

## 图像和视频中基于部件检测器的人体姿态估计

苏延超<sup>\*①</sup> 艾海舟<sup>①</sup> 劳世站<sup>②</sup>

<sup>①</sup>(清华大学计算机与科学技术系 北京 100084)

<sup>②</sup>(欧姆龙公司核心科技中心 京都 619-0283)

**摘要:** 人体姿态估计是计算机视觉领域中的一个关键问题, 可以应用于人体活动分析, 人机交互以及视频监视等方面。该文针对单目图像和视频中的人体姿态估计问题, 基于部件及图推理的方法, 对观测模型和推理方法提出改进。该文设计实现了一种旋转不变的边缘力场特征, 采用基于边缘力场特征的 Boosting 分类器作为观测模型, 并利用一种基于粒子采样和置信度传播的优化算法进行姿态估计。算法的性能和速度在几个数据集上得到了验证。

**关键词:** 图像处理; 人体姿态估计; 边缘力场特征; 置信度传播

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2011)06-1413-07

DOI: 10.3724/SP.J.1146.2010.01042

## Part Detector Based Human Pose Estimation in Images and Videos

Su Yan-chao<sup>①</sup> Ai Hai-zhou<sup>①</sup> Lao Shi-hong<sup>②</sup>

<sup>①</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China)

<sup>②</sup>(Core Technology Center, OMRON Corporation, Kyoto 619-0283, Japan)

**Abstract:** Human pose estimation is an essential issue in computer vision area since it has many applications such as human activity analysis, human computer interaction and visual surveillance. In this paper, 2D human estimation issue in monocular images and videos is addressed. The observation model and the inference method are improved based on part based graph inference method. A rotation invariant edge field feature is designed and based on which a Boosting classifier is learnt as the observation model. The human pose estimation is done with a particle based belief propagation inference method. Experiments show the effectiveness and the speed of the proposed method.

**Key words:** Image processing; Human pose estimation; Edge field feature; Belief Propagation (BP)

### 1 引言

人体姿态估计可以应用于人体活动分析, 人机交互以及视觉监视等领域, 是近期计算机视觉领域中的一个热门问题。

然而, 现阶段单目人体姿态估计尚未有成型的解决方案, 这是因为: 首先, 人体由多个部件组成, 运动情况十分复杂, 且在单目图片和视频中缺少 3 维信息, 使得人体 2 维姿态的变化不能简单的用一个统一的模型描述; 其次, 在不同的场景中, 图像和视频中的背景及光照会有很大的变化, 而人体本身的服饰也会发生变化, 从而导致人体的表现发生剧烈的变化, 很难被统一的模型所描述。

现有的工作大致可以分为基于整体的方法和基于部件的方法。基于整体的方法利用图像中人体的

整体表现或者分割结果, 估计人体的姿态。基于整体的方法又可以分为基于回归的方法和基于样例的方法。基于回归的方法<sup>[1,2]</sup>将姿态估计问题看作从人体表现到人体姿态的直接映射, 根据大量样本学习从图像特征或分割结果到人体姿态的回归函数。基于样例的方法<sup>[3]</sup>则在用一个具有代表性的样例集合来表示姿态空间, 并利用图像特征对这些样例进行编码, 进行姿态估计时提取出待估计图像中的图像特征, 寻找样例集合中最为相似的一个或多个样例作为姿态估计结果。基于整体的方法通常具有较高的效率, 但由于人体姿态变化过于复杂, 利用有限样本或者整体模型难以覆盖整个姿态空间, 基于整体的方法往往只适合于特定姿态的估计。

基于部件的方法将人体分为若干相互关联的部件, 利用图模型表示整个人体, 并利用图推理的方法对人体姿态进行优化。基于部件的方法中有 3 个要点, 分别是图模型, 部件的观测模型和优化算法。图模型用于表示部件之间的约束关系, 其中最常用

2010-09-25 收到, 2011-01-11 改回

国家自然科学基金(61075026)资助课题

\*通信作者: 苏延超 syc02@mails.tsinghua.edu.cn

的是树形模型<sup>[4-9]</sup>, 树形模型根据部件的连接关系定义, 最为直观。文献[10]则在树形模型的基础上, 引入不相邻部件之间的约束, 解决姿态估计中的自遮挡问题。文献[11]则在对称的部件之间引入对称性约束。

