DOI：10.13973/j.cnki.robot.2016.0578

# 基于多级动态模型的 2 维人体姿态估计

马 淼，李贻斌
（山东大学控制科学与工程学院，山东 济南 250061）

**摘 要**：提出一种利用多级动态模型来估计单目视频中的人体姿态的方法．首先，构建了一种多级动态人体姿态模型，该模型将人体姿态视为各部位姿态的铰接组合，用部位姿态的最优估计来逼近整体姿态的最优估计，从而解决了整体姿态估计带来的歧义性问题．其次，提出了一种通过构建虚拟姿态来计算视频相邻帧之间姿态一致性的算法，该算法能够有效利用视频中表观特征及运动特征的连续性，从而提高姿态估计精度．此外，使用粒子群优化算法用较小的姿态样本优化出最优部位姿态，并将最优部位姿态重组为最优的人体姿态．通过实验验证了所提方法的有效性，并与几种前沿方法进行了比较．实验结果表明，本文方法有效提高了人体姿态估计的准确度．

**关键词**：人体姿态估计；多级动态模型；视频理解；人体行为理解
**中图分类号**：TP391 **文献标识码**：A **文章编号**：1002-0446(2016)-05-0578-10

## 2D Human Pose Estimation Using Multi-level Dynamic Model

MA Miao，LI Yibin
(*School of Control Science and Engineering*, *Shandong University*, *Ji'nan* 250061, *China*)

**Abstract:** A human pose estimation algorithm with a multi-level dynamic model for monocular videos is presented. Firstly, a multi-level dynamic model of human pose is constructed to decompose human entire pose into articulated pose parts, and approach optimal human pose candidates by optimizing pose parts candidates. This model solves the ambiguity problem caused by the entire pose estimation method. Secondly, an algorithm for calculating the pose consistency between the adjacent video frames is proposed by constructing virtual poses. This method can make use of the continuity of appearance features and motion features between the adjacent frames to improve the estimation accuracy. Thirdly, particle swarm optimization method is utilized to search for the best pose parts candidates with a small amount of candidates, and then the achieved pose parts are recomposed into the optimal human entire poses. The efficiency of the proposed method is tested and experimentally compared with several related state-of-the-art methods on challenging video sequences, which shows significant improvements.

**Keywords:** human pose estimation; multi-level dynamic model; video understanding; human activity understanding

## 1 引言（Introduction）

视频中的人体姿态估计问题是人机交互中需要解决的一项重要任务，可用于机器人视觉感知、机器学习以及机器人社会等[1-3]．近年来，人体行为识别在机器人领域中日益受到关注．例如，美国马里兰大学近年来致力于研发一种可使机器人通过观看人类烹饪的视频来学习烹饪动作的算法，现已实现用卷积神经网络识别不同的抓取方式[4]；美国康奈尔大学研发出一种算法，可以使机器人通过 RGB-D（RGB-depth）视频识别人的行为[5]．

在当前的研究中，利用多摄像头或 RGB-D 信息进行 3 维人体姿态估计的研究受到了广泛的关注并取得了一定的成果，然而在无标定的单目视频中进行人体姿态检测的有效方法仍在探索当中，特别是提高手臂位姿的估计精度．本文针对无标定的单目视频中人体上半身的姿态估计算法进行研究．

人体姿态估计的核心问题在于目标模型的构造与使用．目前提出的人体检测模型[6-7]大多依赖于 Fischler 和 Elschlager[8] 在 1973 年提出的图形化架构（PS，pictorial structure）．这种图形化结构用树模型进行优化，将人的肢体视为刚体并用矩形来

基金项目：国家 863 计划（2015AA042201）；国家自然科学基金（61233014）．
通信作者：马淼，mamiaosdu@hotmail.com；李贻斌，liyb@sdu.edu.cn 收稿/录用/修回：2016-03-07/2016-06-22/2016-07-18

表示．随着 PS 模型的发展，近年来有些学者对 PS 的刚体性质提出了改进，并构造出一系列的可变形部位模型，如轮廓人模型[9]和可变形结构（DS，deformable structure）模型[10]．这种非刚体模型通常能够更好地利用图像信息，从而得到更精确的姿态估计．

这些可变形部位模型虽然对每个身体部位进行独立的表达，但其姿态估计过程是对姿态进行全局优化．由于铰接人体模型的复杂性，全局优化在遮挡情况下或某个身体部位快速移动的情况下会产生歧义[11-13]．例如一个候选姿态的左臂估计正确而右臂估计错误，另一个候选姿态的右臂正确而左臂错误，这种情况下用全局目标函数无法得到准确的姿态估计结果．

此外，有些学者利用分层推理的方法来优化姿态估计，首先通过背景减除[14]或粗略的人体区域检测[15]来缩小搜索区域，继而在得到的小区域中进行精确的人体姿态检测．然而这种方法在获取小区域时通常无法完整保留下臂，从而使手臂的检测精度下降．还有一些学者对每个身体部位模型单独建模并施加一定的限制[16-17]，但是这些附加的限制使算法不具有通用性．

针对这些问题，本文提出了一种基于多级动态模型的 2D 人体姿态估计方法．在 DS 模型的基础上，构造出一种可分解与重构的 2D 人体上半身的模型．将提出的模型用于 2D 视频中的人体姿态估计，以每个身体部位作为对象进行优化迭代，由每个部位的局部最优来逼近整个身体姿态的全局最优．本文的主要贡献有：(1) 提出一种可分解及重构的多级动态人体姿态模型，在姿态估计过程中逐级优化，从而更有效地估计人体姿态；(2) 提出一种通过构造虚拟姿态来计算候选姿态样本在视频中的损耗的代价函数，提高了姿态估计的鲁棒性；(3) 有效使用粒子群优化算法，利用较小的姿态候选样本集优化出最优人体姿态．

## 2　多级动态模型结构及人体姿态表达（Multi-level dynamic model and human pose expression）

### 2.1　多级动态模型结构

本文的多级动态模型算法共有 5 级：第 1 级为整体姿态模型，本级将人体上半身姿态完整地表达出来，并对姿态样本进行评价；第 2 级为身体部位模型，本级可将上半身完整姿态分解成为身体各部位的姿态并进行优化，也可将最优的身体各部位姿态重组为完整姿态；第 3 级是我们提出的虚拟姿态损耗模型，本级构造出虚拟姿态，并利用虚拟姿态计算相应的代价函数；第 4 级是低水平特征，如颜色直方图、光流场以及轮廓特征，这些特征能够辅助第 3 级生成虚拟姿态损耗，并且辅助第 2 级的优化运算；第 5 级为视频中相邻 2 帧的原始图像，用于生成第 4 级的低水平特征．模型中的各级随着视频中图像帧的变化而动态变化，因此称之为多级动态模型，如图 1 所示．第 3 节中将针对在视频中利用多级动态模型估计人体姿态的应用给出详细的介绍．

