

时空特征局部保持的运动视频关键帧提取

石念峰¹, 侯小静^{1*}, 张平²

(1. 洛阳理工学院 计算机与信息工程学院, 河南 洛阳 471023; 2. 河南科技大学 数学与统计学院, 河南 洛阳 471023)

(* 通信作者电子邮箱 alisha@163.com)

摘要:为提高运动视频关键帧的运动表达能力和压缩率,提出柔性姿态估计和时空特征嵌入结合的运动视频关键帧提取技术。首先,利用人体动作的时间连续性保持建立具有时间约束限制的柔性部件铰接人体(ST-FMP)模型,通过非确定性人体部位动作连续性约束,采用N-best算法估计单帧图像中的人体姿态参数;接着,采用人体部位的相对位置和运动方向描述人体运动特征,通过拉普拉斯分值法实施数据降维,获得局部拓扑结构表达能力强的判别性人体运动特征向量;最后,采用迭代自组织数据分析技术(ISODATA)算法动态地确定关键帧。在健美操动作视频关键帧提取实验中,ST-FMP模型将柔性混合铰接人体模型(FMP)的非确定性人体部位的识别准确率提高约15个百分点,取得了81%的关键帧提取准确率,优于KFE和运动块的关键帧算法。所提算法对人体运动特征和人体姿态敏感,适用于运动视频批注审阅。

关键词:关键帧提取;运动视频;姿态估计;柔性混合铰接人体模型;特征选择

中图分类号:TP391.4 **文献标志码:**A

Key frame extraction of motion video based on spatial-temporal feature locally preserving

SHI Nianfeng¹, HOU Xiaojing^{1*}, ZHANG Ping²

(1. School of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang Henan 471023, China;

2. School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang Henan 471023, China)

Abstract: To improve the motion expression and compression rate of the motion video key frames, a dynamic video frame extraction technique based on flexible pose estimation and spatial-temporal feature embedding was proposed. Firstly, a Spatial-Temporal feature embedded Flexible Mixture-of-Parts articulated human model (ST-FMP) was designed by preserving the spatial-temporal features of body parts, and the N-best algorithm was adopted with spatial-temporal locally preserving of uncertain body parts to estimate the body configuration in a single frame based on ST-FMP. Then, the relative position and motion direction of the human body were used to describe the characteristics of the human body motion. The Laplacian scoring algorithm was used to implement dimensionality reduction to obtain the discriminant human motion feature vector with local topological structure. Finally, the ISODATA (Iterative Self-Organizing Data Analysis Technique) algorithm was used to dynamically determine the key frames. In the key frame extraction experiment on aerobics video, compared to articulated human model with Flexible Mixture-of-Parts (FMP) and motion block, the accuracy of uncertain body parts by using ST-FMP was 15 percentage points higher than that by using FMP, achieved 81%, which was higher than that by using Key Frames Extraction based on prior knowledge (KFE) and key frame extraction based on motion blocks. The experimental results on key frame extraction for calisthenics video show that the proposed approach is sensitive to motion feature selection and human pose configuration, and it can be used for sports video annotation.

Key words: key frame extraction; motion video; pose estimation; articulated human model with Flexible Mixture-of-Parts (FMP); feature selection

随着混合式教学和MOOC(Massive Open Online Course)等在线学习技术在体育教学中的应用^[1],产生了大量的学生动作长视频,通常采用基于视频关键的视频批阅技术提高学生动作长视频的评价效率^[2-3]。在动作视频批阅时,教师依据人体部位的位置、形状等信息预测学生动作的运动轨迹和运动趋势,判断视频中是否存在错误或者不规范动作。因此,它要求关键帧集合帧数要尽可能少,同时人体动作局部拓扑结构表达要更准确。

关键帧技术是模式识别领域的研究热点,广泛地应用在

运动捕获、人体行为及动作识别等方面^[4-5]。但由于运动视频的高复杂性和非线性特征,至今尚未通用的关键帧算法。文献[6]通过相邻视频间颜色直方图的熵值差异比较,借助特定阈值筛选出高压缩率的关键帧集合;但由于需要事先设定阈值,所以当视频中动作变化剧烈时容易造成关键帧冗余或者遗漏,自适应性差。文献[7]采用基于核与局部信息的多维度模糊C均值聚类权衡图像的噪声和细节,自适应性筛选关键帧;但是由于缺乏时空约束,筛选的关键帧集合的运动时序表达能力较差,不适合本文的视频批阅。文献[8]将视

