软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(11):2747−2757 [doi: 10.3724/SP.J.1001.2013.04484]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于局部区域稀疏编码的人脸检测*

张 抒[1], 蔡 勇[1,2], 解 梅[1]

[1](电子科技大学 电子工程学院,四川 成都 611731)

[2](空军工程大学 信息与导航学院,陕西 西安 710071)

通讯作者: 张抒, E-mail: jlu_zhangshu@163.com

摘 要: 提出一种基于局部区域稀疏编码的人脸检测方法.首先提取人脸局部区域作为训练样本;然后学习得到一个具有较强判别性的字典,字典中的每个基与人脸各局部区域有明确的对应关系;接着,基于各检测窗口稀疏编码的响应判断人脸某一局部区域是否出现;最后,利用人脸局部区域的检测结果和位置约束进行投票,完成人脸定位.该方法的创新在于将稀疏编码和基于部件模型的思想相结合,实现人脸检测.在 Caltech 和 BioID 人脸数据库的实验结果表明:该方法适用于小样本问题,且在遮挡、复杂表情、人脸偏转等情况下具有较好的检测效果.

关键词: 人脸检测;稀疏编码;字典学习;部分模型;检测精度;召回率

中图法分类号: TP391 **文献标识码**: A

中文引用格式: 张抒,蔡勇,解梅.基于局部区域稀疏编码的人脸检测.软件学报,2013,24(11):2747−2757. http://www.jos.org.cn/1000-9825/4484.htm

英文引用格式: Zhang S, Cai Yong, Xie M. Face detection based on local region sparse coding. Ruan Jian Xue Bao/Journal of Software, 2013,24(11):2747−2757 (in Chinese). http://www.jos.org.cn/1000-9825/4484.htm

Face Detection Based on Local Region Sparse Coding

ZHANG Shu[1], CAI Yong[1,2], XIE Mei[1]

[1](School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

[2](School of Information and Navigation, Air Force Engineering University, Xi'an 710071, China)

Corresponding author: ZHANG Shu, E-mail: jlu_zhangshu@163.com

Abstract: In this paper, a face detection method based on local region sparse coding is proposed. First, every local face regions are extracted as training sample. Next, a discriminative dictionary whose atoms have explicit relations with local regions is learned. Then the appearance of a particular local region is determined based on the response of its sparse coding for each detection window. Finally, face location is obtained using position constraints and detection results of local regions. The innovation of the proposed method lies in combining sparse coding and part based model for face detection. Experimental results in Caltech and BioID database show that the proposed method is suitable for small sample size problem and has good detection results in case of occlusion, rotation, complex expressions.

Key words: face detection; sparse coding; dictionary learning; part based model; detection accuracy; recall

人脸检测是计算机视觉领域的一个经典问题,是人脸分析的基础,其应用场合包括人脸配准、人脸识别、头部姿态跟踪、表情跟踪/识别以及性别/年龄识别等[1].人脸检测任务首先要判断给定的图像或者视频帧中是否出现人脸:若是,则需指出人脸出现的位置.早期研究中的典型方法主要利用纹理、边缘、颜色等一些低层图像特征[2,3],结合人脸局部器官的几何约束[4,5]定位人脸;或者通过计算图像中每个位置与标准人脸模板之间的关联性检测人脸[6].总的来说,早期的方法是根据一些启发式知识进行检测,很难适应实际环境中的各种变化,且

* 基金项目: 国家自然科学基金(61271288); 广东省自然科学基金(8152840301000009)
收稿时间: 2013-05-28; 修改时间: 2013-07-17; 定稿时间: 2013-08-27

实时性和检测精度难以满足实际应用需求.针对早期研究工作的缺陷,研究人员提出了基于外观的人脸检测方法[7−9].这类方法的基本思想是:利用统计学习理论找到区分人脸和非人脸的特性,并对图像中每个位置进行判断.其中,级联 Haarlike-Adaboost 方法[9]实时性好、检测精度较高,其优异的检测性能使人脸检测从理论研究走向了实际应用.该方法的提出,对于人脸检测问题具有里程碑的意义.但是,Haarlike-Adaboost 方法仍存在两个缺陷:首先,该方法训练需要大量正负样本,且训练过程复杂,耗时长;其次,该方法检测鲁棒性较差,各种人脸全局变化,如遮挡、光照变化、表情变化、姿态变化等,都可能导致漏检.