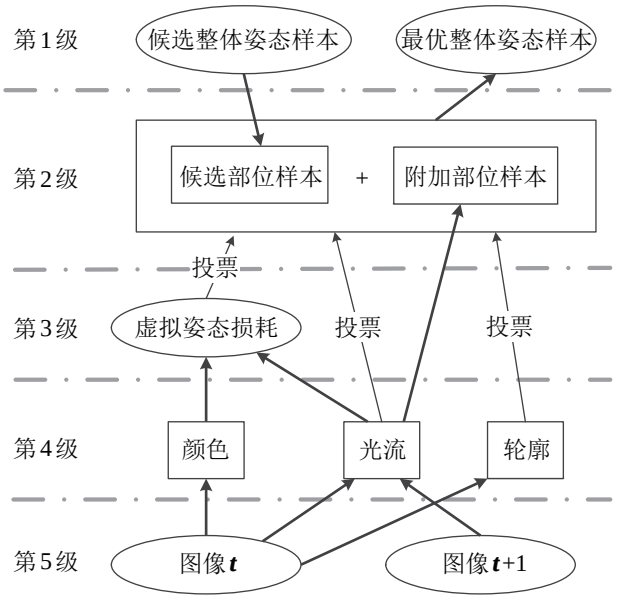



图 1　多级动态模型人体姿态估计算法

Fig.1　Multi-level model based human pose estimation algorithm

### 2.2　人体姿态表达

本文基于 DS 模型构造出一种可分解与重构的轮廓模型，如图 2 所示．DS 模型是根据参数化的铰接人体 3D 模型 SCAPE（shape completion and animation of people）[18] 学习得到的 2D 轮廓模型，其部位的可变形性是由主成分分析（PCA）模型实现的[10]．本文提出的轮廓模型可将完整的人体姿态分解成为独立的身体部位姿态，如图 2(b) 所示；也可将部位姿态重组为完整的身体姿态，如图 2(a) 所示．人体姿态由 $n_{\mathrm{p}}=6$ 个部位构成，如图 2(b) 中的 $\{p_i|i=1:n_{\mathrm{p}}\}$，它们按顺序分别是躯干、头部、左上臂、右上臂、左下臂、右下臂．全局姿态有 $n_{\mathrm{k}}=9$ 个关键点 $\{k_i|i=1:n_{\mathrm{k}}\}$，按顺序分别定义为为肚脐、头、颈、左肩、右肩、左肘、右肘、左腕和右腕，如图 2(a) 所示．其中 $k_{\mathrm{joint}}=\{k_3,k_4,k_5,k_6,k_7\}$ 为铰接连接关节，用于连接相邻的 2 个部位．

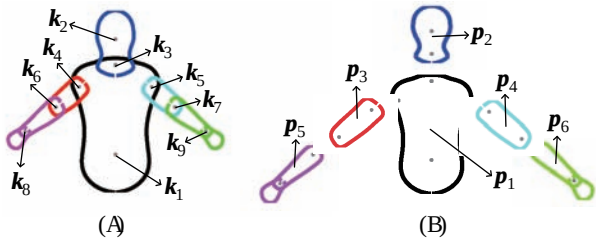



图 2 人体姿态轮廓

Fig.2 Human pose contour

人体部位 $p_i$ 的姿态轮廓 $L_i$ 的计算方法如下：

$$\begin{bmatrix} L_i \\ k_i, s_i \end{bmatrix} = \boldsymbol{B}_i \boldsymbol{z}_i + \boldsymbol{m}_i \tag{1}$$

其中 $k_i$ 是部位 $p_i$ 中的关键点位置，由图 2(b) 可知，除 $p_1$ 有 4 个关键点外，其余部位 $\{p_i | i \neq 1\}$ 都有 2 个关键点．此外，$s_i$ 是部位的尺度参数，$\boldsymbol{B}_i$ 和 $\boldsymbol{m}_i$ 是用来计算部位 $p_i$ 的轮廓的 PCA 模型参数[10]，其中 $\boldsymbol{m}_i$ 是主轮廓向量，$\boldsymbol{B}_i$ 是主要特征向量构成的矩阵．$\boldsymbol{z}_i$ 是一个线性形状系数向量．给定 $k_i$、$s_i$ 以及 PCA 模型，则可由式 (1) 计算出相应的 $\boldsymbol{z}_i$，进而可将固定的 $\boldsymbol{z}_i$ 值代回式 (1) 求得轮廓 $L_i$．完整的人体姿态轮廓 $L$ 可表示为

$$L = \bigcup_{i=1}^{n_{\mathrm{p}}} L_i \tag{2}$$

## 3 利用多级动态模型进行人体姿态估计（Human pose estimation using the multi-level dynamic model）

本节介绍了利用多级动态模型对视频中的人体姿态进行估计及优化的具体实现过程．

### 3.1 图像中的人体姿态

本文将完整的姿态记为 $\boldsymbol{P}$，每个部位姿态记为 $\boldsymbol{P}_i$．假设在单帧图像中有 $n$ 个候选整体姿态，则需要构建一个评价函数来评价每个候选姿态的准确度．在评价过程中主要考虑候选姿态在图像中的表观特征与整个铰接模型的概率特征，选出表观特征的代价函数与模型概率特征代价函数均较小的姿态作为最优全局姿态．将每个全局模型样本在当前图像 $\boldsymbol{I}$ 中的代价函数 $C(\boldsymbol{I}_t, \boldsymbol{P})$ 表示为

$$C(\boldsymbol{I}_t, \boldsymbol{P}) = \lambda_\psi \sum_{i=1}^{n_{\mathrm{p}}} \psi(\boldsymbol{I}_t, \boldsymbol{P}_i) + \lambda_\phi \sum_{(i,j)\in\varepsilon} \phi_{i,j}(\boldsymbol{P}_i, \boldsymbol{P}_j, \Theta_{i,j}) \tag{3}$$

其中 $\varepsilon$ 为模型中相连接的部位对构成的集合，$\psi(\boldsymbol{I}_t, \boldsymbol{P}_i)$ 为部位姿态 $\boldsymbol{P}_i$ 在当前图像中的表观模型，$\phi_{i,j}(\boldsymbol{P}_i, \boldsymbol{P}_j, \Theta_{i,j})$ 为模型中部位成对的势能函数[19]，通过将相连接部位类比为弹簧能量模型来计算．

表观模型 $\psi(\boldsymbol{I}_t, \boldsymbol{P}_i)$ 考虑了图像中部位边缘处的轮廓以及当前部位区域的光流：

$$\psi(\boldsymbol{I}_t, \boldsymbol{P}_i) = \psi_{\mathrm{ct}}(\boldsymbol{I}_t | \boldsymbol{P}_i) + \psi_{\mathrm{fl}}(\boldsymbol{I}_t, \boldsymbol{I}_{t+1} | \boldsymbol{P}_i) \tag{4}$$

其中 $\psi_{\mathrm{ct}}(\boldsymbol{I}_t | \boldsymbol{P}_i)$ 代表轮廓响应，$\psi_{\mathrm{fl}}(\boldsymbol{I}_t, \boldsymbol{I}_{t+1} | \boldsymbol{P}_i)$ 代表光流响应．

采集含有不同姿态的图像样本并注释出其中真实的姿态位置，然后用方向梯度直方图（HOG）检测子沿每个部位姿态 $\boldsymbol{P}_i$ 的轮廓进行检测，得到的特征向量记为 $\boldsymbol{h}_i(\boldsymbol{I} | \boldsymbol{P}_i)$．利用 LIBSVM 工具箱[20] 针对每个部位姿态 $\boldsymbol{P}_i$ 训练支持向量机（SVM），训练结果记为 $\mathrm{svm}_i$，然后将训练得到的 $\mathrm{svm}_i$ 用于计算部位的轮廓响应 $\psi_{\mathrm{ct}}(\boldsymbol{I}_t | \boldsymbol{P}_i)$．

