



(12) 发明专利

(10) 授权公告号 CN 108564012 B

(45) 授权公告日 2022.03.08

(21) 申请号 201810273078.9

G06K 9/62 (2022.01)

(22) 申请日 2018.03.29

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 107609541 A, 2018.01.19

申请公布号 CN 108564012 A

CN 104881685 A, 2015.09.02

CN 106650827 A, 2017.05.10

(43) 申请公布日 2018.09.21

Ke Gong等.Look into Person: Self-

(73) 专利权人 北京工业大学

supervised Structure-sensitive Learning

地址 100124 北京市朝阳区平乐园100号

and A New Benchmark for Human Parsing.

(72) 发明人 杨金福 张京玲 王美杰 李明爱

《2017 IEEE Conference on Computer Vision

许兵兵

and Pattern Recognition (CVPR)》.2017,

(74) 专利代理机构 北京思海天达知识产权代理

审查员 王楠

有限公司 11203

代理人 刘萍

(51) Int.Cl.

G06V 40/10 (2022.01)

G06V 10/762 (2022.01)

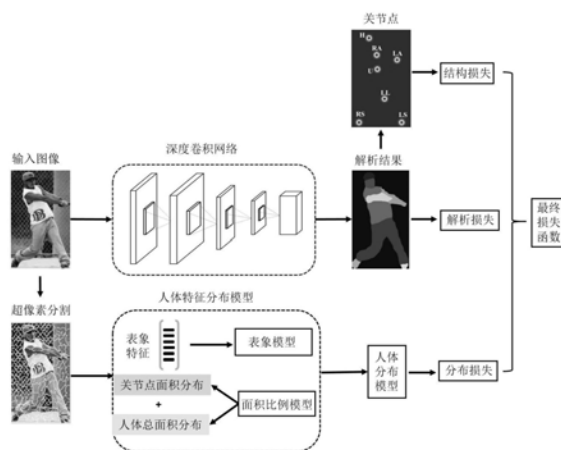
权利要求书2页 说明书6页 附图2页

(54) 发明名称

一种基于人体特征分布的行人解析方法

(57) 摘要

本发明涉及一种基于人体特征分布的行人解析方法,从人体特征分布的角度出发,提出了一种智能的融合人体特征分布模型和自监督结构敏感学习策略的行人解析方法。以自监督结构敏感学习方法为底层框架,首先,通过超像素分割的方法产生候选区域,对每个区域提取颜色和纹理特征建立表象模型,然后再利用高斯函数建立面积比例模型,最后通过将两个模型叠加得到总的人体特征分布模型。并将解析损失函数、关节结构损失函数和人体特征分布模型的特征分布损失函数相叠加的得到最终的损失函数。本发明利用自监督结构敏感学习方法使得生成的解析结果在语义上与人体的结构具有很强的一致性,更符合人体特征分布的特性,并且对遮挡,视角,复杂背景具有不变性。



1. 一种基于人体特征分布的行人解析方法,其特征在于,包括如下步骤:

步骤1:数据预处理

将训练集中图像作为输入图像,对输入图像进行预处理,即对输入图像进行多尺度变换、水平翻转和随机裁剪,得到预处理过的训练集;

步骤2:训练基础人体解析网络

步骤2.1:使用深度残差网络ResNet-101作为基础人体解析网络,其中包括卷积层,池化层和全连接层,共101层;将由步骤1得到的预处理过的训练集作为卷积神经网络的输入,训练卷积神经网络,进而对卷积神经网络的网络结构文件和网络参数配置文件进行修改,并产生初始的行人解析结果;

步骤2.2:使用softmax计算解析损失函数:

$$L_{parse} = \frac{\exp(\theta_i^T x)}{\sum_{i=1}^K \exp(\theta_i^T x)}$$

其中, L_{parse} 为解析损失函数, θ_i^T 为深度残差网络模型参数, x 为深度特征, K 为关节类别数;此处, $K=9$;

