

基于流形学习与隐条件随机场的人体动作识别*

刘法旺⁺, 贾云得

(北京理工大学 计算机科学技术学院 智能信息技术北京市重点实验室 100081)

Human Action Recognition Using Manifold Learning and Hidden Conditional Random Fields

LIU Fa-Wang⁺, JIA Yun-De

(Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China)

+ Corresponding author: E-mail: liufawang@bit.edu.cn, http://www.mcislab.org.cn

Liu FW, Jia YD. Human action recognition using manifold learning and hidden conditional random fields. *Journal of Software*, 2008,19(Suppl.):69-77. <http://www.jos.org.cn/1000-9825/19/s69.htm>

Abstract: This paper presents a probabilistic method of human action recognition based on manifold learning and Hidden Conditional Random Fields (HCRF). A supervised Neighborhood Preserving Embedding (NPE) is employed for dimensionality reduction by preserving the local neighborhood structure on the data manifold. Most existing approaches to action recognition use a Hidden Markov Model or suitable variant to model actions; a significant limitation of these models is the requirements of conditional independence of observations. In addition, generative models are selected to maximize the likelihood of generating all the examples of a given class and may not uncover the distinctive configuration that sets one class uniquely against others. HCRF relaxes the independence assumption and classifies actions in a discriminative hidden-state formulation. Experimental results on a recent database have demonstrated that this approach can recognize human actions accurately with temporal, intra- and inter-person variations even when noise and other factors such as partial occlusion exist.

Key words: action recognition; manifold learning; discriminative model; hidden conditional random fields

摘要: 提出了一种基于流形学习与隐条件随机场(hidden conditional random fields,简称 HCRF)的动作识别方法.算法提取人体剪影作为输入特征,采用有监督的保持邻域嵌入(neighborhood preserving embedding,简称 NPE)的子空间学习算法获得高维运动特征的低维流形表示,基于 HCRF 建模运动特征与动作语义之间的映射关系.在降维过程中,通过保持数据的局部邻接关系,NPE 可以获取动作特征在低维流形空间上的本质分布特性.与 HMM(hidden Markov model)等产生式模型相比,HCRF 侧重从样本数据中抽取共有特征以获取正确的分类边界,不需要假定观测过程条件独立,可以更加自然地对动作的时空邻域关系进行建模.实验结果表明,即便对于特征差异较大或存在噪声干扰的动作序列,算法也能取得较好的识别效果.

关键词: 动作识别;流形学习;判别式模型;隐条件随机场

人体动作识别可以广泛应用于智能人机交互、视频监控和数字娱乐等领域,是目前计算机视觉与模式识别领域的研究热点^[1,2].人体动作识别研究的主要问题有:(1) 动作特征维数较高,对这些高维向量进行存储和运算

* Supported by the National Natural Science Foundation of China under Grant No.60675021 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z120 (国家高技术研究发展计划(863))

Received 2008-05-01; Accepted 2008-11-25

的空间复杂度和运算复杂度较大,降维算法又难以保持原有特征的本质分布特性;(2) 动作具有长相关性(long-range dependence),识别算法需要结合上下文信息.

本文提出了一种基于流形学习与隐条件随机场(hidden conditional random fields,简称 HCRF)^[3]的动作识别方法.算法提取人体剪影作为输入特征,采用有监督的保持邻域嵌入(neighborhood preserving embedding,简称 NPE)^[4]的子空间学习算法对降维数据的局部空间邻接关系进行约束,在保持动作特征本质分布特性的前提下,将特征数据从高维空间投影到低维流形空间.HMM 是目前应用最为广泛的动作识别方法,但作为以满足模型解释要求为目的的产生式模型,HMM 难以学习得到区分动作类别的独特构型;此外,产生式模型需要假定观测过程条件独立,难以表征人体动作的长相关特性.本文使用具有隐变量结构的判别式模型 HCRF 对动作特征的时空邻域关系进行建模,实现动作识别.图 1 所示为算法的主要流程.