相邻两帧图像 $\boldsymbol{I}_t$、$\boldsymbol{I}_{t+1}$ 对应的光流图记为 $\boldsymbol{U}_t$．光流图 $\boldsymbol{U}_t$ 中的每个像素 $(x,y)$ 对应的 $\boldsymbol{U}_t(x,y)$ 含有 2 维数据，分别表示像素 $(x,y)$ 从图 $\boldsymbol{I}_t$ 到图 $\boldsymbol{I}_{t+1}$ 中对应像素的横轴方向与纵轴方向的位移．部位 $p_i$ 的光流响应 $\psi_{\mathrm{fl}}(\boldsymbol{I}_t, \boldsymbol{I}_{t+1} | \boldsymbol{P}_i)$ 为

$$\psi_{\mathrm{fl}}(\boldsymbol{I}_t, \boldsymbol{I}_{t+1} | \boldsymbol{P}_i) = \frac{1}{n} \sum_{(x,y)\subset R(i)} \sqrt{U_t^2(x,y,1) + U_t^2(x,y,2)} \tag{5}$$

其中 $R(i)$ 表示部位姿态 $\boldsymbol{P}_i$ 的轮廓 $L_i$ 所包含的区域，$n$ 表示区域 $R(i)$ 内 $(x,y)$ 像素对的个数．

### 3.2 候选部位姿态的获取

给定视频序列 $\boldsymbol{I} = (\boldsymbol{I}_1, \boldsymbol{I}_2, \cdots, \boldsymbol{I}_T)$，在每一帧图像 $\boldsymbol{I}_t$ 中用文 [6] 提出的方法计算出 $N$ 个最优的候选全局姿态 $\boldsymbol{P}^{t,n}$，并将这些姿态分解为部位姿态 $\boldsymbol{P}_i^{t,n}$，其中 $t$ 为图像索引，$n$ 表示第 $n$ 个候选姿态．在算法实现中本文取 $N = 5$．

在人体姿态中下臂的运动很灵活，是最难检测的部位，很多现有的方法需要产生足够多的候选姿态以保证其中含有较好的姿态样本．而本文方法并不依赖于海量的候选姿态，因此需要针对下臂采集附加的下臂部位姿态．由于下臂是各部位中运动最灵活的部位，而光流图最能反映两帧图像之间的运动情况，因此首先用相邻的两帧图像 $\boldsymbol{I}_t$、$\boldsymbol{I}_{t+1}$ 计算出运动光流图 $\boldsymbol{U}_t$，如图 3(b) 所示．但是光流图中可能包含各种运动，如背景运动、其他物体运动或遮挡等，因此需要对光流图进行进一步处理．本文用迁移学习[21] 的方法训练出手的探测滤波器[22]，并以 15° 为间隔旋转滤波器对光流图像进行滤波．在所得到的一系列响应图中，选择每个位置在所有响应图中的最大值构成手的概率图，见图 3(d)．在所得到的概率图中对可能的手的位置进行采样[22]，则可生成附加的下臂样本．

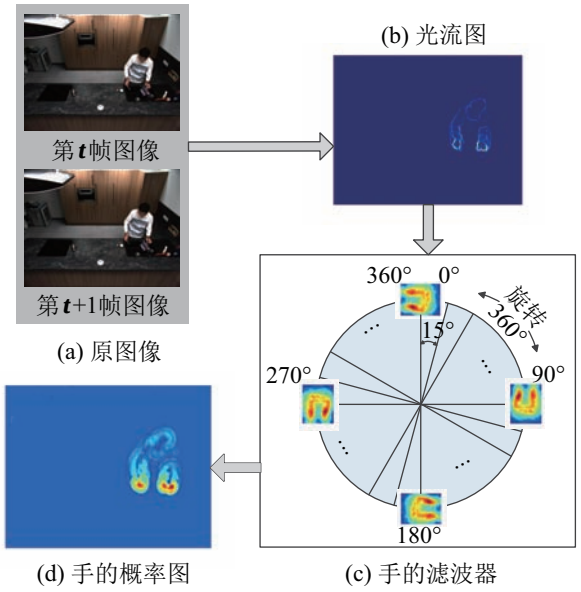



图 3　计算手的概率图

Fig.3　Calculation of the hand probability map

### 3.3　虚拟姿态损耗及目标函数

人体姿态的多变性及视频中人的位置及大小的不确定性使得姿态目标变化多端．为了得到更多的目标信息，一些现有的方法给视频提出限制，例如设置人体身高的像素范围或将场景固定在某个特定范围，如特定实验室或会议室环境[14]．这类方法能有效解决目标不确定的问题，但不具有普适性．本文不对视频环境作限制，而是探索一种能够有效利用视频中各帧之间的一致性信息来提高姿态估计精度的方法．

目前在视频姿态估计中常用的一致性限制方法是给肢体部位之间的角度与位置关系附加一定的动力学限制，如文 [16,18]．然而这种动力学限制通常是针对某一个视频集或某一种行为的，从而导致算法的针对性过强．此外还有一些学者采用精确裁剪及调整的图像，或者用特定数据集进行训练来获取能够表达视频一致性的代价函数，例如对运动幅度与范围进行学习后生成的动力学限制．然而图像裁剪的主要方法是背景减除法，但视频中变化的背景以及多变的肢端部位都会影响背景减除的精确度．前景中的人体姿态不完整会导致这类算法的精度较低．此外，在实际应用中，基于特定数据集训练得到的方法对于不同背景以及不同人物着装的视频普适性较差．针对上述提到的这些问题，本文方法不采用预先附加或训练得到的一致性限制，而是在检测的过程中动态构造出反映视频一致性的代价函数．

