为 $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{T_1}]$ 和 $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{T_2}]$, 其中 $\mathbf{a}_i, \mathbf{b}_j$ 分别是目标侧影图像在张量子空间 $\mathbf{R}^{l_1 \otimes l_2}$ 上的投影, T_1, T_2 分别是

两序列的长度。则 2 个动作序列之间的归一化欧氏距离可以定义为

$$d(\mathbf{A}, \mathbf{B}) = \frac{1}{T_1} \sum_{j=1}^{T_1} \min_{1 < i < T_2} \left\| \frac{\mathbf{a}_j}{\|\mathbf{a}_j\|} - \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|} \right\|^2 \quad (17)$$

为保证对称性, 定义两个序列之间的距离为

$$D(\mathbf{A}, \mathbf{B}) = d(\mathbf{A}, \mathbf{B}) + d(\mathbf{B}, \mathbf{A}) \quad (18)$$

不失一般性, 本文采用最近邻距离方法对人动作序列进行分类, 假设 \mathbf{T} 为一测试序列, \mathbf{R}_i 是第 i 类动作的参考序列, 则将与测试序列距离最小的已知模板所属类别 i 作为被测试序列 \mathbf{T} 的识别结果, 即

$$i = \operatorname{argmin}_j D(\mathbf{T}, \mathbf{R}_j) \quad (19)$$

4 实验分析

为验证本文方法的动作分类效果以及鲁棒性能, 设计了 2 个实验: (1) 动作识别实验。用前面提出的方法来识别不同类的人动作以验证其分类性能; (2) 鲁棒性实验。对目标有形变、遮挡以及不同视角下拍摄的动作序列进行分类实验以验证算法的鲁棒性。

4.1 动作识别实验

本文采用文献[11]和[15]使用的动作数据集^[16]进行动作分类实验, 该数据集由 9 人执行的 90 段动作序列组成(分辨率为 $180 \times 144, 25$ 帧/秒), 其中每人分别执行包括弯腰(bend)、挥手原地跳(jack)、蹦跳(pjump)、双脚向前跳(jump)、跑(run)、走(walk)、单脚跳(skip)、侧跑(side)、挥单手(wave1)和挥双手(wave2)等共 10 种不同的日常动作。

为简单起见, 从动作数据集^[16]获得目标前景掩码实现运动目标分割处理, 并将目标侧影归一化到同样大小(64×48), 同时根据视频中动作执行的次数和周期性, 将 90 段视频分割成 203 段, 使得分割后的每一段视频中包含一个完整的动作周期, 则分割后弯腰(bend)、挥手原地跳(jack)、双脚向前跳(jump)、蹦跳(pjump)、跑(run)、走(walk)、单脚跳(skip)、侧跑(side)、挥单手(wave1)和挥双手(wave2)等 10 种动作序列的样本数分别是 9, 23, 24, 27, 16, 19, 25, 22, 19 和 19。

由于数据集中样本较少, 为获得正确识别率的全面无偏估计, 采用留一法交叉检验方法来对动作进行分类。每次实验时, 首先从样本中挑出一段序

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jack	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pjump	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Side	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Skip	0.00	0.00	0.00	0.00	0.22	0.00	0.67	0.11	0.00	0.00
Walk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Wave1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(a) 混淆矩阵(TPCA)

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jack	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pjump	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.00	1.89	0.00	0.00	0.11	0.00	0.00
Side	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Skip	0.00	0.00	0.00	0.00	0.22	0.00	0.56	0.22	0.00	0.00
Walk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Wave1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(b) 混淆矩阵(TLPP, $K=20$)

	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jack	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Jump	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pjump	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Run	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Side	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Skip	0.00	0.00	0.00	0.00	0.11	0.00	0.89	0.00	0.00	0.00
Walk	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Wave1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

(c) 混淆矩阵(S-TLPP, $K=20$)

图 2 TPCA、TLPP 和 S-TLPP 3 种子空间方法下动作分类的混淆矩阵

Fig. 2 Confusion matrixes of action classification basing TPCA, TLPP and S-TLPP

出来的测试样本与训练样本之间的相似性,并根据序列之间的相似性采用最近邻距离框架分类测试序列的动作类型。图 2 给出了 TPCA、TLPP、S-TLPP 等 3 种不同方法下动作分类实验的混淆矩阵,图中每一行的值表示一种动作序列被分类成其他动作的概率,从图可以看出借助本文提出的算法,大多数动作序列具有极高的正确分类结果,只有少数的动作序列(如跑动作与单脚跳动作)被错分,造成错分的最可能原因是因为 2 种动作序列中某些帧的侧影形状之间具有较大的相似性。表 1 给出留一法交叉检验下的动作正确识别率。