针对 Haarlike-Adaboost 方法训练复杂和耗时较长的问题,文献[10]推广了级联的框架,提出一种用于人脸检测的软级联 Adaboost 算法.不同于传统的级联,该方法仅需训练一个单独的 Adaboost 分类器,然后为每一级的弱分类器训练一个拒绝阈值,最终,在不损失检测精度和检测速度的情况下,极大地降低了训练的复杂度.针对 Haarlike-Adaboost 方法易受光照干扰的问题,文献[11]提出利用 LBP(local binary pattern)特征替代原始的 Haarlike 特征进行人脸检测,LBP 特征具有一定的光照不变性,有效减小了光照变化对人脸检测的干扰.针对 Haarlike-Adaboost 方法易受遮挡、姿态变化干扰的问题,文献[12]提出了一种基于部件的模型,通过目标部件的响应和位置约束来辅助定位目标,在 PASCAL 数据库上的测试结果,验证了基于部件目标检测的鲁棒性.虽然上述方法能够部分地弥补 Haarlike-Adaboost 方法的缺陷,但是这些方法为了获得较好的检测性能,无一例外都需要大量训练样本.

稀疏表达是近年来计算机视觉和机器学习领域中的一大研究热点,已被成功地应用于图像去噪[13]、人脸识别[14,15]、图像分析[16]等场合.但目前,利用稀疏表达完成目标检测的研究还较少,比较相关的工作有:Mairal 等人[17]提出的利用稀疏表达检测目标类的边缘;Song 等人[18]在可变形部件模型的基础上引入稀疏表达,从而提高多类目标检测的效率.

本文借鉴基于部件模型的思想,将稀疏表达应用于人脸检测,提出了一种基于局部区域稀疏编码的人脸检测方法.该方法在保证检测精度和召回率的前提下,较好地解决了小样本情况下的人脸检测问题,系统框图如图 1 所示.

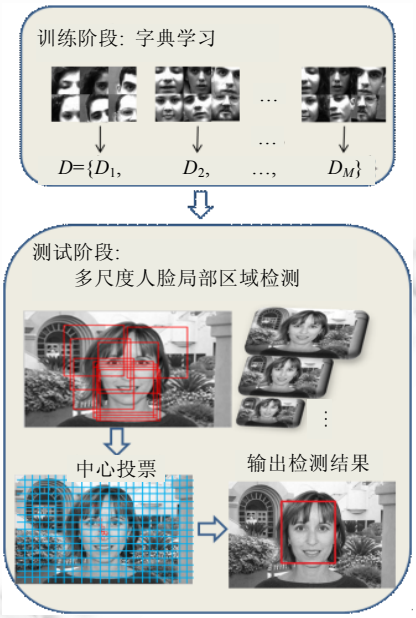


Fig.1 Diagram of the proposed face detection method

图 1 本文提出的人脸检测方法框图

在训练阶段,通过训练各人脸局部区域得到具有较强判别力的字典,其中,不同的人脸局部区域对应生成字

典中不同的基(atom);在测试阶段,基于训练生成的字典计算图像金字塔中各待检测窗口的稀疏编码,并根据稀疏编码中对不同字典项的响应值来判断待检测窗口是否属于人脸某一局部区域;最后,利用人脸各局部区域的相对位置关系进行目标中心投票,完成人脸定位.

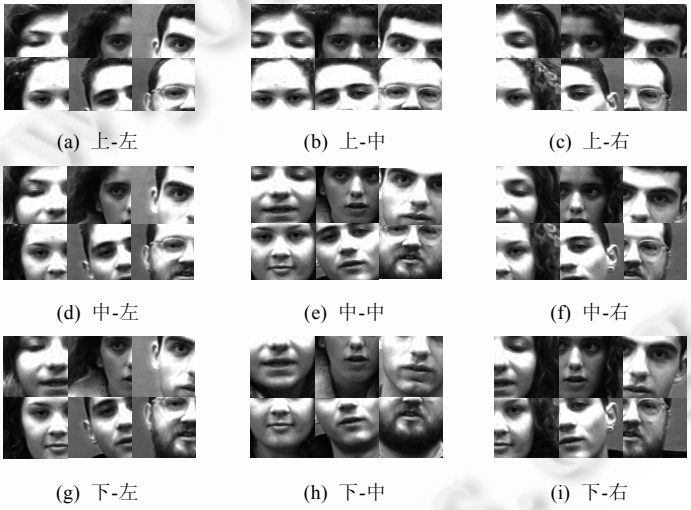
本文第 1 节对文中的方法进行详细描述.第 2 节给出在 Caltech 和 BioID 两个标准人脸数据集上的实验结果,并讨论字典参数选择问题.最后是结论与展望.

