

基于 poselets 的特定位置人物多姿势提取^①

王维兰, 刘秉瀚

(福州大学 数学与计算机科学学院, 福州 350108)

摘要: 目前, 对于人物识别的研究依然是一个非常具有挑战性的难题, 结合多姿势来进行人物识别则是一个新的课题, 因此准确提取多姿势样本是人物识别关键的一步. Poselets 算法可以检测出图像中的所有人物及其相应的姿势, 但是无法对特定位置的人物进行定位. 因此本文提出了一种基于 poselets 的特定位置人物姿势提取的方法: 首先根据特定位置人物头部标定框设置过滤模型, 通过过滤模型对图像中由 poselets 算法检出的人物框进行筛选, 并对筛选结果进行排序, 然后结合排序得分利用二分图最大权值匹配算法对筛选结果进行匹配, 找到特定位置的目标人物, 提取对应的姿势. 实验表明, 本文算法能有效精确的检测特定位置的人物, 并提取出相应的人物姿势.

关键词: poselets; 特定位置; 过滤模型; 二分图; 多姿势

Multi-Pose Extraction of Person in Specific Location Based on Poselets

WANG Wei-Lan, LIU Bing-Han

(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

Abstract: At present, the study on people recognition is still a quite challenging problem, and it is a new subject to combine with multi-pose for person recognition. Therefore, it is a significant step for people recognition to capture multi-pose samples accurately. Poselets can detect all the people and the corresponding position in the image, but it cannot locate the people on a particular location. Therefore, this paper presents a method of pose extraction which based on poselets for the specific location: First, we should set filtering model according to the head calibration form of the human in specific location, and then using the filtering model, we can screen the figure box detected by the poselets algorithm and sort the results of the screening. Then, we can find the target person of the specific location by combining the sorting score with the maximum weight bipartite graph matching algorithm. Finally, it is easy to extract the corresponding pose. Experimental results show that the algorithm mentioned in this paper can detect the person of the specific location effectively and extract the corresponding pose.

Key words: poselets; specific location; filtering model; bipartite graph; multi-pose

人物识别在一个缺乏明确的姿势特点的情况下, 只能依靠一些其他的线索, 包括人的侧脸, 手臂, 背影等方面特征来进行^[1]. 因此, 对特定目标人物进行检测, 分析对应人物的动作姿态, 并提取出多姿态特征, 是人物识别的关键.

目前, 目标检测依然是计算机视觉领域的研究热点, 由于不同目标的形状的不确定性, 应用场景的复

杂性, 目标与目标之间以及目标与背景的相互遮挡都使得其难度大大增加.

现有的目标检测算法都是先对图像提取特征^[2-5], 再对目标进行检测. 文献[6]使用导向梯度(HOG)^[7]的直方图迭代进行快速人物检测, 但准确率不高; 文献[8]利用导向梯度(HOG)提取特征, 采用SVM训练对人物和车辆进行检测; 文献[9]采用二维 Gabor 小波变换

^① 基金项目:福建省科技厅项目(2013J01186,JK2010056);福建省教育厅项目(JB10160)

收稿时间:2016-07-13;收到修改稿时间:2016-09-05 [doi:10.15888/j.cnki.csa.005681]

对图像进行特征提取,再利用 Adaboost 算法训练特征,能够对人体目标进行检测;文献[10]提出一种多尺度方向特征描述子,包含图像方块特征和 Gabor 小波变换提取的特征,分别用于级联 Adaboost 算法和 SVM 进行训练,能够检测图像中相互遮挡的人体目标.而 poselets 算法^[11]是将人物姿势分别提取导向梯度(HOG)特征,每类姿势特征进行 SVM 训练,结合训练好的所有 SVM 分类器,对人物目标进行检测,并可以提取相应人物的姿势.

其他目标检测算法都只能检测出人物目标的存在,并不能对其相应目标人物的姿势进行提取. Poselets 算法不仅可以对人物目标进行检测,还能够提取出对应人物的姿势.但是, poselets 算法是对图像中所有目标人物进行检测,无法对特定位置的人物进行定位.因此,针对此问题,本文在 poselets 算法的基础上,提出根据特定头部框位置初步筛选 poselets 算法检测出的人物框并结合排序得分利用二分图最大权值匹配算法^[12]对人物框进行最佳匹配的方法,精确定位特定人物框,提取其相应的姿势.