部件的观测模型对部件的表现进行建模, 用于度量每个部件的图像似然度, 确定部件的位置。文献[7]中利用前景图覆盖区域和边缘特征定义部件的图像似然度, 文献[9]只利用前景图覆盖区域作为图像特征, 需要有较好的前景分割结果。文献[4]利用背景差分的结果定义部件的图像似然度, 适用于背景简单的视频序列。文献[4]中采用基于边缘的部件模板, 文献[12]中则采用颜色直方图部件模板, 可以应用在单幅图像上。文献[6,8,13]中则引入了行人检测中被广泛使用的基于 HOG(Histogram of Oriented Gradient)特征的 Boosting 分类器, 取得了较好的效果。

优化算法则根据图模型和观测模型估计最终的姿态。基于信息传递的置信度传播(Belief Propagation, BP)算法是一种常用的图模型优化算法。然而在人体姿态估计问题中, 部件状态向量维数较高, 直接利用 BP 算法并不现实。文献[10]中则利用非参数化置信度传播算法, 利用混合高斯函数模拟图中节点之间传递的信息, 文献[9]中采用整数规划的优化算法在候选部件中选择最好的部件。文献[4]中根据树形模型的特点, 提出了基于信息传递的图案结构(pictorial structure)算法, 加速了图优化的过程, 但图案结构算法限制了部件之间的势函数的形式。而文献[8]则在文献[4]的基础上, 提出自适应的势函数和观测模型, 得到了更好的结果。

本文针对单目图像和视频中的人体姿态估计问题, 采用基于部件的方法, 将人体分为 10 个部件, 利用树形模型表示人体, 提出一种旋转不变的边缘力场特征, 利用基于边缘力场特征的 Boosting 分类器作为部件的观测模型。边缘力场特征计算简单且富有表现性, 可以用于快速精确计算部件的观测模型。另外本文借鉴了粒子滤波器的思想, 提出一种基于粒子采样和 BP 算法的优化算法, 采用迭代搜索的方法估计最终的姿态。与文献[10]中的方法相比, 这种算法利用粒子模拟部件的状态分布, 计算更为简便。

## 2 问题定义

本文将人体分为 10 个相互关联的部件(头部、躯干、左右上臂、左右前臂、左右大腿和左右小腿), 并表示为一个树形马尔科夫网络, 见图 1。其中每

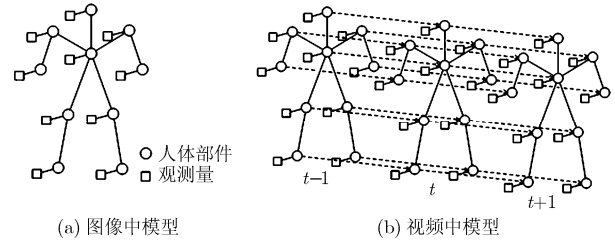


图 1 人体的马尔科夫网络模型

个节点表示一个人体部件, 其状态为  $\mathbf{l}_i$ , 每个部件都有自己的观测模型, 记为  $P(I_i | \mathbf{l}_i)$ , 相邻的部件之间存在几何约束关系, 记为  $P(\mathbf{l}_i, \mathbf{l}_j)$ , 人体的姿态可表示为每个部件的状态集合:  $\mathbf{L} = \{\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_n\}$ , 其中每个部件  $i$  的状态由位置  $\mathbf{p}_i = (x_i, y_i)$ 、大小  $s_i$  和方向  $\theta_i$  构成:  $\mathbf{l}_i = \{\mathbf{p}_i, s_i, \theta_i\}$ 。

给定图像表现  $I$ , 人体姿态估计可以形式化为如下优化问题:

$$\begin{aligned} \mathbf{L} &= \arg \max P(\mathbf{L} | I) = \arg \max P(I | \mathbf{L}) P(\mathbf{L}) \\ &= \prod_i P(I_i | \mathbf{l}_i) \prod_{(i,j) \in E} P(\mathbf{l}_i, \mathbf{l}_j) \end{aligned} \quad (1)$$

式中  $P(I_i | \mathbf{l}_i)$  为每个部件的观测模型, 将在第 2 节中描述。  $P(\mathbf{l}_i, \mathbf{l}_j)$  表示相邻部件之间的几何约束, 可假设位置, 大小及方向的约束相互独立, 定义如下:

$$\begin{aligned} P(\mathbf{l}_i, \mathbf{l}_j) &= P(\mathbf{p}_i, \mathbf{p}_j) P(s_i, s_j) P(\theta_i, \theta_j) \\ &\propto N(\|\mathbf{p}_i - \mathbf{p}_j\|; 0, \sigma_p) N(s_i - s_j; 0, \sigma_s) \\ &\quad \cdot e^{k \cos(\theta_i - \theta_j - \mu_{ij})} \end{aligned} \quad (2)$$