收稿日期:2017-04-25;修回日期:2017-06-10。 基金项目:河南省科技攻关项目(152102210329, 172102310635)。

作者简介:石念峰(1976—),男,河南洛阳人,副教授,博士,CCF高级会员,主要研究方向:计算机协同工作、模式识别; 侯小静(1975—),女,河南洛阳人,讲师,硕士,主要研究方向:模式识别; 张平(1976—),男,黑龙江牡丹江人,副教授,博士,主要研究方向:网络安全、模式识别。

频分割成运动物体和运动背景,通过分析物体运动和形状变化,采用无监督聚类算法确定关键帧;由于从语义层面分析和理解原始视频,所以提取的关键帧集合紧致且运动特征表达准确,能够满足本文的视频批阅需要。

采用文献[8]进行关键提取时通常需要根据应用背景设计物体识别及运动特征描述模型,因此如何建立运动视频中的人体运动模型是关键。文献[4]采用 3D 卷积神经网络(Convolutional Neural Network, CNN),通过执行 3D 卷积从多个连续视频帧中提取空间和时间维度上的运动特征,借助多通道信息融合确定关键帧,较好地保持了人体动作的时空特征,可以准确识别人体动作。然而,本文研究的运动视频中通常存在大量错误或者不规范动作,导致动作特征变化较大,所以这种人体运动特征模型不适合运动视频关键帧提取。文献[9]采用梯度方向直方图(Histogram of Gradients, HoG)人体分类器确定图像中的人体包围盒,利用人体模板将人体包围盒划分为 16 个不同权重的运动区域,通过比较运动区域的运动方向差异确定关键帧。由于受人体分类器的比例和人体模板尺寸限制,在复杂场景或者运动景深变化剧烈时,容易造成人体动作识别错误,降低关键帧提取准确率。

综上所述,迫切需要一种自适应、能准确反映人体动作特征的运动视频关键帧提取技术。考虑到刚性人体模型的姿态识别方法的良好性能^[10-11],本文提出在柔性混合铰接人体模型(articulated human model with Flexible Mixture-of-Parts, FMP)嵌入人体部位运动时空特征,提高人体及动作识别的鲁棒性,利用人体姿态参数及动作特征确定运动视频关键帧。

1 理论模型

1.1 嵌入时空特征的人体姿态估计模型

在 FMP 模型中,每个人体部位的不同仿射变形(如旋转或弯曲)被称为该部位的混合类型,简称混合类型^[10]。由于同一个人体部位对应若干个混合类型,所以一幅图像 I 中的人体姿态参数由各个部位的位置信息及其混合类型共同确定。一般用一个 K 关系图 $G = (V, E)$ 描述 I 中的某个姿态,其中顶点集合 V 表示人体部位(如头、上肢和躯干等),边集合 $E \subseteq V \times V$ 表示不同人体部位之间的一致性约束关系。

根据 FMP 定义,一幅图像 I 中人体姿态 p 的参数估计问题可形式化为成本最小化问题,其成本函数 $C(I, p)$ 为:

$$C(I, p) \propto \sum_{u \in V} \varphi_u(I, p^u) + \sum_{(u, v) \in E} \psi_{u, v}(p^u - p^v) \quad (1)$$

其中: $\varphi_u(I, p^u)$ 是一个外观模型,表示在图像 I 的位置 p^u 处识别出人体部位 u 的成本; $\psi_{u, v}(p^u - p^v)$ 是一个变形模型(通常假设为弹簧能量模型),表示两个人体部位 u 和 v 之间的变形成本。