步骤3:计算关节结构损失函数

步骤3.1:定义9个关节点,分别为头部、上身、左臂、右臂、下身、左腿、右腿、左脚和右脚,并定义如下关节点列表:

$$C_i^p = \{c_i^p \mid i \in [1, K]\}$$

$$C_i^{GT} = \{c_i^{GT} \mid i \in [1, K]\}$$

其中, c_i^p 为步骤2的初始解析结果所对应的第*i*个关节点, c_i^{GT} 是步骤1中训练数集的标签图像对应的第*i*个关节点, I 为输入图像;

步骤3.2:计算关节结构损失函数为

$$L_{joint} = \frac{1}{2K} \sum_{i=1}^K \|c_i^p - c_i^{GT}\|_2^2$$

表示预测的解析结果对应关节点和训练集中标签图像对应的9个关节点的差值的均值;差值越大,预测越不准确;差值越小,预测越准确,此处 $K=9$;

步骤4:构建人体特征分布模型

首先利用超像素分割方法将人体分割成大小不同的候选区域,然后分别对每个区域建立表象模型和面积比例模型;

步骤4.1:基于超像素分割方法产生人体候选区域

步骤4.1.1:按照设定的超像素个数,在图像内均匀分配种子点;假设图像中共有*M*个像素点,预分割为*N*个相同尺寸的超像素,则每个超像素的大小为*M/N*,则相邻种子点的距离为 $S = \sqrt{M/N}$;

步骤4.1.2:在种子点的*S*S*邻域内,计算该邻域内所有像素点的梯度值,将种子点移到该邻域内梯度最小的地方;对该种子点所在邻域内的每个像素点,分别计算它与各种子点的颜色距离和空间距离,并取距离最小的为该像素点的聚类中心;

步骤4.1.3:对步骤4.1.2进行15次以上的迭代,得到最终的聚类中心和候选区域;

步骤4.2:建立表象概率模型

步骤4.2.1:对每个候选区域j,利用L*a*b*颜色空间模型,提取11维颜色特征向量,每一维向量描述了像素或某一图像区域中的颜色属于11种颜色的概率;

步骤4.2.2:对每个候选区域j,计算LBP纹理特征;并与步骤4.2.1提取的颜色特征串联得到候选区域的表象特征 H_j ,最后利用逻辑斯蒂回归函数建立表象概率模型:

$$P(Y_j = k | H_j) = \max_{k \in K} \frac{1}{1 + e^{\|H_j - G_k\|^2}}$$

其中,e为指数函数, $\|\cdot\|^2$ 表示L2范数; G_k 是标签图像中第k类关节点对应的表象特征, $Y_j = k$ 表示区域j所属的关节点类别为k,k取值为1-9;区域j所属关节点类别,即为表象特征差值最小时对应的标签图像对应的关节点类别;

步骤4.3:建立面积比例模型

步骤4.3.1:对给定的输入图像I,分别计算每种关节区域的面积 A_k 和图像中人体所占的总面积 S_I ,面积比例概率模型计算如下:

$$P(Y_j = k | A_k, S_I) = \max_{k \in K} \psi_k(S_I)^{\frac{1}{\psi_k(A_k)}}, \quad k \in [1, 9]$$

$$\psi_k(A_k) = \exp\left(-\frac{\|A_j - \mu_k\|^2}{\sigma_k}\right), \quad k \in [1, 9]$$

$$\psi_l(S_I) = \exp\left(-\frac{\|S_I - \mu_l\|^2}{\sigma_l}\right)$$

其中面积大小用像素总数来表示, $\psi_k(A_k)$ 表示第k种关节点的面积高斯分布, $\psi_l(S_I)$ 表示人体总面积 S_I 的高斯分布; μ_k 表示第k种关节点面积分布的均匀程度, σ_k 表示第k种关节点面积分布的平滑程度; μ_l 表示人体总面积分布的均匀程度, σ_l 人体总面积分布平滑程度;对每个候选区域计算属于不同类别关节点的概率,并认为概率最大对应的就是所属类别;