1 基于局部区域稀疏编码的人脸检测方法

本方法的基本思想是利用稀疏表达对人脸局部区域进行检测,并结合局部区域的位置约束完成人脸定位.具体细节如下文所述.

1.1 人脸局部区域提取

人脸局部区域提取[19]主要包括局部器官法和局部区域法.局部器官法注重根据人脸中具有明确物理意义的器官,例如眉、眼、口、鼻等进行图像划分;划分的区域语义信息明确,但需要繁重的人工标注或者复杂度较高的特征检测方法.局部区域法只需要根据具体坐标位置进行人脸划分,操作上简单易行.文献[19]采用不重叠相等大小的矩形区域对人脸进行划分,而本文将人脸图像划分为 M 个互有重叠相等大小的局部区域.更具体地说,通过上、中、下和左、中、右的 9 种不同组合方式,本文提取人脸正样本中 9 个不同局部区域作为训练样本.为使中心局部区域可以包含眼睛、鼻子、嘴巴几个重要五官,从而有利于提升字典的判别能力,文中每个局部区域块大小设为 70×70 像素.人脸局部区域构造方式如图 2 所示.



(a) 上-左 (b) 上-中 (c) 上-右

(d) 中-左 (e) 中-中 (f) 中-右

(g) 下-左 (h) 下-中 (i) 下-右

Fig.2 Examples of face local region

图 2 人脸局部区示例

1.2 基于局部区域的字典学习

1.2.1 字典学习和稀疏编码

本文定义数据观测矩阵 $Y\in\mathbb{R}^{m\times N}$,其中,$Y$ 的任意列 $y_j\in\mathbb{R}^m(j=1,\ldots,N)$代表一个 m 维信号.设 $D\in\mathbb{R}^{m\times K}(K>N)$是过完备字典,由一组归一化基向量 $\{d_j\}_{j=1}^K$ 构成.在训练阶段,如何设计出过完备字典 D 使得 y_j 仅仅由少量字典基向量的线性组合较好重构,是字典学习需要考虑的问题,如公式(1)所示:

$$\min_{D,X}\frac{1}{2}\|Y-DX\|_2^2 \quad \text{s.t. } \forall i,\|x_i\|_0\leqslant T \tag{1}$$

其中,X 的任意列 $x_i\in\mathbb{R}^K(i=1,\ldots,N)$为对应观测数据 y_i 的稀疏表达;$\|\cdot\|_0$ 是 l^0 范数,表示稀疏向量 x_i 中非零元的个

数;T为x_i满足的稀疏度阈值.由公式(1)可知,重构性是字典学习算法需要优先考虑的性质.K-SVD[20]是一种通过迭代求解字典的经典解法,迭代速度快,且训练得到的字典对信号的重构误差较小.

在测试阶段,固定 D,求解测试样本 y^{test} 对应的稀疏向量 x^{test},使得 $y^{test}\approx Dx^{test}$ 成立,称为 y^{test} 在字典 D 下的稀疏编码,如公式(2)所示:

$$\min_X \frac{1}{2}\| y^{test}-Dx^{test}\|_2^2 \quad \text{s.t. } \forall i,\| x^{test}\|_0 \leqslant T \tag{2}$$

稀疏编码问题最典型的解法是利用追踪算法去估计最优解,包括匹配追踪(MP)、正交匹配追踪(OMP)[21]以及基追踪(BP)[22].此类方法计算复杂度低,收敛速度快,能够很好地估计全局最优解,因此,本文采用 OMP 算法对信号进行稀疏编码.

1.2.2 字典学习模型的建立

本文需要根据窗口的稀疏编码判别其隶属于哪一类人脸局部区域,所以要求字典具有如下一些性质:

(1) 重构性.利用字典能够很好地表达和重构人脸各局部区域.