当采用 FMP 模型估计运动视频中人体姿态时,为保持人体姿态参数的时间和空间连续性,本文在相邻帧 I_t 和 I_{t+1} 中的相同人体部位 p_t^u 和 p_{t+1}^u 之间增加了一个时序特征边,形成具有时空特征的柔性混合铰接人体模型(Spatial-Temporal feature embedded Flexible Mixture-of-Parts articulated human model, ST-FMP)。在 ST-FMP 中,时序特征边限定的人体姿态连续误差采用 p_t^u 和 p_{t+1}^u 之间的光流差异来计算,即:

$$\theta(p_t, p_{t+1}, I_t, I_{t+1}) = \sum_{u \in V} \|p_{t+1}^u - p_t^u - f(p_t^u)\|_2^2 \quad (2)$$

其中: $f(p_t^u)$ 是在 p_t^u 处估计的从 I_t 到 I_{t+1} 的光流值。

假设一段运动视频的帧图像集合为 $I = \{I_1, I_2, \dots, I_T\}$,

估计的姿态参数序列为 $P = \{p_1, p_2, \dots, p_T\}$, 那么采用 ST-FMP 模型从 I 中得到 P 的成本为:

$$C(I_T, p_T) + \sum_{t=1}^{T-1} C(I_t, p_t) + \lambda_1 \theta(p_t, p_{t+1}, I_t, I_{t+1}) \quad (3)$$

其中: $C(I_t, p_t)$ 表示根据式(1)得到的从图像 I_t 中估计人体姿态 p_t 的成本, T 是运动视频帧数, λ_1 是一个规范化常量, $\theta(\cdot)$ 是式(2)表达的时空连续性误差。

1.2 基于非确定性部位优化的 ST-FMP 求解

FMP 是一个刚性人体模型,可用一个马尔可夫随机场(Markov Random Field, MRF)表示,通过机器学习方法来确定人体部位参数^[12]。在单帧图像中采用 FMP 估计人体姿态参数时, MRF 被看作树状或星形图结构,通过置信度传播(Belief Propagation, BP)进行求解^[10]。引入时间约束后, ST-FMP 中会产生大量回路,需采用循环置信度传播(Loopy Belief Propagation, LBP)等近似算法通过最小化式(3)求解。然而 LBP 算法是图的最大团问题,具有指数级复杂度,在进行长视频姿态估计时效率很低。因此,本文设计了一种基于非确定性人体部位的两段式 ST-FMP 求解算法。

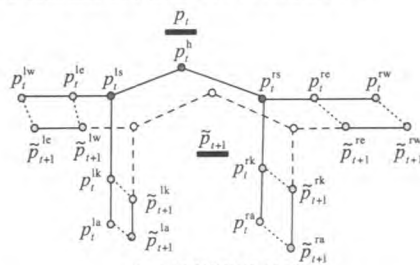
第一步,产生候选人体姿态集合。首先不考虑人体姿态的时空连续性约束,利用文献[12]的 N-best 算法和式(1)在 $O(K_b^2 T)$ 时间内从单帧图像中产生 K_b 个人体姿态集合。由于运动模糊和自遮挡等原因, K_b 中的某些人体部位,如图 1(a) 中 8 个白色点标识的肘部(le, re)、手腕(lw, rw)、膝盖(lk, rk)和踝关节(la, ra)等(称为非确定性部位,记作 \tilde{p}_{t+1}),很难估计准确。为此,本文将 I_{t+1} 中的非确定性部位 \tilde{p}_{t+1} 引入 I_t 的姿态估计中,借助人体部位局部时间连续性保持(图 1(a) 中用四条虚线边表示)提高人体姿态估计的精度,则式(3)即修改为:

$$C(I_t, p_t) + \tilde{C}(I_{t+1}, \tilde{p}_{t+1}) + \bar{\lambda}_1 \sum_{u \in W} \|\tilde{p}_{t+1}^u - p_t^u - f(p_t^u)\|_2^2 \quad (4)$$

其中: $W \subset V$ 代表非确定性部位的集合; $\bar{\lambda}_1$ 是一个规范化常量;除了 $\tilde{C}(\cdot)$ 仅计算 W 中非确定性部位的外观成本和变形成本外, $\tilde{C}(\cdot)$ 和式(1)中的 $C(\cdot)$ 含义相同。



(a) 非确定性部位示意图(图中用圆点标识)