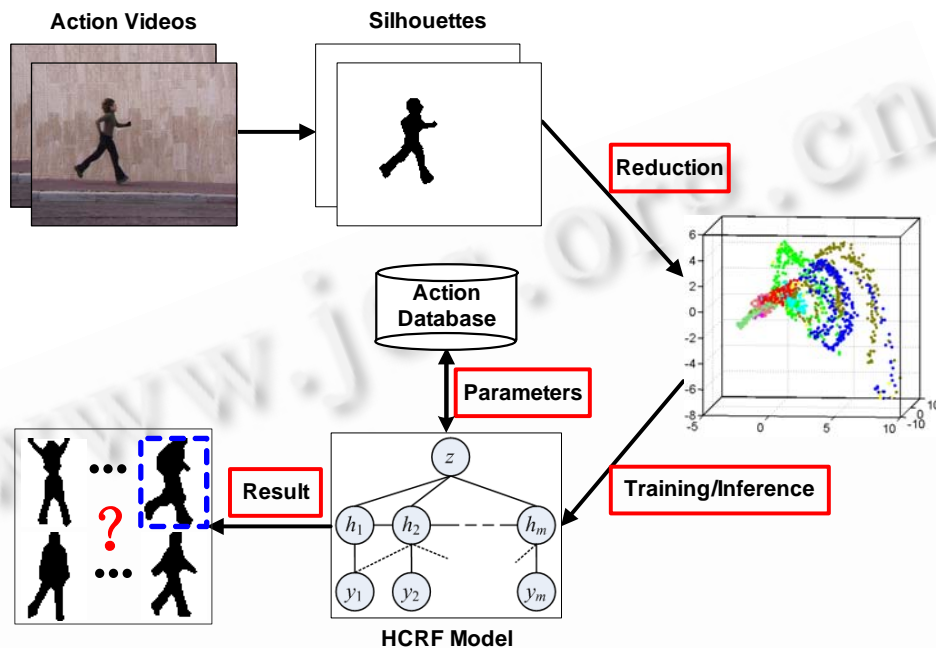


Fig.1 Flowchart of the proposed method of human action recognition

图 1 人体动作识别算法的流程图

1 相关研究

人体运动是非刚体运动,鲁棒的三维人体姿态跟踪非常困难,因此,光流^[5]、关键姿态^[6]、剪影^[7]、轨迹^[8]和包容形状^[9]等基于表观的特征被广泛采用.不过,光流法难以处理平滑表面和低分辨率的视频序列;关键姿态法没有考虑动作特性的时序变化信息;当动作类别逐渐增多时,基于剪影和二维轨迹进行动作识别时容易产生歧义;包容形状损失了一些视角变化的信息.本文采用剪影序列描述人体动作,剪影序列不仅包括每帧图像中人体剪影这样的空间信息,还包含剪影沿时序的变化信息.

常见的特征降维算法主要包括 PCA^[10],KPCA^[11],LLE^[13]和 Isomap^[14]等.Ahmad 和 Lee^[10]采用 PCA 对高维人体剪影特征进行降维,但 PCA 只是寻找线性重建误差最小意义下的最优子空间,旨在保存全局的欧式结构,存在局部最优问题.Sminchisescu 等人^[11]采用 kernel PCA(KPCA)将特征约束到低维的状态子空间.KPCA 是非线性的降维方法,但没有考虑数据在低维流形空间中的本质分布特性.与 PCA 和 KPCA 相比,NPE 具有非线性特性,有利于获取动作内在的局部流形结构.Elgammal 和 Lee^[12]将 LLE^[13]和 Isomap^[14]应用到姿态估计的低维动作流形学习中.LLE 和 Isoma 可以保持数据之间的本质结构,但计算代价较高,而且只基于训练数据产生映射关系,

对于训练集之外数据的映射关系尚无定论.因此,Elgammal 和 Lee^[12]需要先学习得到高维图像特征的低维流形表示,然后再获取由高维图像特征到低维流形特征的映射关系.与 LLE 和 Isomap 相比,NPE 可以同时学习动作的低维流形表示以及由高维图像特征空间到低维流形空间的映射关系.