(2) 判别性.基于该字典的稀疏编码可分性较好,即每一类人脸局部区域的稀疏编码与其他类尽量满足线性可分.

(3) 可解释性.字典中的基与人脸的某个局部区域有明确的对应关系.

虽然 K-SVD 算法生成的字典重构性较好,但是其字典不具备判别性以及可解释性,所以不满足本文字典学习的要求.本文采用 LC-KSVD[23]方法学习字典.该方法定义的目标函数中不仅考虑重构误差,而且把稀疏编码的判别能力和分类误差也引入到目标函数中,因此训练得到的字典既能适应训练样本的潜在结构,又能使得稀疏编码具有较强的判别力.过完备字典的形式如公式(3)所示:

$$D=[D_1,D_2,\ldots,D_M]=[d_{1,1},d_{1,2},\ldots,d_{1,L},\ldots,d_{k,1},d_{k,2},\ldots,d_{k,L},\ldots,d_{M,1},d_{M,2},\ldots,d_{M,L}]\in\mathbb{R}^{m\times K} \tag{3}$$

其中,字典 D 是由 M 类人脸局部区域对应的 M 个子字典组成,D_k 是第 k 类人脸局部区域对应的子字典,$L=K/M$ 为各个子字典中基的数量,$\{d_{kj}\}_{j=1}^{L}$ 是属于第 k 个子字典的基.人脸局部区域、子字典和字典基三者的对应关系如图 3 所示.本文 9 类人脸局部区域所属的样本对应于 9 个子字典.

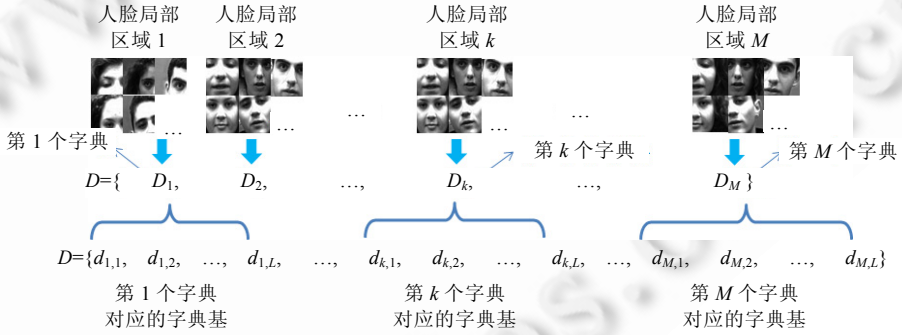


Fig.3 Corresponding relation among face local region, sub-dictionary, and the dictionary atoms

图 3 人脸局部区域、子字典和字典基的对应关系

将字典学习建模为最优化问题,如公式(4)所示:

$$\langle D,W,A,X\rangle=\arg\min_{D,W,A,X}\| Y-DX\|_2^2+\alpha\| Q-AX\|_2^2+\beta\| H-WX\|_2^2 \quad \text{s.t. } \forall i,\ \| x_i\|_0\leqslant T \tag{4}$$

其中,Y 为数据观测矩阵,其每一列对应一个人脸局部区域提取的特征向量,为了提高检测的速度,本文直接把人脸局部区域拉成向量,并利用 PCA 方法对特征进行降维;D 是待学习的过完备字典,线性变换矩阵 A 和分类器权重 W 也是需要学习的参数;稀疏矩阵 $X\in R^{K\times N}$ 的列对应为样本特征的稀疏编码;类别标示矩阵 $H=[h_1,\ldots,h_N]\in R^{m\times N}$ 的每列对应为一个类别标示向量 $h_i=[0,\ldots,0,1,0,\ldots,0]^t$;$Q=[q^1,\ldots,q^N]\in R^{K\times N}$ 是训练样本对应稀疏编码的判别

矩阵,若第 i 个训练样本属于第 k 类人脸局部区域,则

$$q^i=[q^i_{1,1},q^i_{1,2},...,q^i_{1,L},...,q^i_{k,1},q^i_{k,2},...,q^i_{k,L},...,q^i_{M,1},q^i_{M,2},...,q^i_{M,L}]^t=[0,0,...,0,...,1,1,...,1,...,0,0,...,0]^t\in R^K.$$

α,β是标量,为目标函数中第 2 项和第 3 项的权重.α,β太大,会造成训练过拟合,检测过程中人脸局部区域的任何较小变化都有可能被判断成背景;α,β太小,会造成字典的判别性变差,违背了我们利用 LC-KSVD 训练字典的初衷.实验中,我们设定α=4,β=2.