(b) 时间连续性约束

图 1 非确定性部位局部动作时间连续性保持

Fig. 1 Spatial-temporal feature locally preserving of uncertain body parts

第二步,确定最优解。当得到每幅帧图像的 K_b 个姿态参

数后,通过最小化式(3)采用动态规划算法在 $O(K_0^2 T)$ 时间内确定每帧图像的最优姿态参数,即:

$$\min_{p_i \in P_i, \forall i} C(I_T, p_T) + \sum_{i=1}^{T-1} C(I_i, p_i) + \lambda_1 \theta(p_i, p_{i+1}, I_i, I_{i+1}) \quad (5)$$

由于 ST-FMP 方法的两个步骤均可以在 $O(K_0^2 T)$ 时间内完成,所以基于非确定性人体部位的两段式 ST-FMP 求解算法显著降低了基于 LBP 的 ST-FMP 算法的时间复杂度。

1.3 人体部位运动特征描述

由于无法获得关节角速度、位移速度等精确的人体部位运动参数导致现有运动捕获和动作识别的人体运动模型无法直接使用,为此,本文设计了一个基于人体部位相对位置特征和运动方向的人体运动特征描述模型。

如果以单帧图像的左上角为坐标原点,宽度和高度方向为 X 和 Y 轴,像素为单位建立坐标系,并假设 $C^u = (C_x^u, C_y^u)$ 是人体部位 p_u 的中心位置坐标, (C_x, C_y) 是人体重心位置坐标。那么,本文将某个人体部位 p_u 的位置信息 $l_u = (x^u, y^u)$ 定义为 (C_x, C_y) 和 (C_x^u, C_y^u) 之间的相对位置,即:

$$l_u = (x^u, y^u) = (C_x^u - C_x, C_y^u - C_y) \quad (6)$$

同时,定义 t 时刻 p_u 的运动方向 V_t^u 为 p_u 内所有运动点的运动方向的合成矢量,即:

$$V_t^u = \frac{1}{|\xi_t^u|} \sum_{i=1}^{|\xi_t^u|} v_t^{u,i} \quad (7)$$

其中, $v_t^{u,i}$ 是 t 时刻 p_u 内第 i 个像素的运动方向, ξ_t^u 是 t 时刻 p_u 内所有运动点的集合。本文通过比较相邻帧 I_t 和 I_{t+1} 的稠密光流^[13]估计 $v_t^{u,i}$ 。

借助式(6)和式(7),可以采用运动方向和人体部位相对位置信息构成的特征矩阵来表示一幅帧图像中的人体运动特征。假设 $j_u = (x^u, y^u, V_t^u) \in \mathbf{R}^3$ 是 t 时刻 p_u 的运动特征, $J_t = [j_{t1}, j_{t2}, \dots, j_{td}]^T \in \mathbf{R}^{3d}$ 是 t 时刻包含 $d(d=26)$ 个人体部位的图像 I_t 的运动特征,那么一段帧数为 T 的运动视频的运动特征即可表示为 $f = [J_1, J_2, \dots, J_T] \in \mathbf{R}^{3d \times T}$ 。

根据上述定义,人体姿态特征 J_t 是一个 78 维运动的向量,运动视频的运动特征 f 是一个 $78 \times T$ 的向量矩阵。实验证明,在 $78 \times T$ 的高维运动特征空间中进行关键帧提取具有较高时间复杂度,而且大量的数据冗余和噪声信息^[14]将直接影响关键帧提取的准确性。为提高运动向量的局部特征表达能力,本文采用拉普拉斯分值法(Laplacian Scoring, LS)^[15]对运动向量实施数据降维,确定更具判别性的人体动作特征。首先构建一个 k 近邻图 G_k ;然后采用热核函数计算 G_k 中相连两节点的相似度,获得第 r 个运动特征的拉普拉斯分值 L_r ;最后将前 $n(1 \leq n \leq 3d)$ 个 L_r 较小的运动特征确定为该运动视频的人体姿态特征向量。

