

导引概率图与显著特征相结合的行人目标检测^①刘 琼^②

(北京信息科技大学自动化学院 北京 100192)

摘 要 研究了仿生人眼视觉注意机制,采用目标导引概率图作为自上而下的信息,通过调制基于目标显著特征的自下而上信息,实现行人目标检测的方法。首先,对相似场景的目标样本图像提取尺度不变特征变换(SIFT)特征,基于贝叶斯公式,采用高斯混合模型(GMM)建立目标导引概率模型,利用期望最大(EM)算法和狄利克雷过程(DP)自动估计模型参数;进而,对一副待检测图像,采用已估概率模型计算图像中每一像元的目标似然性,形成导引概率图作为自上而下的信息;同时,针对行人目标,模拟中央-外周机制计算多尺度的肤色特征和竖直方向特征,形成基于目标显著特征的自下而上信息;最后,将两者结合得到候选目标区域,再通过提取候选区域的积分梯度直方图和等价的局部二值模式(LBP)特征,输入到级联支持向量机(SVM)分类器,验证并得到目标检测结果。基于实拍数据库和复旦大学-宾夕法尼亚大学行人数据库的大量实验表明,对概率模型的这种改进能显著提升行人目标预测效果,且检测算法在整体上优于传统检测算法。

关键词 视觉注意(VA), 概率引导图, 显著特征, 行人目标检测(PD)

0 引 言

人类视觉注意(visual attention)机制具有高效、准确的优点,实现该机制的仿生计算是计算机视觉研究领域追求的目标,尤其对无人驾驶等情况下实现该功能应用的目标更为迫切^[1]。视觉注意来源于感知驱动和任务驱动。来源于感知驱动,称为感知信息驱动的自下而上的视觉注意^[2],来源于任务驱动,称为任务信息驱动的自上而下的视觉注意^[3]。前者是指人们在无观测任务或观测期望的情况下,被场景中的显著物体或区域所吸引的视觉注意,该注意的机制是由底层感知数据驱动,与观察任务、情感、记忆等因素无关,因而人们无法有意识地控制其信息处理过程,且在该过程中所做出的反应是实时的。后者则是在有意识的条件下,寻求期

望目标的一种视觉行为,它与大脑高层的感知过程相关,存在空间搜索和特征匹配两条视觉通路,且显著区域不受感受野的范围与大小控制,通过显性的注意转移实现目标检测^[4-6]。本文着眼于对自上而下任务信息驱动的视觉注意的过程模拟,通过模拟空间搜索和特征匹配两条视觉通路,应用于行人目标检测(pedestrian detection)。

空间搜索基于目标与场景中其他物体之间的关系,因此,如何建模物体之间的关系是实现仿生计算的关键。回归人类视觉注意心理,假设搜索目标为电脑,很显然目光停留在书桌附近的概率要明显大于在沙发周围的概率,因此,采用概率推理的方式模拟空间搜索过程符合仿生机理要求。文献[7]通过构建贝叶斯模型,采用全局 gist 特征预测目标出现的区域,在该区域中再采用显著特征定位目标。由于 gist 特征是全局特征,适合于对场景做粗水平归

① 北京市教委 2014 年度科研面上基金(KM201411232008)和北京信息科技大学 2016 年度大学生科技创新资助项目。

② 女,1984 年生,博士,讲师;研究方向:模式识别,认知计算;联系人,E-mail:liuqiongq@126.com
(收稿日期:2016-03-23)

类,如室内或户外等,因此,该方法适合于远距离拍摄所获取的大场景图像。与此同时,预测的区域仅为图像水平划分的区域,搜索尺度太大。本文主要在该模型的工作基础之上,做如下改进:(1)采用尺度不变特征变换(SIFT)特征作为预测特征,估计场景图像每一个像元的目标概率似然度,使其适用于任意距离拍摄的图像;(2)采用 Dirichlet 过程对高斯混合模型(Gaussian mixture model, GMM)的聚类

中心个数进行估计,提高模型预测的准确率。(3)通过提取更符合目标特性的 IHOG-LBP 特征,并经过 SVM 分类判别,进一步验证候选区域是否含有目标,提高检测准确率,更符合工程实际的要求。

1 导引概率图计算

全文整体算法流程如图 1 所示。

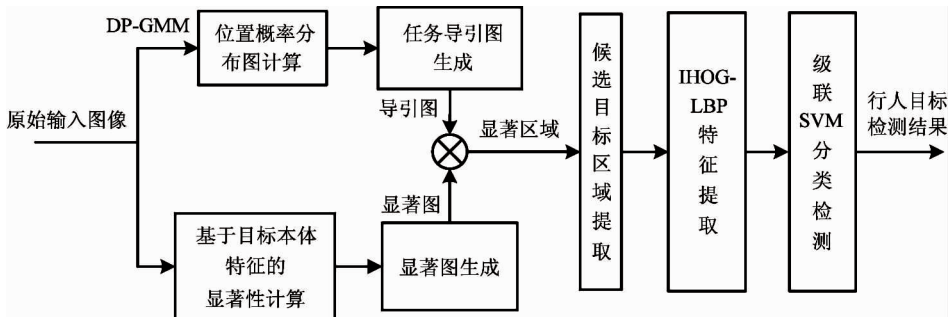


图 1 基于导引概率图与显著特征的行人目标检测计算流程图

1.1 基于 SIFT 特征的混合高斯概率模型构建与参数求解

在给定图像的观测数据集 v 下,计算目标物体 O 的似然函数,可在经典的概率论框架下依据贝叶斯原理构建:

