



(12)发明专利申请

(10)申请公布号 CN 110033007 A
(43)申请公布日 2019.07.19

(21)申请号 201910321093.0

(22)申请日 2019.04.19

(71)申请人 福州大学

地址 350108 福建省福州市闽侯县上街镇
福州大学城学院路2号福州大学新区

(72)发明人 柯道 李振达

(74)专利代理机构 福州元创专利商标代理有限公司 35100

代理人 蔡学俊 陈明鑫

(51)Int.Cl.

G06K 9/46(2006.01)

G06K 9/62(2006.01)

G06N 3/04(2006.01)

G06T 5/00(2006.01)

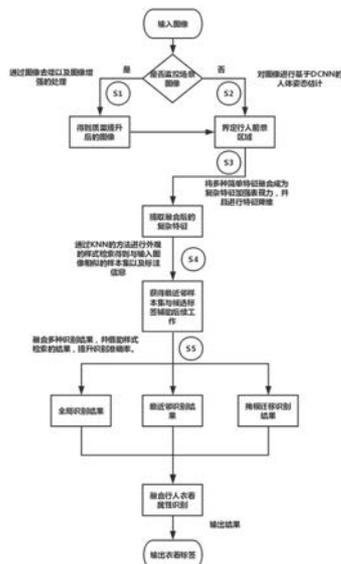
权利要求书4页 说明书9页 附图1页

(54)发明名称

基于深度姿态预估和多特征融合的行人衣着属性识别方法

(57)摘要

本发明涉及一种基于深度姿态预估和多特征融合的行人衣着属性识别方法。该方法首先通过外观特征匹配,选取部分检索结果用于后续的属性识别;随后通过基于SSD的深度人体姿态估计方法,能有效定位出图像中属于行人的前景区域,并较好地排除背景因素干扰;最后融合多种方式的解析结果,并结合迭代平滑过程,采取最大后验概率分配的方式,加强属性标签与像素之间的相关性,得到最终的属性解析识别结果。本发明解决了单一解析方式下标签识别不准、像素解析区域偏差等问题。该方法简单灵活,具有较强的实际应用性。



CN 110033007 A

1. 一种基于深度姿态预估和多特征融合的行人衣着属性识别方法,其特征在于,包括如下步骤:

步骤S1、对于监控场景下的输入图像通过图像去噪以及图像增强方式进行预处理,提升图像质量;

步骤S2、对预处理后的输入图像进行基于深度卷积神经网络DCNN的姿态预估,界定图像中的前景与背景区域,将姿态特征作为融合特征的其中一部分;

步骤S3、对经步骤S2处理后的图像的前景区域提取融合特征,并通过PCA进行特征降维;

步骤S4、利用一个公共数据集对输入图像进行样式检索,检索得到包括:相似图像样本、图像样本本身的服装标签标注;

步骤S5、将以上得到的不同形式的融合特征,输入设计的行人衣着属性识别框架,获得最终的衣着属性识别结果。

2. 根据权利要求1所述的基于深度姿态预估和多特征融合的行人衣着属性识别方法,其特征在于,所述步骤S1具体实现如下:

步骤S11、通过基于盲去卷积的图像去噪方法,解决监控场景下运动模糊带来的干扰,得到复原图像:

假设一个初始的复原图像 f_0 和退化函数 g_0 ,且图像各部分的模糊函数相同,通过如下迭代公式得到模糊干扰前的输入图像:

$$g_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{g_i^k(x) \otimes f^{k-1}(x)} \right] \otimes f^{k-1}(-x) \right\} g_i^k(x)$$

$$f_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{f_i^k(x) \otimes g_i^k(x)} \right] \otimes g_i^k(-x) \right\} f_i^k(x)$$

其中, g_i^k 表示第*i*轮第*k*次迭代时的退化函数, f_i^k 表示第*i*轮第*k*次迭代时的复原图像, $c(x)$ 为退化图像,即最初输入图像, \otimes 为卷积运算;

步骤S12、通过带色彩恢复的多尺度Retinex算法进行图像增强,加强监控场景下图像的色彩表现:

定义入射光为 $L(x,y)$,物体的反射图像为 $R(x,y)$,反射后的图像为 $S(x,y)$,那么有: $S(x,y) = L(x,y) \cdot R(x,y)$,将三个分量都映射到对数域,则有对应的对数结果为

$$\log S(x,y), \log L(x,y), \log R(x,y)$$

那么上式可以转化为:

$$\log S(x,y) = \log L(x,y) + \log R(x,y)$$

在多尺度的情况下引入颜色通道权重 ω_i ,则有:

$$\log R_{MSR}(x,y) = \sum_k^K \omega_k \{ \log S(x,y) - \log F_k(x,y) \otimes S(x,y) \}$$

其中, K 表示中心环绕函数 F 的个数;

定义 C_i 表示三个通道中其中一个通道的色彩恢复因子,用于平衡三个通道之间的比例,将相对较暗的区域凸显出来,达到消除失真的目的,那么MSRCR的模型可以表示为:

$$\log \text{RMSRCri}(x, y) = C_i(x, y) \log \text{RMSri}(x, y)$$

最后将其映射到实数域即为最终增强结果。

3. 根据权利要求1所述的基于深度姿态预估和多特征融合的行人衣着属性识别方法, 其特征在于, 在所述步骤S2中, 对预处理后的输入图像进行基于深度卷积神经网络DCNN的姿态预估的具体方式如下:

步骤S21、构建图像模型 $G = (V, E)$ 直观表示一个人体模型, 其中 V 表示人体的关节点或者身体某部位, E 为节点之间联结的边, $E \in V \times V$, 表示相邻节点之间的空间关系, 且令 $K = |V|$ 表示关节数量; 定义 I 表示图像, i 表示该图像中第 i 个节点, l 表示节点的像素坐标, t 表示节点的混合空间关系, 由不同姿态实例聚类抽象并组合, 那么根据图像模型中的定义, 有 $i \in \{1, \dots, K\}$, 定义 $l_i \in \{1, \dots, L\}$, $t_i \in \{1, \dots, T\}$, l_i 表示节点 i 的像素坐标 $\{(x_i, y_i)\}$, t_i 表示节点 i 与其相邻节点的空间关系类型集, 即该节点处的 T 个姿态类型;