在动作识别算法中,基于全局时空特征的方法和基于时序建模的方法占据着主导地位.全局时空特征法把整个动作序列表示成特征向量,采用简单的度量规则进行识别.例如,Bobick 和 Davis^[15]将每个动作描述成二值运动能量图(motion energy image,简称 MEI)和二值运动历史图(motion history image,简称 MHI),然后基于时间模板进行动作识别.Weinland 等人^[16]引入运动历史量(motion history volumes,简称 MHV)对人体动作进行自由视角描述,提出通过对不同动作的 MHV 进行对齐和比较来学习和识别基本动作.Mokhber 等人^[17]将运动人体轮廓表示成二元材积(binary volume),采用三维几何距向量对其进行描述,最后使用最近邻分类器基于归一化几何距的马氏距离进行识别.Blank 等人^[7]则把整个序列的剪影表示成时空形状,通过求解时空形状的泊松方程获得如局部时空显著性、形状方向等运动特征,然后基于全局特征向量来描述行为.基于全局时空特征的方法便于获取整体运动信息,但常常需要对运动时长进行限定,容易受噪声干扰,多应用在相对简单的场景.

时序建模方法将动作序列看作是局部特征的时序集合,采用相对复杂的算法进行动作识别.由于 HMM 能够有效地对动作的时空特征进行建模,HMM 及其变化模型被广泛地应用于人体动作识别.Kapoor 和 Piacard^[18]提出了一种基于 HMM 的点头和摇头检测算子.Nguyen 等人^[8]根据人体运动轨迹,采用分层 HMM 模型进行动作检测和识别.不过,为了枚举所有可能的观测序列,HMM 等产生式模型需要假定观测过程条件独立,难以表征人体动作的长相关特性.此外,产生式模型采用联合分布模型来解决条件变分问题,实际上也就是关注如何对既有的观测序列进行建模.如果隐含的产生式模型十分复杂,而各个类别之间的边界非常明显,那么通过产生式模型进行动作识别就体现不出优势.

近年来,条件随机场模型受到了越来越多研究人员的关注.Sminchisescu 等人^[19]采用 CRF 进行动作识别,实验表明,他们的模型能够比 MEMM 取得更好的识别效果.Torralla 等人^[20]提出了 Boosted Random Fields,融合全局和局部图像信息进行物体识别.Wang 和 Suter^[21]采用 KPCA 降维动作特征,基于 factorial CRF 进行动作建模.Quattoni 等人^[3]采用 HCRF 进行人体动作识别.作为一种判别式模型,HCRF 将隐变量引入到条件随机场中,对隐变量和观测建立联合分布.HCRF 并不假定局部变量赋值是完全可以观测的,而是把它们作为隐状态变量.通过引入隐变量,HCRF 增强了对随机变量为结构型数据时的处理能力.当图像序列相邻帧之间存在时空邻域约束关系时,利用 HCRF 对结构型数据的处理能力,可以更加自然地在这种时空邻域关系进行建模.

2 特征提取与表示

图像中跟踪目标的形状位置信息会随着目标运动、目标与视觉系统的距离等发生变化,因此,需要对提取到的剪影序列进行归一化处理.对于一段 m 帧的视频 V ,假定提取的剪影序列为 $o = \{o_1, o_2, \dots, o_m\}$,将所有剪影平移到图像中心,在保持纵横比不变的前提下对整个序列图像进行适当缩放,提取图像中心大小为 $r \times c$ 的图像块作为新的观测序列 $o' = \{o'_1, o'_2, \dots, o'_m\}$.图 2 所示为归一化处理示例,图中第 1~3 行分别对应原始图像、提取的剪影图像和归一化的剪影图像.对归一化图像按行扫描,将 o'_i 表示为 $\mathbb{R}^{r \times c}$ 空间中的向量 x_i ,整个观测序列也就相应可以表示为 $x = [x_1, x_2, \dots, x_m]$.

本文采用有监督的 NPE 流形学习算法进行降维.对于给定的序列 $x = [x_1, x_2, \dots, x_m]$,NPE 算法可以获得低维流形特征 $y = [y_1, y_2, \dots, y_m]$,由 x_i 到 y_i 的映射表示为 $y_i = A_{npe}^T x_i$,其中 $x_i \in \mathbb{R}^D$, $y_i \in \mathbb{R}^d$, $D = r \times c$, $d \ll D$, A_{npe} 是 $D \times d$ 维的转换矩阵.具体流形学习过程见算法 1.