$$P(O|v) = \frac{P(v|O)}{P(v)}P(O) \quad (1)$$

其中, $P(O|v)$ 是在给定图像所有位置观测数据集 v 的条件下,目标 O 存在的条件概率密度函数。 O 表示图像中目标物体的属性: $O = \{o, x, \sigma, \dots\}$, 其中 o 表示目标物体的类别, x 表示目标物体在图像中的位置坐标, σ 表示目标物体的大小。这里并没有列完所有的参数,其他的一些参数也可以用来描述目标物体,比如物体的形态姿势,物体的亮度等。但在空间搜索中,目标物体在图像中的位置 x 是首要关注的参数。考虑到 SIFT 特征具有优良的稳定性^[8],式(1)中的 v , 本文采用图像所有的 SIFT 特征点的特征向量,因此,也将其称之为 SIFT 概率模型。

式(1)是计算目标物体在图像中理想位置 o 的概率密度函数,由于观测数据 v 考虑了图像所有位置,其维度将非常高,从而使建模概率密度函数的问

题变得极为病态。为了降低复杂度,考虑在目标一定的邻域范围内进行观测,并假设图像中其他区域的观测值与目标邻域内的观测值相互独立。于是,式(1)可以改写为

$$P(x|v_c, o) = \frac{P(x, v_c|o)}{P(v_c|o)} \quad (2)$$

其中,视觉特征向量 v_c 变为目标区域的特征,其维度明显低于前述视觉特征向量 v , v_c 为行人目标的 SIFT 特征向量, o 指定为行人目标。 $P(v_c|o)$ 是在行人目标类里,有关 SIFT 特征向量的条件概率分布,可以利用样本库中的行人目标图像事先学习得到。因而求解式(2)的主要问题就是如何得到似然函数 $P(x, v_c|o)$ 。为了能够使式(2)所代表的模型更加具体且方便计算,接下来,采用 GMM 对似然函数 $P(x, v_c|o)$ 进行建模。

由于似然函数 $P(x, v_c|o)$ 求解在给定目标物体类别的情况下,目标物体在图像中的位置坐标 x 与目标物体的视觉特征 v_c 的条件概率。为了计算简便,可以认为目标物体在图像中的位置 x 与目标物体的视觉特征 v_c 为两个随机变量,所以直接应用混合高斯模型不太适宜。为了简化问题,可假设目

标物体在图像中的位置坐标 x 与目标物体的视觉特征 v_c 从统计的角度来看是互相独立的,故可以将混合高斯模型中的联合概率分解为两个高斯分布的乘积,也即

$$p(x, v_c | o) = \sum_{i=1}^M b_i N(x | \mu_i^x, \sum_i^x) N(v_c | \mu_i^v, \sum_i^v) \tag{3}$$

其中第一个高斯分布建模了行人目标在图像中的位置分布,第二个高斯分布建模了每一个聚类的视觉特征的分布。而混合高斯的每一个成分可以看作一个聚类中心,它描述了整个联合概率的分布情况。

得到似然函数 $P(x, v_c | o)$ 模型式(3)后,接下来的工作就是求解模型,完成对模型中未知参数的估计,从而能够利用该模型求得目标物体在图像中

表 1 基于 EM 算法的混合高斯模型参数推导

步骤 1. 初始化模型中的待估计的参数:

混合系数 b_i , 均值向量 μ_i^x, μ_i^v , 协方差矩阵 $\sum_i^x, \sum_i^v$, 偏置量 a_i , 相关矩阵 A_i

步骤 2. E 步:

$$\text{更新后验概率 } h_i(t) = \frac{b_i N(x(t) | \mu_i^x, \sum_i^x) N(v(t) | \mu_i^v, \sum_i^v)}{\sum_{i=1}^M b_i N(x(t) | \mu_i^x, \sum_i^x) N(v(t) | \mu_i^v, \sum_i^v)}$$

步骤 3. M 步:

$$\text{更新混合系数 } b_i = \frac{\sum_{t=1}^N h_i(t)}{\sum_{i=1}^M \sum_{t=1}^N h_i(t)}$$

$$\text{更新均值向量 } \mu_i^v = \frac{\sum_{t=1}^N h_i(t) v(t)}{\sum_{t=1}^N h_i(t)}$$

$$\text{更新协方差矩阵 } \sum_i^v = \frac{\sum_{t=1}^N h_i(t) (v(t) - \mu_i^v) (v(t) - \mu_i^v)^T}{\sum_{t=1}^N h_i(t)}$$

$$\text{更新相关矩阵 } A_i = (\sum_i^v)^{-1} \frac{\sum_{t=1}^N h_i(t) (v(t) - \mu_i^v) x(t)^T}{\sum_{t=1}^N h_i(t)}$$

$$\text{更新偏置量 } a_i = \frac{\sum_{t=1}^N h_i(t) (x(t) - A_i^T v(t))}{\sum_{t=1}^N h_i(t)}$$

$$\text{更新协方差矩阵 } \sum_i^x = \frac{\sum_{t=1}^N h_i(t) (x(t) - a_i - A_i^T v(t))^T (x(t) - a_i - A_i^T v(t))}{\sum_{t=1}^N h_i(t)}$$