1 Poselets算法

Poselets 算法通过寻找姿势候选集,对候选集提取 HOG 特征,并对每个姿势候选集进行 SVM 分类器模型训练,总共训练了 156 个姿势分类器,将其进行组合得到最终的姿态检测器.它能够对输入的图像进行全面扫描,得到整张图像的各部位姿势和检测出图像中存在的所有人物框.作为一个好的姿势特征,这个姿势无论是在二维空间或者是三维空间里面,它都应当是紧簇的^[11]. Poselets 算法定义了在三空间中样例 y 到样例 z 的距离:

$$d_y(z) = \sum_j \omega_y(j) \|x_y(j) - x_z(j)\|_2^2 (1 + \theta_{y,z}(j)) \quad (1)$$

$x_y(j)$ 代表样例 y 的第 j 个关键点的三维坐标, $\omega_y(j)$ 是权重项,这里权值矩阵的分布是与高斯分布等价的. $\theta_{y,z}(j)$ 在这边是代表惩罚参数,如果关键点 j 是可见的或者是在样例 y 和样例 z 中都是不可见的,那么 $\theta_{y,z}(j)=0$, 否则, $\theta_{y,z}(j)=a$, $a>0$. 对于要进行处理的图像会从给定的人物注释中找出一个矩形窗口,做一个滑动窗口对于图片中的人物进行多尺度扫描,并在基于 $d_y(z)$ 的基础上用权重最小二乘法推导得出相似变换信息从而得到最合适的匹配窗口,找出最合适的姿势集

用于训练分类器.为了提高检测的精确度,用每个姿势分类器去检测输入图像中的每个位置,对于每个位置姿势分类器都能够得出一个命中值得分,结合所有姿势分类器的命中值得分去确定人物对象的位置.式(2)即为目标对象 O 在 x 位置上的概率分值:

$$P(O|x) \propto \sum_i \omega_i a_i(x) \quad (2)$$

$a_i(x)$ 是一个姿势分类器对 x 位置上的得分值, ω_i 是这个姿势分类器的权重值.

2 基于 poselets 的特定位置人物多姿势提取

Poselets 算法对每个检测到的人物框都有一个概率分值,而且还有相应每个人物的姿势.但是,它是对图像中所有人物进行扫描,无法检测特定位置的人物姿势,本文针对这一问题,结合给定的特定位置的头部框信息,提出了一种基于 poselets 的特定位置人物多姿势提取方法:首先,利用 poselets 算法检测出图像中的所有人物框;其次,利用给定的特定头部框信息设定了一个过滤模型,对 poselets 算法中检测出的人物框进行初步筛选,并建立一个得分准则对筛选出的人物框进行排序;最后,融合排序得分和二分图最大权值匹配算法^[12]对筛选出的人物框进行最佳匹配,精确定位特定人物框,提取其相应的姿势.

2.1 poselets 算法检测人物框

输入一张 $w*h$ 的图像, w 和 h 分别为图像的宽度和高度.运行 poselets 姿势分类器对输入图像进行扫描,每个姿势分类器对输入图像的每个位置计算得分,最终结合所有姿势分类器的结果检测出图像中所有人物框的位置.设人物框表示为: $O=(lx,ly,rx,ry)$, (lx,ly) 为左上角坐标, (rx,ry) 为右上角坐标,则通过 poselets 算法可以检测出图像中所有的人物框.

2.2 特定位置头部信息的人物框过滤模型

设给定需要寻找的人物的头部框 H 为: $H=(lu,lv,ru,rv)$, 这里 (lu,lv) 为左上角坐标, (ru,rv) 为右下角坐标.

本文只考虑头部位置是在上方的情况(即不考虑一些特殊姿势,例如倒立,侧躺等),根据正常人体特点,头部框与人物框应具有如下关系:①人物框包含头部框;②在水平轴方向上,两个框的中心位置接近;③头部框位于人物框的上部.因此,针对头部框 H 和人物框 O ,过滤模型应满足以下条件:

条件 1: $lx \leq lu \ \&\& \ ru \leq rx \ \&\& \ ly \leq lv \ \&\& \ ru \leq ry$

条件 2: $|lx+rx-lu-ru|/(ru-lu) \leq k1$

条件 3: $\frac{ly-lv}{rv-lv} \leq k2 \ \& \ \& \ \frac{ry-rv}{rv-lv} \geq k3$

由 poselets 算法检测出的目标人物框 O , 通过式(3)过滤模型筛选, 作为疑似目标人物框进行进一步的匹配.