步骤4.4:根据朴素贝叶斯法则得到最终的损失函数:

$$L_{\text{distribution}}(Y_j = k | H_j, A_k, S_I) = P(Y_j = k | H_j) P(Y_j = k | A_k, S_I)$$

步骤5:构建总体分布函数

将步骤2的解析损失函数、步骤3的关节结构损失函数和步骤4.4的人体分布损失函数组合即可得到最终的损失函数:

$$L = L_{\text{parse}} \cdot L_{\text{joint}} + L_{\text{distribution}}$$

步骤6:模型训练

训练平台采用NVIDIA GeForce GTX TITAN X GPU,网络搭建采用Caffe框架,batch-size设置为20,初始学习率为0.001,动量设置为0.9,权重衰减设置为0.0005,优化方法采用梯度下降法;

步骤7:测试

对于给定的输入图像,首先根据解析网络计算初始解析结构,并根据解析结果计算关节结构热图;然后采用超像素分割的方法分割输入图像,并对分割产生的候选区域进行人体分布建模,得到总的损失函数,通过损失函数最小化得到最终的解析结果。

一种基于人体特征分布的行人解析方法

技术领域

[0001] 本发明属于图像行人解析技术领域,以自监督结构敏感学习方法(Self-supervised Structure-sensitive Learning approach)为底层框架,融合符合人体特征分布的人体特征分布模型进行人体解析。该模型首先利用超像素分割方法产生候选区域,然后计算候选区域与数据集中人体特征分布模型的相似得分,从而得到人体各部件关节节点的语义标签。本发明不仅利用自监督结构敏感学习方法使得生成的解析结果在语义上与人体的结构具有很强的一致性,而且由于引入人体特征分布模型,具有更符合人体特征分布的特性,对遮挡、视角变化和复杂背景具有不变性。

背景技术

[0002] 随着科技的发展和人类生活智能化的提高,智能服务机器人逐渐进入到人类生产生活中,并且被广泛地应用到了各个行业。机器人服务人类时,需要建立对服务本体的认知与理解。行人解析是服务机器人对人类最重要的认知方式之一,直接影响机器人智能化水平的发挥,也是完善人机交互的基础。为了使机器人能够准确的完成诸如送餐、握手等交互操作,服务机器人的视觉感知系统不仅需要识别出行人的位置和类别属性,还有需要提供更具体的行人语义部件(如头、躯干、手臂、腿、脚等)信息,因此,需要研究行人解析方法,进而使机器人能够理解人类的行为。实际应用场景中,由于存在如光照变化、局部遮挡、姿态变化、视角变化等复杂因素,因此,行人解析任务具有很高的难度。

[0003] 近年来,行人解析获得了越来越多学者的关注。现有的行人解析大多基于分割、姿态估计和自下而上的区域产生式的方法。例如,2012年,Yamaguchi^[1]等人提出将姿态估计和行人解析相结合的方法进行解析,同时采用基于检索的方法对结果进行微调,取得了较好的性能。与大多数把人体姿态估计问题作为行人解析的前提所不同的是,2013年Dong^[2]等人提出了一种解析部件表达方法(Parselet representation),将人体划分为不同的语义区域,然后建立一个可变形的混合解析模型(DMPM,Deformable Mixture Parsing Model)。该模型有两个特点:(1)被划分的语义区域可以组合成与或(And-Or)形式的树形结构;(2)对树形结构的叶节点的可见属性(visibility property)进行建模,可以解决Parselet存在的遮挡问题。然而,上述传统的方法通常需要手动设计复杂预处理步骤(如姿态估计、模板字典学习等),导致行人解析运行效率和准确率的降低。随着目前深度学习的发展,尤其是卷积神经网络在图像识别领域中取得的优异性能,卷积神经网络被广泛应用到机器视觉领域中。2015年,Liang^[3]等人提出一种上下文卷积神经网络结构(Co-CNN,Contextualized Convolutional Neural Network),将交叉层上下文信息(cross-layer context)、全局图像级上下文信息(global image-level context)、以及相邻超像素上下文信息(cross-super-pixel neighborhood context)加入到网络中,实现了端到端的像素级分类。该方法能有效提高分割精度,但其设计也容易导致局部最优。2017年,Gong^[4]等人提出一种自监督结构敏感学习的行人解析方法,利用丰富的关节结构信息来监督行人解析。针对姿态估计中关节节点的定义和行人解析中关节定义的不一致性,作者定义了9种关节