步骤 4. 重复步骤 2 和步骤 3 直至对数似然函数 $\ln p(X | \pi, \mu, \sum) = \sum_{n=1}^N \ln \{ \sum_{k=1}^K \pi_k N(x_n | \mu_k, \sum_k) \}$ 收敛到局部极值,或小于设定阈值。

的位置的后验概率,即 $P(x | v_c, o)$ 。在此,本文通过期望最大化(expectation maximization, EM)算法^[9]求解模型参数。由于式(2)所示模型中出现了统计上相互独立的2个随机变量:目标物体在图像中的位置坐标 x 与目标物体的视觉特征 v_c 。在求偏导时,为了简化,可以假设目标物体在图像中的位置坐标 x 与目标物体的视觉特征 v_c 之间具有线性关系,即

$$x_i = a_i + A_i(v_c - v_i) \quad (4)$$

其中 a_i 为偏置量, A_i 为相关矩阵。此时模型的参数即变为 $(b_i, a_i, A_i, \sum_i^x v_i, \sum_i^v v_i)$, $i = 1, \dots, M$ 。

接下来,便可按照对似然函数求偏导的方法,推导出模型参数,从而得到物体在图像中的位置。具体的基于 EM 算法的参数推导过程如表 1 所示。

1.2 基于狄利克雷过程的 GMM 聚类中心数估计

此前所建立的概率模型是基于高斯模型的混合,并用 EM 算法求解未知参数。但在求解过程中,如何确定混合高斯的个数是一个很棘手的问题。如果设置混合高斯的个数过多,则整个模型的复杂度将会很高,在学习过程中也容易出现过学习现象。反之,如果混合高斯分布的个数过少,则整个模型的复杂度可能偏低,模型将不会有好的泛化和推广能力,而且会出现欠学习现象。如图 2 所示,其中图 2(a)是原始的数据点的分布,但它们由几个混合高斯分布来表示是未知的。如果预先指定 5 个聚类中心,则混合高斯模型学习后得到的结果如

图 2(b)所示。而自组织情况下,数据点的分布只有 3 个聚类中心,其聚类结果如图 2(c)所示。由此可以看出,预先指定聚类中心个数的方式不合理,而且会使得模型的预测和泛化能力较低。

为了解决这个问题,本文引入非参数贝叶斯模型中的狄利克雷过程(Dirichlet Process, DP)^[10]对混合高斯分布中的高斯模型个数进行估计,并将其称之为 DP-GMM。

Dirichlet 分布用于描述 n 维 ($n > 2$) 随机变量 X 的分布,是 Beta 分布的多维泛化形式,一般记做 $Dir(\alpha)$, 其概率密度函数表达式为

$$p(X | \alpha) = \frac{\Gamma(\sum_{n=1}^N \alpha_n)}{N!} \prod_{n=1}^N x_n^{\alpha_n - 1} \quad (5)$$

其中 $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ 为伽玛函数, $\sum_{n=1}^N x_n = 1$ 且 $x_n > 0, \alpha_n > 0$ 。由上面的定义可知, n 维随机变量 X 是一个离散概率分布,所以 Dirichlet 是分布的分布。

给定一个狄利克雷过程 $DP(H, \alpha)$, 其中基准分布 H 可以认为是一个任意的分布,而浓度系数 α 是一个正实数。对狄利克雷过程进行一次采样,得到的是一个随机分布,这个随机分布的取值是从基准分布 H 上采样得到的。也就是说,狄利克雷过程的输出分布的支撑集和狄利克雷过程的基准分布的支撑集是一样的。狄利克雷过程的输出分布是离散的,这意味着即使狄利克雷过程的基准分布是连续分布(连续分布也就意味着从这个分布上采样的任意两个采样值互不相同的概率为 1),从狄利克雷过程的输出分布上采样得到的单个样本值有可能会重复。其中,狄利克雷过程采样值重复程度的大小由浓度系数 α 决定,当浓度系数 α 很大时,采样值重复的次数越少。如果狄利克雷过程的基准分布是连续分布,则每次对这个基准分布的采样值都是不一样的,而狄利克雷过程的输出分布返回的每一个可能采样值的频率对应的是一个无限的概率集合,则这个无限的概率集合是按照破棍过程^[11]分布的。

由于从狄利克雷过程采样得到的是离散的样本值,它的一个重要的应用是将其作为对无限混合模型的先验概率分布的一个估计。此时,整个生成模

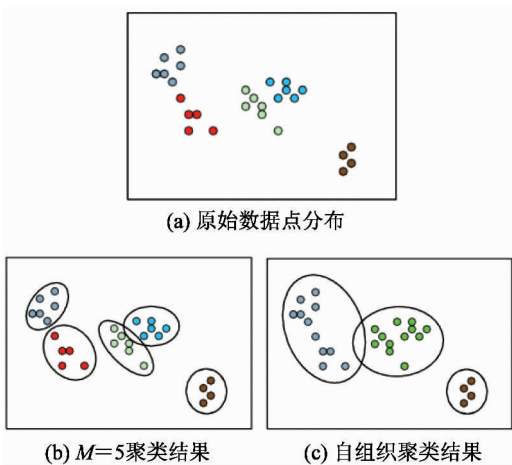


图 2 同一数据集在指定分布个数与自组织情况下的聚类效果比较

型的过程如下:首先从狄利克雷过程采样得到一个样本分布,然后对应每一个数据点,从这个样本分布上逐一采样,再将其作为这个数据点对应某一混合成分的概率分布。由于模型最后生成的不同的混合成分的个数没有限制,所以在无预先定义混合成分个数时,狄利克雷过程是一个较为合适的模型。运用到对混合高斯分布中的高斯模型个数进行估计,本文首先从狄利克雷过程中进行采样,得到一个似然函数 $P(x, v_c | o)$ 的先验概率分布,也即通过狄利