步骤S22、由步骤S21, 人体关节部分的外观模型可以表示为:

$$F(l_i, t_i | I; \theta) = \sum_{i \in V} \phi(l_i, t_i | I; \theta)$$

且有

$$\phi(l_i, t_i | I; \theta) = \log p(l_i, t_i | I; \theta)$$

其中, $p(l_i, t_i | I; \theta)$ 是通过DCNN中前向传播Softmax函数最终计算得到的得分结果映射得到的概率域, 来预测在图像 I 中节点部分 i 的混合姿态类型为 t_i 并且像素坐标是位于 l_i 的概率, θ 是模型的参数;

关节间空间关系可以表示为:

$$F(l_i, l_j, t_i, t_j | I; \omega_{i,j}^{t_i, t_j}) = \sum_{i, j \in E} \omega_{i,j}^{t_i, t_j} \cdot \phi(l_i, l_j | I)$$

$$\phi(l_i, l_j | I) = \langle d(l_i - l_j) \rangle$$

对空间关系模型加入标准二次变化, 定义 $\langle d(l_i - l_j) \rangle = [dx \ dx^2 \ dy \ dy^2]^T$, 且 $dx = x_i - x_j$, $dy = y_i - y_j$, 表示节点 i 关于节点 j 的相对像素定位;

最后可得

$$F(l, t | I; \theta, \omega) = \sum_{i \in V} \phi(l_i, t_i | I; \theta) + \sum_{i, j \in E} \phi(l_i, l_j | I, \omega_{i,j}^{t_i, t_j})$$

上式表示人体姿态估计模型, 并以树形结构表示加速计算上的高效性;

步骤S23、通过K-means聚类的方法, 根据中心关节与相邻关节的空间相对位置, 对预处理得到的局部图像块进行聚类操作, 得到预训练模型;

步骤S24、利用DCNN进行训练, 通过得分函数模型将图像映射到不同的姿态类型, 并且根据标注信息, 通过损失函数调整权重、参数, 使映射的得分结果与实际类别吻合, 完成对姿态类型的分类, 得到DCNN多分类模型。

4. 根据权利要求1所述的基于深度姿态预估和多特征融合的行人衣着属性识别方法, 其特征在于, 所述步骤S3具体实现方式如下:

步骤S31、对输入图像提取包括颜色、梯度、纹理的简单特征并根据不同属性识别阶段融合成不同的融合复杂特征;

步骤S32、设经过特征描述器之后得到m条n维的数据,将原始数据按列组成m行n列的矩阵X;

步骤S33、将X的每一行减去这一行的均值;

步骤S34、求出协方差矩阵;

步骤S35、求出协方差矩阵的特征值及对应的特征向量r;

步骤S36、将特征向量r按对应特征值大小从上到下按行排列成矩阵,取前n'行组成新的矩阵P;

步骤S37、矩阵P即为降维到n'维后的数据。

5. 根据权利要求1所述的基于深度姿态预估和多特征融合的行人衣着属性识别方法,其特征在于,所述步骤S4具体实现方式如下:

步骤S41、首先在一个公共数据集上建立KD-tree索引树;

步骤S42、选取融合特征作为样本特征,对比输入图像以及数据集中样本的外观融合特征,结合L2-distance,通过KNN的聚类方法,搜索KD-tree;

步骤S43、选取前25个结果组成最近邻样本集,并且这些样本本身标注好的服装标签信息也组成候选标签,为后续衣着属性识别提供帮助。

6. 根据权利要求5所述的基于深度姿态预估和多特征融合的行人衣着属性识别方法,其特征在于,所述步骤S5具体实现方式如下:

步骤S51、定义样本中图像的像素为i,该像素的预测服装标签为 l_i ,该像素的复杂特征为 f_i ;定义通过最近邻检索得到的最近邻样本集为D,样本集中带有的已标注的标签集定义为 $\tau(D)$,t表示一个服装类目标签;定义每种解析都分别带有一个混合参数 $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 从而进行最终的置信度组合;

步骤S52、融合基于逻辑回归的全局识别,基于最近邻样例的近似识别,基于掩模转换的迁移识别的识别结果,提升识别准确率:

全局识别像素-标签置信度的函数模型如下:

$$C_{global}(l_i|f_i, D) \equiv P(l_i = t | f_i, \theta_t^g) \cdot 1[t \in \tau(D)]$$

P表示的是给定复杂特征 f_i 以及模型参数 θ_t^g 的逻辑回归结果,表示其中一个标签t在这个样本中存在的概率值, $1[\cdot]$ 是一个指示函数,表示标签t是最近邻样本的标签集中的一员,模型参数 θ_t^g 使用The Fashionista Dataset作为正样本进行训练;

近似识别像素-标签置信度的函数模型如下:

$$C_{nearest}(l_i|f_i, D) \equiv P(l_i = t | f_i, \theta_t^n) \cdot 1[t \in \tau(D)]$$

模型参数 θ_t^n 使用最近邻样本集D作为训练集训练;

基于掩模转换的迁移解析像素-标签置信度的函数模型为:

$$C_{transfer}(l_i|f_i, D) \equiv \frac{1}{Z} \sum_{d \in D} \frac{M(l_i, f_i, d)}{1 + \|h(s_i) - h(s_{i,d})\|}$$

$$M(l_i, s_{i,d}) \equiv \frac{1}{|s_{i,d}|} \sum_{j \in s_{i,d}} P(l_j = t | f_j, \theta_t^g) \cdot 1[t \in \tau(d)]$$