结构并提出直接从标注的解析数据集中产生近似的关节标签,最后利用结构敏感损失函数评估预测的关节结构和标签关节结构的一致性。这种自监督的框架具有一定的普适性,可以应用在任何的网络结构中,但没有充分考虑相同对象不同语义部件之间的分布关系。

[0004] 本发明旨在提出一种融合人体特征分布和关节结构信息的行人解析方法,以自监督结构敏感学习方法为底层框架,融合符合人体特征分布的模型进行人体解析。人体特征分布模型充分考虑了相同部件的分布关系和部件间的分布关系,包含2个主要部分:一是基于颜色和纹理特征的表象模型,二是基于关节点面积和人体总面积的面积比例模型。

[0005] 参考文献:

[0006] 1.Yamaguchi K,Kiapour M H,Ortiz L E,et al.Parsing clothing in fashion photographs[C]//Computer Vision and Pattern Recognition.IEEE,2012:3570-3577.

[0007] 2.Dong J,Chen Q,Xia W,et al.A Deformable Mixture Parsing Model with Parselets[C]//IEEE International Conference on Computer Vision.IEEE,2014:3408-3415.

[0008] 3.Liang X,Xu C,Shen X,et al.Human Parsing with Contextualized Convolutional Neural Network[C]//IEEE International Conference on Computer Vision.IEEE,2015:1386-1394.

[0009] 4.Ke Gong,Xiaodan Liang,Dongyu Zhang,et al.Look into Person:Self-Supervised Structure-Sensitive Learning and a New Benchmark for Human Parsing [J].2017:6757-6765.

发明内容

[0010] 行人解析旨在将行人图像进行细粒度的分割,得到不同人体部件的语义信息。该技术可以广泛应用于图像检索、行人再识别、人体行为分析等领域。然而由于真实场景中不可避免地存在诸如光照变化、局部遮挡、姿态变化、视角变化等复杂因素,导致行人解析研究具有挑战性。本发明提出了一种融合人体特征分布和关节结构信息的行人解析方法,以自监督结构敏感学习方法为底层框架,融合符合人体特征分布的人体特征分布模型进行人体解析。利用颜色和纹理特征建立表象模型,利用各个部件面积关系建立面积比例模型,形成最终的人体特征分布模型。本发明方法充分考虑了人体特征分布的关系,增强了预测结果和实际人体特征分布关系的一致性,在一定程度上克服了遮挡、光照和姿态变化的影响。

[0011] 此外,该发明采用Penn-Fudan数据集训练改进后的行人解析网络,该数据集中共有169张行人图像,9个部件标签,分别是头发、人脸、上衣、下衣、左臂、右臂、左腿、右腿、左脚和右脚。

[0012] 本发明是采用以下技术手段实现的:

[0013] 一种基于人体特征分布的行人解析方法,其特征在于,包括如下步骤:

[0014] 步骤1:数据预处理

[0015] 将训练集中图像作为输入图像,对输入图像进行预处理,即对输入图像进行多尺度变换、水平翻转和随机裁剪,得到预处理过的训练集。

[0016] 步骤2:训练基础人体解析网络

[0017] 步骤2.1:使用深度残差网络ResNet-101作为基础人体解析网络,其中包括卷积

层,池化层和全连接层,共101层。将由步骤1得到的预处理过的训练集作为卷积神经网络的输入,训练卷积神经网络,进而对卷积神经网络的网络结构文件和网络参数配置文件进行修改,并产生初始的行人解析结果。