其中  $N(x; m, \sigma)$  表示变量  $x$  服从以  $m$  为均值,  $\sigma$  为方差的正态分布, 参数  $\sigma_p, \sigma_s, k, \mu$  可根据数据学习得到。

在视频中, 各帧中的人体部件可组成一个时序马尔科夫网络, 如图 1。

给定当前帧图像  $I_t$  及前一帧姿态估计结果  $\mathbf{L}_{t-1}$ , 当前帧的姿态  $\mathbf{L}_t$  可由下式给出:

$$\begin{aligned} \mathbf{L}_t &= \arg \max \prod_i P(I_{t,i} | \mathbf{l}_{t,i}) \prod_{(i,j) \in E} P(\mathbf{l}_{t,i}, \mathbf{l}_{t,j}) \\ &\quad \times \prod_i P(\mathbf{l}_{t,i} | \mathbf{l}_{t-1,i}) \end{aligned} \quad (3)$$

其中  $P(\mathbf{l}_{t,i} | \mathbf{l}_{t-1,i})$  部件  $i$  的运动模型。定义为

$$\begin{aligned} P(\mathbf{l}_{t,i} | \mathbf{l}_{t-1,i}) &= N(\|\mathbf{p}_{t,i} - \mathbf{p}_{t-1,i}\|; 0, \sigma_{p,i}) \\ &\quad \cdot N(s_{t,i} - s_{t-1,i}; 0, \sigma_{s,i}) \\ &\quad \cdot \exp(k_i \cos(\theta_{t,i} - \theta_{t-1,i} - \mu_i)) \end{aligned} \quad (4)$$

其中参数  $\sigma_{p,i}, \sigma_{s,i}, k_i, \mu_i$  可根据数据学习得到。

### 3 部件检测器

#### 3.1 边缘力场特征

在针对人体的计算机视觉研究中, 基于图像梯度的特征是最常用的图像特征。其中最常用的是统计特定方形区域内的梯度直方图的 HOG 特征<sup>[14]</sup>。HOG 特征在行人检测和人体姿态估计中都有广泛的应用。然而, HOG 特征的计算需要对图像进行大量的预处理工作, 且对于物体的方向和尺寸都比较敏感, 仅适合于快速检测图像中方向大致确定的物体(如行人), 而人体部件在图像中的尺寸和方向变化都十分剧烈, 计算 HOG 特征时间和空间消耗都很大。

另外一种基于梯度的特征是 Edgelet 特征<sup>[15]</sup>, 它统计位于特定曲线段上的边缘响应。Edgelet 特征便于计算, 但仅利用局部梯度特征, 易受噪声的影响。

在人体姿态估计中, 由于人体运动姿态和服饰的影响, 人体部件不但具有多变的方向和尺寸, 同时也具有丰富的纹理。因此我们需要设计一种稳定的旋转不变且便于计算的特征。

借鉴基本物理知识, 我们可以将梯度图像看作一个充满了带“电”粒子的场, 其中每个像素  $p$  都看作一个带“电”粒子, 称为梯度粒子, 其“电”量可以定义为该点图像梯度强度  $|d_p|$ 。而这些粒子之间存在相互作用力, 与电场力定义类似, 本文定义梯度“粒子” $q$ 对 $p$ 的作用力为

$$f(p, q) = \frac{d_p \cdot d_q}{\|r\|^3} r \quad (5)$$

其中  $r=p-q$  为从  $q$  指向  $p$  的矢量。该式表示两个梯度粒子之间的作用力与它们的梯度幅值成正比, 与梯度方向差的余弦值成正比, 与距离的平方成反比。在梯度图像  $I$  中位置  $p$  放入一个梯度粒子  $d_p$ , 其受到的作用力为

$$f(p, I) = \sum_{q \in \Gamma(p)} \frac{d_p \cdot d_q}{\|r_q\|^3} r_q \quad (6)$$