其中j表示存在于最近邻样本超像素块中的像素,参数 θ_i^s 为全局解析中的模型参数,因为 $M(l_i, s_i, d)$ 表示的是,针对最近邻样本超像素块区域进行全局解析得到的逻辑回归结果的平均值;

步骤S53、单一的置信度并不足以保证标签分配结果的精确率,因此,考虑将三种解析融合分析, $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 分别定义三种解析的权重比率,通过一个融合模型计算各个服装标签-像素的置信度:

$$\begin{aligned} C(l_i | f_i, D) &\equiv C_{global}(l_i | f_i, D)^{\lambda_1} \\ &\quad C_{nearest}(l_i | f_i, D)^{\lambda_2} \\ &\quad C_{transfer}(l_i | f_i, D)^{\lambda_3} \end{aligned}$$

步骤S54、结合迭代平滑的处理过程:定义所有像素的标签分配情况为 $L \equiv \{l_i\}$,即对图像中各像素标注服装标签的情况和服装项目外观模型 $\Theta \equiv \{\theta_i^c\}$,其中 θ_i^c 是一个服装类目标签t的融合外观模型,那么最终的优化结果是找到最优的像素-标签分配 L^* 以及外观模型集 Θ^* ;迭代过程开始时,定义初始的像素-标签分配情况为 \hat{L}_0 ,表示通过MAP分配结合首次经过融合解析置信度得到的像素-标签结果,初始服装类目标签的融合外观模型集合为 $\hat{\Theta}_0$,表示服装类目逻辑回归模型集合,使用 \hat{L}_0 作为训练数据训练;定义迭代的过程用常数k表示进行迭代的次数,E表示相邻的像素对,那么在第k次迭代时,其像素-标签分配情况为 \hat{L}_k ,融合外观模型 $\hat{\Theta}_k$,为则优化像素-标签分配情况模型为:

$$\hat{L}_k \in \arg \max_L \prod_i \Phi(l_i | f_i, C, \hat{\Theta}_k) \prod_{i,j \in E} \Psi(l_i, l_j | f_i, f_j)$$

其中: $\Phi(l_i | f_i, C, \hat{\Theta}_k) \equiv C(l_i | f_i)^\lambda \cdot P(l_i | f_i, \theta_i^c)^{1-\lambda}$,得到最终的识别结果。

基于深度姿态预估和多特征融合的行人衣着属性识别方法

技术领域

[0001] 本发明属于计算机视觉,深度学习,图像处理领域,并将其应用到智能监控、行人重识别等场景,具体为一种基于深度姿态预估和多特征融合的行人衣着属性识别方法。

背景技术

[0002] 现实世界中在监控视频获取的监视图像中进行行人属性识别具有较大的挑战性,其原因如下:(1)成像质量差,通常分辨率低,而且容易受运动模糊影响;(2)属性可能因为行人所佩戴或穿着的衣物的外观影响,并且不同的图像中也因为行人姿态不同,对应的属性也就处于图像中的不同空间位置;(3)来自监控视频图像中的标记属性数据难以收集并且仅可以少量获得。以上这些因素都使得通过训练学习一个行人属性模型变得非常困难。早期的属性识别方法主要依赖于手动提取的特征,例如该物品的颜色或者文本注释。而近些年,基于深度学习的行人属性识别模型开始吸引越来越多的人进行研究,这是由于深度学习得到的模型在大规模数据集下具有更加强大且稳定的学习能力,能够获得可以表示复杂特征的通用模型。同时由于通过监控视频获得的图像质量差,分辨率通常很低以及监控场景中的复杂的衣着外观变化,这些因素都无疑提高了深度学习在行人属性识别中的地位。

[0003] 行人衣着属性识别相当于一种多标签图像分类(MLIC-Multiple Label Image Classification)问题。现有的方法已经探索了顺序多标签预测。这些方法基于CNN-RNN模型设计。至关重要的是,这些现有的MLIC模型假定(1)大规模标记训练数据的可用性(2)具有足够好的图像质量。这两种假设对于监视图像中的行人属性识别都是无效的。最近的一种多人图像标注方法通过结合额外的人际关系和场景背景来推进这种连续的MLIC范式。该方法特别利用家庭成员和以朋友为中心的高分辨率照片图像的背景,但不能扩展到不良图像数据的开放世界监视场景。此外,需要强大的属性级别标签,而行人属性在图像级别大多是弱标签。

[0004] 监控场景下得到的图像中行人属性标签处于弱级别的原因,且现有的属性识别方法在进行识别时存在的因为环境因素影响出现的定位偏差。

[0005] 针对以上这些问题,本文提出利用行人姿态估计方法进一步界定图像中行人前景区域以及背景区域,排除背景因素的干扰。且通过图像处理的方式,提升监控场景下的图像质量,并且依托融合特征,以及融合属性识别的方法,加强识别方法的准确性。

发明内容

[0006] 本发明的目的在于提供一种基于深度姿态预估和多特征融合的行人衣着属性识别方法,以克服现有技术中存在的缺陷,并解决单一解析方式下标签识别不准、像素解析区域偏差的问题。

[0007] 为实现上述目的,本发明的技术方案是:一种基于深度姿态预估和多特征融合的行人衣着属性识别方法,包括如下步骤:

[0008] 步骤S1、对于监控场景下的输入图像通过图像去噪以及图像增强方式进行预处理,提升图像质量;

[0009] 步骤S2、对预处理后的输入图像进行基于深度卷积神经网络DCNN的姿态预估,界定图像中的前景与背景区域,将姿态特征作为融合特征的其中一部分;

[0010] 步骤S3、对经步骤S2处理后的图像的前景区域提取融合特征,并通过PCA进行特征降维;

[0011] 步骤S4、利用一个公共数据集对输入图像进行样式检索,检索得到包括:相似图像样本、图像样本本身的服装标签标注;

[0012] 步骤S5、将以上得到的不同形式的融合特征,输入设计的行人衣着属性识别框架,获得最终的衣着属性识别结果。

[0013] 在本发明一实施例中,所述步骤S1具体实现如下:

[0014] 步骤S11、通过基于盲去卷积的图像去噪方法,解决监控场景下运动模糊带来的干扰,得到复原图像:

[0015] 假设一个初始的复原图像 f_0 和退化函数 g_0 ,且图像各部分的模糊函数相同,通过如下迭代公式得到模糊干扰前的输入图像:

$$[0016] \quad g_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{g_i^k(x) \otimes f^{k-1}(x)} \right] \otimes f^{k-1}(-x) \right\} g_i^k(x)$$

$$[0017] \quad f_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{f_i^k(x) \otimes g_i^k(x)} \right] \otimes g_i^k(-x) \right\} f_i^k(x)$$

[0018] 其中, g_i^k 表示第i轮第k次迭代时的退化函数, f_i^k 表示第i轮第k次迭代时的复原图像, $c(x)$ 为退化图像,即最初输入图像, \otimes 为卷积运算;

[0019] 步骤S12、通过带色彩恢复的多尺度Retinex算法进行图像增强,加强监控场景下图像的色彩表现:

[0020] 定义入射光为 $L(x,y)$,物体的反射图像为 $R(x,y)$,反射后的图像为 $S(x,y)$,那么有: $S(x,y) = L(x,y) \cdot R(x,y)$,将三个分量都映射到对数域,则有对应的对数结果为

$$[0021] \quad \log S(x,y), \log L(x,y), \log R(x,y)$$

[0022] 那么上式可以转化为:

$$[0023] \quad \log S(x,y) = \log L(x,y) + \log R(x,y)$$

[0024] 在多尺度的情况下引入颜色通道权重 ω_i ,则有:

$$[0025] \quad \log R_{MSR}(x,y) = \sum_k^K \omega_k \{ \log S(x,y) - \log F_k(x,y) \otimes S(x,y) \}$$

[0026] 其中,K表示中心环绕函数F的个数;

[0027] 定义 C_i 表示三个通道中其中一个通道的色彩恢复因子,用于平衡三个通道之间的比例,将相对较暗的区域凸显出来,达到消除失真的目的,那么MSRCR的模型可以表示为:

$$[0028] \quad \log R_{MSRCR_i}(x,y) = C_i(x,y) \log R_{MSR_i}(x,y)$$

[0029] 最后将其映射到实数域即为最终增强结果。

[0030] 在本发明一实施例中,在所述步骤S2中,对预处理后的输入图像进行基于深度卷

积神经网络DCNN的姿态预估的具体方式如下：

[0031] 步骤S21、构建图像模型 $G=(V,E)$ 直观表示一个人体模型，其中 V 表示人体的关节点或者身体某部位， E 为节点之间联结的边， $E \in V \times V$ ，表示相邻节点之间的空间关系，且令 $K=|V|$ 表示关节数量；定义 I 表示图像， i 表示该图像中第 i 个节点， l 表示节点的像素坐标， t 表示节点的混合空间关系，由不同姿态实例聚类抽象并组合，那么根据图像模型中的定义，有 $i \in \{1, \dots, K\}$ ，定义 $l_i \in \{1, \dots, L\}$ ， $t_i \in \{1, \dots, T\}$ ， l_i 表示节点 i 的像素坐标 $\{(x_i, y_i)\}$ ， t_i 表示节点 i 与其相邻节点的空间关系类型集，即该节点处的 T 个姿态类型；

[0032] 步骤S22、由步骤S21，人体关节部分的外观模型可以表示为：

$$[0033] \quad F(l_i, t_i | I; \theta) = \sum_{i \in V} \phi(l_i, t_i | I; \theta)$$

[0034] 且有

$$[0035] \quad \phi(l_i, t_i | I; \theta) = \log p(l_i, t_i | I; \theta)$$

[0036] 其中， $p(l_i, t_i | I; \theta)$ 是通过DCNN中前向传播Softmax函数最终计算得到的得分结果映射得到的概率域，来预测在图像 I 中节点部分 i 的混合姿态类型为 t_i 并且像素坐标是位于 l_i 的概率， θ 是模型的参数；

[0037] 关节间空间关系可以表示为：

$$[0038] \quad F(l_i, l_j, t_i, t_j | I; \omega_{i,j}^{t_i, t_j}) = \sum_{i, j \in E} \omega_{i,j}^{t_i, t_j} \cdot \phi(l_i, l_j | I)$$

$$[0039] \quad \phi(l_i, l_j | I) = \langle d(l_i - l_j) \rangle$$

[0040] 对空间关系模型加入标准二次变化，定义 $\langle d(l_i - l_j) \rangle = [dx \ dx^2 \ dy \ dy^2]^T$ ，且 $dx = x_i - x_j$ ， $dy = y_i - y_j$ ，表示节点 i 关于节点 j 的相对像素定位；

[0041] 最后可得

$$[0042] \quad F(l, t | I; \theta, \omega) = \sum_{i \in V} \phi(l_i, t_i | I; \theta) + \sum_{i, j \in E} \phi(l_i, l_j | I, \omega_{i,j}^{t_i, t_j})$$

[0043] 上式表示人体姿态估计模型，并以树形结构表示加速计算上的高效性；

[0044] 步骤S23、通过K-means聚类的方法，根据中心关节与相邻关节的空间相对位置，对预处理得到的局部图像块进行聚类操作，得到预训练模型；

[0045] 步骤S24、利用DCNN进行训练，通过得分函数模型将图像映射到不同的姿态类型，并且根据标注信息，通过损失函数调整权重、参数，使映射的得分结果与实际类别吻合，完成对姿态类型的分类，得到DCNN多分类模型。

