计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(12):109−115 [doi: 10.15888/j.cnki.csa.006665]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于光流的快速人体姿态估计①

周文俊[1], 郑新波[2], 卿粼波[1], 熊文诗[1], 吴晓红[1]

[1](四川大学 电子信息学院, 成都 610065)
[2](东莞前沿技术研究院, 东莞 523000)
通讯作者: 卿粼波, E-mail: qing_lb@scu.edu.cn

摘　要: 针对目前深度学习领域人体姿态估计算法计算复杂度高的问题, 提出了一种基于光流的快速人体姿态估计算法. 在原算法的基础上, 首先利用视频帧之间的时间相关性, 将原始视频序列分为关键帧和非关键帧分别处理(相邻两关键帧之间的图像和前向关键帧组成一个视频帧组, 同一视频帧组内的视频帧相似), 仅在关键帧上运用人体姿态估计算法, 并通过轻量级光流场将关键帧识别结果传播到其他非关键帧. 其次针对视频中运动场的动态特性, 提出一种基于局部光流场的自适应关键帧检测算法, 以根据视频的局部时域特性确定视频关键帧的位置. 在OutdoorPose 和 HumanEvaI 数据集上的实验结果表明, 对于存在背景复杂、部件遮挡等问题的视频序列中, 所提算法较原算法检测性能略有提升, 检测速度平均可提升 89.6%.

关键词: 人体姿态估计; 深度学习; 光流; 自适应关键帧

引用格式: 周文俊,郑新波,卿粼波,熊文诗,吴晓红.基于光流的快速人体姿态估计.计算机系统应用,2018,27(12):109–115. http://www.c-s-a.org.cn/1003-3254/6665.html

Fast Human Pose Estimation Based on Optical Flow

ZHOU Wen-Jun[1], ZHENG Xin-Bo[2], QING Lin-Bo[1], XIONG Wen-Shi[1], WU Xiao-Hong[1]

[1](College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)
[2](Dongguan Institute of Advanced Technology, Dongguan 523000, China)

Abstract: Aiming at the problem of high computational complexity of human pose estimation algorithm in deep learning field, a fast human pose estimation algorithm based on optical flow is proposed. Based on the original algorithm, using the time correlation between video frames, the original video sequence is divided into key frames and non-key frames, which are processed respectively (the images between two adjacent key frames and the forward key frame compose a video frame group, which is similar to the frames in the same video frame group), the human pose estimation algorithm is applied only to the key frames, and the key frame recognition result is propagated to other non-key frames through the lightweight optical flow field. Secondly, aiming at the dynamic characteristics of the video field, this study proposes an adaptive key frame detection algorithm based on local optical flow to determine the position of the key frame of video according to the local time-domain characteristics of the video. The experimental results in OutdoorPose and HumanEvaI data sets show that the detection performance of the proposed algorithm is slightly higher than the original algorithm in the video sequences with complex background and component occlusion. The detection speed is increased by 89.6% in average.

Key words: human pose estimation; deep learning; optical flow; adaptive key frame

① 基金项目: 东莞市社会科技发展项目 (2017507102428)
Foundation item: Social Science and Technology Development Project of Dongguan City (2017507102428)
收稿时间: 2018-04-11; 修改时间: 2018-05-24; 采用时间: 2018-06-05; csa 在线出版时间: 2018-12-03

基于视觉的人体姿态估计问题是指根据图像特征估计人体各个部位的位置与关联信息[1]. 人体姿态估计是图像处理、计算机视觉、模式识别、机器学习、人工智能等多个学科的交叉研究课题, 在视频监控、视频检索、人机交互、虚拟现实、医疗看护等领域, 具有深远的理论研究意义和很强的实用价值[2,3].