其中  $\Gamma(p) \subset I$  是  $I$  中  $p$  的邻域,  $r_q$  为从  $q$  指向  $p$  的向量。

显而易见,  $f$  一般指向  $I$  中  $p$  附近边缘响应最强的方向。对  $f$  正交分解可得

$$\begin{aligned} f(p, I) &= (f(p, I) \cdot e_x, f(p, I) \cdot e_y)^T \\ &= \left( d_p \cdot \sum_{q \in \Gamma(p)} (r_q \cdot e_x) / \|r_q\|^3, d_p \cdot \sum_{q \in \Gamma(p)} (r_q \cdot e_y) / \|r_q\|^3 \right)^T \\ &= \begin{pmatrix} d_p \cdot F_x(p) \\ d_p \cdot F_y(p) \end{pmatrix} = d_p F(p) \end{aligned} \quad (7)$$

其中  $e_x, e_y$  分别为表示  $x$  和  $y$  轴正方向的单位向量。而  $F(p) = (F_x(p)^T, F_y(p)^T)^T$  是一个  $2 \times 2$  的矩阵, 只与位置  $p$  有关, 本文称之为边缘力场。图 2(a), 2(b) 分别是边缘力场特征的 4 个分量及其等高线的图示。

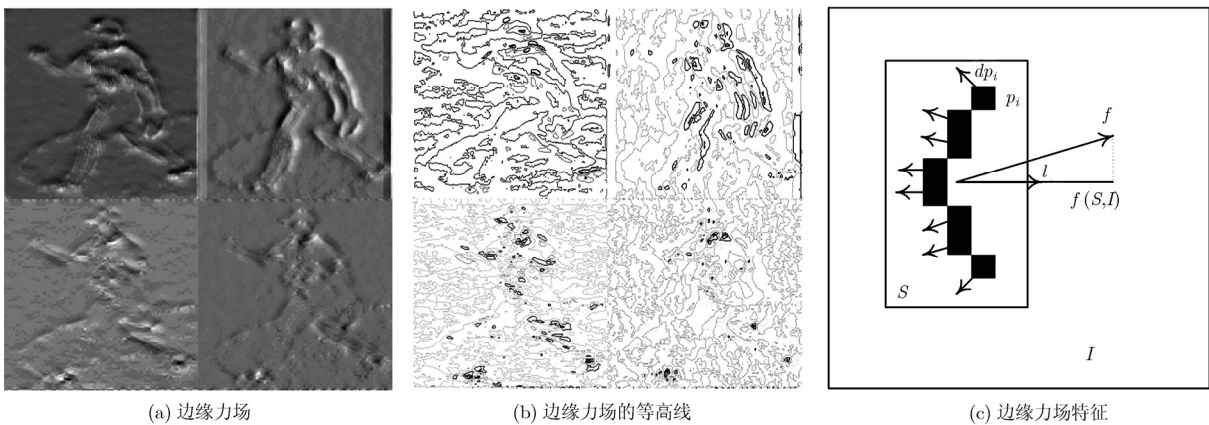
实际上, 这个作用力可以看作位置  $p$  邻域内图像梯度在  $d_p$  方向投影的加权和。如图 2(c) 所示, 本文的边缘力场特征定义为一条由若干梯度粒子组成的曲线段  $S$ 。其特征值定义为所有粒子在边缘力场中受到的作用力总和在单位向量  $l$  上的投影。

$$f(S, I) = \sum_{p \in S} F(p, I) \cdot l = \sum_{p \in S} d_p^T F(p) l \quad (8)$$

由于特征值为矢量的内积结果, 因此特征值的计算是旋转不变的。而边缘力场  $F(p)$  与具体的特征无关, 可以根据原始图像利用卷积预先计算。

$$F_{ij} = I \otimes D_i \otimes T_j, \quad i, j \in \{x, y\} \quad (9)$$

其中  $D_x, D_y$  是计算图像梯度的卷积核, 而  $T_x$  和  $T_y$  则是强度按半径递减的卷积核。根据边缘力场特征, 我们可以简单地设计一个基于分段线性函数的弱分