克雷过程确定高斯聚类中心的个数。然后再采用期望最大化算法对混合高斯模型的参数进行最大似然估计。

为了验证基于狄利克雷过程非参数贝叶斯模型对聚类效果的改进,本文对比了对于随机生成的一组数据,在预先指定混合高斯聚类中心个数与采用狄利克雷过程确定混合高斯聚类中心个数后的聚类情况,如图 3 所示。

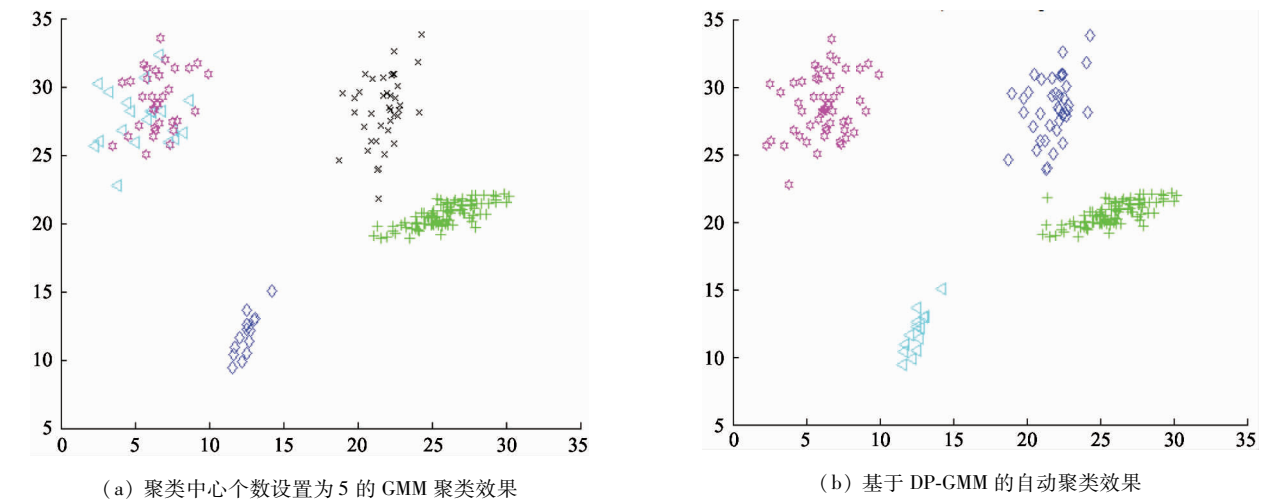


图 3 GMM 与 DP-GMM 对同一组数据的聚类效果对比

可以看出,图 3(b)中聚类个数为 4,相对图 3(a)中设定聚类个数为 5 的效果更为合理。因此,利用狄利克雷过程的非参数贝叶斯模型对聚类个数估计能够减少因聚类数错误而对后续计算带来较大的误差与影响。

1.3 基于 DP-GMM 的导引注意区域提取算法

通过大量的样本图像对式(2)的 $P(x, v_c | o)$ 建立 DP-GMM 模型。对于一幅新待检测图像,依赖学习得到的高斯模型以及提取到的整幅图像的 SIFT 特征点,得到图像中对该目标关注的中心,也即 x ,实现对图像中目标区域的预注意。详细算法描述如表 2 所示。

从概率模型的计算结果得到了行人目标在图像中的位置概率分布,因此,将目标的位置分布作为一种自上而下的任务导引特征。任务导引图保持与原始图像相同的拓扑结构,其中每一个像素点的像素

值与目标在该点出现的概率成正比。关注区域中具有较大概率值的位置,其显著值也相应的较高。而对于注意关注区域之外的其他区域,显著值均设置为一个极小值,如 10^{-6} 。

2 基于行人目标显著检测计算的集中注意

2.1 候选区域提取

基于概率模型得到了行人目标在图像中位置的概率分布,完成了视觉选择性注意的第一步——空间搜索。进一步地,为了找到各个目标的具体位置,需要在关注区域中,通过特征匹配找到目标候选区域。针对行人目标,本文主要选择肤色和竖直方向作为行人的两种本体特征,将特征计算的结果作为特征图。其中竖直方向特征的计算通过原始图像的灰度图与二维 Gabor 卷积核得到,此处不再赘述。

表 2 基于 DP-GMM 的预注意关注区域提取算法

步骤 1. 建立高斯混合模型

(1) 标定样本图像中的目标 ROI 区域,并提取 SIFT 特征向量;

(2) 采用 Dirichlet 过程估计 GMM 的聚类个数 k ;