目前的人体姿态估计算法主要分为两类: 一类是基于深度图像, 另一类是基于可见光图像. 基于深度图像的算法, 主要利用如 Kinect[4]等深度传感器获取代表着人体外貌和几何信息的颜色和深度 (RGBD) 数据, 进而分析人体的姿态. 但深度传感器等硬件设备的配置及数量有限, 导致其无法分析如监控视频, 网上的海量视频等数据. 而基于可见光图像的算法, 只需要获取图片中人体的表观特征, 如人体姿态各部分的 HOG 特征[5]、人体轮廓特征[6]及视频中上下文 (Context) 关系[7]. 但以上特征都需要手动提取, 且不具有鲁棒性. 直到近几年, 深度学习被广泛运用到图像处理领域, 促进了人体姿态检测进一步的发展. 其中, Cao 等[8]从图像底层出发, 对人体姿态关节点进行回归分析, 同时运用并行网络提取人体关节点间的亲和力场, 确定多人姿态关节点之间的联系. Pfister 等[9]利用光流信息将前后帧的人体姿态热力图扭曲到当前帧, 然后赋予不同时刻热力图不同的权值, 综合得到当前帧人体姿态. He 等[10]首先采用神经网络提取人体候选区域, 然后在候选区域上用两个并行的网络分别进行目标检测和人体姿态关节点检测. Charles 等[11]利用已有的人体姿态估计模型初始化视频帧, 然后在相邻帧中进行空间匹配, 时间传播及人体姿态关节点的自我评估, 不断地重复上述过程得到准确的人体姿态关节点. 上述基于深度学习的方法能够很好的解决基于单帧图像的人体姿态估计问题. 然而, 对于大量现有的视频数据, 大多数视频分析任务是通过直接将识别网络应用到视频的所有帧, 这一方法将消耗大量的计算资源, 且未考虑到视频帧之间的时间相关性.

综上所述, 基于深度学习的人体姿态检测算法虽然已经在单帧图像上取得了较好的效果, 但它们往往依赖于强大的计算机硬件平台, 一般需要多个 GPU 进行加速. 而在计算资源受限的移动终端上对视频数据运用上述基于深度学习的人体姿态检测算法时, 终端的计算能力往往无法达到需求, 所以如何降低或转移人体姿态估计算法的计算复杂度[12], 是该领域的一个重要研究方向. 另外, 由于人和相机具有运动连续性的特点, 相邻帧之间的人体姿态也将表现出运动的连续性[13], 即时间相关性, 因此本文提出了一种基于光流的快速人体姿态估计算法, 该算法利用视频帧之间的时间相关性实现人体姿态估计的加速. 在一个视频帧组内, 首先根据人体姿态估计算法对关键帧进行人体姿态检测. 而对于其他的非关键帧, 计算它与前向关键帧之间的光流场信息 (时间相关性), 然后根据光流场将关键帧的检测结果传播到非关键帧上, 避免了在每一帧上运行人体姿态估计算法. 当光流场计算复杂度低于人体姿态估计算法时, 本文框架可以有效提高人体姿态检测算法的检测速度.

1 基于光流的快速人体姿态估计

1.1 视频帧姿态相关性分析

视频相邻帧之间存在极强的时空相关性[14], 这种相关性是由运动的连续性决定的, 因此视频相邻帧中运动目标及人体姿态信息具有更强的时空相关性. 如图 1 所示为视频帧间相关性及人体姿态相关性效果图, 第一行 *Frame* 为原始视频帧, 第二行 *Pose* 为原始视频帧对应的真实姿态信息, 第三行 *Flow* 为第 i 帧 (i=2, …, 5) 图像与第一帧图像之间的真实光流场, 第四行 *Dsp* 为第 i 帧 (i=2, …, 5) 图像中人体关键点与第一帧图像中人体关键点之间的运动矢量场.

如图 1 中 Dsp 反映了视频序列中 $Frame_i$ (i=2, …, 5) 与 $Frame_1$ 对应关键点之间的运动矢量, 而视频帧间的运动矢量为视频帧之间对应相似块的运动信息, 同时通过运动矢量可将第一帧图像的人体姿态信息传播到后续视频帧中. $Flow$ 为视频序列中 $Frame_i$(i=2, …, 5) 与 $Frame_1$ 之间的光流信息, 而光流就是在图像灰度模式下, 图像间的亚像素级运动矢量, 被广泛用于估计两个连续帧之间的像素点的运动[15]. 因此可以通过视频帧间的光流信息及 $Frame_1$ 中的人体姿态信息预测 $Frame_i$(i=2, …, 5) 中的人体姿态信息. 另外, 由图 1 中 *Frame* 可知, $Frame_1$ 到 $Frame_5$ 相邻帧之间人体姿态变化较为平缓. 而随着时间的推移, 当前帧 $Frame_i$(i=2, …, 5) 与 $Frame_1$ 的人体姿态信息变化越来越大, 相关性也越来越低.