(a) 边缘力场

(b) 边缘力场的等高线

(c) 边缘力场特征

图 2 边缘力场图

类器，将对应特征值的值域等间隔地划分为若干互不相交的子空间 $\{(v_i, v_{i+1}]\}$ ，并对同一子区间内的所有样本输出一个固定的预测结果  $\text{lut}_i$ ，如式(10)所示。

$$h(S, I) = \text{lut}_i, \quad v_i < f(S, I) \leq v_{i+1} \quad (10)$$

### 3.2 基于边缘力场特征的 Boosting 分类器

基于边缘力场特征的 Boosting 分类器是一系列弱分类器的加权和。

$$H(I) = \sum_i w_i h(S_i, I) \quad (11)$$

在经典 AdaBoost 学习算法中，每次迭代枚举所有特征，并根据当前样本权重选取一个最好的弱分类器。然而采用边缘力场特征时，一个图像子窗口中可能的边缘力场特征数量巨大。以  $30 \times 30$  的子窗口为例：不同结构的边缘力场特征的数量达到  $10^8$  的规模，而每一种结构的又可以根据不同的投影方向分为不同的特征，因此不能利用穷举搜索的方法选择弱分类器。我们采用启发式搜索的方法进行弱分类器选择，其算法如表 1 所示。

表 1 算法 1 弱分类器选择

输入：样本集合 $\{I_i\}$ ，样本权重 $\{D_i\}$
(1)初始化
初始化特征列表 $OL = \{S_1, S_2, \dots, S_n\}$ 并计算分类损失， $S_i$ 为只包含一个梯度粒子的特征
扩展完成特征列表 CL 初始为空集
(2)扩展搜索过程
for $i = 1$ to $T$
从 OL 中取出分类损失最小的特征 $S^*$ ：
$S^* = \arg \min_{s \in OL} (Z(S))$ ，其中
$Z(S) = \sum_i D_i \exp(-h(S, I_i))$
$OL = OL - \{S^*\}$ ， $CL = CL \cup \{S^*\}$
根据规则扩展 $S^*$ 得到若干新特征并计算分类损失并放入 OL 中
end for
(3)输出 从 OL 和 CL 中选出分类损失最小的特征

这个算法首先将特征列表初始化为所有只包含一个梯度粒子的特征，每次迭代时选出特征列表中具有最小分类损失的特征，根据扩展规则生成一系列新的特征加入特征列表，若干轮迭代之后选出所有特征中具有最小分类损失的特征作为弱分类器的特征。特征的每一次扩展都会增加一个粒子，为保证扩展过程中不生成相同的特征，并保证曲线的平滑，每次扩展的粒子的位置都有一定限制，如图 3 所示。

给定特征中每个粒子的位置 $\{p\}$ 后，我们需要确定投影方向  $l$  和每个粒子的梯度 $\{d_p\}$ ，使得训练样本可以尽量分开。由于式(8)中 $\{d_p\}$ 和  $l$  不是线性关系，因此我们采用迭代的方法交替优化它们：

(1)给定  $l$  优化 $\{d_p\}$ ：给定  $l$  后，式(8)可转化为

$$f(S, I) = \sum_{p \in S} d_p^T F(p) l = (d_{p_1}^T, d_{p_2}^T, \dots, d_{p_n}^T) \cdot (F(p_1)l, F(p_2)l, \dots, F(p_n)l)^T \quad (12)$$

这与线性鉴别分析(LDA)中的问题类似，可以用相同方法求出 $\{d_p\}$ 。

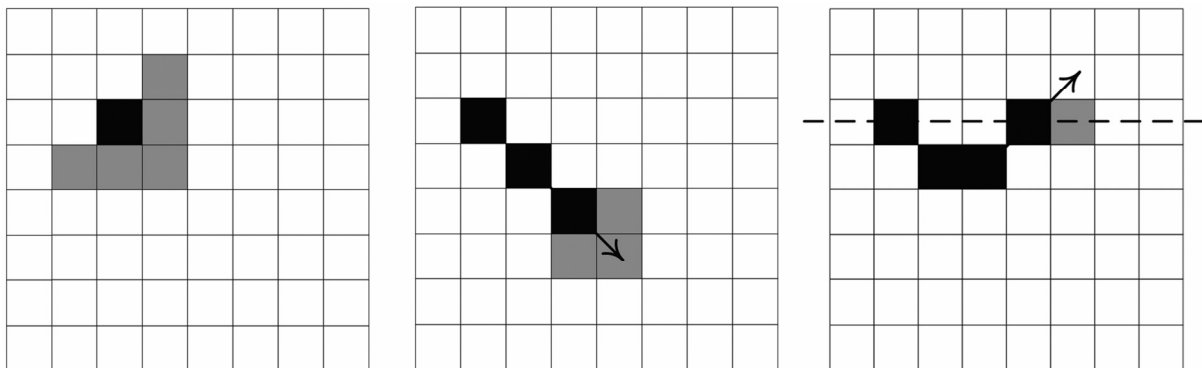
(2)给定 $\{d_p\}$ 优化  $l$ ：给定 $\{d_p\}$ 后， $l$  也可用 LDA 的方式求出。