公式(4)中,目标函数的第 1 项为重构误差,第 2 项为稀疏编码判别误差,第 3 项为分类误差.因为第 2 项的存在,使训练得到的字典具有了判别性和可解释性,从而使得样本的稀疏编码具有更好的分类能力,即每一类人脸局部区域稀疏编码之后,类内距离更小,类间距离更大.需要强调的是,本文的方法要求字典必须具有可解释性,否则定义的判别准则会失效.目标函数的第 3 项进一步提升了字典的判别性.所以,尽管没有用到分类器参数 W,但第 3 项与前两项联合优化求得的字典判别性变得更好.基于该字典的稀疏表达不仅能够较好地重构信号,而且会更好地兼顾稀疏表达的分类性能,有利于正确判断人脸局部区域.之所以没有用到分类器 W,是因为分类器只有对正样本分类的能力,而没有拒绝负样本的能力,所以,本文在测试阶段重新定义了判别准则来检测人脸各类局部区域.

1.2.3 字典学习模型的求解

文献[16]利用依次反复迭代更新 D,A,W 的方式来求解公式(4),但是该方法收敛速度慢,且可能收敛到局部最优解.本文并不直接求解公式(4),而是根据矩阵范数的性质,把公式(4)作变形,如公式(5)所示:

$$\langle D_{new},X\rangle=\arg\min_{D_{new},X}\|Y_{new}-D_{new}X\|_2^2\quad\text{s.t. }\forall i,\|x_i\|_0\leqslant T \tag{5}$$

其中,$Y_{new}=\left(Y^T,\sqrt{\alpha}Q^T,\sqrt{\beta}H^T\right)^T$ 为已知量,$D_{new}=\left(D^T,\sqrt{\alpha}A^T,\sqrt{\beta}W^T\right)^T$ 是需要训练的未知参数.易知:通过巧妙变换,优化问题(5)就可以利用 K-SVD 算法[16]同时获得所有参数的最优解.利用 K-SVD 求解,首先需要获得各未知参数的初始值 D_0,A_0,W_0.从 M 类人脸局部区域中随机抽取样本,利用 K-SVD 算法得到 M 类局部区域各自的初始字典 $\{D_i^*\}_{i=1}^M$,从而构造出初始字典 $D_0=|D_1^*,D_2^*,...,D_M^*|$.根据每个字典基的标号以及训练样本的类标号,就可以
确定判别矩阵 Q.再利用 OMP 算法[21]得到基于字典 D_0 的训练样本初始稀疏矩阵 X.根据公式(6)初始化 A_0:

$$A_0=\arg\min_A\|Q-AX\|^2+\lambda_2\|A\|_2^2 \tag{6}$$

由岭回归易知,公式(6)的闭合形式解为

$$A_0=(XX^T+\lambda_2 I)^{-1}XQ^T \tag{7}$$

同理可以得到初始值 W_0 为

$$W_0=(XX^T+\lambda_1 I)^{-1}XH^T \tag{8}$$

公式(6)~公式(8)中,λ_1,λ_2 是正则项系数.初始化 D_0,A_0,W_0 后,即可依照算法 1 中的迭代优化方法求解公式(5).

算法 1. 迭代优化方法.

步骤 1. 初始化.

根据 D_0,A_0,W_0 得到初始化的 D_{new}^0,并初始化迭代次数 J=1.

步骤 2. 稀疏编码.

判断是否满足停止条件:如果是,则跳出循环;如果不是,则利用正交匹配追踪算法(OMP)计算 Y_{new} 中每一个样本 y_i 的稀疏编码 x_i,即

$$X=\arg\min_X\|Y_{new}-D_{new}^{J-1}X\|_2^2\quad\text{s.t. }\forall i,\|x_i\|_0\leqslant T.$$

步骤 3. 字典更新.