算法 1. 动作流形学习算法.

(1) 构造邻接图

假定 G 表示具有 m 个节点的图, t 和 s 对应图像在动作序列中的序号,按照如下方式构造邻接图:

(a) 如果 x_t 和 x_s 属于同一动作类别,计算二者之间的欧氏距离 $\text{dist}(t, s) = \|x_t - x_s\|$;否则, $\text{dist}(t, s) = C$, C 是预定义常量;

(b) 如果 x_s 位于 x_i 的 k -近邻范围内,在 x_i 到 x_s 之间建立有向连接线.

(2) 计算权重矩阵

假设每个数据由其时序邻近点的线性组合重构而成,在满足约束 $\sum_s W_{is} = 1$ 的前提下最小化目标函数 $\sum_i \|x_i - \sum_s W_{is} x_s\|^2$, 得到最优的代表局部邻近线性关系的权重矩阵 W . 其中, W_{is} 表示 x_i 由其邻近点 x_s 重构的系数,概括了 x_i 和 x_s 之间的空间近邻关系.

(3) 计算投影矩阵

在选择投影矩阵时,最小化成本函数 $\Phi(Y) = \sum_i (y_i - \sum_s W_{is} y_s)^2 = a^T X M X^T a$, $M = (I - W)^T (I - W)$, I 是单位矩阵.为了消除投影中的随机尺度因子,施加约束 $y^T y = 1 \Rightarrow a^T X X^T a = 1$. 转换向量 a 是通过求解广义特征方程 $X M X^T a = \lambda X X^T a$ 的最小特征值得到.假定列向量 a_1, a_2, \dots, a_d 是按照特征值 $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_d$ 排序的对应解,映射关系表示为 $x_i \rightarrow y_i = A_{npe}^T x_i$, $A_{npe} = (a_1, a_2, \dots, a_d)$ 是 $D \times d$ 矩阵.

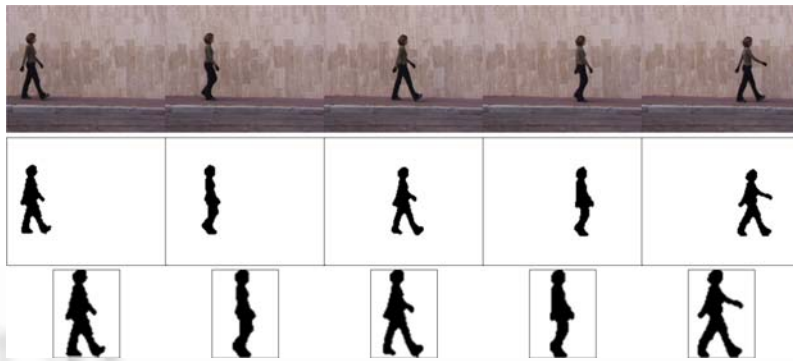


Fig.2 Examples of silhouette extraction and normalization

图2 人体剪影提取与归一化处理示例

图3给出了对 Blank 等人^[7]数据库的流形学习结果.归一化处理时提取图像中心大小为 48×80 的图像块作为处理结果.流形学习算法根据 10-近邻构建邻接图.通过流形降维,图像特征由 3 840 维降至 6 维.为了便于描述,图3仅选择前三维用于输出显示.

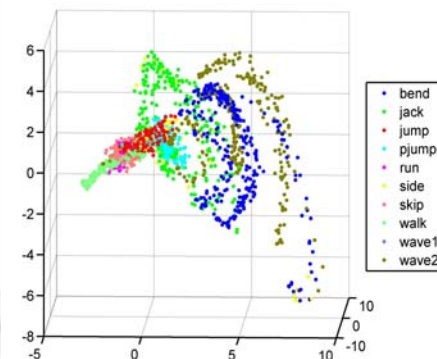


Fig.3 Low-Dimensional representation of action features in the manifold space

图3 动作特征在流型空间中的低维表示

3 动作建模和识别

HMM 等产生式模型是以满足模型解释要求为目的,旨在通过联合概率函数模拟随机观测数据和隐藏状态之间的关系;判别式模型是以识别为学习目的,侧重从样本数据中抽取共有特征以获取正确的分类边界.本文采

用 HCRF 判别式模型对人体动作序列进行建模和识别.