## 4 图像和视频帧中人体姿态估计

置信度传播(Belief Propagation, BP)算法是一种常用的图推理算法，它通过迭代的方法在图中相邻节点之间传播信息来估计每个节点的边际概率分布。以式(1)中的优化问题为例，每次迭代时从节点  $s$  传播到节点  $t$  的信息为

$$m_{st}^n(l_t) = \alpha \int_{l_s} p(l_s, l_t) p(l_s | I) \prod_{u \in T(t)} m_{us}^{n-1}(l_s) dl_s \quad (13)$$

而每个节点的边际概率可以根据该节点的观测模型和相邻节点的信息估计。



(a) 只有 1 个粒子 (黑色块) 时，可向 5 个方向 (灰色块) 扩展

(b) 有多个粒子时，向当前前进的方向及相邻两个方向扩展

(c) 扩展的粒子只能在位于第 1 个粒子高度上的水平线的同一侧

图 3 特征扩展规则

$$\hat{p}(\mathbf{l}_s | I) = \beta p(I | \mathbf{l}_s) \prod_{u \in \Gamma(s)} m_{us}^n(\mathbf{l}_s) \quad (14)$$

其中  $\alpha, \beta$  为归一化因子。

对于树形图, BP 算法能够收敛到真实的边际分布。然而, 由于每个节点的观测模型只能离散地测量, 而且节点的状态向量维度较高, 因此在整个状态空间中计算式(13)中的积分非常耗时。

借鉴粒子滤波器中粒子采样的思想, 本文设计了一种基于粒子采样的置信度传播推理方法。在图 1 中的马尔科夫网络中, 对于每一个节点  $i$ , 我们利用一组带权重的粒子  $\{(l_i^{(0)}, \pi_i^{(0)}), \dots, (l_i^{(n)}, \pi_i^{(n)})\}$  来模拟节点  $i$  的状态分布, 通过不断重新采样来寻找问题的最大后验概率解。算法如表 2 所示。

表 2 算法 2 基于粒子采样的推理方法

(1)初始化
根据部件检测结果或前一帧信息初始化每个部件的粒子
(2)迭代优化
利用 BP 算法求出每个部件在其粒子集合上的后验边际分布, 并更新权重: $\pi_i^{(k)} = p(l_i^{(k)}   I)$
根据权重对每个部件的粒子重新采样
选取具有最大后验边际概率的粒子作为最终状态:
$l_i = \arg \max(p(l_i^{(k)}   I))$
若与上一轮迭代结果差异小于一定阈值, 则收敛, 否则继续迭代
(3)输出

这个算法本质上是一种随机的变步长搜索算法。

## 5 实验

### 5.1 实验数据

实验中采用了 4 个不同的数据集合, 其中第 1 个数据集包括 350 张拍摄于足球或篮球场中的图片, 其中人体的姿态经过手工标定。第 2 个数据集是文献[4]中提到的 Buffy 数据集, 来自美剧《Buffy the Vampire Slayer》5 集中共 748 帧图像, 其中人

体上半身姿态经过手工标定。第 3 个数据集是文献[4]中提到的人体姿态估计数据集, 包括 305 张拍摄于自然场景中的人体图片, 人体姿态经过手工标定。第 2 和第 3 这两个数据集中人体姿态变化较大, 且背景和光照条件也有剧烈的变化, 是两个具有挑战性的数据集。而第 4 个数据集是 HumanEVA-II 数据集, 包含两个人从 4 个角度拍摄的共 8 段视频。

在前 3 个数据集中分别取出 200, 472 和 100 张图像作为训练集, 剩余的 150, 276 和 205 张图像作为测试集, 分别进行了部件检测器和人体姿态估计的实验。