[0046] 在本发明一实施例中，所述步骤S3具体实现方式如下：

[0047] 步骤S31、对输入图像提取包括颜色、梯度、纹理的简单特征并根据不同属性识别阶段融合成不同的融合复杂特征；

[0048] 步骤S32、设经过特征描述器之后得到 m 条 n 维的数据，将原始数据按列组成 m 行 n 列的矩阵 X ；

[0049] 步骤S33、将 X 的每一行减去这一行的均值；

[0050] 步骤S34、求出协方差矩阵；

[0051] 步骤S35、求出协方差矩阵的特征值及对应的特征向量 r ；

[0052] 步骤S36、将特征向量 r 按对应特征值大小从上到下按行排列成矩阵,取前 n' 行组成新的矩阵 P ;

[0053] 步骤S37、矩阵 P 即为降维到 n' 维后的数据。

[0054] 在本发明一实施例中,所述步骤S4具体实现方式如下:

[0055] 步骤S41、首先在一个公共数据集上建立KD-tree索引树;

[0056] 步骤S42、选取融合特征作为样本特征,对比输入图像以及数据集中样本的外观融合特征,结合L2-distance,通过KNN的聚类方法,搜索KD-tree;

[0057] 步骤S43、选取前25个结果组成最近邻样本集,并且这些样本本身标注好的服装标签信息也组成候选标签,为后续衣着属性识别提供帮助。

[0058] 在本发明一实施例中,所述步骤S5具体实现方式如下:

[0059] 步骤S51、定义样本中图像的像素为 i ,该像素的预测服装标签为 l_i ,该像素的复杂特征为 f_i ;定义通过最近邻检索得到的最近邻样本集为 D ,样本集中带有的已标注的标签集定义为 $\tau(D)$, t 表示一个服装类目标签;定义每种解析都分别带有一个混合参数 $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 从而进行最终的置信度组合;

[0060] 步骤S52、融合基于逻辑回归的全局识别,基于最近邻样例的近似识别,基于掩模转换的迁移识别的识别结果,提升识别准确率:

[0061] 全局识别像素-标签置信度的函数模型如下:

$$[0062] \quad C_{global}(l_i | f_i, D) \equiv P(l_i = t | f_i, \theta_t^g) \cdot 1[t \in \tau(D)]$$

[0063] P 表示的是给定复杂特征 f_i 以及模型参数 θ_t^g 的逻辑回归结果,表示其中一个标签 t 在这个样本中存在的概率值, $1[\cdot]$ 是一个指示函数,表示标签 t 是最近邻样本的标签集中的一员,模型参数 θ_t^g 使用The FashionistaDataset作为正样本进行训练;

[0064] 近似识别像素-标签置信度的函数模型如下:

$$[0065] \quad C_{nearest}(l_i | f_i, D) \equiv P(l_i = t | f_i, \theta_t^n) \cdot 1[t \in \tau(D)]$$

[0066] 模型参数 θ_t^n 使用最近邻样本集 D 作为训练集训练;

[0067] 基于掩模转换的迁移解析像素-标签置信度的函数模型为:

$$[0068] \quad C_{transfer}(l_i | f_i, D) \equiv \frac{1}{Z} \sum_{d \in D} \frac{M(l_i, f_i, d)}{1 + \|h(s_i) - h(s_i, d)\|}$$

$$[0069] \quad M(l_i, s_i, d) \equiv \frac{1}{|s_i, d|} \sum_{j \in s_i, d} P(l_j = t | f_j, \theta_t^g) \cdot 1[t \in \tau(d)]$$

[0070] 其中 j 表示存在于最近邻样本超像素块中的像素,参数 θ_t^g 为全局解析中的模型参数,因为 $M(l_i, s_i, d)$ 表示的是,针对最近邻样本超像素块区域进行全局解析得到的逻辑回归结果的平均值;

[0071] 步骤S53、单一的置信度并不足以保证标签分配结果的精确率,因此,考虑将三种解析融合分析, $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 分别定义三种解析的权重比率,通过一个融合模型计算各个服装标签-像素的置信度:

$$C(l_i | f_i, D) \equiv C_{global}(l_i | f_i, D)^{\lambda_1}.$$

$$[0072] \quad C_{nearest}(l_i | f_i, D)^{\lambda_2}.$$

$$C_{transfer}(l_i | f_i, D)^{\lambda_3}.$$

[0073] 步骤S54、结合迭代平滑的处理过程：定义所有像素的标签分配情况为 $L \equiv \{l_i\}$ ，即对图像中各像素标注服装标签的情况和服装项目外观模型 $\Theta \equiv \{\theta_t^c\}$ ，其中 θ_t^c 是一个服装类目标签 t 的融合外观模型，那么最终的优化结果是找到最优的像素-标签分配 L^* 以及外观模型集 Θ^* ；迭代过程开始时，定义初始的像素-标签分配情况为 \hat{L}_0 ，表示通过MAP分配结合首次经过融合解析置信度得到的像素-标签结果，初始服装类目标签的融合外观模型集合为 $\hat{\Theta}_0$ ，表示服装类目逻辑回归模型集合，使用 \hat{L}_0 作为训练数据训练；定义迭代的过程用常数 k 表示进行迭代的次数， E 表示相邻的像素对，那么在第 k 次迭代时，其像素-标签分配情况为 \hat{L}_k ，融合外观模型 $\hat{\Theta}_k$ ，为则优化像素-标签分配情况模型为：

$$[0074] \quad \hat{L}_k \in \arg \max_L \prod_i \Phi(l_i | f_i, C, \hat{\Theta}_k) \prod_{i,j \in E} \Psi(l_i, l_j | f_i, f_j)$$

[0075] 其中： $\Phi(l_i | f_i, C, \hat{\Theta}_k) \equiv C(l_i | f_i)^{\lambda} \cdot P(l_i | f_i, \theta_t^c)^{1-\lambda}$ ，得到最终的识别结果。

[0076] 相较于现有技术，本发明具有以下有益效果：本发明通过结合基于深度学习的人体姿态估计进一步界定图像中行人前景区域以及背景区域，排除背景因素的干扰。同时，通过图像去噪和图像增强的方法提升监控场景下的图像质量。再者，融合多种简单特征形成复杂描述特征，加强衣着属性的表现力，并融合单一识别方式的结果，且通过迭代平滑处理提升识别的准确率。本专利融合深度人体姿态估计结果与多重融合特征，可以准确地识别行人衣着属性。该方法简单，实现灵活，实用性较强。