If(人物框 O 同时满足条件 1、2、3)
then 保留
else 删除 (3)

2.3 筛选结果排序

在进行下一步匹配之前, 为了提高算法的运行效率, 可以对筛选出来的人物框进行排序. 大部分人体姿势其头部位置在 x 轴上的中心点值与其整体在 x 轴上的中心点值应当是一致或者相距较近的, 而 $k1$ 能够代表二者的距离值, 因此本文采用通过 poselets 算法对检测到的人物框的得分值以及 $k1$ 对筛选出的人物框按式(4)进行排序得分:

$$S(O|H) = (P(O|H) + k1(O|H))/2 \quad (4)$$

$S(O|H)$ 是人物框 O 总的得分值, $P(O|H)$ 是 poselets 算法中根据式(2)得到的人物框 O 的得分值, $k1(O|H)$ 是人物框 O 对应于 H 的中心距离相对值. 这里, $P(O|H)$ 和 $k1(O|H)$ 是经过归一化处理后的值.

2.4 基于特定位置人物信息的二分图匹配

由于大部分图像中都需要检测多个特定人物, 就需要多对多进行匹配. 二分图最大权值匹配可以剔除二分图中多余的点, 找到一个最佳匹配, 这个最佳匹配中的任意两条边都不依附于同一个顶点, 且是权值最大的边, 使得不同类的两个点集中的点能够一一对应. 因此本文把特定人物头部框与由 poselets 检测出的人物框的匹配问题转化为二分图最大权值匹配问题.

2.4.1 二分图

二分图是图论中的一种特殊模型. 设 $G=(V,E)$ 是一个无向图, E 是每一条边 (i,j) 的集合, 如果顶点 V 可以分为两个互不相交的子集 (X,Y) , 且 (i,j) 所关联的两个顶点 i 和 j 分别属于这两个互不相交的子集, i 属于 X, j 属于 Y , 那么称图 G 为二分图.

若给定一个二分图 G , 在 G 的子图 M 中, M 的边集 $\{E\}$ 中的任意两条边都不依附于同一个顶点, 则 M 是 G 的一个匹配. 而二分图的最大匹配就是包含边数最多的一个匹配. 不包含权值的二分图最大匹配算法, 最经典的就是匈牙利算法^[12], 该算法由 Edmonds 于

1965 年提出. 如果边上带有权值, 求最大权值二分图匹配的最常用算法则是 Kuhn-Munkres 算法^[12]. 本文则是参考二分图这一特殊模型, 将其应用到筛选出的人物框, 与特定位置的人物进行匹配, 从筛选出的结果中找出最佳匹配值, 得到对应特定位置的人物框信息.

2.4.2 基于特定人物头部框的二分图最大权值匹配

为了更好的进行相应的匹配, 本文将特定位置的头部框 H 根据正常人体比例: 八头身, 除去头部宽度, 左右肩宽分别为头部宽度的 0.8 倍. 按式(5)做一个估算, 得到特定位置人物框 $O'=(lu',lv',ru',rv')$.

$$\begin{cases} lu' = \max(0, lu - 0.8 * (ru - lu)) \\ lv' = lv \\ ru' = \min(w, ru + 0.8 * (ru - lu)) \\ rv' = \min(h, rv + 7 * (rv - lv)) \end{cases} \quad (5)$$

将筛选出的人物框集合 $SO=\{O_1, O_2, \dots, O_n\}$ 与特定头部位置估算出的特定位置人物框集合 $SO'=\{O'_1, O'_2, \dots, O'_m\}$ 进行匹配. 令 SO 和 SO' 分别构成加权二分图的 2 个点集, 若 O_i 和 O'_j 人物框之间的面积区域有重叠, 则两者之间是存在着相应的边 (i,j) , 其边的权重值 ω_{ij} 按式(6)设置.

$$\omega_{ij} = \lambda_1 \frac{a_{ij}}{ao_i} + \lambda_2 \frac{a_{ij}}{ao'_j} + \lambda_3 S(O_i | H_j) \quad (6)$$

a_{ij} 为 O_i 和 O'_j 重叠区域的面积, ao_i 为 O_i 的面积, ao'_j 为 O'_j 的面积, $S(O_i | H_j)$ 是由 2.3 方法计算出的人物框 O_i 对应 H_j 的得分值. 将 $\frac{a_{ij}}{ao_i}$ 和 $\frac{a_{ij}}{ao'_j}$ 、 $S(O_i | H_j)$ 这三个因素通过层次分析法^[13] 计算得到对应三因素的特征向量值作为 λ_1 、 λ_2 和 λ_3 的值, 并且这三个因素的值都经过归一化处理. 通过以上步骤建立好二分图, 从而将原始匹配问题转化为二分图最大权匹配问题. 采用 KM 算法求解其最大权匹配, 得到特定人物框与筛选后的人物框匹配的结果值. 如图 1 所示.