3.1 HCRF模型

HCRF 是判别式模型,可以根据输入序列 y 推断动作类别 z . y 是局部观测向量 $y = \{y_1, y_2, \dots, y_m\}$, 每个局部观测 y_i 对应一个特征向量. $z \in Z$, Z 表示动作类别集合. 对于任意观测序列 y , 存在对应向量 $h = \{h_1, h_2, \dots, h_m\}$, 其中 $h_i \in H$, H 表示所有可能出现的隐藏状态集合.

对于给定的观测数据, HCRF 按照如下方式对动作类别的条件概率进行建模:

$$P(z | y, \theta, \omega) = \sum_h P(z, h | y, \theta, \omega) = \frac{\sum_h e^{\Psi(z, h, y; \theta, \omega)}}{\sum_{z' \in Z, h \in H^m} e^{\Psi(z', h, y; \theta, \omega)}} \quad (1)$$

其中, 参数为 θ 和 ω 的势能函数 $\Psi(z, h, y; \theta, \omega)$ 表示隐藏状态结构、观测数据和动作类别之间的相互影响程度, ω 是窗口大小.

$$\Psi(z, h, y; \theta, \omega) = \sum_{(j,k) \in E} \theta_e [z, h_j, h_k] + \sum_{j=1}^m \theta_z [z, h_j] + \sum_{j=1}^m \varphi(y, j, \omega) \cdot \theta_h [h_j] \quad (2)$$

图 E 是链状结构, 每个点都对应 t 时刻的一个隐藏状态, $\varphi(y, j, \omega)$ 可以表示观测窗口中的任意特征.

3.2 训练与识别

参数向量 θ 由 3 部分组成, $\theta = [\theta_e, \theta_z, \theta_h]$. θ_e 度量的是连续状态 j 和 k 与动作类别 z 之间的兼容度, θ_z 度量的是隐藏状态 j 与动作类别 z 之间的兼容度, θ_h 表示对应状态 $h_j \in H$ 的参数 θ_h . 训练中, 采用目标函数:

$$L(\theta) = \sum_{i=1}^n \log P(z_i | y_i, \theta, \omega) - \frac{1}{2\sigma^2} \|\theta\|^2 \quad (3)$$

其中, n 表示训练样本序列的总个数.

给定测试序列 y 和学习得到的参数 θ^* , 测试序列 y 的动作类别由如下方式确定:

$$\tilde{z} = \arg \max_{z \in Z} P(z | y, \omega, \theta^*) \quad (4)$$

4 实验结果与分析

本文采用 Blank 等人^[7]的数据库对算法效果进行验证. 该数据库共计包含 93 个低分辨率的视频序列 (180×140 像素, 25 帧/秒), 动作包括 bend, jack, jump-forward-on-two-legs(jump), jump-in-place-on-two-legs(pjump), run, gallop-sideways(side)等. 所有动作分别由 9 名实验人员完成. 如图 4 所示, 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.

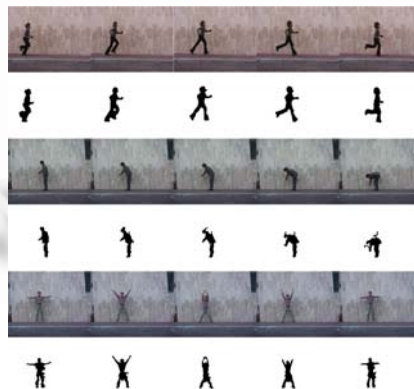


Fig.4 Example images and the corresponding silhouettes extracted from the action dataset

图 4 数据库的示例图像及对应的人体剪影

4.1 动作识别

采用“循环调度(round-robin scheduling)”方式验证算法的识别效果.实验提取 171 个动作序列组成动作数据库,然后将所有序列分别划分到 9 个分离的集合中,每个集合包含 1 个 bend 动作实例和其余每种动作的 2 个动作实例.循环过程中,依次选择一个集合作为测试数据,然后基于其他 8 个集合训练一个“多类分类(multi-class)”的 HCRF 模型,最后根据训练得到的 HCRF 模型进行动作识别测试,选择概率最高的动作类别为识别结果.图 5 所示为当窗口参数 $\omega=1$ 、隐藏状态数等于 10 时的动作识别结果统计.由图 5 可知,只有少量的 run,skip 和 walk 动作出现了识别错误.