附图说明

[0077] 图1为本发明基于深度姿态预估和多特征融合的行人衣着属性识别方法的流程图。

具体实施方式

[0078] 下面结合附图，对本发明的技术方案进行具体说明。

[0079] 如图1所示，本发明提供了一种基于深度姿态预估和多特征融合的行人衣着属性识别方法。针对现有属性识别方法存在环境因素干扰进而影响定位准确率等问题，提出基于行人姿态估计与多特征融合的行人属性识别方法。该方法首先通过外观特征匹配，选取部分检索结果用于后续的属性识别。随后通过基于SSD的深度人体姿态估计方法，能有效定位出图像中属于行人的前景区域，并较好地排除背景因素干扰。最后融合多种方式的解析结果，并结合迭代平滑过程，采取最大后验概率分配的方式，加强属性标签与像素之间的相关性，得到最终的属性解析识别结果。本发明解决了单一解析方式下标签识别不准、像素解析区域偏差等问题。具体步骤如下：

[0080] 步骤S1、对于监控场景下的输入图像通过经典的图像去噪以及图像增强方式提升

图像质量。

[0081] 步骤S2、对输入图像进行基于深度卷积神经网络 (DCNN) 的姿态预估界定图像中的前景与背景区域,且姿态特征也作为融合特征的其中一部分,后续识别工作将基于图像中的前景区域展开。

[0082] 步骤S3、对输入的图像提取颜色,梯度,纹理等多种简单特征融合成为复杂特征,并通过PCA进行特征降维。

[0083] 步骤S4、利用一个公共数据集对输入图像进行样式检索,最终检索得到的结果用于帮助后续的属性识别。结果包括a.相似图像样本.b.图像样本本身的服装标签标注。

[0084] 步骤S5、将以上得到的不同形式的融合特征,输入设计的行人衣着属性识别框架,获得最终的衣着属性识别结果。

[0085] 进一步的,在本实施例中,在所述步骤S1中,通过如下步骤提升图像质量:

[0086] 步骤S11、通过基于盲去卷积的图像去噪方法,一定程度上解决监控场景下运动模糊带来的干扰,得到复原图像。假设一个初始的复原图像 f_0 和退化函数 g_0 ,且图像各部分的模糊函数相同,通过如下迭代公式得到模糊干扰前的输入图像:

$$[0087] \quad g_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{g_i^k(x) \otimes f^{k-1}(x)} \right] \otimes f^{k-1}(-x) \right\} g_i^k(x)$$

$$[0088] \quad f_{i+1}^k(x) = \left\{ \left[\frac{c(x)}{f_i^k(x) \otimes g_i^k(x)} \right] \otimes g_i^k(-x) \right\} f_i^k(x)$$

[0089] 其中 g_i^k 表示第i轮第k次迭代时的退化函数, f_i^k 表示第i轮第k次迭代时的复原图像, $c(x)$ 为退化图像,即最初输入图像, \otimes 为卷积运算。

[0090] 步骤S12、通过带色彩恢复的多尺度Retinex (Multiple Scale Retinex with Color Restoration MSRCR) 算法进行图像增强,加强监控场景下图像的色彩表现。定义入射光为 $L(x, y)$,物体的反射图像为 $R(x, y)$,反射后的图像为 $S(x, y)$,那么有: $S(x, y) = L(x, y) \cdot R(x, y)$,将三个分量都映射到对数域,则有对应的对数结果为 $\log S(x, y)$, $\log L(x, y)$, $\log R(x, y)$,那么上式可以转化为: $\log S(x, y) = \log L(x, y) + \log R(x, y)$,在多尺度的情况下引入颜色通道权重 ω_i ,则有:

$$[0091] \quad \log R_{MSR}(x, y) = \sum_k^K \omega_k \{ \log S(x, y) - \log F_k(x, y) \otimes S(x, y) \}$$

[0092] K表示中心环绕函数F的个数,取值为3。

[0093] 定义 C_i 表示三个通道中某个通道的色彩恢复因子,用于平衡三个通道之间的比例,将相对较暗的区域凸显出来,达到消除失真的目的,那么MSRCR的模型可以表示为: $\log R_{MSRCR}(x, y) = C_i(x, y) \log R_{MSR}(x, y)$ 。最后将其映射到实数域即为最终增强结果。

[0094] 进一步的,在本实施例中,在所述步骤S2中,通过如下步骤进行基于DCNN的人体姿态估计:

[0095] 步骤S21、构建图像模型 $G = (V, E)$ 直观表示一个人体模型,其中V表示人体的关节点或者身体某部位,E为节点之间联结的边, $E \in V \times V$,表示相邻节点之间的空间关系,且令 $K = |V|$ 表示关节数量。定义I表示图像,i表示该图像中第i个节点,l表示节点的像素坐标,t表示节点的混合空间关系(由不同姿态实例聚类抽象并组合),那么根据图像模型中的定

义,有 $i \in \{1, \dots, K\}$,定义 $l_i \in \{1, \dots, L\}$, $t_i \in \{1, \dots, T\}$, l_i 表示节点 i 的像素坐标 $\{(x_i, y_i)\}$, t_i 表示节点 i 与其相邻节点的空间关系类型集(即该节点处的 T 个姿态类型)。

[0096] 步骤S22、那么人体关节部分的外观模型可以表示为:
$$F(l_i, t_i | I; \theta) = \sum_{i \in V} \phi(l_i, t_i | I; \theta),$$