[0018] 步骤2.2:使用softmax计算解析损失函数:

$$[0019] \quad L_{\text{parse}} = \frac{\exp(\theta_i^T x)}{\sum_{i=1}^K \exp(\theta_i^T x)}$$

[0020] 其中, L_{parse} 为解析损失函数, θ_i^T 为深度残差网络模型参数, x 为深度特征, K 为关节点类别数;此处, $K=9$ 。

[0021] 步骤3:计算关节结构损失函数

[0022] 步骤3.1:定义9个关节点,分别为头部、上身、左臂、右臂、下身、左腿、右腿、左脚和右脚,并定义如下关节点列表:

$$[0023] \quad C_i^P = \{c_i^P \mid i \in [1, K]\}$$

$$[0024] \quad C_i^{GT} = \{c_i^{GT} \mid i \in [1, K]\}$$

[0025] 其中, c_i^P 为步骤2的初始解析结果所对应的第*i*个关节点, c_i^{GT} 是步骤1中训练数集的标签图像对应的第*i*个关节点, I 为输入图像。

[0026] 步骤3.2:计算关节结构损失函数为

$$[0027] \quad L_{\text{joint}} = \frac{1}{2K} \sum_{i=1}^K \|c_i^P - c_i^{GT}\|_2^2$$

[0028] 表示预测的解析结果对应关节点和训练集中标签图像对应的9个关节点的差值的均值。差值越大,预测越不准确。差值越小,预测越准确,此处 $K=9$ 。

[0029] 步骤4:构建人体特征分布模型

[0030] 首先利用超像素分割方法将人体分割成大小不同的候选区域,然后分别对每个区域建立表象模型和面积比例模型。

[0031] 步骤4.1:基于超像素分割方法产生人体候选区域

[0032] 步骤4.1.1:按照设定的超像素个数,在图像内均匀分配种子点。假设图像中共有 M 个像素点,预分割为 N 个相同尺寸的超像素,则每个超像素的大小为 M/N ,则相邻种子点的距离近似为 $S = \sqrt{M/N}$ 。

[0033] 步骤4.1.2:在种子点的 $S*S$ 邻域内,计算该邻域内所有像素点的梯度值,将种子点移到该邻域内梯度最小的地方。对该种子点所在邻域内的每个像素点,分别计算它与各种种子点的颜色距离和空间距离,并取距离最小的为该像素点的聚类中心;

[0034] 步骤4.1.3:对步骤4.1.2进行15次以上的迭代,得到最终的聚类中心和候选区域;

[0035] 步骤4.2:建立表象概率模型

[0036] 步骤4.2.1:对每个候选区域*j*,利用 $L*a*b*$ 颜色空间模型,提取11维颜色特征向量,每一维向量描述了像素或某一图像区域中的颜色属于11种颜色的概率。

[0037] 步骤4.2.2:对每个候选区域*j*,计算LBP纹理特征。并与步骤4.2.1提取的颜色特征串联得到候选区域的表象特征 H_j ,最后利用逻辑斯蒂回归函数建立表象概率模型:

$$[0038] \quad P(Y_j = k | H_j) = \max_{k \in K} \frac{1}{1 + e^{\|H_j - G_k\|^2}}$$

[0039] 其中, e 为指数函数, $\|\cdot\|^2$ 表示L2范数; G_k 是标签图像中第 k 类关节点对应的表象特征, $Y_j = k$ 表示区域 j 所属的关节点类别为 k , k 取值为 $0 \sim 9$ 。区域 j 所属关节点类别, 即为表象特征差值最小时对应的标签图像对应的关节点类别。