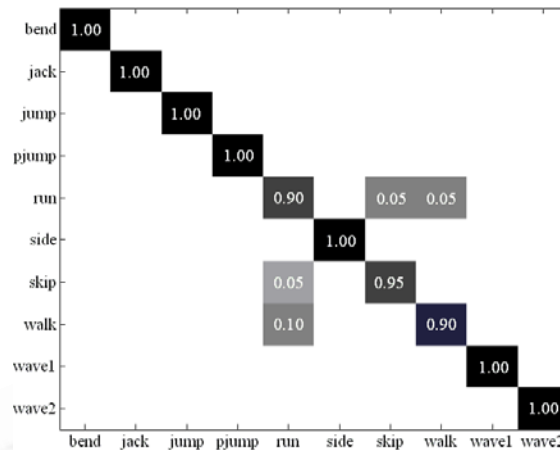


Fig.5 Confusion matrix for action recognition when $\omega=1$ and the number of hidden states is 10

图 5 当窗口参数 $\omega=1$ 且隐藏状态个数等于 10 时的混淆矩阵

4.2 算法对比

为了测试不同算法和模型结构对识别结果的影响,在相同数据库上对多重判别分析(multiple discriminant analysis,简称 MDA),HMM,CRF,HCRF 等不同算法的识别效果进行了对比,同时也对不同程度长相关的 CRF 模型和 HCRF 模型的识别效果进行了比较.

多重判别分析:(1) 通过线性插值,将所有动作序列规整为 N 帧,并将流形降维后的动作特征表示为 $f = [y_1^T, y_2^T, \dots, y_N^T]^T$, $f \in \mathbb{R}^{N \times d}$; (2) 使用多重判别分析寻找最小均方意义下,最能够分开各类动作特征的投影矩阵 A_{mda} ; (3) 计算投影后各类动作特征的中心位置; (4) 根据测试序列投影结果到各类动作特征中心位置的最小距离判定动作类别.

HMM 模型:为每类动作训练一个 HMM 模型.每个 HMM 模型具有 6 个隐藏状态和 1 个单高斯观测模型.识别测试时,分别在每类动作模型上进行动作识别,选择概率最大的那个模型类别为识别结果.

CRF 模型:训练一个线性链状 CRF 模型,每类动作对应一种状态.在训练过程中,状态序列 $\{z_1, z_2, \dots, z_m\}$ 定义为动作类别 z 的 m 次常量.对于所有的测试序列,CRF 模型为序列中每帧图像预测状态,最后根据状态序列中出现次数最多的状态判定动作类别.

HCRF 模型:分别采用“一对多(one-vs.-all)”和“多类分类(multi-class)”两种模式训练 HCRF 模型.在“一对多”实验中,为每类动作训练一个 HCRF 模型,每个模型包含 6 个隐藏状态;识别测试时,分别向每类模型输入测试序列,选择概率最大的那个模型类别为识别结果.在“多类分类”实验中,训练一个具有 10 个隐藏状态的 HCRF 模型;识别测试时,选择概率最大的模型类别为识别结果.

表 1 对比了在给定数据库上,利用不同算法模型进行动作识别的实验结果.由表 1 可知:(1) HMM 模型、CRF 模型和 HCRF 模型的识别效果普遍优于 MDA.这可能是因为 MDA 没有考虑动作特征的时序变化信息;(2) CRF 模型和 HCRF 模型的识别率普遍高于 HMM 模型.这从一定程度上验证了对有限动作集合进行动作识别时,判别式模型要优于产生式模型;(3) 适当地加入长相关约束有利于提高算法的识别率,但是识别率并不会随着窗

口参数 ω 的不断增加而相应地一直提高.例如,实验中将参数 ω 由 0 提高到 2 时,HCRF 模型的识别率会相应提高,但是进一步加大 ω 并没有提高识别效果;(4)“多类分类”模式训练的 HCRF 模型的识别效果要优于“一对多”模式训练的 HCRF 模型,采用联合训练的方式学习判别式结构是有益的.