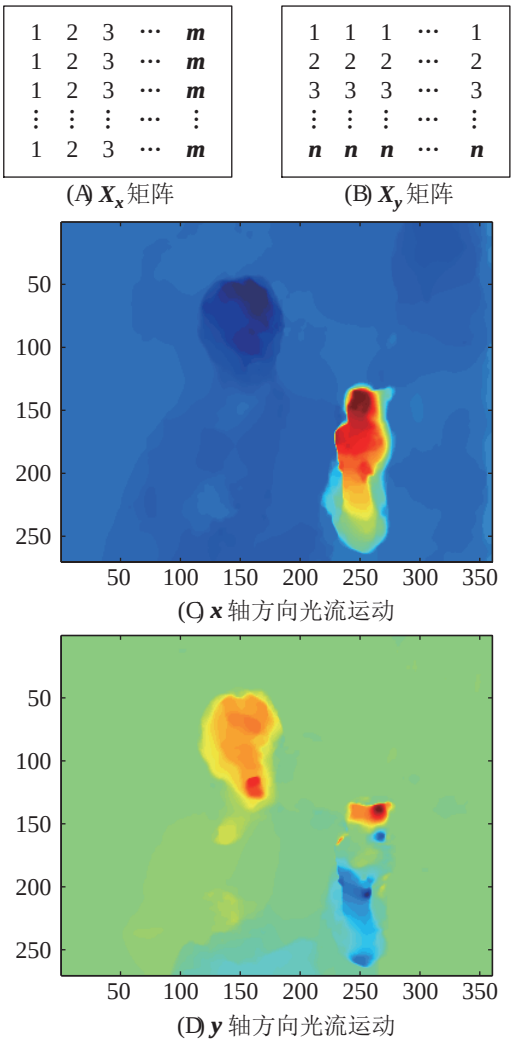



图 4　计算仿射运动模型

Fig.4　Calculation of the affine motion model

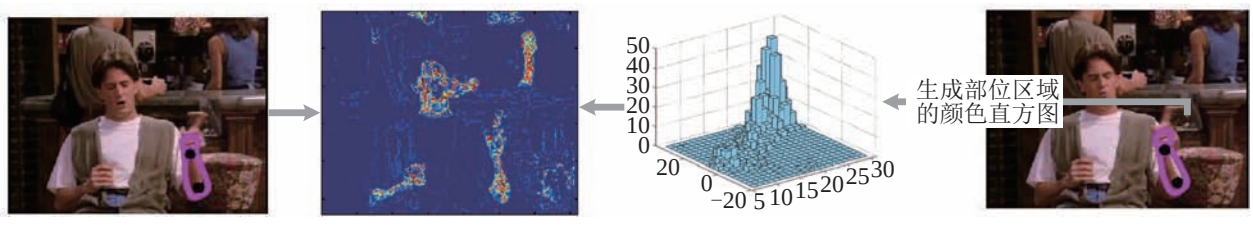



图 5　虚拟姿态颜色损耗的计算

Fig.5　Calculation of the virtual pose color cost

针对每个部位构造一个仿射运动模型，建立像素位置与像素运动之间的映射关系．给出相邻图像帧，如图 5(a)、(d) 所示，假设图像的尺寸为 $m\times n$，构造出如图 4(a) 所示的矩阵 $\boldsymbol{X}_x$ 和如图 4(b) 所示的矩阵 $\boldsymbol{X}_y$．2 帧相邻图像之间 $x$ 轴及 $y$ 轴方向的光流运动分别如图 4(c)、(d) 所示．

计算第 $t$ 帧中每个部位 $p_i$ 的仿射运动模型 $\boldsymbol{A}_i$ 时，对部位轮廓内的区域进行膨胀处理，屏蔽掉膨胀后的部位 $p_i$ 以外的区域，仅保留膨胀后的部位 $p_i$ 内的区域作为掩膜，记为 $\Omega_i$．然后将 $\boldsymbol{X}_x$ 中对应 $\Omega_i$ 区域内的数据转化为行向量 $\boldsymbol{x}_i$，将 $\boldsymbol{X}_y$ 对应 $\Omega_i$ 区域内的数据记为行向量 $\boldsymbol{y}_i$．此外，将 $\Omega_i$ 区域对应的 $x$ 轴及 $y$ 轴方向光流运动也分别排列为行向量，记为 $\boldsymbol{u}_{x,i}$ 与 $\boldsymbol{u}_{y,i}$．则第 $t$ 帧中部位 $p_i$ 的仿射运动模型 $\boldsymbol{A}_i$ 定义为

$$\boldsymbol{A}_i=\begin{bmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{bmatrix}=\begin{bmatrix}\boldsymbol{u}_{x,i}\\\boldsymbol{u}_{y,i}\\1\cdots\end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{x}_i\\\boldsymbol{y}_i\\1\cdots\end{bmatrix}^{-1}\tag{6}$$

对于任意一点 $(x_0,y_0)$，可用仿射运动模型 $\boldsymbol{A}_i$ 计算出相应 $x$ 轴与 $y$ 轴方向的运动 $u(x_0)$ 与 $u(y_0)$：

$$\begin{cases}u(x_0)=a_{11}\cdot x_0+a_{12}\cdot y_0+a_{13}\\u(y_0)=a_{21}\cdot x_0+a_{22}\cdot y_0+a_{23}\end{cases}\tag{7}$$

用式 (7) 将部位姿态 $\boldsymbol{P}_i^t$ 的关键点 $k_i$ 从第 $t$ 帧传递到第 $t+1$ 帧，记为 $\tilde{k}_{i+1}$．再利用式 (1) 计算 $t+1$ 帧中相应的部位轮廓 $\tilde{L}_{i+1}$，从而得到由第 $t$ 帧中的部位姿态 $\boldsymbol{P}_i^t$ 传递到第 $t+1$ 帧后的部位姿态 $\tilde{\boldsymbol{P}}_i^{t+1}$．

由于 $\tilde{\boldsymbol{P}}_i^{t+1}$ 是用仿射运动模型得到的，而不是在第 $t$ 帧中检测得到的，因此将 $\tilde{\boldsymbol{P}}_i^{t+1}$ 称为虚拟姿态．尽管此时没有对虚拟姿态 $\tilde{\boldsymbol{P}}_i^{t+1}$ 在图像 $t+1$ 中的准确度进行评价，但其反映了图像 $t$ 与 $t+1$ 的连续性，仍然可以通过 $\tilde{\boldsymbol{P}}_i^{t+1}$ 剔除图像 $t$ 中那些由于背景变换或光照强度变化而产生的违反视频一致性的误估计候选部位姿态．4.2 节实验分析部分验证了虚拟姿态项的有效性．

利用虚拟姿态 $\tilde{\boldsymbol{P}}_i^{t+1}$ 构建出一个反映姿态连续性的损耗项，称为虚拟姿态损耗 $V_{\rm cost}$（virtual pose cost）．虚拟姿态损耗的目的是惩罚那些违反视频一致性的误检测部位姿态样本，而对正确估计的部位样本不产生明显影响．在构建虚拟姿态损耗项时，主要考虑颜色一致性损耗和运动一致性损耗：

$$\begin{aligned}V_{\rm cost}(\boldsymbol{I}_t|\boldsymbol{P}_i^t)=&\lambda_{\rm c}\cdot f_{\rm c}(\boldsymbol{I}_t,\boldsymbol{I}_{t+1}|\boldsymbol{P}_i^t,\tilde{\boldsymbol{P}}_i^{t+1})+\\&\lambda_{\rm m}\cdot f_{\rm m}(\boldsymbol{I}_t,\boldsymbol{I}_{t+1}|\boldsymbol{P}_i^t,\tilde{\boldsymbol{P}}_i^{t+1})\end{aligned}\tag{8}$$

其中，第 1 项中的 $f_{\rm c}(\boldsymbol{I}_t,\boldsymbol{I}_{t+1}|\boldsymbol{P}_i^t,\tilde{\boldsymbol{P}}_i^{t+1})$ 为利用虚拟姿态计算的颜色一致性损耗，简称颜色损耗；第 2 项中的 $f_{\rm m}(\boldsymbol{I}_t,\boldsymbol{I}_{t+1}|\boldsymbol{P}_i^t,\tilde{\boldsymbol{P}}_i^{t+1})$ 为利用虚拟姿态计算的运动一致性损耗，简称运动损耗．$\lambda_{\rm c}$ 与 $\lambda_{\rm m}$ 分别为颜色损耗和运动损耗的权重．下面详细说明虚拟姿态损耗的计算过程．

1) 颜色损耗．将图像 $\boldsymbol{I}_{t+1}$ 由 RGB 颜色空间转换成为 CIE $l*a*b$ 颜色空间．CIE $l*a*b$ 颜色空间中 $l$ 维表示亮度，$a$ 与 $b$ 维表示颜色．为了排除光线较弱时产生的色度畸变，剔除掉 $l<0.3$ 的像素，然后针对每个部位的区域，利用剩余像素计算出颜色直方图并存储，如图 5(c) 所示．在计算第 $t$ 帧中部位 $p_i$ 的颜色损耗时，用 $\tilde{\boldsymbol{P}}_i^{t+1}$ 区域内的像素计算出的颜色直方图（图 5(c)）计算出第 $t$ 帧的颜色概率图（图 5(b)）．所得颜色概率图中部位 $\boldsymbol{P}_i^t$ 对应的区域内像素的均值即为颜色损耗值．