(3) 采用 EM 算法估计混合系数 b_i , 均值向量 μ_i^x, μ_i^y , 协方差矩阵 $\sum_i^x, \sum_i^y$, 偏置量 a_i , 相关矩阵 A_i 。

步骤 2. 提取待检测图像的 n 个 SIFT 特征点;

步骤 3. 根据已训练好的 GMM 模型以及待注意图像的 n 个 SIFT 特征点的位置与特征值,计算每一个特征点

$$\text{的概率值 } P(x | v_i, o) = \frac{P(x, v_i | o)}{P(v_i | o)} \quad i = 1, 2, \dots, n;$$

$$\text{步骤 4. 计算图像每个 SIFT 特征点的 } P(x | v_i, o) \text{ 值的份额 } p(i) = \frac{P(x | v_i, o)}{\sum_{i=1}^n P(x | v_i, o)};$$

$$\text{步骤 5. 计算图像关注区域的中心 } c = \sum_{i=1}^n \text{cor}_{i_sift} \times p(i);$$

$$\text{步骤 6. 计算图像关注区域的范围,区域长短轴的长度分别为 } \max(\text{mean}(\text{cor}_{i_sift_x} - c_x), \text{mean}(\text{cor}_{i_sift_y} - c_y)), \\ \min(\text{mean}(\text{cor}_{i_sift_x} - c_x), \text{mean}(\text{cor}_{i_sift_y} - c_y))$$

肤色的计算公式主要基于文献[12]中给出的肤色高斯分布型概率计算公式

$$h(c) = \exp\left[-\frac{1}{2}(c - \mu)^T \Sigma^{-1}(c - \mu)\right] \quad (6)$$

通过该式计算各个像素点的肤色似然性大小。式(6)中 $c = (r', g')$, $\mu = (\mu_{r'}, \mu_{g'})$, $\Sigma =$

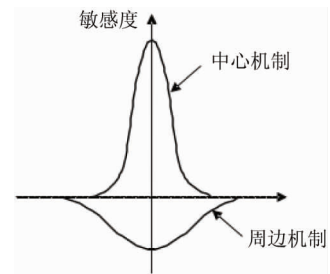
$$\begin{pmatrix} \sigma_{r'}^2 & \frac{\sigma_{r'}\sigma_{g'}}{\rho} \\ \frac{\sigma_{r'}\sigma_{g'}}{\rho} & \sigma_{g'}^2 \end{pmatrix} \quad \sigma_{r'}^2, \sigma_{g'}^2 \text{ 分别为 } r', g' \text{ 分量的方差,}$$

ρ 为 r', g' 分量的相关系数, Σ 为 $\sigma_{r'}, \sigma_{g'}$ 的协方差矩阵。将原始图像由 RGB 颜色空间转换到 $R'G'B'$ 空间后(任意一个给定的像素点 P , 其在 $R'G'B'$ 空间中的各颜色通道值为原始 RGB 颜色空间单通道值占有所有通道值的比例)。

由于 $r' + g' + b' = 1$, 因而式中的向量 c 采用 r', g' 和 b' 中的任意两个分量均可计算得到结果。将 c, μ, Σ 的值分别代入式(6)中, 即可得到像素点 P 的肤色似然性 $SL(P)$ 公式

$$SL(P) = \exp\left[-\frac{1}{2}\left(\frac{(r' - \mu_{r'})^2}{\sigma_{r'}^2} + \frac{(g' - \mu_{g'})^2}{\sigma_{g'}^2} - \frac{\rho(r' - \mu_{r'})(g' - \mu_{g'})}{\sigma_{r'}\sigma_{g'}}\right)\right] \quad (7)$$

考虑人类视觉系统感受野中的中心-外周机制, Rodieck 于 1965 年提出了同心圆拮抗式(Homocentric Opponent)感受野的数学模型^[13], 如图 4 所示。它由一个兴奋作用强的中心机制和一个作用较弱但面积更大的抑制性周边机制构成。这两个具有相互拮抗作用的机制, 都具有高斯分布的特性, 但中心机制具有更高的峰敏感度, 而且彼此方向相反, 两者为相减关系, 通常称为高斯差分(Difference of Gaussian, DoG)模型。

图 4 同心圆拮抗式感受野的数学模型^[13]

因此, 本文综合显著性的直观感受与感受野的特性, 采用如下基于高斯尺度空间的显著特征图计算公式:

$$F(c, s) = |F(c) \ominus F(s)| \quad (8)$$

式中, $F(c)$ 与 $F(s)$ 分别表示中心(Center)与外周

(Surrounding)尺度下的特征图;外周尺度比中心尺度大,通过对中心尺度做进一步的采样得到,因此,外周尺度下的图像是中心尺度下图像结构的简化。 $F(c,s)$ 表示作差后的结果,体现中心外周尺度图像之间的差异大小,模拟“中心-外周”竞争;式中符号“ $\ominus$ ”表示不同尺度图像之间的一种做差运算,通过将插值后的外周尺度图像与中心尺度图像的像素点之间进行一一对应作差而实现。在此基础上,归一化运算不同特征分量的特征图,并在不同的尺度间合并,得到各个特征分量的显著图,最后通过线性相加运算得到最终的显著图。本文共有 6 对尺度图像进行了作差运算:其中中心尺度 $c = \{2,3,4\}$,环绕尺度 $\delta = \{3,4\}$, $s = c + \delta$ 。