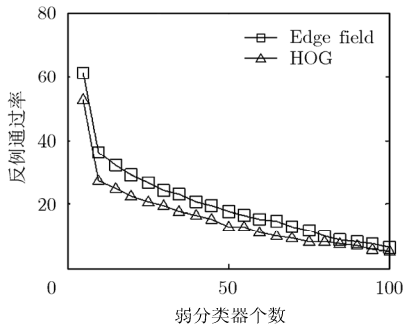
### 5.2 部件检测器实验

给定一张图片, 对于特定的部件, 在其标定的状态量邻域内均匀采样 9 个正例样本, 邻域外随机采样 9 个反例样本。使用 5.1 节中介绍的训练图片集, 用 HOG 特征和本文提出的 Edge field 特征对 6 个不同的部件训练部件检测器, 对比如图 4: 图 4(a)为躯干检测器正例通过率为 90%时的反例通过率(纵坐标)随弱分类器数量(横坐标)变化图, 图中可以看出, 由于单个 Edge field 特征更具有区分性, 使用 Edge field 特征的 AdaBoost 算法较使用 HOG 特征收敛更快。图 4(b)为最终躯干检测器的准确率-召回率曲线, 从图中可见使用 Edge field 特征的 AdaBoost 分类器性能更好。

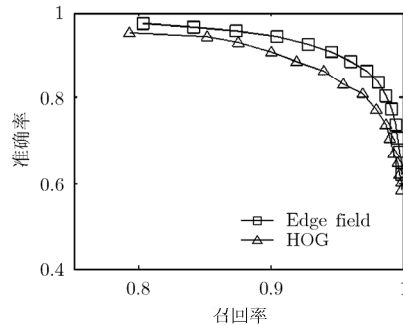
图 4(c)是各个部件分类器中的前两个 Edge field 特征, 从图中可见 Edge field 特征能够很好地刻画各个部件中主要的边缘特征。

### 5.3 人体姿态估计对比试验

利用 5.2 节中学习得到的部件检测器, 根据第 3 节中的基于粒子采样的置信度传播的方法对人体姿态进行估计。对于静止图像和视频的第 1 帧图像, 在部件检测器响应较大的地方随机采样初始粒子; 对于视频中后续帧, 根据上一帧估计结果随机采样初始粒子。每个部件分别有 400 个粒子。



(a) AdaBoost 算法收敛趋势



(b) 部件检测器的 PRC 曲线



(c) 各部件检测器的前两个 Edge field 特征示例

图 4 HOG 特征与本文提示的 Edge field 特征比较

在第2和第3数据集中,我们与文献[4]中的方法以及文献[6]中的方法进行了对比。利用前3个数据集中的训练集训练了部件检测器,并对数据集2,数据集3进行姿态估计,姿态估计结果的各个部件的准确率如表3,表4所示。可见,本文方法各个部件的准确率都较文献[4,6]中的方法有一定的提升,且速度提升超过10倍,效率明显更高。

图5是4个数据集中人体姿态估计的示例。

## 6 结论

本文针对单目图像和视频中的人体姿态估计问题,采用基于部件的方法,利用包含10个部件树形

模型表示人体。准备每个部件的观测模型,本文提出一种旋转不变的边缘力场特征,并基于这种特征学习 Boosting 分类器。另外本文借鉴了粒子滤波器的思想,利用状态空间中的一组粒子模拟部件的状态分布,并用置信度传播算法计算各个部件在这组粒子上的边际分布并更新粒子位置。边缘力场特征计算简单且富有表现性,可以用于快速精确计算部件的观测模型。而基于粒子采样的置信度传播算法利用粒子模拟状态分布,减少了计算负担。实验证明本文的方法准确率较前人算法有所提高,而其速度较快。

表3 数据集2中各部件估计正确率对比

方法	躯干(%)	上臂(%)	前臂(%)	头(%)	合计(%)	处理时间(s)
文献[4]方法	-	-	-	-	57.9	133.3
文献[6]方法	90.7	80.6 82.1	44.2 47.9	95.5	73.5	122.6
本文方法	94.4	87.7 87.0	61.7 64.7	96.7	82.0	8.5

表4 数据集3中各部件估计正确率对比

方法	躯干(%)	上臂(%)	前臂(%)	大腿(%)	小腿(%)	头(%)	合计(%)	处理时间(s)
文献[4]方法	52.1	17 18	14.6 12.6	30.2 31.7	27.8 30.2	37.5	27.2	142.3
文献[6]方法	81.4	47.3 47.8	31.2 32.1	67.3 59.0	63.9 46.3	75.6	55.2	125.1
本文方法	85.8	51.7 53.2	47.8 41.0	72.1 68.3	63.9 62.4	77.1	62.3	10.5