1.2 基于光流的快速人体姿态估计框架

如上所述, 视频帧间人体姿态信息存在极强的时间相关性, 而充分利用视频帧间的相关性及运动信息可将已检测的人体姿态信息传播到随后相关性较高的相邻帧中, 从而避免对每帧图像进行复杂的人体姿态

检测. 因此本文提出了基于光流的快速人体姿态估计算法, 首先将视频帧分割成多个视频帧组确定视频关键帧 (每个视频帧组的第一帧为该视频帧组的关键帧, 其余视频帧为非关键帧). 然后采用 Cao 等[8]的人体姿态估计算法确定关键帧人体姿态信息, 该算法可有效地检测图片中的人体姿态信息; 最后利用轻量级光流算法 Flownet2-c[16]计算关键帧与非关键帧之间的光流信息, 将关键帧的检测结果传播到非关键帧 (如图 2).

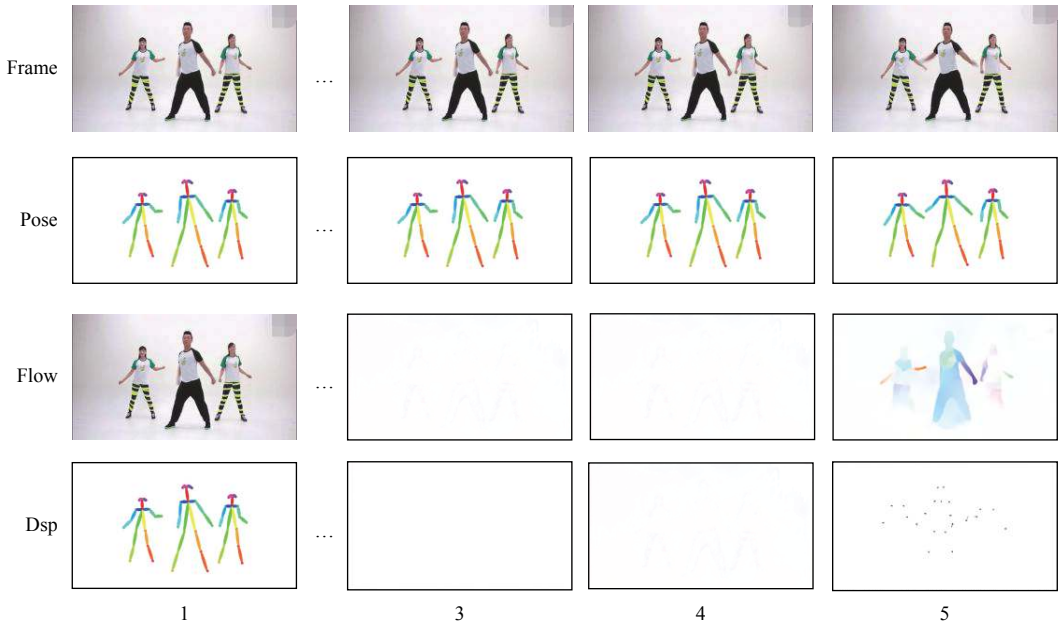


图 1 视频帧间相关性及人体姿态相关性效果图

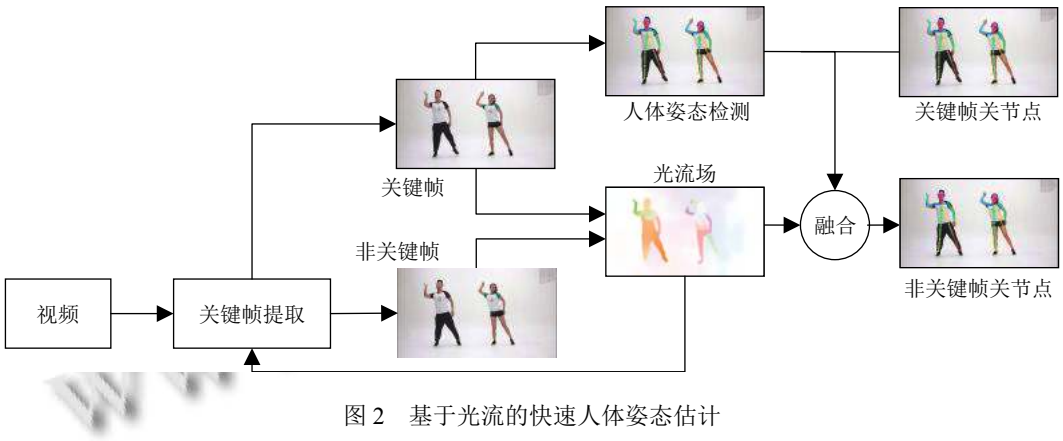


图 2 基于光流的快速人体姿态估计

具体定义如下:

$$\begin{cases} Flow_i = flow(Frame_1, Frame_i) \\ Pose_i' = add(Pose_1, Flow_i) \end{cases} \quad (1)$$