根据人的一般比例布局与分布,且考虑到行人皮肤裸露的位置主要在脸部,因此,假设所有显著区域位于人的脸部,确定候选区域目标的大致包围盒(bounding box)。

2.2 候选区域验证与目标检测

方向梯度直方图(histogram of oriented gradient, HOG)是由 Dalal 在 2005 年提出,在 SIFT 特征的基础之上发展起来的用于描述物体形状的一种特征,最初即用于行人检测^[14]。局部二值模式(local binary pattern, LBP)通过比较图像像素点之间的灰度值差异,有效描述物体纹理。两者在行人检测领域均应用广泛^[15]。本文在候选区域的验证主要通过这两类特征,并引入积分图思想^[16],构建积分 IHOG-LBP 特征即 IHOG-LBP 特征,最后采用三级级联 SVM 分类器对检测区域进行分级检测。若三级分类检测均通过,则标记为行人目标。由于 HOG、积分图思想和 LBP 特征都比较常见,本文只对特征计算过程中参数的设置和个别重要的公式加以说明。

本文 HOG 特征的计算细胞单元分别设置为 $32 \times 32, 16 \times 16, 8 \times 8$ 个像素大小;每 2×2 个细胞单元组成一个方块。为了消弱阴影、光照以及边缘变化给特征向量带来的影响,采用下式

$$L' = \frac{L}{\sqrt{\|L\|_2^2 + \varepsilon^2}} \tag{9}$$

归一化区间块内的 HOG 特征向量。式中 L 是特征向量, $\|L\|_k$ 是它的 k 范数, $k = 1,2,\cdots$, ε 是一个无穷小的正数常量,为了防止归一化时出现分母为零的情况。在 HOG 特征的提取过程中,由于方块之间的重叠造成大量的重复计算,因此,我们引入积分图思想,对各个区域 HOG 特征计算采用积分图累加的方式,以提高算法效率。积分图方法是加速特征计算的一种有效方式,以类似数学中积分的方式,图像中的每一个像素点存储以原点为矩形左上角点,该点为右下角点的矩形区域所有像素点的特征值,以便在对图像只进行一次遍历的情况下完成对图像中所有像素点的特征值计算,从而提高算法运行效率。此外,考虑到后续分类中使用级联 SVM 分类器中需多次提取不同大小细胞单元的 HOG 特征,采用积分的 HOG 特征(简称为 IHOG 特征)后,只需对该细胞单元的四个顶点进行简单的加减运算即可得到该单元的 HOG 特征,而不需要重新累加细胞单元内各个像素点的特征值。

本文 LBP 特征计算选用等价模式 LBP。采用等价模式主要是为了在不丢失任何信息的条件下,减少二进制模式的种类^[17]。

将目标位置概率图作为任务导引图调制自下而上的显著性注意图,并将调制结果得到的显著区域作为集中注意的目标区域,并选择合适的候选区域,通过提取候选区域的 IHOG-LBP 特征,得到目标分类检测的结果。图 5 描述了该方法对一幅图像具体的检测流程。

从图 5 中可以看出,基于 SIFT 概率模型计算的位置概率作为任务导引特征可以有效地调制基于目标显著性计算的结果。场景中的背景墙部分由于具有与肤色较为近似的颜色特征,单纯通过显著性注意图容易将其作为候选区域而有可能发生错检且浪费时间,而通过任务导引图的调制能够有效地得到原始图像中行人所在的集中注意区域。进而,通过基于 IHOG-LBP 特征对候选区域进行验证,得到了理想的目标检测结果。其中,4 号候选区域中由于行人脸部未能完整覆盖,因此该区域被认为没有覆盖目标。

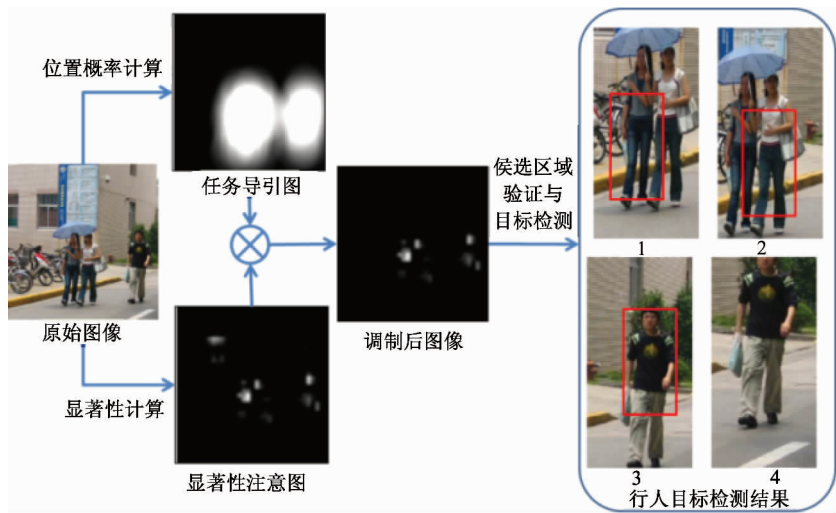


图 5 实例化行人目标检测流程

3 实验结果分析与评价

实验分成两个部分,第一部分将对基于 DP-GMM 改进的图像预注意区域提取结果与文献[7]中基于混合高斯模型的提取效果进行比对,第二部分将对行人目标检测的结果与传统的目标检测方法相比较。实验所采用的数据库主要是复旦大学-宾夕法尼亚大学行人数据库和校园实拍图像。当 *CoverRate* 大于 50% 时,检测结果视为正确。考虑到图像中存在相互之间有遮挡或部分重叠的多个行人,在基于显著性计算结果生成候选区域时,容易造成一个候选区域覆盖多个行人目标,因此,在检测过程中,当一个外围框包含有多个行人时,可作为多次检测的结果,但外围框的个数不能少于显著性计算得到的显著区域个数。此外,当一个行人有多个外围框包围时,只有一个结果被认为正确,其他所有结果都列为误检结果。图 6 所示为复旦大学-宾夕法尼亚大学行人数据库部分图像示例。