2 关键帧提取

本文提出一个基于 ST-FMP 模型和动态聚类的运动视频关键帧提取流程,简称 ST-FMP 算法。

首先,采用 ST-FMP 模型通过式(5)估计第 r 帧图像的人体姿态参数获得人体部位集合 $p = \{p_u | u \in V\}$,并利用式(6)和(7)得到人体运动向量,即:

$$f_r = (f_{r1}, f_{r2}, \dots, f_{rK}) \quad (8)$$

其中: K 等于 $3d$,表示运动特征数; f_{ri} 表示 I_r 中第 i 个人体部位的运动特征。

然后,将运动视频的所有运动向量进行组合得到动作特征的组合向量,即:

$$f_{\text{com}} = (f_1, f_2, \dots, f_{KT}) \quad (9)$$

其中: T 表示运动视频的帧数。

接着,利用 LS 算法计算 f_{com} 的 L_r 分值,获得判别性运动特征向量,记作 f_{sub} 。

最后,利用 L_r 分值向量构建 f_{sub} 的特征曲线,通过极值判定原理选择特征函数上的局部极值点作为候选关键帧集合,记作 \tilde{f}_c 。

在实践中发现,当 \tilde{f}_c 中部分候选关键帧之间时间间隔短或姿态参数差异小时,会导致 \tilde{f}_c 中存在大量冗余数据;反之,则会降低 \tilde{f}_c 的运动表达能力。因此,需要一个关键帧精选过程筛选出关键帧。考虑到动态聚类方法 ISODATA 算法^[16]的良好自适应性,本文选择该算法对 \tilde{f}_c 进行聚类精选。

3 实验结果与分析

本文以健美操动作视频的关键帧为例进行仿真实验,实验结果与人工提取以及最新的运动视频关键帧算法^[9,17]的提取结果进行了对比分析。

3.1 实验数据样本及特征训练

首先邀请 3 位学生每人做两遍 120 s 的大众二级健美操动作,并以 20 帧/s 的采样频率采用普通网络摄像头录制成分辨率为 640×480 的视频;然后从第 10 个韵律节拍开始从 6 套视频中各选 300 帧图像作为实验数据;最后在 1800 帧图像中按照图 1(a)所示人工标记出每个健美操动作的 13 个关节位置。仿真实验时,选择第一、二遍动作的各 900 帧图像作为训练样本数据集和测试样本数据集。

在训练过程中,为了丰富正样本的人体动作特征,首先采用旋转(按 -15° 、 -7.5° 、 7.5° 和 15° 等四个角度)和镜像等图像运算,将 900 帧训练样本扩展成 10×900 帧正样本集合,记作 Φ_p^* ;然后通过图 1(a)中相邻两个关节位置的中点处增加关键点,计算出 Φ_p^* 中每幅图像的 26 个柔性人体部位坐标;最后以 5×5 像素大小的 HoG 单元计算每个人体部位的 HoG 特征,通过线性支持向量机确定式(3)中的 $\varphi(I, l_u)$ 和 $\psi(l_u - l_v)$ 等参数。

3.2 评价标准

人工提取出测试样本集中的所有关键帧,并以常用的关键帧提取准确率^[18]作为算法性能评价标准,即:

$$\text{Accuracy} = \left(\sum_{i=1}^m \delta(f_i, r_i) \right) / n \quad (10)$$

其中: n 和 m 分别代表人工和算法提取的关键帧帧数; f_i 和 r_i 表示算法和人工提取的关键帧; $\delta(\cdot)$ 是 f_i 和 r_i 之间的相似函数,当 f_i 和 r_i 相同时 $\delta(\cdot)$ 值为 1,否则为 0。

3.3 实验结果分析

3.3.1 非确定性部位时空特征嵌入的有效性比较

为了检验非确定性部位时空特征嵌入的人体模型在视频中的人体部位准确率,分别采用 3 种不同的 ST-FMP 模型实现,按照不同的误差像素阈值进行了肘部和膝盖两个人体部位的比较实验,实验结果如图 2 所示。

从图 2 可以看出,和 FMP 模型相比,采用 ST-FMP 算法进行运动视频中人体姿态估计时,在一定像素误差范围内非确定性部位的准确率显著提高。以误差阈值为 20 像素为例,

ST-FMP 算法得到的肘部和膝盖两个部位的准确率分别比 FMP 模型大约提高了 15 和 19 个百分点。但是当像素误差阈值较大(例如大于 40 像素)或者较小(例如大于 10 像素)时准确率差异不显著。图 2 还可以看出,当仅保持上肢(下肢)非确定性部位的时间连续性时,肘部和手腕(膝盖和脚踝)的识别准确率比直接采用 FMP 模型高,但比 ST-FMP 算法低。实验结果表明,通过人体部位局部时间连续性约束来优化人体部位识别结果,ST-FMP 算法显著提高了运动视频中非确定性部位的识别性能。