且有 $\phi(l_i, t_i | I; \theta) = \log p(l_i, t_i | I; \theta)$,其中 $p(l_i, t_i | I; \theta)$ 是通过DCNN中前向传播Softmax函数最终计算得到的得分结果映射得到的概率域,来预测在图像 I 中节点部分 i 的混合姿态类型为 t_i 并且像素坐标是位于 l_i 的概率, θ 是模型的参数。关节间空间可以表示为:

$$F(l_i, l_j, t_i, t_j | I; \omega_{i,j}^{t_i, t_j}) = \sum_{i,j \in E} \omega_{i,j}^{t_i, t_j} \cdot \phi(l_i, l_j | I), \phi(l_i, l_j | I) = \langle d(l_i - l_j) \rangle,$$
对空间关系模型加入标准二次

变化,定义 $\langle d(l_i - l_j) \rangle = [dx \ dx^2 \ dy \ dy^2]^T$,且 $dx = x_i - x_j, dy = y_i - y_j$,表示节点 i 关于节点 j 的相对像素定位。

[0097] 最后得到

[0098]
$$F(l, t | I; \theta, \omega) = \sum_{i \in V} \phi(l_i, t_i | I; \theta) + \sum_{i,j \in E} \phi(l_i, l_j | I, \omega_{i,j}^{t_i, t_j})$$

[0099] 上式表示人体姿态估计模型,并以树形结构表示加速计算上的高效性。

[0100] 步骤S23、通过K-means聚类的方法,根据中心关节与相邻关节的空间相对位置,对预处理得到的局部图像块进行聚类操作,得到预训练模型,为后续训练提供帮助。

[0101] 步骤S24、利用DCNN网络进行训练,通过得分函数模型将图像映射到不同的姿态类型,并且根据标注信息,通过损失函数调整权重、参数,使映射的得分结果与实际类别吻合,完成对姿态类型的分类,得到DCNN多分类模型。

[0102] 进一步的,在本实施例中,在所述步骤S3中,通过如下步骤提取图像融合特征并进行特征降维,降低计算复杂度:

[0103] 步骤S31、对输入图像提取颜色,梯度,纹理等简单特征并根据不同属性识别阶段融合成不同的融合复杂特征。

[0104] 步骤S32、设经过特征描述器之后得到 m 条39168维的数据,将原始数据按列组成 m 行39168列的矩阵 X ;

[0105] 步骤S33、将 X 的每一行(代表一个属性字段)减去这一行的均值;

[0106] 步骤S34、求出协方差矩阵;

[0107] 步骤S35、求出协方差矩阵的特征值及对应的特征向量 r ;

[0108] 步骤S36、将特征向量按对应特征值大小从上到下按行排列成矩阵,取前441行组成新的矩阵 P ;

[0109] 步骤S37、矩阵 P 即为降维到441维后的数据。

[0110] 进一步的,在本实施例中,在所述步骤S4中,通过如下步骤进行样式检索以及候选标签的获取,辅助后续衣着属性识别:

[0111] 步骤S41、首先在一个公共数据集上建立KD索引树。

[0112] 步骤S42、选取融合特征作为样本特征,对比输入图像以及数据集中样本的外观融合特征,结合L2-distance,通过KNN(KNearestNeighbors)的聚类方法,搜索KD-tree。

[0113] 步骤S43、选取前25个结果组成最近邻样本集,并且这些样本本身标注好的服装标签信息也组成候选标签,为后续衣着属性识别提供帮助。

[0114] 进一步的,在本实施例中,在所述步骤S5中,通过融合深度人体姿态估计得到的结果,尽可能排除背景因素的干扰,且融合多种识别方式的结果,结合迭代平滑的处理,加强识别准确率:

[0115] 步骤S51、定义样本中的像素为*i*,该像素的预测服装标签为 l_i ,该像素的复杂特征为 f_i 。定义通过最近邻检索得到的最近邻样本集为*D*,样本集中带有的已标注的标签集定义为 $\tau(D)$, t 表示一个服装类目标签。定义每种解析都分别带有一个混合参数 $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 从而进行最终的置信度组合。

[0116] 步骤S52、融合基于逻辑回归的全局识别,基于最近邻样例的近似识别,基于掩模转换的迁移识别的识别结果,提升识别准确率。全局识别像素-标签置信度的函数模型如下:

[0117] $C_{global}(l_i | f_i, D) \equiv P(l_i = t | f_i, \theta_t^g) \cdot 1[t \in \tau(D)]$, P 表示的是给定复杂特征 f_i 以及模型参数 θ_t^g 的逻辑回归结果,表示某个标签 t 在这个样本中存在的概率值, $1[\cdot]$ 是一个指示函数,表示标签 t 是最近邻样本的标签集中的一员,模型参数 θ_t^g 使用The FashionistaDataset作为正样本进行训练。近似识别像素-标签置信度的函数模型如下:

$C_{nearest}(l_i | f_i, D) \equiv P(l_i = t | f_i, \theta_t^n) \cdot 1[t \in \tau(D)]$, P 表示的是给定复杂特征 f_i 以及模型参数 θ_t^n 的逻辑回归结果,模型参数 θ_t^n 使用最近邻样本集*D*作为训练集训练。基于掩模转换的迁移解析像素-标签置信度的函数模型为:

$$[0118] \quad C_{transfer}(l_i | f_i, D) \equiv \frac{1}{Z} \sum_{d \in D} \frac{M(l_i, f_i, d)}{1 + \|h(s_i) - h(s_i, d)\|}$$

$$[0119] \quad M(l_i, s_i, d) \equiv \frac{1}{|s_i, d|} \sum_{j \in s_i, d} P(l_j = t | f_j, \theta_t^g) \cdot 1[t \in \tau(d)]$$

[0120] 其中*j*表示存在于最近邻样本超像素块中的像素,参数 θ_t^g 为全局解析中的模型参数,因为 $M(l_i, s_i, d)$ 表示的是,针对最近邻样本超像素块区域进行全局解析得到的逻辑回归结果的平均值。