Table 1 Comparisons of action recognition performance
表 1 不同模型的动作识别率对比

Models	Average accuracy (%)
MDA	79.4
Hidden Markov model	82.9
CRF ($\omega = 0$)	85.3
CRF ($\omega = 1$)	88.2
HCRF (one-vs.-all, $\omega = 0$)	90.6
HCRF (multiclass, $\omega = 0$)	92.3
HCRF (multiclass, $\omega = 1$)	97.1
HCRF (multiclass, $\omega = 2$)	97.6
HCRF (multiclass, $\omega = 3$)	95.3

4.3 鲁棒性测试

为了测试遮挡、动作形变等因素对识别结果的影响,对 9 种复杂背景下的人体动作进行鲁棒性测试.图 6 列举了一些复杂背景下的动作图像和相应的人体剪影,由上至下、从左到右分别对应于动作 Walking in 45° , Swinging a bag, Walking with a dog, Knees up, Limp, Moonwalk, Occluded legs, Normal walk 和 Walking in a skirt.

表 2 为不同场景下算法的鲁棒性测试结果.由此表可知,在 9 种复杂背景下有 6 种动作的识别结果是正确的.这在一定程度上证明了即便受遮挡、动作形变等因素的影响,算法仍然能够取得较好的识别效果.

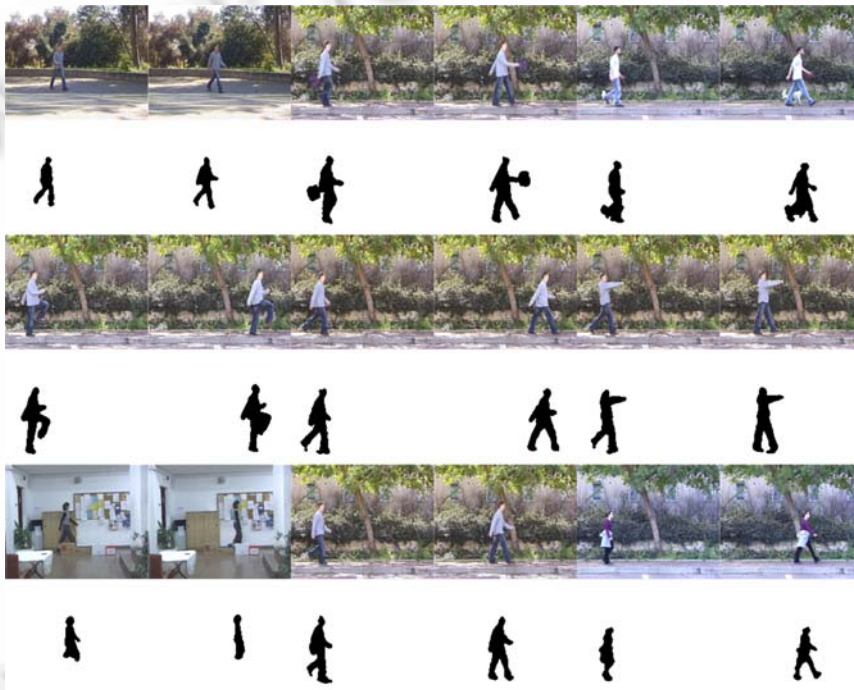


Fig.6 Example images and the corresponding silhouettes for robustness test
图 6 算法鲁棒性测试采用的图像及对应的人体剪影

Table 2 Robustness evaluation with respect to other factors
表 2 算法鲁棒性测试的实验结果

Test sequences	The best match
Walking in 45°	Walk
Swinging a bag	Walk
Walking with a dog	Walk
Knees up	Skip
Limp	Walk
Moonwalk	Skip
Occluded legs	Pjump
Normal walk	Walk
Walking in a skirt	Walk

5 结 论

本文提出了一种基于人体剪影序列的动作识别算法,算法采用有监督的 NPE 获得高维运动特征的低维流形表示,基于 HCRF 模型进行动作识别.实验表明,即便对于特征差异较大或存在噪声干扰的动作序列,算法仍然能够取得较好的识别效果.此外,考虑到在基于模型的人体行为分析过程中,模型结构通常是由研究人员根据个人经验定义,很少考虑动作本身的结构特性,这难免会对行为分析结果产生负面影响.本文对 HMM,CRF,HCRF 等不同模型的实验结果进行了对比分析,结果表明,对于有限动作集合进行动作识别时,判别式模型要优于产生式模型,适当加入长相关约束和采用联合训练方式学习判别式结构有助于提高算法的识别率.基于这些认识,未来将会在如何合理选择和建立模型方面展开更加深入的研究.