2) 运动损耗．计算出部位 $p_i$ 的每个关键点在 $\boldsymbol{U}_t$ 中的 $r$ 邻域内的 $x$ 及 $y$ 方向的光流最大值，记为 $\boldsymbol{u}(x,y)$；计算出部位姿态 $\boldsymbol{P}_i^t$ 的关键点与 $\tilde{\boldsymbol{P}}_i^{t+1}$ 的关键点之间的位移，记为 $\boldsymbol{v}(x,y)$．则运动损耗为

$$\begin{aligned}\widehat{m}&=\sum\|\boldsymbol{u}(x,y)-\boldsymbol{v}(x,y)\|_2\\f_{\rm m}(\boldsymbol{I}_t,\boldsymbol{I}_{t+1}|\boldsymbol{P}_i^t,\tilde{\boldsymbol{P}}_i^{t+1})&=\begin{cases}\widehat{m}, & \text{if } \widehat{m}<\delta_{\rm f}\\\delta_{\rm f}, & \text{otherwise}\end{cases}\end{aligned}\tag{9}$$

式 (9) 表示运动损耗为部位 $p_i$ 中各关键点的 $\boldsymbol{u}(x,y)$ 与 $\boldsymbol{v}(x,y)$ 之间欧氏距离之和．其中 $\delta_{\rm f}$ 表示常量阈值，设置 $\delta_{\rm f}=40$．

自此，得到了多级动态模型中第 3 级的虚拟姿态损耗的计算方法．对于第 2 级中的每个候选部位姿态，用计算出的虚拟姿态损耗以及第 4 级中的轮廓、光流特征来计算候选部位姿态的得分 $P_{\rm cost}$（part cost）：

$$P_{\rm cost}(\boldsymbol{I}_t|\boldsymbol{P}_i^t)=V_{\rm cost}(\boldsymbol{I}_t|\boldsymbol{P}_i^t)+\psi(\boldsymbol{I}_t,\boldsymbol{P}_i^t)\tag{10}$$

其中 $V_{\rm cost}(\boldsymbol{I}_t|\boldsymbol{P}_i^t)$ 为虚拟姿态损耗（式 (8)），$\psi(\boldsymbol{I}_t,\boldsymbol{P}_i^t)$ 为部位 $p_i$ 的表观模型（式 (4)）．利用式 (10) 可计算并记录每个部位姿态的得分，并用于后续的优化．

### 3.4 部位姿态的优化及最优整体姿态的生成

现有的姿态估计算法大多依赖于庞大的候选样本集，用构造出的合适的目标函数，从海量候选姿态中选出最优姿态，如文 [10-11]．而本文方法仅产生少量候选姿态（如 3.2 节所述），为得到最优姿态值，本文采用粒子群优化算法 [25] 对产生的候选姿态样本进行优化．粒子群优化算法是一种基于种群

的随机算法，优化过程中不需要选择或剔除粒子，而是利用所有种群个体的移动来选择最优位置.

本文为每个部位 $p_i$ 定义一个粒子 $\boldsymbol{X}_i$，表示为

$$\boldsymbol{X}_i = [k_i, s_i] \tag{11}$$

其中 $k_i$ 与 $s_i$ 分别为式 (1) 中定义的关键点的位置及尺度参数. 通过式 (10) 计算出每个粒子的得分，并在优化过程中记录下每一个粒子所经历过的得分最高的位置，记为 $\boldsymbol{X}^{\mathrm{BL}}$，和所有粒子所经历的得分最高的位置，记为 $\boldsymbol{X}^{\mathrm{BG}}$. 对于状态空间中的每一维，粒子的运动速度 $\boldsymbol{V}_{i,d}$ 的迭代公式为

$$\begin{aligned}\boldsymbol{V}_{i,d} = w\cdot\boldsymbol{V}_{i,d} + c_1\cdot r_1\cdot(\boldsymbol{X}^{\mathrm{BL}}_{i,d} - \boldsymbol{X}_{i,d}) + \\ c_2\cdot r_2\cdot(\boldsymbol{X}^{\mathrm{BG}}_{i,d} - \boldsymbol{X}_{i,d})\end{aligned} \tag{12}$$

其中，下标 $d$ 表示第 $d$ 维，$w$ 称为惯性权重，$c_1$ 与 $c_2$ 为两个加速常数，随机数 $r_1, r_2 \in [0,1]$. 粒子位置的迭代更新公式为

$$\boldsymbol{X}_{i,d} = \boldsymbol{X}_{i,d} + \boldsymbol{V}_{i,d} \tag{13}$$

算法进行 $M$ 次迭代，在迭代前和每次迭代后分别将得到的最优部位姿态记录为一组. 4.3 节将对迭代次数 $M$ 进行分析. 对于每组部位中相铰接的部位 $p_i$ 与 $p_j$，用由上述优化方法得到的铰接关键点位置 $k_i$ 与 $k_j$ 计算得到铰接关节 $k_0$：

$$k_0 = k_i + \frac{1}{10}(k_j - k_i);\quad k_i = k_0,\ k_j = k_0 \tag{14}$$

其中 $k_i$ 表示部位 $p_i$ 中对应的铰接关键点，$k_j$ 表示部位 $p_j$ 中对应的铰接关键点.

本文将部位连接的先后顺序定义为：躯干、头部、左上臂、右上臂、左下臂、右下臂. 在利用式 (14) 更新两个铰接部位 $p_i$ 与 $p_j$ 的铰接关键点时，将顺序靠前的部位视为 $p_i$，靠后的视为 $p_j$.

对于每组部位中的每个部位姿态 $\boldsymbol{P}^t_i$，将优化得到的关键点位置 $k_i$ 与尺度 $s_i$ 代入式 (1) 计算得到部位轮廓 $L_i$，进而通过式 (2) 获得整体姿态 $\boldsymbol{P}^t$ 的轮廓 $L$. 然后利用式 (3) 中的代价函数 $C(\boldsymbol{I}_t, \boldsymbol{P}^t)$ 选出所有姿态组中得到的最优整体姿态 $\boldsymbol{P}^t$ 作为第 $t$ 帧图像中的人体姿态估计结果.

得到第 $t$ 帧的最优整体姿态 $\boldsymbol{P}^t$ 后，依据光流运动 $\boldsymbol{U}_t$ 将 $\boldsymbol{P}^t$ 传递到 $t+1$ 帧中，作为第 $t+1$ 帧中附加的整体候选姿态 $\boldsymbol{P}^{t+1}_i$，并用于后续的分解与优化.

## 4　实验分析（Experiments and analysis）

本文采用马克斯 • 普朗克计算机科学研究所发布的具有挑战性的身体姿态数据集 MPII Cooking Activities[26] 来验证所提出的人体姿态估计算法. 该数据集中包含不同性别、不同年龄的多种上半身人体数据，并含有多达 65 种操作行为的姿态.

### 4.1　对比实验

为验证本文提出的基于多级动态模型的人体姿态估计算法，本节采用 MPII Cooking Activities 数据集中的 20 段视频来进行验证实验，其中每段视频包含约 100 帧图像. 同时用 6 个相关的前沿算法进行对比，用对比实验证明本文算法有效地提高了现有方法的姿态估计精度.