实验 1:基于 DP-GMM 改进的预注意区域提取较文献[7]中给出模型,其主要改进的地方体现在采用了 Dirichlet 过程(DP)对聚类个数进行估计,此外将方向特征改成了 SIFT 特征作为图像的观测数据。



图 6 复旦大学-宾夕法尼亚大学行人数据库部分图像示例

实验中采用表 2 中给出的算法,将不同拍摄距离下得到的图像一起训练,220 幅图像用于训练得到 DP-GMM 模型,剩余的 92 幅图像用于测试。部分图像基于 GMM 和 DP-GMM 对部分图像的预注意区域的提取效果如图 7 所示,其中估计得到 $M = 10$ 。

从图 7 中可以看出,对于偏近距离拍摄的图片(如图 7 左边的图像),预注意区域提取效果比较接近,都能较好地覆盖行人的范围,且关注的区域与人眼直观观测较为符合。但在图 7 右边所示的近距离拍摄图片中,文献[7]得到的预注意区域仅能覆盖图像中位于中心区域附近的两个行人目标,而本文模型能够覆盖图像中所有的行人目标,更好地捕捉面向行人的关注区域。



(a) 文献[7]中方法得到的关注区域



(b) 本文算法得到的关注区域

图 7 预注意区域提取效果比较

为了反映在整个图像数据库中面向行人目标的关注区域提取效果,本文通过准确率 (Precision)-召回率 (Recall) 曲线即 PR 曲线来进行描述。由于任务引导图中每一个像素点的像素值即为目标在该点出现的概率,因此调节目标似然的概率阈值,并且根据该阈值对图像进行二值化,可以得到不同的关注区域的大小。对于整个数据库而言,随着阈值的改变,可以得到反映算法鲁棒特性的 PR 曲线。在上述数据库中,两种方法得到的 PR 曲线如图 8 所示,从图中我们可以比较得出,采用 DP 对 GMM 进行改进,并采用 SIFT 特征明显优于文献[7]基于 GMM 和方向特征的方法。

实验 2:行人目标检测是当前目标检测领域研究较多的一类问题,但是大多数方法都是采用基于图像处理的方法,即从目标的底层特征分析着眼来

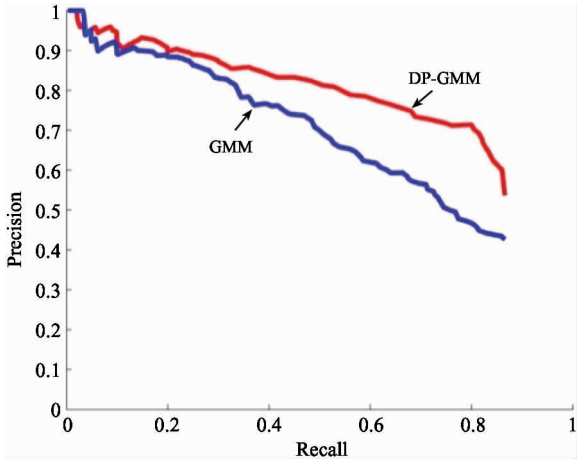


图 8 衡量本文方法与文献[7]的行人目标关注区域提取效果的 PR 曲线

解决问题,比较有名的有联合局部与全局特征的方法(combination of local and global, CLG)^[18]以及基于霍夫变换(Hough transform, HT)投票(Voting)^[19]的方法,这两类方法代表了概率计算框架下以像素级特征为底层元素和以图像子块(Patch)为底层元素的行人检测最为经典的方法,具有典型的代表性,因此,本文主要与这两种方法进行比较。

实验过程中采用与本节实验 1 中同样的方式调整概率阈值,得到的 PR 曲线如图 9 所示。从 PR 曲线中可以看出,在召回率小于 0.3 时,本文方法的精度介于 CLG 与 HT 之间,略逊于 CLG。当召回率大于 0.3 时,该方法较之 CLG 和 HT 均有优势。这说明在概率阈值较小的情况下,综合考虑基于像素级的