致谢 向对本文工作给予热情帮助的吴心筱、赵友东等同学表示衷心的感谢.

References:

- [1] Moeslund TB, Hilton A, Kruger V. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006,104(23):90–126.
- [2] Wang L, Hu WM, Tan TN. A survey of visual analysis of human motion. *Chinese Journal of Computers*, 2002,25(3):353–360 (in Chinese with English abstract).
- [3] Quattoni A, Wang S, Morency LP, Collins M, Darrell T. Hidden conditional random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007,29(10):1848–1852.
- [4] He X, Cai D, Yan S, Zhang H. Neighborhood preserving embedding. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2005, 2:1208–1213.
- [5] Efros A, Berg A, Mori G, Malik J. Recognizing action at a distance. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2003, 2:726–733.
- [6] Carlsson S, Sullivan J. Action recognition by shape matching to key frames. In: *Proc. of the IEEE CS Workshop on Models Versus Exemplars in Computer Vision*. 2002. 263–270.
- [7] Gorelick L, Blank M, Shechtman E, Irani M, Basri R. Action as space-time shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007,29(12):2247–2253.
- [8] Nguyen N, Phung D, Venkatesh S, Bui H. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov models. In: *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2005.
- [9] Huang FY, Xu GY. Viewpoint independent action recognition. *Journal of Software*, 2008,19(7):1623–1634 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/19/1623.htm>
- [10] Ahmad M, Lee SW. HMM-Based human action recognition using multiview image sequences. In: *Proc. of the IEEE Int'l Conf. on Pattern Recognition*. 2006,1:263–266.
- [11] Sminchisescu C, Kanaujia A, Li Z, Metaxas D. Conditional visual tracking in kernel space. In: *Advances in Neural Information Processing Systems*. 2006.

- [12] Elgammal A, Lee CS. Inferring 3D body pose from silhouettes using activity manifold learning. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2004,2:681-688.
- [13] Roweis S, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323-2326.
- [14] Tenenbaum JB, Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290(5500):2319-2323.
- [15] Bobick A, Davis J. The recognition of human movement using temporal templates. IEEE Trans. on Pattern Recognition and Machine Intelligence, 2001,23(3):257-267.
- [16] Weinland D, Ronfard R, Boyer E. Weinland D, Ronfard R, Boyer E. Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding, 2006,104(2):249-257.
- [17] Mokhber A, Achard C, Milgram M. Recognition of human behavior by space-time silhouette characterization. Pattern Recognition Letters, 2008,29(1):81-89.
- [18] Kapoor A, Picard R. A real-time head nod and shake detector. In: Proc. of the Workshop on Perspective User Interfaces. 2001.
- [19] Sminchisescu C, Kanaujia A, Li Z, Metaxas D. Conditional models for contextual human motion recognition. In: Proc. of the 10th IEEE Int'l Conf. on Computer Vision, Vol.2. 2005. 1808-1815.
- [20] Torralba A, Murphy K, Freeman W. Contextual models for object detection using boosted random fields. In: Advances in Neural Information Processing Systems. 2004.
- [21] Wang L, Suter D. Recognizing human activities from silhouettes: Motion subspace and factorial discriminative graphical model. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2007. 1-8.

附中文参考文献:

- [2] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述.计算机学报,2002,25(3):353-360.
- [9] 黄飞跃,徐光佑.视角无关的动作识别.软件学报,2008,19(7):1623-1634. <http://www.jos.org.cn/1000-9825/19/1623.htm>



刘法旺(1981-),男,河南信阳人,博士生,
主要研究领域为模式识别,计算机视觉.



贾云得(1962-),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为计算机视觉,媒体计算,智能人机交互.