[0040] 步骤4.3: 建立面积比例模型

[0041] 步骤4.3.1: 对给定的输入图像 I , 分别计算每种关节区域的面积 A_k 和图像中人体所占的总面积 S_I , 面积比例概率模型计算如下:

$$[0042] \quad P(Y_j = k | A_k, S_I) = \max_{k \in K} \psi_k(S_I)^{\frac{1}{\psi_k(A_k)}}, \quad k \in [1, 9]$$

$$[0043] \quad \psi_k(A_k) = \exp\left(\frac{\|A_k - \mu_k\|^2}{\sigma_k}\right), \quad k \in [1, 9]$$

$$[0044] \quad \psi_I(S_I) = \exp\left(\frac{\|S_I - \mu_I\|^2}{\sigma_I}\right)$$

[0045] 其中面积大小用像素总数来表示, $\psi_k(A_k)$ 表示第 k 种关节点的面积高斯分布, $\psi_I(S_I)$ 表示人体总面积 S_I 的高斯分布; μ_k 表示第 k 种关节点面积分布的均匀程度, σ_k 表示第 k 种关节点面积分布的平滑程度; μ_I 表示人体总面积分布的均匀程度, σ_I 对每个候选区域计算属于不同类别关节点的概率, 并认为概率最大对应的就是所属类别。

[0046] 步骤4.4: 根据朴素贝叶斯法则得到最终的损失函数:

$$[0047] \quad L_{\text{distribution}}(Y_j = k | H_j, A_k, S_I) = P(Y_j = k | H_j) P(Y_j = k | A_k, S_I)$$

[0048] 步骤5: 构建总体分布函数

[0049] 将步骤2的解析损失函数、步骤3的关节结构损失函数和步骤4.4的人体分布损失函数组合即可得到最终的损失函数:

$$[0050] \quad L = L_{\text{parse}} \cdot L_{\text{joint}} + L_{\text{distribution}}$$

[0051] 步骤6: 模型训练

[0052] 训练平台采用NVIDIA GeForce GTX TITAN X GPU, 网络搭建采用Caffe框架, batch-size设置为20, 初始学习率为0.001, 动量设置为0.9, 权重衰减设置为0.0005, 优化方法采用梯度下降法。

[0053] 步骤7: 测试

[0054] 对于给定的输入图像, 首先根据解析网络计算初始解析结构, 并根据解析结果计算关节结构热图。然后采用超像素分割的方法分割输入图像, 并对分割产生的候选区域进行人体分布建模, 得到总的损失函数, 通过损失函数最小化得到最终的解析结果。

[0055] 本发明与现有技术相比, 具有以下明显的优势:

[0056] 传统的基于姿态估计的方法所产生的关节点与行人解析任务中的所需要关节点不完全一致, 容易造成人体部件类别标签被误分配的情况。本发明提出的基于人体特征分布的行人解析方法, 不需要额外的进行姿态估计, 从解析结果中即可获得关节点, 大大减少了计算时间, 并由于人体特征分布模型的引入, 使得解析结果更符合视觉认知, 也在一定程度上克服了遮挡、光照和姿态变化的影响。

附图说明

[0057] 图1基于人体特征分布的行人解析方法的总体结构图；

[0058] 图2超像素分割流程图；

[0059] 图3部分实验结果；

具体实施方式

[0060] 为了使本领域的人员能够更好的理解和使用本发明，下面将结合附图和具体实施案例对本发明的技术方案进行进一步描述。

[0061] 1. 该发明主要采用Penn-Fudan数据集训练改进后的行人解析网络，该数据集中共有169张行人图像和9个关节标签，分别是头发、人脸、上衣、下衣、左臂、右臂、左腿、右腿、左脚、右脚。本发明提出的基于人体特征分布的行人解析方法结构图如图1所示，主要包括三个部分：(1) 根据深度卷积神经网络产生初步解析结果；(2) 依据解析结果得到关节点热图；(3) 运用超像素分割方法产生候选区域，并对候选区域建立人体特征分布模型；最后将各部分的损失函数叠加，得到我们最终的损失函数。