为了进一步验证本文算法的有效性,我们采用本文 3 种算法以及 PCA、非监督的 LPP(LPP)和监督的 LPP(S-LPP)共 6 种方法进行统计分类性能比较。为避免同一人同种动作的影响,每次从总的 90 段样本序列中随机挑选出 50 段序列(保证每类动作至少有 2 段序列),用其所对应的已经分割好的序列作为训练样本,剩余的 40 段序列作为测试序列进行统计显著性分类测试,重复同样的分类实验 100 次(取置信度为 0.95)。其平均识别结果如表 1 所示,由表 1 可知:无论采用留一法交叉检验还是统计分类实验,本文方法的动作分类性能都比以前方法均有较大的改善,尤其是 TPCA 方法,由于不需要对所有样本求协方差矩阵,只需求对单个样本的方差矩阵,其子空间学习速度不仅比 PCA 方法快,而且分类效果要好。而 TLPP 和 S-TLPP 方法首先需要构造权值矩阵,并且式(11)和式(12)相互依赖,不能独立求解,需要通过迭代方法求解 U, V , 计算量取决于迭代次数,所以这两种方法计算消耗时间相对较长。

4.2 鲁棒性实验

为验证本文算法的鲁棒性,采用不同运动方式和衣着以及不同视角条件下拍摄的走动作的运动视频^[16]进行分类实验。这 20 段视频中 10 段是包含不同运动方式下(包括遮挡、不同衣着等)的视频,其余 10 段则是在不同视角(角度为运动方向与摄像机平面的夹角)下拍摄的视频(如图 3)。依旧用前述 6 种子空间方法对这 20 段走动作进行分类测试,实验结果如表 2 所示。由表可以看出,TPCA 和 S-TLPP 方法能够正确分类前 17 种不同条件下的动作,而 PCA、LPP 和 S-LPP 方法对有些序列不能正确分类。实验结果表明本文算法对动作分类具有较强的鲁棒性。

列作为测试样本,考虑到同一人的同一动作变化性很小,所以把被挑选的动作序列所在原始视频分割出来的其他动作序列也剔出,则将剩下所有序列作为训练样本用于张量子空间学习并将每帧中目标图像投影到低维子空间,然后在子空间中度量被挑选

表 1 不同子空间学习方法下动作正确识别率及子空间学习所耗时间比较

Tab. 1 Comparison of correct recognition rates and time consuming basing different subspace learning methods

	PCA $D = 20$	LPP $K = 20, D = 14$	S-LPP $D = 20$	TPCA $D_U = 5, D_V = 4$	TLPP $K = 20, D_U = 5, D_V = 4$	S-TLPP $D_U = 5, D_V = 4$
留一法	93.33%	92.22%	95.56%	96.67%	94.44%	98.89%
统计方法	90.40%	90.00%	93.27%	93.90%	92.63%	96.07%
[置信区间%]	[90.26, 90.54]	[89.91, 90.09]	[92.12, 92.43]	[93.83, 93.97]	[92.53, 92.73]	[96.86, 96.29]
耗时(s) (留一法)	19.062 5	269.984 4	229.234 4	2.640 6	584.859 4	569.328 1

图 3 不同条件下走动作的测试序列

Fig. 3 Test waking sequences under different conditions

表 2 不同子空间学习方法的鲁棒测试结果

Tab. 2 Result of robustness test

	PCA	LPP	S-LPP	TPCA	TLPP	S-TLPP
挥包	walk	walk	walk	walk	walk	walk
提包	walk	walk	walk	walk	walk	walk
遛狗	walk	run	walk	walk	walk	walk
高抬腿	walk	walk	walk	walk	walk	walk
跛行	walk	walk	walk	walk	walk	walk
蹦行	run	jack	run	walk	jack	walk
脚被遮挡	walk	walk	walk	walk	walk	walk
正常走	walk	walk	walk	walk	walk	walk
身体被遮挡	walk	walk	walk	walk	walk	walk
披衣服	walk	walk	walk	walk	walk	walk
角度 0°	walk	walk	walk	walk	walk	walk
角度 9°	walk	walk	walk	walk	walk	walk
角度 18°	walk	walk	walk	walk	walk	walk
角度 27°	walk	walk	walk	walk	walk	walk
角度 36°	walk	walk	walk	walk	walk	walk
角度 45°	run	walk	walk	walk	walk	walk
角度 54°	walk	walk	walk	walk	walk	walk
角度 63°	run	run	run	jack	side	side
角度 72°	jack	jack	jack	jack	jack	jack
角度 90°	jack	jack	jack	jack	jack	jack