用 MPII Cooking Activities 数据集提供的姿态关节真实值注释作为评价基准，并用头、颈、左肩、右肩、左肘、右肘、左腕以及右腕这 8 个关键点的精度来说明姿态估计的精度. MPII Cooking Activities 数据集[26] 提出的评价方法为：当估计得到的部位中所有关键点位置（躯干含有 4 个关键点，其他部位各含有 2 个关键点，如图 2(b) 所示）到真实值之间的距离均小于此部位的长度的一半时，则判断此部位为准确估计，反之视为误估计. 对于某个部位 $p_i$，用视频中准确估计的数量占总数的百分比作为此部位的估计精度. 这种评价方法不设定明确的像素值作为阈值，而是用部位长度的一半作为阈值，从而使评价结果不受图像中人物尺寸的像素值或图片大小的影响，并能用于有效评价多个不同视频序列的姿态估计精度. 本文采用文 [26] 的评价算法，但由于我们更关注关键点的准确度，而有些关键点（如肘关节）属于 2 个部位，因此对于铰接关键点，采用 2 个铰接部位的估计精度的平均值作为对应关键点的精度.

所采用的对比方法有：1) 基于灵活的 PS 模型变体的姿态估计方法[26]；2) 基于级联模型的姿态估计算法[12]；3) 基于图形模型的姿态估计算法[13]；4) 基于 $N$ 个最优解码器的姿态估计算法[11]；5) 基于流动木偶（flowing puppet）的姿态估计算法[19]；6) 基于灵活混合部位的姿态估计算法[6]. 表 1 中列出了各方法的人体姿态估计精度及比较. 其中文 [26] 使用了与本文相同的数据集进行实验验证，并且在同样的数据集下与文 [6,12-13] 中的方法进行了对比. 表 1 中文 [6,12-13,26] 方法的数据来源于文 [26]. 文 [11,19] 中使用的数据集与本文不同，因此本文使用其公开的算法源代码，依据原文的参数调节与运行方法对本文使用的验证数据集进行计算，并用本文的评价方法得到关节的姿态估计精度，如表 1 所示.

表 1 姿态估计精度比较

Tab.1 Comparison of pose estimation precision

| 姿态估计方法 | 姿态关键点的估计精度 /% | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 头 | 颈 | 左肩 | 右肩 | 左肘 | 右肘 | 左腕 | 右腕 |
| 本文 | **81.6** | **82.3** | **77.3** | **77.6** | **69.2** | **69.7** | **66.1** | **66.4** |
| 文 [26] | 79.4 | 78.5 | 71.3 | 70.2 | 62.6 | 62.2 | 61.0 | 62.4 |
| 文 [12] | - | 67.1 | 57.9 | 60.3 | 42.8 | 50.4 | 37.0 | 47.3 |
| 文 [13] | 80.0 | 80.1 | 74.9 | 74.0 | 59.6 | 58.4 | 49.6 | 48.9 |
| 文 [11] | 74.2 | 79.3 | 68.3 | 68.1 | 54.4 | 54.2 | 51.6 | 51.4 |
| 文 [19] | 76.3 | 81.6 | 72.4 | 72.2 | 59.5 | 59.6 | 55.9 | 56.4 |
| 文 [6] | 67.7 | 79.6 | 70.2 | 70.1 | 55.6 | 55.4 | 50.3 | 50.1 |

从表 1 可以看出，与其他相关的前沿算法相比，本文的算法能够更有效地估计头、颈及肩关节的位置，并且对肘以及腕的估计精度有显著提高．由表 1 可知，本文的算法与文 [13, 26] 的方法相比，肘关节精度分别提高了约 10%与 7%，腕关节分别提高了 17%与 5%．这是由于文 [13] 采用了基于图形模型的结构，文 [26] 对基于图形模型的结构进行了改进，但这两种方法在计算姿态样本的代价函数时，都将人体姿态作为整体进行计算．在这种情况下，假若有两个姿态样本，一个样本的左臂正确而右臂错误，另一个与之相反，则用文 [13] 与文 [26] 的代价函数无法选出最优姿态．然而本文的可分解与重构的多级模型结构将整体姿态样本分解为部位姿态样本，对每个部位姿态样本单独计算代价函数，从而解决了上述问题．

另外，如表 1 所示，本文算法的肘关节精度比文 [6]、文 [12] 和文 [9] 的方法分别提高了 14%、22%和 10%，腕关节精度分别提高了 16%、24%和 10%．这是由于本文提出的通过构建虚拟姿态并计算其损耗来剔除姿态估计样本中的不一致样本的方法，惩罚了运动变化大于光流值以及颜色不一致的部位姿态，与文 [6]、文 [12] 和文 [9] 相比有效保持了相邻帧之间姿态的连续性，从而使姿态估计结果更有效．此外，从表 1 中可读出本文算法的肘关节和腕关节的姿态精度与文 [11] 方法相比均提高了约 15%．这是由于本文有效使用粒子群优化算法，用较少的样本优化出最优姿态，与产生大量姿态样本再用非极大值抑制进行处理的文 [11] 方法相比，优化对象的针对性更强且优化效果更好．

表 1 中的实验结果表明，本文提出的基于多级动态模型的姿态估计算法有效提高了人体姿态估计的准确度．

本文针对部位姿态的估计与优化提出的关键算法有：1) 构建虚拟姿态并计算虚拟姿态损耗的代价函数；2) 采用粒子群优化算法对部位姿态进行优化．下面将针对这两个关键算法分别进行实验，用实验证明所提出的算法的有效性．

### 4.2 虚拟姿态损耗项的有效性实验

本文利用连续图像间的一致性构建虚拟姿态来计算候选部位姿态的附加损耗，从而得到更有效的目标函数．为验证所构建的虚拟姿态损耗项的有效性，从 MPII Cooking Activities 数据集中选出具有挑战性的 202 帧图像进行验证，并主要比较灵活性最高的部位，即左、右下臂的估计精度．分别在有、无虚拟姿态损耗项，以及只有一项虚拟姿态损耗项的情况下进行验证实验，估计图像中肘关节及腕关节的位置．

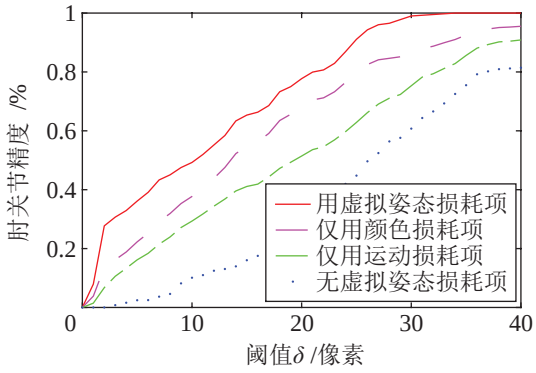
为更明显地观测出姿态估计的精度及虚拟姿态损耗项的有效性，首先用 0 ～ 40 个像素的距离分别作为阈值 $\delta$，计算每帧图像中每个关键点 $k_j$ 与数据集中注释的关键点位置 $\boldsymbol{g}_j$ 之间的欧氏距离，记为 $d_j$，然后用小于阈值 $\delta$ 的 $d_j$ 个数占总数的百分比作为关键点 $j$ 的姿态估计精度，记为 $a_j$:

$$a_j = \frac{\text{count}(d_j < \delta)}{\text{count}(d_j)} \cdot 100\% \tag{15}$$