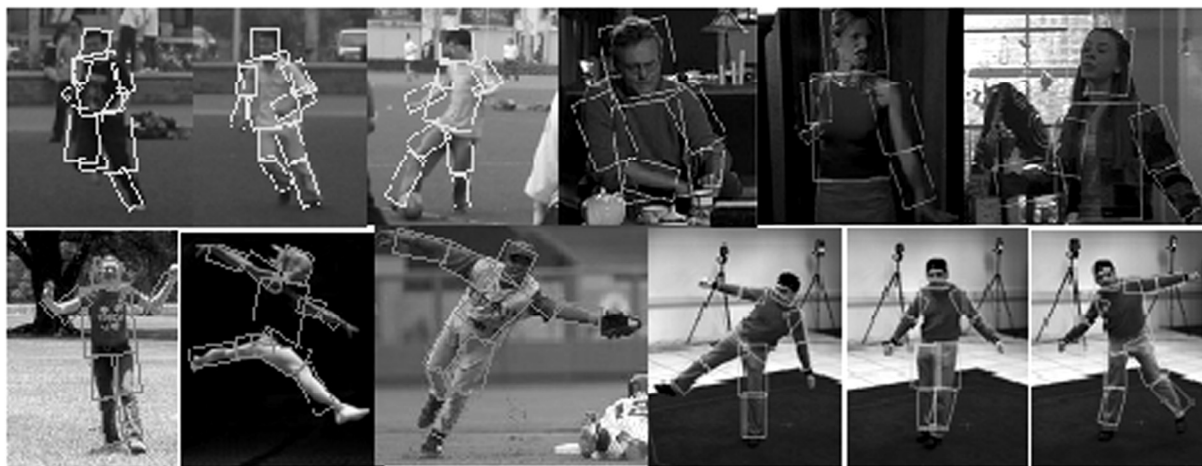


图5 人体姿态估计示例

## 参考文献

- [1] Agarwal A and Triggs B. Recovering 3D human pose from monocular images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(1): 44-58.
- [2] Bissacco A, Yang Ming-hsuan, and Soatto S. Fast human pose estimation using appearance and motion via multi-dimensional boosting regression. [C] CVPR 2007, Minneapolis, US, 2007: 1-8.
- [3] Poppe R and Poel M. Comparison of silhouette shape descriptors for example-based human pose recovery[C]. Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition. Southampton, 2006: 541-546.

- [4] Felzenszwalb P F and Huttenlocher D P. Pictorial structures for object recognition[J]. *International Journal of Computer Vision*, 2004, 61(1): 55–79.
- [5] Ramanan D. Learning to parse images of articulated bodies[C]. NIPS 2006, Cambridge, 2006: 1129–1136.
- [6] Andriluka M, Roth S, and Schiele B. Pictorial structures revisited: people detection and articulated pose Estimation[C]. CVPR 2009, Miami, FL, USA, 2009: 1014–1021.
- [7] Zhang Xiao-qin and Li Chang-cheng, *et al.* Efficient human pose estimation via parsing a tree structure based human model[C]. ICCV 2009, Kyoto, 2009: 1349–1356.
- [8] Sapp B, Jordan C, and Taskar B. Adaptive pose priors for pictorial structures[C]. CVPR 2010, San Francisco, 2010: 422–428.
- [9] Jiang Hao. Human pose estimation using consistent max-covering[C]. ICCV 2009, Kyoto, 2009: 1357–1364.
- [10] Sigal L and Black M J. Measure locally, reason globally: occlusion-sensitive articulated pose estimation[C]. CVPR 2006, New York, NY, USA, 2006: 2041–2048.
- [11] Jiang Hao and Martin D R. Global pose estimation using non-tree Models[C]. CVPR 2008, Anchorage, AK, 2008, DOI: 10.1109/CVPR.2008.4587457.
- [12] Ferrari V, Marin-Jimenez M, and Zisserman A. Progressive search space reduction for human pose estimation[C]. CVPR 2008, Anchorage, AK, 2008, DOI: 10.1109/CVPR.2008.4587468.
- [13] Andriluka M, Roth S, and Schiele B. Monocular 3D pose estimation and tracking by detection[C]. CVPR 2010, San Francisco, CA, 2010: 623–630.
- [14] Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]. CVPR 2005, San Diego, CA, USA, 2005: 886–893.
- [15] Wu B and Nevatia R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors[C]. ICCV 2005, Beijing, China, 2005: 90–97.
- 苏延超: 男, 1984年生, 博士生, 研究方向为模式识别、人脸和人体的配准以及姿态估计.
- 艾海舟: 男, 1964年生, 教授, 研究方向为计算机视觉.
- 劳世玆: 男, 1962年生, 高级研究员, 研究方向为人脸图像处理.