注:角度为运动方向与摄像机图像平面的夹角

5 结 论

形状特征是人行为识别研究中最主要的特征之一,然而目前基于形状识别的方法大多是将 2 维图像展开成一个高维向量进行处理,但这种矩阵到向量的转换方法会丢失其空间信息。本文利用张量子空间方法在对高维图像数据进行有效降维的同时,又保持其像素之间的空间结构信息的特点,采用张量子空间学习方法对动作序列中人的静态二值侧影图像进行降维处理来识别人动作。实验结果表明了本文算法是有效的并且具有较强的鲁棒性。

尽管本文算法在动作识别上是有效的且对不同动作方式具有鲁棒性,但对视角变化并不稳定,而且对于 TensorLPP 方法计算量较大。另外,本文算法只在小的数据集上进行验证,基于大数据集的算法

评估与比较等亟待下一步工作。

参考文献 (References)

- 1 Gavrilu D M. The visual analysis of human movement: A survey [J]. Computer Vision and Image Understanding, 1999, **73**(1):82-98.
- 2 Hu Wei-ming, Tan Tie-niu, Wang Liang, *et al.* A survey on visual surveillance of object motion and behaviours [J]. IEEE Transactions on System Man and Cybernetics, Part C: Applications and Reviews, 2004, **34**(3):334-352.
- 3 Zhu Guang-yu, Xu Chang-sheng, Huang Qing-ming, *et al.* Action recognition in broadcast tennis video [A]. In: Proceedings of the 18th IEEE International Conference on Pattern Recognition [C], Hong Kong, 2006, **1**:251-254.
- 4 Bobick Aaron F, Davis James W. The recognition of human movement using temporal templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, **23**(3), 257-267.
- 5 Masoud O, Papanikolopoulos N. A method for human action recognition [J]. Image and Vision Computing, 2003, **21**(8):729-743.
- 6 Li Hong, Greenspan Michael. Multi-scale gesture recognition from time-varying contours [A]. In: Proceedings of the 10th IEEE International Conference on Computer Vision [C], Beijing, 2005:236-243.
- 7 Jin Ning, Farzin Mokhtarian. Human motion recognition based on statistical shape analysis [A]. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance [C], Como, Italy, 2005:4-9.
- 8 Veeraraghavan A, Roy-Chowdhury A K, Chellappa R. Matching shape sequences in video with applications in human movement analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, **27**(12):1896-1909.
- 9 Johansson G. Visual Perception of Biological Motion and a Model for Its Analysis [J]. Perception and Psychophysics, 1973, **14**(2):201-211.
- 10 Veeraraghavan A, Chowdhury Amit R, Chellappa R. Role of shape and kinematics in human movement analysis [A]. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C], Washington, DC, USA 2004, **1**:730-737.
- 11 Wang Liang, David Suter. Learning and Matching of Dynamic Shape Manifolds for Human Action Recognition [J]. IEEE Transactions on Image Processing, 2007, **16**(6), 1646-1661.
- 12 Cai Deng, He Xiao-fei, Han Jia-wei. Subspace learning based on tensor analysis [R]. 2005-2572, Illinois State, USA; University of Illinois at Urbana-Champaign, 2005.
- 13 He Xiao-fei, Cai Deng, Niyogi Partha. Tensor subspace analysis [A]. In: Proceeding of the Conference on Advances in the Neural Information Processing Systems [C], Vancouver, Canada, 2005: 249-254.
- 14 He Xiao-fei, Niyogi P. Locality preserving projections [A]. In: Proceeding of the Conference on Advance in Neural Information Processing Systems [C], Vancouver, Canada, 2003: 153-160.
- 15 Blank M, Gorelick L, Shechtman E, *et al.* Actions as space-time shapes [A]. In: Proceedings of the International Conference on Computer Vision [C], Beijing, 2005, **2**:1395-1402.
- 16 动作数据集 [EB/OL]. <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>.