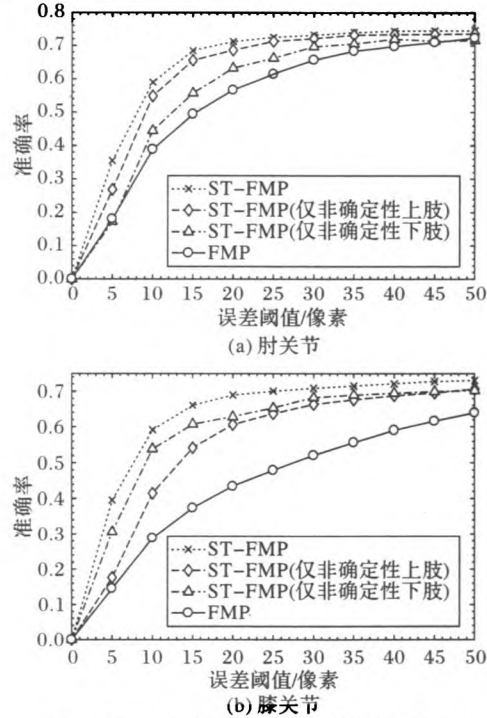


图2 不同 ST-FMP 实现下人体部位识别准确率比较

Fig. 2 Recognition accuracy comparison of body parts of different implements of ST-FMP

本文还对 FMP 模型和 ST-FMP 模型及其不同实现在运动视频关键帧提取方面的性能进行了比较。实验结果如图 3 所示。从图 3 可以发现:1)当精度误差小于 30 像素时, ST-FMP 算法的准确率较高且比较稳定,比采用 FMP 的关键帧算法平均提高约 11 个百分点,而且仅增加上肢(下肢)非确定性部位的时间连续性约束时, FMP 模型的关键帧提取准确率仍提高了约 3 个百分点;2)当误差精度大于 35 像素时, ST-FMP 算法性能仍比采用 FMP 模型有所提高,但准确率降低大约 15 个百分点。

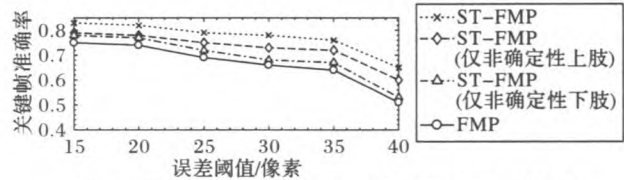


图3 基于 FMP 和 ST-FMP 的关键帧提取比较实验结果

Fig. 3 Accuracy comparison of keyframe extraction of FMP and ST-FMP

同时,当精度误差为 30 个像素,以不同运动特征数通过 ST-FMP 算法提取关键帧时,算法准确率曲线在运动特征数为 15~60 的区间内波动不剧烈、性能稳定,如图 4 所示。

上述实验结果表明,ST-FMP 算法对人体姿态估计结果和

人体局部拓扑结构敏感,非确定性部位的时空约束保持对关键帧提取性能作用明显。

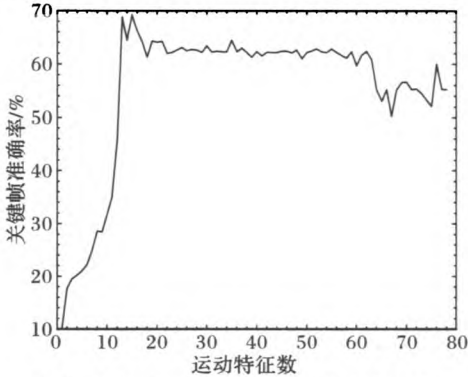


图4 不同运动特征数的关键帧准确率比较

Fig. 4 Keyframe extraction recognition accuracy comparison with different number of motion features

3.3.2 关键帧算法性能比较

为了比较关键帧算法的性能,ST-FMP 算法的仿真实验结果和文献[9]的基于先验的动作视频关键帧提取算法(简称 KFE 算法)及文献[17]的基于运动块关键帧提取算法(简称运动块算法)的运行结果进行了对比实验,如表 1 所示。