其中, $Flow_i$ 为第 i 帧图像与对应关键帧 $Frame_1$ 之间的光流信息, $Pose_1$ 为关键帧的真实人体姿态信息, $Pose_i$'为将关键帧的真实人体姿态信息通过第 i 帧图像与关键帧之间的光流场融合后的人体姿态信息.

基于上述算法原理, 本文算法中关键帧的选取, 以及关键帧人体姿态信息与光流信息的融合效果直接影响非关键帧的人体姿态估计精度. 而由 1.1 节分析可知视频帧间相关性随着时间推移而降低. 因此, 视频中关键帧的位置应该根据视频中帧间相对运动程度的不同而重新设置, 以适应视频序列帧间相关性的改变. 针对上述问题本文提出一种自适应关键帧检测算法. 同时也对融合过程中光流计算算法对图像中噪声过于敏感的问题进行优化.

1.2.1 自适应关键帧检测算法

本文算法主要利用光流信息将关键帧的姿态信息传播到非关键帧, 当同一视频帧组内关键帧与非关键帧中同一关节点之间有较大的位移时, 光流信息就无法准确的描述关节点的运动, 从而导致非关键帧人体姿态预测失败. 因此本文提出一种自适应关键帧检测算法, 其中为了不引入多余的计算量, 通过已有的光流场, 判断两视频帧之间是否出现剧烈位移运动, 从而划分关键帧与非关键帧, 达到自适应关键帧检测的目的, 具体算法如下:

步骤 1. 第 i 帧与前向关键帧 P_K 之间的光流信息 $f_i(x,y)$. 计算局部光流信息模的累加和 $Local\_sum(f)$ 和局部光流信息的最大值 $Local\_max(x,y)$, 具体定义如下:

$$f_i(x,y)=(v_{\bar{x}}(x,y),v_{\bar{y}}(x,y)) \tag{2}$$

$$Local\_sum(f)=\underset{(x,y)\in mask}{add}\sqrt{v_{\bar{x}}(x,y)^2+v_{\bar{y}}(x,y)^2} \tag{3}$$

$$Local\_max(x,y)=\underset{(x,y)\in s}{max}\sqrt{v_{\bar{x}}(x,y)^2+v_{\bar{y}}(x,y)^2} \tag{4}$$

其中, (x,y) 为像素坐标, $v_{\bar{x}}(x,y)$为光流场在 x 方向上的分量, $v_{\bar{y}}(x,y)$为光流场在 y 方向上的分量, $mask$ 为图像中每个人的矩形掩模框并集 (如图 3 所示, 恰好覆盖所有人的关节点), s 为关键帧所有关节点处像素点的集合.

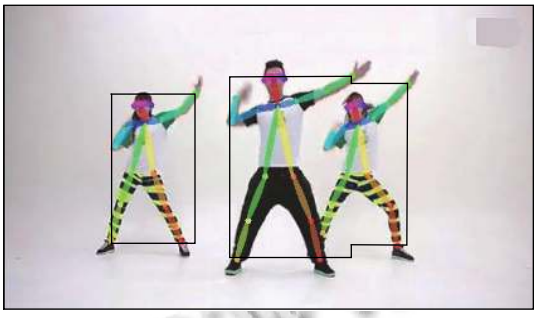


图 3 矩形掩模区域

步骤 2. 确定判决阈值:

$$\begin{cases} Local\_sum\_T=mask\_sum*m \\ Local\_max\_T=10 \end{cases} \tag{5}$$

其中, $Local\_sum\_T$ 为局部光流信息模的累加和的阈值, $Local\_max\_T$ 为局部光流信息最大值的阈值. 若固定 $Local\_sum\_T$ 则无法适应视频中由人距离相机的远近不同, 而引起的光流信息不同的问题. 因此本文自适应关键帧检测算法会在每一帧自适应阈值. 计算关键帧中每个人的矩形掩模框并集的总面积 $mask\_sum$, 将 $mask\_sum*m$(m 为掩模系数) 作为 $Local\_sum(f)$ 的阈值. 而对于局部光流信息的最大值, 我们通过大量实验发现, 当两帧图像关节点处光流信息的模大于 10 个像素时, 光流场无法准确的预测关节点的位移; 当模小于等于 10 个像素时, 光流场可以有效的预测关节点的位移. 所以 $Local\_max\_T$ 的取值为 10.