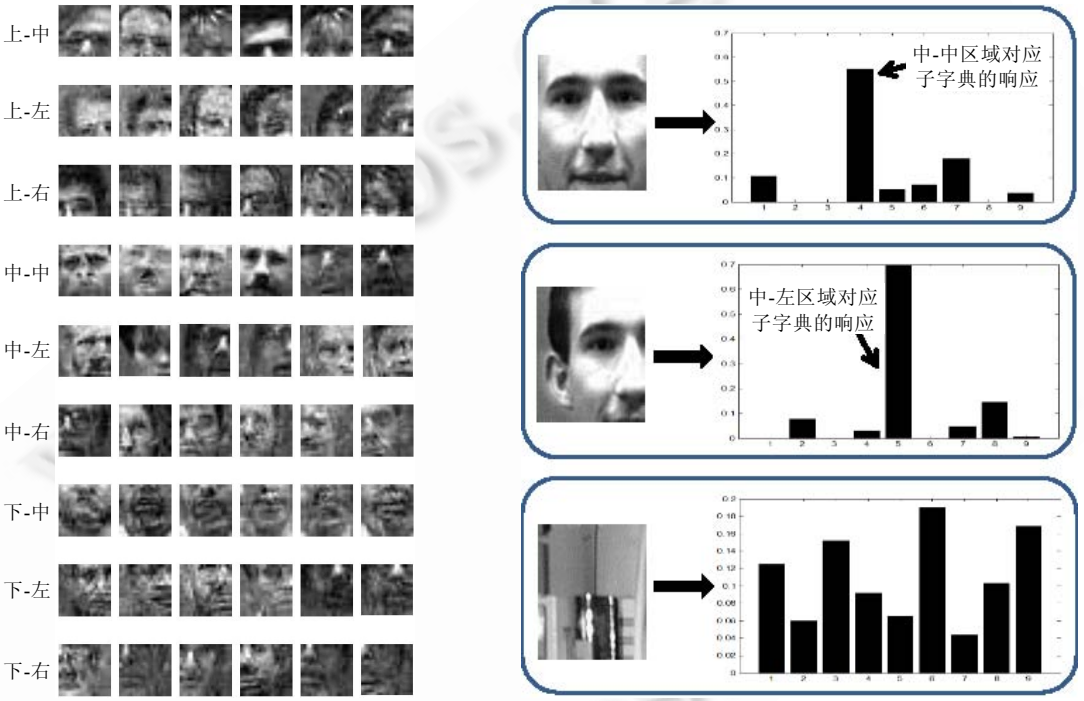
更新字典 D_{new}^{J-1} 的每一列,其中,第 k 列更新步骤如下:

- 定义训练样本的子集 $\{y_i\}_{i\in\omega_k}$,即子集中的样本都利用了字典的第 k 列(第 k 个基)来重构信号,

$$\omega_k=\{i|1\leqslant i\leqslant N,x_i(k)\neq 0\}.$$

- 计算总的重构误差矩阵 $E_k = Y_{new} - \sum_{j \neq k} d_j X(j)$, 其中,$d_j$ 为字典的第 j 个基,$X(j)$为矩阵 X 的第 j 行.
- E_k 的每一列对应每个训练样本的重构误差, E_k^R 为 E_k 中对应 ω_k 的列.
- 利用 SVD 做低秩估计,对 E_k^R 做 SVD 分解, $E_k^R = U\Delta V^T$.选择更新字典的第 k 列为 U 的第 1 列,并归一化新的基.更新系数为 V 的第 1 列乘以 $\Delta(1,1)$.
- 若 D_{new}^J 的每一列都更新完毕则依次执行 $D_{new}^J = D_{new}^{J-1}$,J=J+1,并返回步骤(2);否则,继续执行步骤(3).

通过上面的步骤,即可求解字典 D_{new},从而很容易可以获得问题(4)中的最优参数 D,A,W.为了更直观地分析训练得到的矩阵 D,我们利用 PCA 降维矩阵把字典的基恢复到高维空间中,图 4(a)在每一个子字典中随机选择 6 个基显示.可以看出,每一个子字典包含的基都明确对应一类人脸局部区域,这说明字典具有可解释性.可推知:当规定的这 9 类区域出现时,该字典能够很好地对其进行重构.图 4(b)给出了 3 个测试样本稀疏表达在各个子字典上的响应.可以看出,人脸中-中和中-左区域的稀疏表达在各自对应的子字典上有较强的响应,而背景样本在各个子字典上响应比较均匀.利用这一特性能够正确区分各个人脸部分和背景,说明字典具有很好的判别性.