表 1 的实验结果显示 ST-FMP 算法的准确率和召回率均优于其他两种算法。首先,从表 1 中可以看出,ST-FMP 算法的准确率比 KFE 算法和运动块算法分别高约 18 和 26 个百分点。KFE 算法采用预先定义的 16 个区块的运动方向表示人体运动特征,而 ST-FMP 算法采用每套动作视频前 15 个 LS 人体运动姿态特征值表示人体运动。因此,ST-FMP 算法采用的运动特征向量冗余小、噪声少,人体部位局部运动表达准确,利于提高关键帧和动作识别的准确率^[19]。

表 1 不同关键帧提取算法性能比较

Tab. 1 Performance comparison of different keyframe extraction algorithms

算法	准确率	召回率
ST-FMP 算法	0.81	0.82
KFE 算法	0.63	0.59
运动块算法	0.55	0.69

其次,从表 1 还可以看出,ST-FMP 算法的召回率也明显优于其他两种算法,平均分别高约 23 和 13 个百分点。KFE 算法和运动块算法均属于基于图像底层特征差异的关键帧技术,它们通过比较图像不同区域内运动变化,采用特定阈值筛选关键帧。而 ST-FMP 算法描述的是人体部位的局部运动特征,本质上是一个语义模型,可以从人体运动参与部位及其运动变化趋势等更高层面来分析和理解运动视频中的人体动作,利用人体姿态相似等语义规则来筛选关键帧,可以获得符合人们认知过程的、更准确的关键帧结果。

上述实验结果说明,由于不仅具有较强的人体部位动作局部拓扑结构表达能力,而且还具有支持基于语义规则进行关键帧筛选的能力,所以 ST-FMP 算法更接近于人工提取结果,更适合基于关键帧的动作视频批阅。同时,由于 ST-FMP 将人体部位拆分成不同的柔性部件,通过柔性部件的局部拓扑结构识别人体姿态,借助时序特征边约束降低人体姿态估计连续误差,所以在复杂场景中具有较强的鲁棒性^[10-11]。

4 结语

本文提出在柔性混合人体模型中嵌入人体部位运动时间约束,通过关注非确定性人体部位的时空连续性和优化人体姿态估计参数提高运动视频关键帧提取算法的准确率和召回率。首先,为保持 FMP 模型在运动视频中人体姿态估计的时空连续性,通过在相邻帧图像的人体部位顶点对间建立时间连续性约束得到具有时空连续特征的 ST-FMP 模型;然后,在相邻视频帧中的非确定性人体部位之间嵌入时空约束简化 ST-FMP 模型,采用 N-best 优化算法估计人体姿态参数,提高运动视频中人体姿态参数求解效率;接着,利用人体部位的相对位置特征和运动方向等特征描述人体部位运动特征,采用拉普拉斯分值算法实施特征选择,形成具有局部判别性的人体运动特征向量;最后,通过 ISODATA 动态聚类算法确定运动视频关键帧。对比实验表明,非确定性人体部位的时空连续性保持和局部动作时空特征嵌入等显著地提高了 ST-FMP 模型的人体部位识别准确率和人体姿态估计性能,获得的关键帧集合更符合人们认知过程。

由于本文采用 ST-FMP 算法从单帧图像中估计人体姿态参数,所以当运动视频中自遮挡或者运动模糊较多时会导致人体部位识别误差显著增加,从而影响关键帧提取的性能。因此,下一步可以考虑采用稠密光流轨迹或 CNN 技术通过提高人体部位识别的鲁棒性改善算法性能。

参考文献 (References)