步骤 3. 将局部累加和 $Local\_sum(f)$ 与 $Local\_sum\_T$ 比较, 局部光流信息最大值 $Local\_max(x,y)$ 与 $Local\_max\_T$ 比较, 以避免局部剧烈运动.

$$\begin{cases} Local\_sum(f)<=Local\_sum\_T \\ Local\_max(x,y)<=Local\_max\_T \end{cases} \tag{6}$$

式 (6) 成立时第 i 帧为非关键帧, 否则结束该视频帧组, 第 i 帧为下一视频帧组的关键帧.

1.2.2 关键点局部融合优化

通过人体姿态估计阶段对关键帧进行人体姿态估计, 得到关键帧的人体关节点. 然后利用密集光流来预测关节点应如何在时间上流动到下一帧[13].

本文使用的 Flownet2-c 算法可求得关键帧与非关键帧之间的光流信息. 但视频中阴影或噪点在运动物体周围尤其明显, 如图 4 所示, 图 4(c) 为图 4(a) 和图 4(b) 利用 Flownet2-c 算法计算的光流信息. 由图可知, 运动物体周围的光流信息分布十分不均匀. 因此若在融合关键帧姿态信息和光流信息时, 只使用关键帧关节处的光流信息作为非关键帧关节点的运动信息, 则会因光流信息计算不准确导致关节点信息预测失败. 针对这个问题本文利用邻域特性, 根据邻域像素点光流信息决定该关节点的运动矢量. 采用关节点处 5x5 邻域的光流信息代替关节点的运动信息, 以提高融合预测的准确率.

Flownet2-c 算法效果

具体使用式 (7) 计算得到非关键帧的关节点.

$$\begin{cases} Df(x_i,y_i)=\dfrac{1}{25}\displaystyle\sum_{l=-2}^{2}\sum_{n=-2}^{2}f(x_i+l,y_i+n) \\ P'(x_i,y_i)=add(P(x_K,y_K)+Df(x_i,y_i)) \end{cases} \tag{7}$$

其中, $Df(x_i,y_i)$ 为关键帧关节点处 5×5 邻域的光流信息的均值, $P(x_k,y_k)$ 为关键帧关节点坐标, $P'(x_i,y_i)$ 为非关键帧关节点坐标.

2 实验结果及分析

2.1 实验设置

本文主要利用 Caffe[17]框架搭建基于光流的快速

人体姿态估计算法框架与 Cao 等[8](Caffe) 的算法对比 (在 Intel i5, 8 G 内存, 单张 GTX 1070 的机器上测试).

(a) 图1

(b) 图2

(c) 图1与图2之间的光流信息

图 4 Flownet2-c 算法效果

实验对两个公开数据集进行测试, 分别为: (1) OutdoorPose 数据集, 该数据集由 Ramakrishna 等[18]提出, 共包含 6 段视频序列, 约 1000 帧已标注人体各部件真实值的图像. (2)HumanEvaI 数据集, 该数据集由 Sigal 等[19]提出, 本文采用 S1_Jog_1_(C1), S1_Walking_1_(C1), S2_Jog_1_(C1) 中各 150 帧进行验证. 以上两个数据集中包含丰富的人体自遮挡. 对于 1.2.1 中掩模系数取 0.4, 该系数越大姿态估计速度越快, 准确度相对越低; 反之, 姿态估计速度越慢, 准确度相对越高.