(a) 字典中基的图像表示 (b) 样本稀疏表达在各个子字典上的响应

Fig.4 Illustration of learned dictionary

图 4 学习得到的字典的说明

1.3 人脸局部区域检测

已有的人脸局部区域检测方法大多是基于人脸部件的分类器[5,12],所以需要大量的正负样本参与训练.与之不同的是:本文提出的方法没有用到负样本,同样能够有效地排除图像中的复杂背景.通过大量观察发现:当某人脸局部区域出现时,稀疏编码仅仅在该人脸局部区域对应的基上有较强的响应;当背景窗口出现时,其稀疏编码在字典的所有基上响应比较均匀.基于此,本文提出人脸局部区域检测准则如下:记待检测窗口的稀疏编码为 $x^{test} \in \mathbb{R}^{K \times 1}$,当 x^{test} 在字典上的响应满足公式(9)时,则判定该测试窗口为人脸的第 k 个局部区域:

$$label\_x^{test}=\begin{cases}\arg\max\limits_k x_k^{test}, & 若\dfrac{fm-sm}{sm}>\tau\\ 0, & 其他\end{cases} \tag{9}$$

其中, $x_k^{test}=\sum_{l=1}^{L}x_{k,l}^{test}$ 表示待检测窗口的稀疏编码对应第 k 个子字典上的响应,fm 和 sm 分别为集合 $x_{set}^{test}=\{x_k^{test}\}_{k=1}^{M}$ 中最大和次大的值.$label\_x^{test}$ 为 k,表示该测试窗口为人脸的第 k 个局部区域;$label\_x^{test}$ 为 0,表示该测试窗口为背景.

已有的基于人脸部件的检测方法[5,12]需要对每个人脸部件分别建模,然后独立搜索每个部件.本文的方法只需建立一个共用的人脸部件模型(字典),且在搜索过程中能够直接判断每个窗口是否属于规定的 9 个人脸局部区域.简而言之,本文的方法对于测试图片仅需进行 1 次多尺度窗搜索,即可获得 9 个规定局部区域的检测结果,而不用对 9 类区域单独进行检测,减少了计算开销.这里需要强调的是:本文的方法只能检测训练时所规定的 9 种人脸局部区域,任何其他人脸局部区域(比如眼睛和鼻子周围的局部区域)都会被当作背景丢掉.

1.4 投票策略

直觉上,人脸各部分和人脸中心的相对位置不会有较大偏差,因此,本文利用人脸局部区域检测结果来估计人脸的位置:

首先,我们利用训练样本来估计每个人脸局部区域与人脸中心的相对位置.假设第 k 部分的 N 个训练样本保留了样本中心与人脸中心的相对位置,记为 $\{(x_i^k,y_i^k)\}_{i=1}^N$, 则取其均值作为该部分与目标中心的相对位置,即

$$x^k=\frac{1}{N}\sum_{i=1}^{N}x_i^k,\ y^k=\frac{1}{N}\sum_{i=1}^{N}y_i^k \tag{10}$$

其次,在每一个尺度下,把图像分成较小的晶格,并根据人脸部分的检测和得到的相对位置关系对人脸中心进行投票,人脸中心投票结果如图 5 所示.最后统计每个晶格内目标中心的投票数,保留大于阈值的目标中心,目标窗口的尺寸由投票所在的图像尺度决定.由图 5 可以看出,虽然人脸中心位置仍然存在误检,但绝大部分投票是正确的,这也正是本方法鲁棒性强的原因.