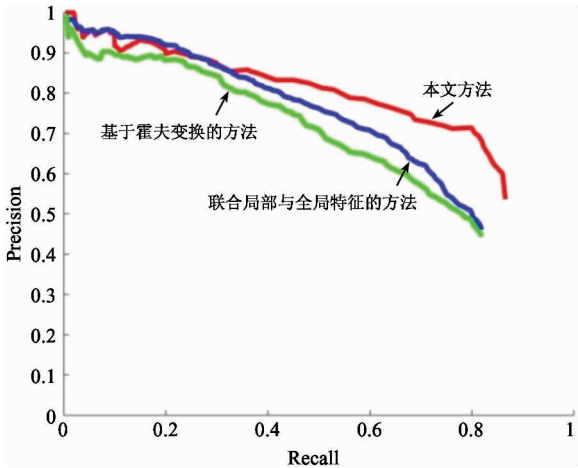


图 9 衡量传统方法与本文方法检测效果的 PR 曲线

全局与局部特征更有利于行人目标检测,但随着概率阈值的增大,本文方法采用 SIFT 特征生成的概率导引图更为鲁棒,且综合显著特征与层级筛选机制,能更好地检测到行人目标,且算法复杂性低于 CLG 方法。

4 结 论

针对任意距离拍摄场景中的行人目标检测,本文模拟人眼任务导引视觉注意机制的空间搜索与特征匹配实现过程,采用基于 Dirichlet 改进的 SIFT 混合高斯概率模型计算图像中每个像素点的目标似然概率作为空间搜索的启发式信息,调制基于目标本体显著特征计算结果,在图像数据库与实拍图像中均取得了较为优良的检测精度。本文方法适用于任意距离拍摄的场景图片,且具有较好的鲁棒性。但显著性的检测依赖于行人目标的特性是竖直且有部分皮肤的裸漏。

然而,本文方法也存在一定的局限性,主要表现在考虑到算法的快速性要求时,候选区域验证时采用的检测窗口大小恒定,忽略了尺度因素的影响。此外,传统的基于图像处理的行人目标检测方法也在不断地改进与更新,同时也采用了各种方式与方法来加速检测过程,本文作为对基于生物启发式视觉注意机制的模型化计算理论研究与应用研究,仍然需要在快速注意机制上寻求更好的工程化建模方法。

参考文献

- [1] Major Research plan of National Natural Science Foundation of China: cognitive computing of visual and auditory information[Online], available: <http://ccvai.xjtu.edu.cn/mes.do?method=getoverview:xjtu>, 2015
- [2] Yantis S. To See is to attend. *Science*, 2003, 299 (5603): 54-56
- [3] Fecteau J H, Bell A H, Munoz D P. Neural correlates of the automatic and goal-driven biases in orienting spatial attention. *Journal of Neurophysiology*, 2004, 92 (3): 1728-1737
- [4] Corbetta M. Frontoparietal cortical networks for directing

- attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences of the United States of America* (PNAS). 1998, 95(3): 831-838
- [5] Findlay J M, Gilchrist I D. Visual Attention: The Active Vision Perspective. Vision and attention, New York: Springer, 2000. 83-103
- [6] Yarbus A L. Eye Movements and Vision. New York: Plenum Press, 1967
- [7] Torralba A, Oliva A, Castelano M S. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 2006. 113(4): 766-786
- [8] Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. 60(2): 91-110
- [9] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B (Methodological)*. 1977, 39(1): 1-38
- [10] Ferguson T S. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1973. 1(2): 209-230
- [11] Ishwaran H, James L F. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 2001, 96(453): 161-173
- [12] Walther D. Interactions of Visual Attention and Object Recognition: Computational Modeling, Algorithms, and Psychophysics: [Ph. D dissertation]. Pasadena, California: California Institute of Technology, 2006
- [13] Rodieck R W. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 1965. 5 (12): 583-601
- [14] Dala I N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, 2005. 886-893
- [15] Dollar P, Wojek C, Schiele B, et al. Pedestrian detection: an evaluation of the state of the art. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 2012, 34(4): 743-761
- [16] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and

- Pattern Recognition (CVPR), Kauai, USA, 2001. 511-518
- [17] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987
- [18] Leibe B, Seemann E, Schiele B. Pedestrian detection in crowded scenes. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, 2005. 878-885
- [19] Barinova O, Lempitsky V, Kohli P. On detection of multiple object instances using Hough transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(9): 1773-1784

Pedestrian detection by combining guided probability map and saliency features

Liu Qiong

(Department of Automation, Beijing Information Science and Technology University, Beijing 100192)

Abstract

The pedestrian detection mechanism characterized by Abstract mimicking human visual attention and using the guided probability map as the top-down information to mediate the bottom-up information based on saliency features was studied. Firstly, the guided probability map was built based on the Gaussian Mixture Model (GMM) to extract the scale-invariant feature transform (SIFT) features of similar scene images, and the expectation maximization (EM) algorithm and the Dirichlet Processing (DP) were used to estimate the parameters of the GMM automatically. Then, to a new image, the likelihood probability of every pixel was computed by the obtained GMM so as to form the top-down information. Meanwhile, the center-surround mechanism was mimicked to compute the multi-scale skin feature and the vertical direction feature to from the bottom-up information of the pedestrian's saliency feature. Finally, through combining the two-way information, the candidate target areas were obtained, and the integral histogram of oriented gradients and local binary pattern (IHOG-LBP) features of the candidate areas were extracted and then inputted to the cascade support vector machine (SVM) classifier to verify and obtain the pedestrian detection result. The results of the experiment based on the Penn-Fudan pedestrian database and realistic images show that the proposed probability map can improve the prediction detection and the whole detection algorithm outperforms other traditional object detection methods.

Key words: visual attention (VA), guided probability map, saliency feature, pedestrian detection (PD)