本文采用每秒处理的帧数 (帧率: fps) 评估算法检测速度, 利用 PCP 评价标准[20]来评估算法对于人体各部件的估计准确度. 相关定义如下:

$$Fps = nFrame/\sum_{i=1}^{nFrame} t_i \tag{8}$$

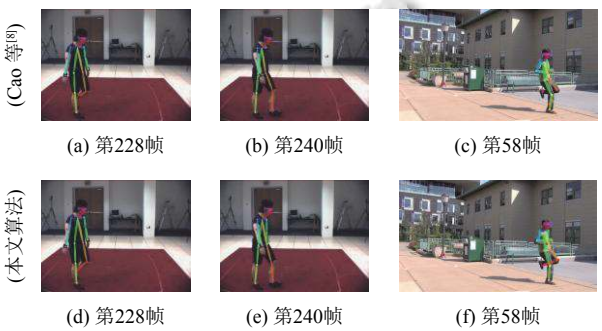
$$PCP = \frac{pose_{\text{true}}}{pose_{\text{all}}} \times 100\% \tag{9}$$

其中, Fps 为每秒处理的帧数 (帧率), $nFrame$ 为测试视频的帧数, t_i 为第 i 帧的检测用时, 其中包括利用 Flownet2-c 计算两张图片 (分辨率: 640×380) 之间的光流信息 15 ms. $pose_{\text{ture}}$ 为检测正确的关节点数量, $pose_{\text{all}}$ 为测试视频中所有的关节点数量. PCP 评价标准规定当估计的所有部件端点到其对应真实值端点的距离小于部件长度的一半时,则认为该部件被正确定位. 其中, PCP 值越大表示对人体各部件的估计准确度越高.

2.2 结果分析

表 1 为本文算法与 Cao 等[8]的算法在不同场景下检测帧率及准确度的比较表. 从表 1 可以看出本文基于光流的快速人体姿态估计算法与 Cao 等[8]的算法在检测准确度差异不大的情况下, 有效的提升检测速度. 其中在 OutdoorPose 数据集上, 本文算法较 Cao 等[8]的算法在检测准确度上提升 1.3%, 检测帧率提升 87.5%; 在 HumanEvaI 数据集上, 本文算法较 Cao 等[8]的算法在检测准确度下降 1% 的情况下, 检测帧率提升 91.8%.

如图 5 所示为 Cao 等[8]的算法与本文算法在上述两个数据集上复杂环境 (包含静态复杂背景、肢体遮挡等) 下的部分姿态估计效果图, 其中第一列图 5(a)、图 5(b)、图 5(c) 为 Cao 等[8]的算法的部分检测结果, 第二列图 5(d)、图 5(e)、图 5(f) 为本文算法的部分检测结果.

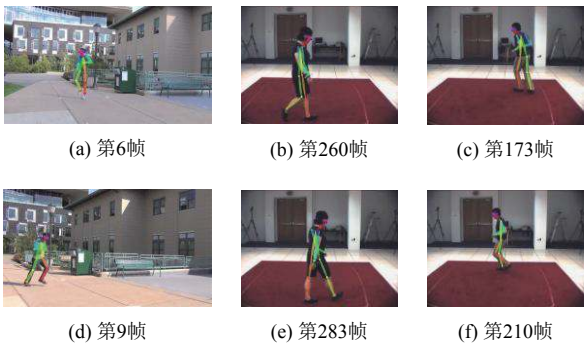


(a) 第228帧 (b) 第240帧 (c) 第58帧

(d) 第228帧 (e) 第240帧 (f) 第58帧

图 5 部分姿态估计效果图

如图 6 所示为上述两数据集的部分图片及本文算法的检测效果图. 由图 5 的测试图片可知, Cao 等[8]的算法在复杂环境下可能会检测失败, 而本文算法中非关键帧的姿态信息由关键帧姿态信息及两帧之间的光流信息预测得到, 由图 5 可知在复杂环境下本文算法较原算法在一定程度上可增加人体姿态检测的检测性能. 上述结果说明作者提出的加速算法较原算法在平均检测准确度略有提升的情况下, 能够利用视频帧间的时间相关性, 有效的提升处理速度, 降低计算复杂度.



(a) 第6帧 (b) 第260帧 (c) 第173帧

(d) 第9帧 (e) 第283帧 (f) 第210帧

图 6 数据集部分效果图

3 结论与展望

为了降低深度学习领域人体姿态估计算法的计算复杂度, 本文提出了一种基于光流的快速人体姿态估计算法. 该算法将视频分为关键帧和非关键帧分别处理, 利用光流场将关键帧人体姿态信息传播到非关键帧, 将高计算复杂度的人体姿态估计算法的计算复杂度转移到低计算复杂度的光流信息的计算过程中, 同时提出自适应关键帧检测算法及融合算法, 确定关键帧位置, 防止光流预测不准确等问题. 实验表明当光流算法的计算复杂度低于对关键帧的人体姿态估计算法时, 本文方法可以在检测效果与原算法差异不大的情况下, 有效地降低人体姿态估计的计算复杂度, 提升检测速度. 在今后的工作中应该进一步考虑如何高效的选取关键帧, 进一步对算法进行加速.