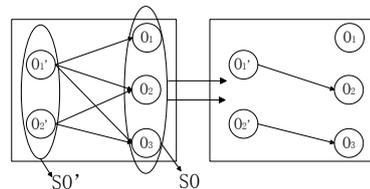


图 1 二分图匹配

图 1 左侧为根据是否有边的存在以及权值而建立的原始二分图, 右侧为通过二分图最大权匹配算法, 即 KM 算法得到的最终匹配结果图.

3 实验结果

为了验证本文提出的算法的有效性, 本文选取的样本来源是 Lubomir Bourdev 等人从 FLICKR(一个公开的相册专辑平台)上收集的图片, 其对图片上的部分人物进行人的头部位置的标注(注意: 是头部位置并不是脸部位置), 被称为 PIPA(People In Photo Albums)数据库^[1], 并使用样本中的图片进行结果展示.

3.1 特定位置人物检测

以下图像中红色框为 poselets 算法检测出的人物框, 黄色框为给定的特定人物的头部框.

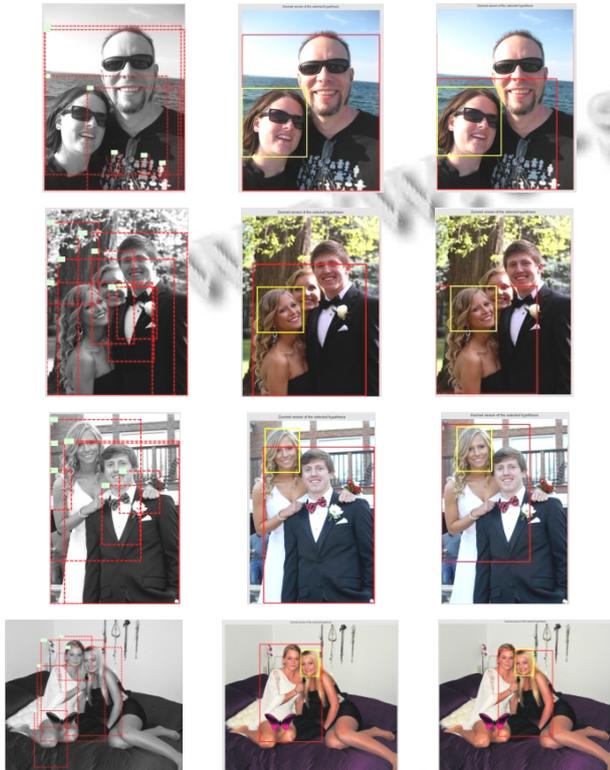


图 2 人物框提取效果图

如图 2(a)所示, 为 poselets 算法对输入图像的检测结果, 它检测出了图像中的所有人物框. 而过滤模型只是进行初次筛选, 其对于输入图像为单个人物时效果较好, 但是若输入的图像包含多个人物时, 得到的匹配结果就不够准确. 因此, 当不使用二分图进行下一步精确匹配时, 检测结果就会出现错误, 特别是针对人物较为密集以及人物之间相距较近的情况. 如图 2(b)所示为单独使用过滤模型且人物间距较近时的错误检测结果. 图 2(b)中根据红色边界框对应的人物提取的姿势发现其与黄色框中的人物不同, 因此, 对特

定位置人物的匹配错误.

当加入二分图最大权匹配算法后, 能够有效的提高对特定位置人物定位的精准度, 特别是对于一张图像中的多个人物相距较近的情况, 其效果是非常明显的. 图 2(c)为过滤模型加上 2.3 和 2.4 方法之后的定位准确的图示.