[0062] 2. 基于超像素分割产生人体候选区域，如图2所示，其步骤主要包括：(1) 初始化种子点；(2) 在种子点的 $S \times S$ 邻域内重新选择种子点；(3) 在邻域内为每个像素分配类别标签；(4) 对于每个搜索到的像素点，分别计算它和该种子点的颜色距离和空间距离，取最小值对应的种子点作为该像素点的聚类中心。(5) 判断是否收敛，若聚类中心不再变化，则认为收敛。否则，不收敛。

[0063] 3. 对每个候选区域通过提取颜色和纹理特征建立表象模型，步骤如下：(1) 提取颜色特征；本发明采用 $L \times a \times b \times \text{颜色空间模型}$ ，并通过统计每个区域内中颜色分布直方图的方法提取11维颜色特征，其中每一维向量描述了像素或某一图像区域中的颜色属于11种颜色的概率。(2) 提取纹理特征；纹理特征则采用LBP方法，首先将区域划分成 16×16 的小区域(cell)，对每个cell中的中心像素点与周围的8个像素点进行比较。如果中心像素比邻点像素值大，则将邻点赋值为1，否则赋值为0，这样每个点都会获得一个8位二进制数(可转换为一个十进制数)。然后计算每个cell的直方图，并对直方图进行归一化处理。最后将得到的每个cell的统计直方图进行连接，就得到了整个候选区域的LBP纹理特征。(3) 建立表象模型。将颜色特征和纹理特征串联即可建立表象模型。

[0064] 4. 建立人体特征分布模型。对给定的输入图像，计算每种关节点的高斯分布和人体总面积的高斯分布，从而建立面积比例模型，然后将表象模型和面积比例模型合并即可得到总的人体特征分布模型。对于每个超像素分割候选区域，通过人体特征分布模型即可求得人体特征分布损失函数。

[0065] 5. 最后将基础解析网络的softmax损失函数、关节结构损失函数和人体特征分布损失函数结合起来得到最终的损失函数，这种结合关节结构分布和人体特征分布的做法能够有效降低错分的概率。

[0066] 6. 本发明的部分实验结果如图3所示，其中，图3(a)所示为数据集的部分示例图，包含了光照、姿态等因素的影响。图3(b)所示为相应的初始解析结果，图3(c)所示为相应的超像素分割图。

[0067] 最后应说明的是：以上示例仅用以说明本发明而并非限制本发明所描述的技术方

案;因此,尽管本说明书参照上述的示例对本发明已进行了详细的说明,但是本领域的普通技术人员应当理解,仍然可以对本发明进行修改或等同替换;而一切不脱离发明的精神和范围的技术方案及其改进,其均应涵盖在本发明的权利要求范围当中。