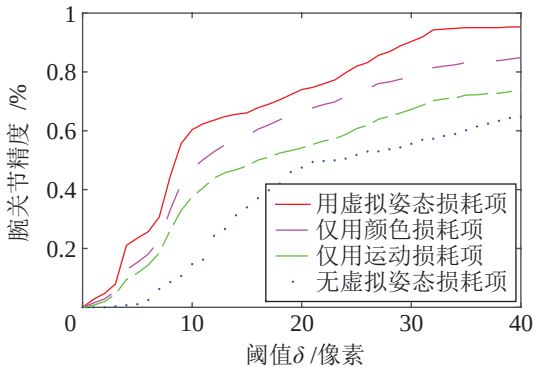
图 6(a) 为肘关节精度，是左右肘关节估计精度的平均值；图 6(b) 为腕关节精度，是左右腕关节估计精度的平均值．图 6 中蓝色点线表示未使用本文提出的虚拟损耗项的姿态估计结果．粉色的点划线与绿色的虚线分别表示仅使用颜色损耗与仅使用运动损耗时的姿态估计精度，而红色实线表示使用本文提出的完整的虚拟姿态损耗项进行人体姿态估计得到的精度．从图中可看出，对于每个阈值 $\delta$，本文提出的虚拟姿态损耗项中的颜色项及运动项都能有效提高姿态估计的精度，并且同时使用虚拟姿态损耗中的颜色项及运动项能够更大地提高姿态估计精度．

以图 6 中的阈值 $\delta = 20$ 为例，在未应用虚拟姿态损耗项的情况下对姿态进行估计时，肘关节误差小于 20 像素的姿态估计结果仅占约 27%，腕关节误差小于 20 像素的结果仅占 45%．使用运动损耗项后，算法能够惩罚那些相邻帧中运动幅度大于光流运动的姿态样本，从而提高姿态估计的准确度．从图 6 中的绿色虚线可见，使用运动损耗项后肘关节误差小于 20 像素的姿态估计结果约占 50%，腕关节误差小于 20 像素的结果约占 53%．使用颜色损耗项后，算法能够惩罚相邻帧中部位颜色不一致的姿态样本，从而剔除产生的误估计姿态样本．从图 6 中的粉色点划线可见，使用颜色损耗项后肘

关节误差小于 20 像素的结果约占 66%，腕关节误差小于 20 像素的结果约占 65%．而使用完整的虚拟姿态损耗项，即同时考虑运动损耗及颜色损耗进行姿态估计时，准确度大幅提高，如图 6 中红色实线所示，肘关节误差小于 20 像素的姿态估计结果约占 75%，腕关节误差小于 20 像素的结果约占 72%．




(A) 肘关节精度




(B) 腕关节精度

图 6　虚拟姿态损耗项有效性实验

Fig.6　Efficiency of the virtual pose cost items

上述实验证明，本文提出的虚拟姿态损耗项能够有效提高视频中人体姿态估计的精度．

### 4.3　粒子群优化算法的有效性实验

粒子群优化算法是一种迭代优化算法，可以通过对较小的样本集进行迭代优化来得到最优的估计值．为了验证在多级动态模型中使用粒子群优化算法的有效性，本文用约 500 帧图像在未使用粒子群优化算法的情况下以及使用不同迭代次数的粒子群优化算法的情况下分别进行实验．我们选择最难估计的关节，即肘关节与腕关节，来展示不同迭代次数的粒子群优化的实验效果．本实验采用 4.1 节中介绍的姿态精度评价方法来计算关节位置估计的准确度．

实验效果如图 7 所示，其中横轴为粒子群优化算法的迭代次数 $M$，纵轴为关节精度的百分数．从图 7 中可以看出，由于本文算法并不是简单的基于大量候选姿态的选择[24]，因此当未使用粒子群优化算法时，算法只能从已生成的有限候选部位姿态中选择最优姿态，从而导致姿态估计精度不理想．而使用粒子群优化算法后，算法可记录并指导粒子群的运动，从而向最优姿态的关节位置逼近．另外从图 7 中还可以看出，随迭代次数的增加，姿态的精度会提高，但在迭代次数大于 5 后精度提升不明显，因此本文算法取迭代次数 $M = 5$.

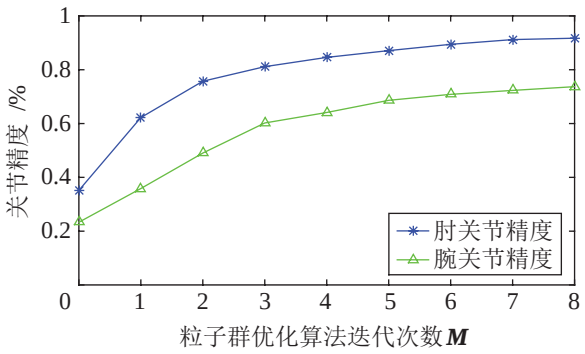



图 7　粒子群优化算法的有效性实验

Fig.7　Efficiency of the particle swarm optimization

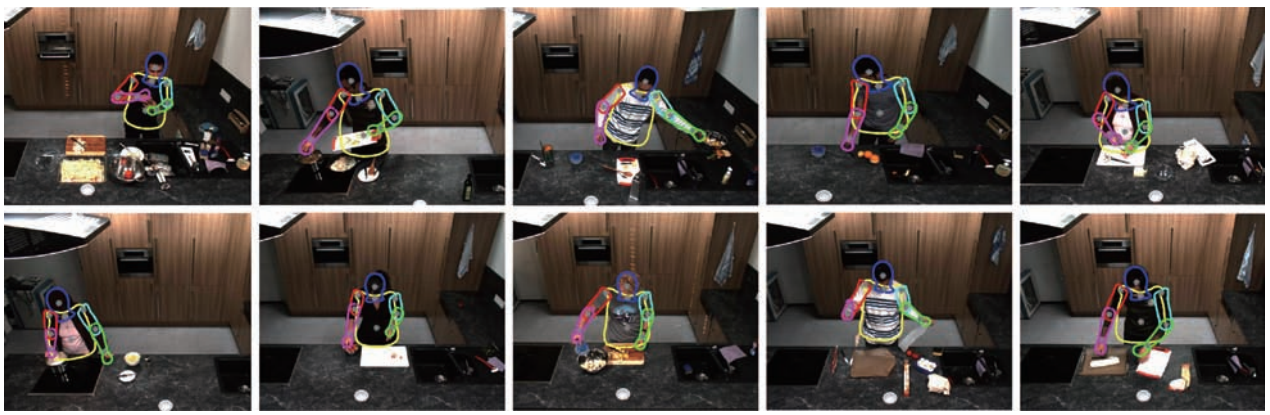


图 8　人体姿态估计结果实例

Fig.8　Samples of human pose estimation results

### 4.4 算法复杂度分析

算法 1 列出了本文算法的关键步骤. 算法的复杂度由生成的候选姿态的数目 $N_1$、附加下臂姿态数目 $N_2$ 与身体部位数 $n_{\mathrm{p}}$ 所决定. 由算法 1 可计算出本文提出的基于多级动态模型的 2D 人体姿态估计算法的复杂度为 $O((N_1+N_2+1)\times n_{\mathrm{p}})$. 为实现与验证本文算法，在内核为英特尔 i5 3570，主频为 3.4 GHz 的计算机上使用 MATLAB R2012b 进行编程，得到计算速度平均约为 0.25 帧/秒.