3.2 特定位置人物姿势提取

在 poselets 算法中, 当输入一张图片, 会对整张图片进行扫描, 得到无数个姿势组成一个集合, 并把属于不同人物的姿势归类到相应人物中. 因此, 本文在 poselets 检测出的人物框中找出特定位置人物框后, 可以直接利用 poselets 算法提取出分类于此特定人物对象下的姿势. 图 3(a)左半部的红色框为红色框, 绿色框为检测出的姿势框, 右半部为对应人物框提取的所有姿势, 绿色区为绿色框对应的姿势, 其得分排名在所有姿势中位于第四. 但是分类于此人物下的多个姿势可能存在错误的情况, 如图 3(b)图示, 检测出的绿色框为肩部姿势, 而对应的绿色区域姿势为人物右侧面. 因此, 需要对姿势进行筛选, 提取出人物的正确姿势. Poselets 算法中对每个姿势都有一个得分值, 通过多次实验可得, 优先姿势得分最高的前 15 位姿势作为候选姿势, 若对应人物的姿势不足十五位者则入选其对应的全部姿势.

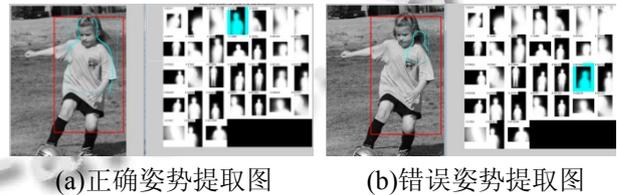


图 3 人物姿势提取效果图

本文还对提取出的姿势进行了特征提取, 将提取出的姿势特征进行了匹配, 如图 4 所示, 第一个为原人物, 后五个为根据第一个特征得到的匹配图. 这些姿势特征可以用于训练人物识别中各个姿势部位的 SVM 分类器, 扩大其应用面.

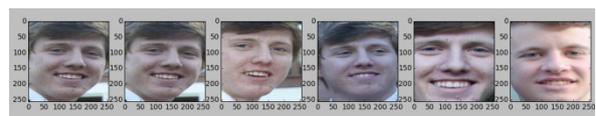


图 4 姿势特征图匹配

4 结语

本文提出了一种基于 poselets 算法的特定位置人

物多姿势提取的方法. 结合给定的特定位置人物的头部框, 通过过滤模型筛选 poselets 算法检测出的所有人物框, 对其再进行排序, 提高算法性能, 然后使用二分图最大权匹配算法进行精确匹配, 从多个人物框中找出特定位置人物框. 最后, 提取出找到的特定位置人物对应的前十五位的姿势. 在大量样本集上的实验表明, 本文的方法能够有效的提取出特定位置人物的姿势, 并提高了算法的精确度. 本文对特定位置的人物提取出的姿势可以应用于不仅仅只是针对脸部的人物识别^[5,14].

参考文献

- 1 Zhang N, Paluri M, Taigman Y, et al. Beyond frontal faces: Improving person recognition using multiple cues. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2015. 4804–4813.
- 2 Tan M, Hu Z, Wang B, et al. Robust object recognition via weakly supervised metric and template learning. *Neuro-computing*, 2016, 181: 96–107.
- 3 Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection. Proc. 2002 International Conference on Image Processing. 2002. IEEE. 2002, 1. 1-900-1-903.
- 4 Felzenszwalb PF, Girshick RB, McAllester D, et al. Object detection with discriminatively trained part-based models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627–1645.
- 5 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. 580–587.
- 6 Zhu Q, Yeh MC, Cheng KT, et al. Fast human detection using a cascade of histograms of oriented gradients. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE. 2006, 2. 1491–1498.
- 7 Dalal N, Triggs B. Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE. 2005, 1. 886–893.
- 8 Oren M, Papageorgiou C, Shinha P, et al. A trainable system for people detection. Proc. of Image Understanding Workshop. 1997, 24.
- 9 梁英宏. 基于 Gabor 变换和 Adaboost 算法的人体目标检测分类器. *计算机工程与设计*, 2009, (24): 5790–5792.
- 10 叶齐祥, 焦建彬, 蒋树强. 基于多尺度方向特征的快速鲁棒人体检测算法. *软件学报*, 2011, 22(12): 3004–3014.
- 11 Poselets: Body part detectors trained using 3D human pose annotations. *International Conference on Computer Vision (ICCV)*. 2009. 1365–1372.
- 12 Lovász L, Plummer MD. *Matching theory*. American Mathematical Soc, 2009.
- 13 邓雪, 李家铭, 曾浩健, 等. 层次分析法权重计算方法分析及其应用研究. *数学的实践与认识*, 2012, 42(7): 93–100.
- 14 Russell BC, Torralba A, Murphy KP, et al. LabelMe: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 2008, 77(1-3): 157–173.
- 15 Yao B, Yang X, Zhu SC. Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks. *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer Berlin Heidelberg. 2007. 169–183.