[0121] 步骤S53、单一的置信度并不足以保证标签分配结果的精确率,因此,考虑将三种解析融合分析, $\Lambda \equiv [\lambda_1, \lambda_2, \lambda_3]$ 分别定义三种解析的权重比率,通过一个融合模型计算各个服装标签-像素的置信度:

$$[0122] \quad C(l_i | f_i, D) \equiv C_{global}(l_i | f_i, D)^{\lambda_1} \cdot C_{nearest}(l_i | f_i, D)^{\lambda_2} \cdot C_{transfer}(l_i | f_i, D)^{\lambda_3}.$$

[0123] 步骤S54、结合迭代平滑的处理过程。定义所有像素的标签分配情况为 $L \equiv \{l_i\}$ (即对图像中各像素标注服装标签的情况)和服装项目外观模型 $\Theta \equiv \{\theta_t^c\}$,其中 θ_t^c 是一个服装类目标签 t 的融合外观模型,那么最终的优化结果是找到最优的像素-标签分配 L^* 以及外观模型集 Θ^* 。迭代过程开始时,定义初始的像素-标签分配情况为 \hat{L}_0 ,表示通过MAP分配结合首次经过融合解析置信度得到的像素-标签结果,初始服装类目标签的融合外观模型集合为

$\hat{\Theta}_0$ (服装类目逻辑回归模型集合), 使用 \hat{L}_0 作为训练数据训练。定义迭代的过程用常数 k 表示进行迭代的次数, E 表示相邻的像素对, 那么在第 k 次迭代时, 其像素-标签分配情况为 \hat{L}_k , 融合外观模型 $\hat{\Theta}_k$, 为则优化像素-标签分配情况模型为:

$$[0124] \quad \hat{L}_k \in \arg \max_L \prod_i \Phi(l_i | f_i, C, \hat{\Theta}_k) \prod_{i,j \in E} \Psi(l_i, l_j | f_i, f_j)$$

[0125] 其中: $\Phi(l_i | f_i, C, \hat{\Theta}_k) \equiv C(l_i | f_i)^\lambda \cdot P(l_i | f_i, \theta_i^c)^{1-\lambda}$, 得到最终的识别结果。

[0126] 以上是本发明的较佳实施例, 凡依本发明技术方案所作的改变, 所产生的功能作用未超出本发明技术方案的范围时, 均属于本发明的保护范围。

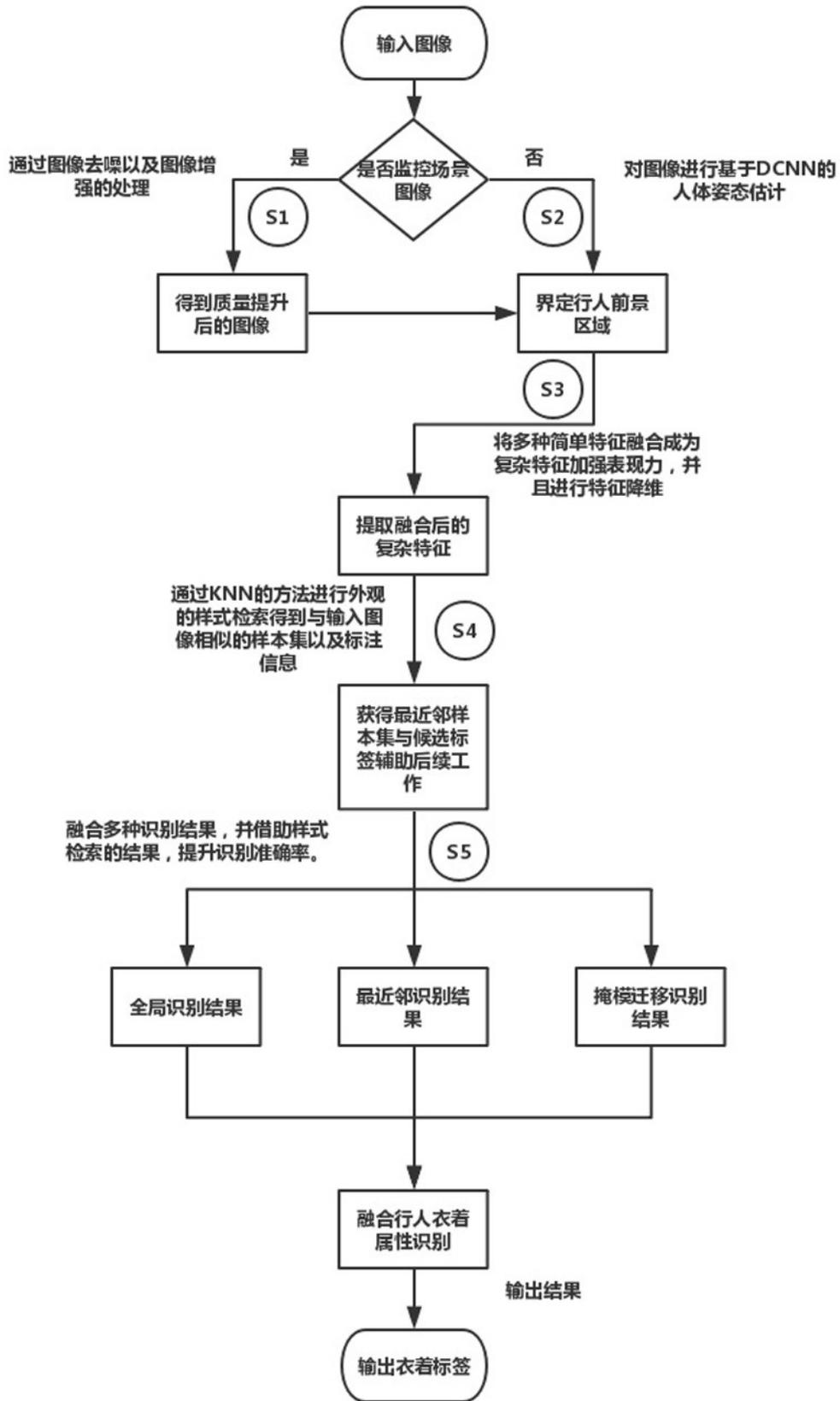


图1