表 1 人体姿态估计帧率及估计准确度比较

算法	数据集	躯干	头部	腿上端	腿下端	胳膊上端	胳膊下端	平均	帧率 (fps)
Cao 等[8]	OutdoorPose	1	1.00	0.81	0.70	0.94	0.73	0.863	4.0
本文算法		1	0.97	0.83	0.70	0.93	0.83	0.876	7.5
Cao 等[8]	HumanEvaI	1	1.00	1.00	0.94	0.91	0.70	0.925	3.7
本文算法		1	1.00	0.97	0.93	0.90	0.69	0.915	7.1

参考文献

1 代钦, 石祥滨, 乔建忠, 等. 结合遮挡级别的人体姿态估计方法. 计算机辅助设计与图形学学报, 2017, 29(2): 279–289. [doi: 10.3969/j.issn.1003-9775.2017.02.009]

2 田国会, 尹建芹, 韩旭, 等. 一种基于关节点信息的人体行为识别新方法. 机器人, 2014, 36(3): 285–292.

3 韩贵金, 朱虹. 基于 HOG 和颜色特征融合的人体姿态估计. 模式识别与人工智能, 2014, 27(9): 769–777. [doi: 10.3969/j.issn.1003-6059.2014.09.001]

4 Zhang ZY. Microsoft kinect sensor and its effect. IEEE Multimedia, 2012, 19(2): 4–10. [doi: 10.1109/MMUL.2012.24]

5 范国娟, 范国卿, 柳絮青. HOGG: 基于 Gabor 变换与 HOG 特征的人体检测. 微型机与应用, 2016, 35(21): 14–15, 19.

6 薄一航, Hao J. 视频中旋转与尺度不变的人体分割方法. 自动化学报, 2017, 43(10): 1799–1809.

7 徐建强, 陆耀. 一种基于加权时空上下文的鲁棒视觉跟踪算法. 自动化学报, 2015, 41(11): 1901–1912.

8 Cao Z, Simon T, Wei SE, *et al*. Realtime Multi-Person 2D pose estimation using part affinity fields. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1302–1310.

9 Pfister T, Charles J, Zisserman A. Flowing convnets for human pose estimation in videos. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1913–1921.

10 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2980–2988.

11 Charles J, Pfister T, Magee D, *et al*. Personalizing human video pose estimation. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 3063–3072.

12 Han S, Mao HZ, Dally WJ. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv:1510.00149, 2016.

13 Zuffi S, Romero J, Schmid C, *et al*. Estimating human pose with flowing puppets. Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, NSW, Australia. 2014. 3312–3319.

14 Kang D, Emmons J, Abuzaid F, *et al*. NoScope: Optimizing neural network queries over video at scale. Proceedings of the VLDB Endowment, 2017, 10(11): 1586–1597. [doi: 10.14778/3137628]

15 Mabrouk AB, Zagrouba E. Spatio-temporal feature using optical flow based distribution for violence detection. Pattern Recognition Letters, 2017, 92: 62–67. [doi: 10.1016/j.patrec.2017.04.015]

16 Ilg E, Mayer N, Saikia T, *et al*. Flownet 2.0: Evolution of

optical flow estimation with deep networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1647–1655.

17 Jia YQ, Shelhamer E, Donahue J, *et al*. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM international conference on Multimedia. Orlando, FL, USA. 2014. 675–678.

18 Ramakrishna V, Kanade T, Sheikh Y. Tracking human pose by tracking symmetric parts. Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA. 2013. 3728–3735.

19 Sigal L, Balan AO, Black MJ. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. International Journal of Computer Vision, 2010, 87(1–2): 4–27.

20 Ferrari V, Marin-Jimenez M, Zisserman A. Progressive search space reduction for human pose estimation. Proceedings of 2018 IEEE Computer Vision and Pattern Recognition. Anchorage, AK, USA. 2008. 1–8.