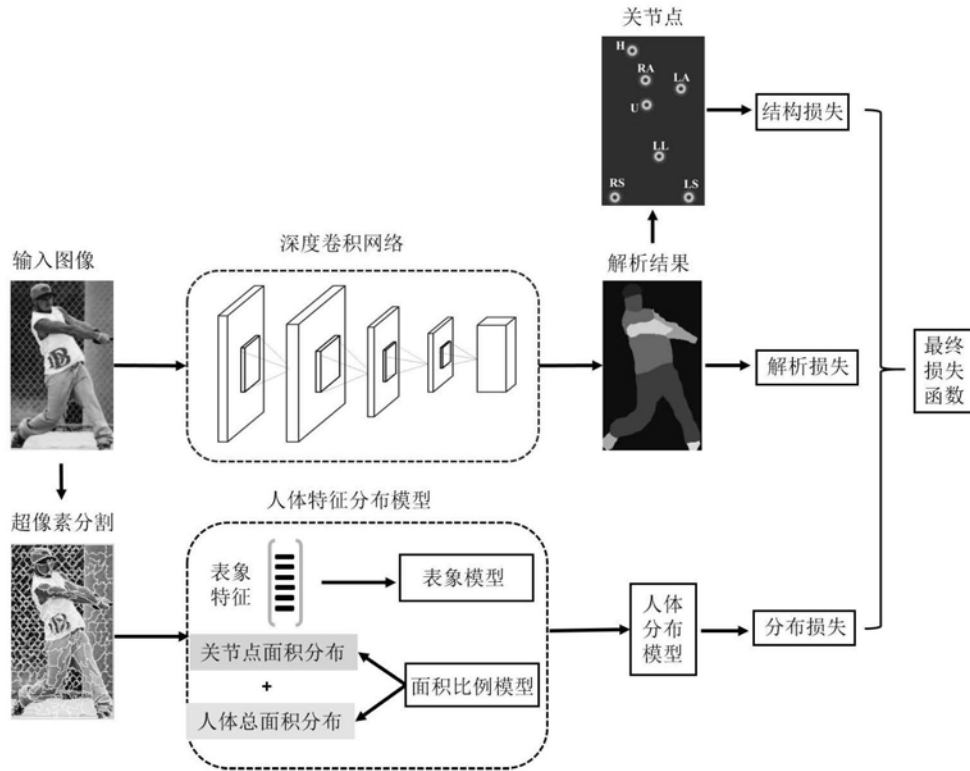


图1

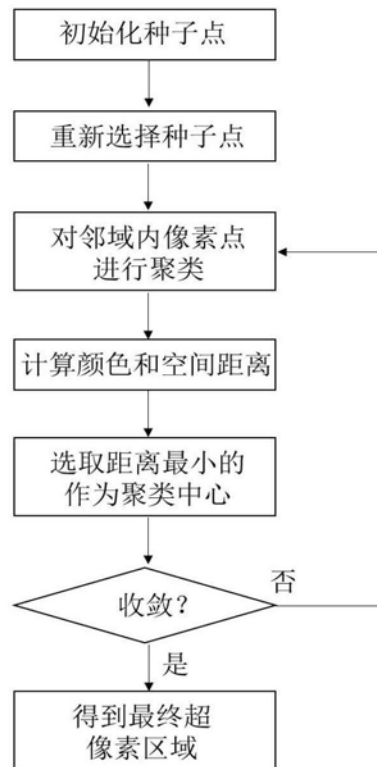


图2



图3 (a)

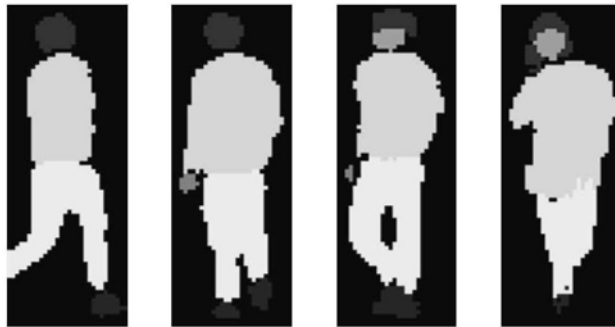


图3 (b)

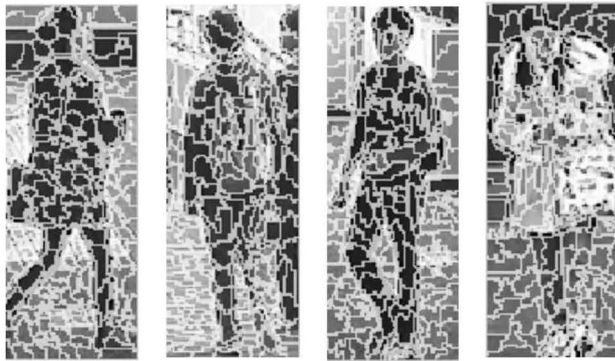


图3 (c)