**算法 1：基于多级动态模型的 2D 人体姿态估计算法**

给定含有 $T$ 帧图像的视频序列 $\{t=1,2,\cdots,T\}$；
为每帧计算出 $N_1$ 个候选全局姿态样本 $\boldsymbol{P}^t$；
**for** $t=1:T-1$ **do**
  将全局姿态 $\boldsymbol{P}^t$ 分解为部位姿态 $\boldsymbol{P}_i^t$；
    ▷ $i$ 为部位的索引号，$i=1,2,\cdots,n_{\mathrm{p}}$
  生成 $N_2$ 个附加的下臂姿态样本；
  **for** $i=1,2,\cdots,n_{\mathrm{p}}$ **do**
    构造出第 $t+1$ 帧的虚拟姿态 $\tilde{\boldsymbol{P}}_i^{t+1}$；
    求出每个 $\boldsymbol{P}_i^t$ 的虚拟姿态损耗；  ▷ 式 (8)
    计算出每个 $\boldsymbol{P}_i^t$ 的得分；  ▷ 式 (10)
    利用粒子群优化算法对 $\boldsymbol{P}_i^t$ 进行 $M$ 次迭代优化，得到 $M$ 组最优部位姿态；
  **end for**
  将 $M$ 组 $\boldsymbol{P}_i^t$ 重组为 $\boldsymbol{P}^t$；  ▷ 式 (1) 和 (2)
  选出最优 $\boldsymbol{P}^t$ 作为姿态估计结果；  ▷ 式 (3)
  用光流运动将第 $t$ 帧的最优 $\boldsymbol{P}^t$ 传递至 $t+1$ 帧作为附加 $\boldsymbol{P}^{t+1}$ 样本.
**end for**
**if** $t==T$ **then**
  从所有 $\boldsymbol{P}^t$ 候选样本中选出最优 $\boldsymbol{P}^t$ 作为姿态估计结果.  ▷ 式 (3)
**end if**

### 4.5 姿态估计结果实例

图 8 展示了用本文算法得到的姿态估计结果的实例. 从图中可以看出，本文算法对视频中不同服装颜色、不同身高、不同性别、不同行为等的人体姿态都能作出有效的估计.

## 5 结论及展望（Conclusions and future works）

针对 2D 视频图像中人体上半身姿态的自动识别问题，提出了一种基于多级动态模型的姿态估计算法. 首先，本文方法将完整的人体姿态模型分解成为单个的身体部位模型，并对每个身体部位单独进行优化. 这种方法解决了用目标函数估计全局姿态时容易产生歧义性的问题. 此外，针对视频序列中姿态的连续性提出了一种利用虚拟姿态计算一致性损耗的方法，实验证明虚拟姿态损耗项能够有效剔除误估计的候选姿态. 另外，有效利用了粒子群优化算法，因而可以用有限的候选姿态逼近最优姿态. 本文用挑战性的人体姿态估计数据集与 6 种前沿的姿态估计算法进行实验对比，实验证明本文算法有效提高了视频中人体姿态估计的准确度.

在未来的工作中，本文提出的人体姿态估计算法将应用于液压四足仿生机器人[27-28]，协助其理解人类的行为并实现与人员的交互.

## 参考文献（References）


[1] Dautenhahn K. Socially intelligent robots: Dimensions of human-robot interaction[J]. Philosophical Transactions of the Royal Society of London, B: Biological Sciences, 2007, 362(1480): 679-704.

[2] Atkeson C G, Hale J G, Pollick F E, et al. Using humanoid robots to study human behavior[J]. IEEE Intelligent Systems and Their Applications, 2000, 15(4): 46-55.

[3] 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[J]. 机器人，2014，36(3)：285-292.
Tian G H, Yin J Q, Han X, et al. A novel human activity recognition method using joint points information[J]. Robot, 2014, 36(3): 285-292.

[4] Yang Y Z, Li Y, Fermüller C, et al. Robot learning manipulation action plans by "watching" unconstrained videos from the World Wide Web[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. 2015: 3686-3693.

[5] Koppula H S, Gupta R, Saxena A. Learning human activities and object affordances from RGB-D videos[J]. International Journal of Robotics Research, 2013, 32(8): 951-970.

[6] Yang Y, Ramanan D. Articulated human detection with flexible mixtures of parts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2878-2890.

[7] Dantone M, Gall J, Leistner C, et al. Human pose estimation using body parts dependent joint regressors[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2013: 3041-3048.

[8] Fischler M A, Elschlager R A. The representation and matching of pictorial structures[J]. IEEE Transactions on Computers, 1973, 22(1): 67-92.

[9] Freifeld O, Weiss A, Zuffi S, et al. Contour people: A parameterized model of 2D articulated human shape[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2010: 639-646.

[10] Zuffi S, Freifeld O, Black M J. From pictorial structures to deformable structures[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2012: 3546-3553.

[11] Park D, Ramanan D. N-best maximal decoders for part models [C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2627-2634.

[12] Sapp B, Toshev A, Taskar B. Cascaded models for articulated pose estimation[C]//11th European Conference on Computer Vision. Berlin, Germany: Springer, 2010: 406-420.

[13] Andriluka M, Roth S, Schiele B. Pictorial structures revisited: People detection and articulated pose estimation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2009: 1014-1021.

[14] Ramanan D, Forsyth D A, Zisserman A. Strike a pose: Tracking people by finding stylized poses[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2005: 271-278.

[15] Lee M W, Nevatia R. Human pose tracking using multi-level structured models[C]//9th European Conference on Computer Vision. Berlin, Germany: Springer, 2006: 368-381.

[16] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition[J]. International Journal of Computer Vision, 2005, 61(1): 55-79.

[17] Sigal L, Black M J. Measure locally, reason globally: Occlusion-sensitive articulated pose estimation[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2006: 2041-2048.

[18] Anguelov D, Srinivasan P, Koller D, et al. SCAPE: Shape completion and animation of people[J]. ACM Transactions on Graphics, 2005, 24(3): 408-416.

[19] Zuffi S, Romero J, Schmid C, et al. Estimating human pose with flowing puppets[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2013: 3312-3319.

[20] Chang C C, Lin C J. LIBSVM: A library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): No.27.

[21] Pan S J, Yang Q. A survey on transfer learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[22] Sapp B, Weiss D, Taskar B. Parsing human motion with stretchable models[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2011: 1281-1288.

[23] Sigal L, Black M J. Predicting 3D people from 2D pictures[C]//4th International Conference on Articulated Motion and Deformable Objects. Berlin, Germany: Springer, 2006: 185-195.

[24] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-parts[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2011: 1385-1392.

[25] Poli R, Kennedy J, Blackwell T. Particle swarm optimization: An overview[J]. Swarm Intelligence, 2007, 1(1): 33-57.

[26] Rohrbach M, Amin S, Andriluka M, et al. A database for fine grained activity detection of cooking activities[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2012: 1194-1201.

[27] 柴汇，孟健，荣学文，等. 高性能液压驱动四足机器人 SCalf 的设计与实现[J]. 机器人，2014，36(4)：385-391.
Chai H, Meng J, Rong X W, et al. Design and implementation of SCalf, an advanced hydraulic quadruped robot[J]. Robot, 2014, 36(4): 385-391.

[28] 张慧，荣学文，李贻斌，等. 四足机器人地形识别与路径规划算法[J]. 机器人，2015，37(5)：546-556.
Zhang H, Rong X W, Li Y B, et al. Terrain recognition and path planning for quadruped robot[J]. Robot, 2015, 37(5): 546-556.

## 作者简介：

马　淼（1989–），女，博士生. 研究领域：机器视觉，智能机器人，模式识别与智能系统.

李贻斌（1960–），男，博士，教授. 研究领域：智能机器人，特种机器人，智能车辆.