- [1] ZHOU M. Chinese university students' acceptance of MOOCs: a self-determination perspective [J]. *Computers & Education*, 2016, 92/93: 194 - 203.
- [2] LEHMANN R, SEITZ A, BOSSE H M, et al. Student perceptions of a video-based blended learning approach for improving pediatric physical examination skills [J]. *Annals of Anatomy—Anatomischer Anzeiger*, 2016, 208: 179 - 182.
- [3] PANG Y J. Techniques for enhancing hybrid learning of physical education [C]// *International Conference on Hybrid Learning*, LNCS 6248. Berlin: Springer, 2010: 94 - 105.
- [4] JI S, XU W, YANG M, et al. 3D convolutional neural networks for human action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221 - 231.
- [5] 姬晓飞, 左鑫孟. 基于关键帧特征库统计特征的双人交互行为识别[J]. *计算机应用*, 2016, 36(8): 2287 - 2291. (JI X F, ZUO X M. Human interaction recognition based on statistical features of key frame feature library [J]. *Journal of Computer Applications*, 2016, 36(8): 2287 - 2291.)
- [6] HANNANE R, ELBOUSHAKI A, AFDEL K, et al. An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram [J]. *International Journal of Multimedia Information Retrieval*, 2016, 5(2): 89 - 104.
- [7] 王少华, 狄岚, 梁久桢. 基于核与局部信息的多维度模糊聚类图像分割算法[J]. *计算机应用*, 2015, 35(11): 3227 - 3231. (WANG S H, DI L, LIANG J Z. Multi-dimensional fuzzy clustering image segmentation algorithm based on kernel metric and local information [J]. *Journal of Computer Applications*, 2015, 35(11): 3227 - 3231.)
- [8] JANWE M N J, BHOYAR K K. Video key-frame extraction using unsupervised clustering and mutual comparison [J]. *International Journal of Image Processing*, 2016, 10(2): 73 - 84.
- [9] 庞亚俊. 基于先验的动作视频关键帧提取[J]. *河南理工大学学报(自然科学版)*, 2016, 35(6): 862 - 868. (PANG Y J. Key frames extraction of motion video based on prior knowledge [J]. *Journal of Henan Polytechnic University (Natural Science)*, 2016, 35(6): 862 - 868.)
- [10] YANG Y, RAMANAN D. Articulated human detection with flexible mixtures of parts [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(12): 2878 - 2890.
- [11] 胡琼, 秦磊, 黄庆明. 基于视觉的人体动作识别综述[J]. *计算机学报*, 2013, 36(12): 2512 - 2524. (HU Q, QIN L, HUANG Q M. A survey on visual human action recognition [J]. *Chinese Journal of Computers*, 2013, 36(12): 2512 - 2524.)
- [12] PARK D, RAMANAN D. N-best maximal decoders for part models [C]// *Proceedings of the 2011 International Conference on Computer Vision*. Washington, DC: IEEE Computer Society, 2011: 2627 - 2634.
- [13] SUN D, ROTH S, BLACK M J. Secrets of optical flow estimation and their principles [C]// *Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2010: 2432 - 2439.
- [14] FU Y. *Human Activity Recognition and Prediction* [M]. Berlin: Springer, 2016: 462 - 471.
- [15] HE X, CAI D, NIYOGI P. Laplacian score for feature selection [C]// *Proceedings of the 18th International Conference on Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2005: 507 - 514.
- [16] BALL G H, HALL J. ISODATA, a novel method of data analysis and pattern classification [R]. Springfield: NTIS, 1965.
- [17] 应锐, 蔡瑾, 冯辉, 等. 基于运动块及关键帧的人体动作识别[J]. *复旦学报(自然科学版)*, 2014, 53(6): 815 - 822. (YING R, CAI J, FENG H, et al. Human action recognition based on motion blocks and key frames [J]. *Journal of Fudan University (Natural Science)*, 2014, 53(6): 815 - 822.)
- [18] XIA G, SUN H, NIU X, et al. Keyframe extraction for human motion capture data based on joint kernel sparse representation [J]. *IEEE Transactions on Industrial Electronics*, 2017, 64(2): 1589 - 1599.
- [19] LIU Z, ZHU J, BU J, et al. A survey of human pose estimation [J]. *Journal of Visual Communication & Image Representation*, 2015, 32(C): 10 - 19.

This work is partially supported by the Key Science and Technology Program of Henan Province (152102210329, 172102310635).

SHI Nianfeng, born in 1976, Ph. D., associate professor. His research interests include computer supported cooperative work, pattern recognition.

HOU Xiaojing, born in 1975, M. S., lecturer. Her research interests include pattern recognition.

ZHANG Ping, born in 1976, Ph. D., associate professor. His research interests include network security, pattern recognition.