Fig.5 Voting results of face center

图 5 人脸中心投票结果

2 实验结果与分析

2.1 训练和测试数据库

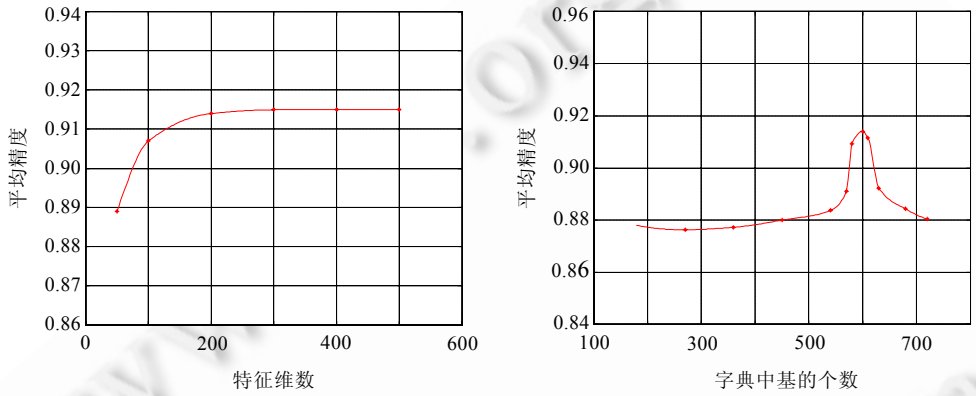
实验采用了 Yale,Face94 和 FERET 这 3 个标准数据库作为训练样本,其中,Yale 数据库仅包含 165 张灰度图片,用于训练本文提出的方法;实验中还选择与 Yale 数据库采集方式差别很大的 Face94 人脸数据库作为训练样本,目的是验证文中模型受训练样本影响的大小.此外,实验比较了本文方法与其他 3 种经典方法——Haarlike-Adaboost、LBP-Adaboost 和软级联[10]的性能.为了训练这 3 种方法,实验中把 Yale 数据库、Face94 数据库和 FERET 数据库中的样本缩放成统一大小的灰度图片作为正样本,另外收集了 3 000 张不包含人脸的图片,负样本是从中截取得到的.最终,用于训练分类器的正样本数量为 4 625,负样本数量超过 10 万.相对于几种级联方法,本文提出的方法在训练阶段仅需要少量正样本,这极大地简化了训练的工作量.

实验中选择 Caltech 和 BioID 人脸数据库来做测试,其目的在于测试模型对真实场景中的人脸检测性能.Caltech 人脸数据库包含 450 张彩色图片,每张图片中有且只有 1 张人脸.所有人脸都正对镜头,平面内偏转较小,但采集该人脸的环境比较复杂,包括室内室外、不同光照、各种背景干扰和尺度变换.BioID 人脸数据库包含 1 521 张灰度图片,图像分辨率为 384×286.在数据库中,人脸的尺度、表情、视角和光照都有较大的变化.

2.2 字典参数设置

字典训练包括两个重要参数:样本特征维数(字典的行数)和基的个数(字典列数).本文通过实验寻求最优的参数,从而实现检测精度和检测效率的一个很好的折中.图 6(a)给出了当字典基的个数为 603 时,特征维数对检测精度的影响.可以看出:当训练样本的维数小于 200 时,随着维数的增加,平均检测精度迅速提高;当训练样本的特征大于 200 维时,平均精度趋于平缓.从平均检测精度和运算效率综合来考虑,实验中训练样本维数取 200.

图 6(b)给出了当特征维数为 200 时,字典中基的个数对平均精度的影响.图 6(b)中显示:字典基过大或者过小,均可能导致系统性能的下降.当字典中基的个数在 600 左右时,平均精度达到最大值,因此,实验中字典基的个数取 603(每类人脸局部区域对应 67 个基).综上,文中训练阶段得到的字典尺寸为 200×603.



(a) 特征维数对平均精度的影响 (b) 字典中基的个数对平均精度的影响

Fig.6 Influence of feature dimension and the number of dictionary atoms on the average precision

图 6 特征维数和字典中基的个数对平均精度的影响

2.3 检测精度和速度

实验 PC 机主频 2.2GHz,内存 4G,实验软件 Matlab2010b,本方法和 15 级 Adaboost 级联分类器分别训练 5 次.训练的字典大小选择为 200×603,本方法训练平均耗时为 4.96s.在同样的实验条件下,训练一个 15 级 Adaboost 级联分类器平均耗时约 80 个小时.由此可见,本方法在训练时间上的优势比较明显.此外,对于 384×286 大小的图像,本文的方法检测每帧约为 400ms,虽然稍逊于级联 Haarlike-Adaboost 方法,但足以满足实际应用.

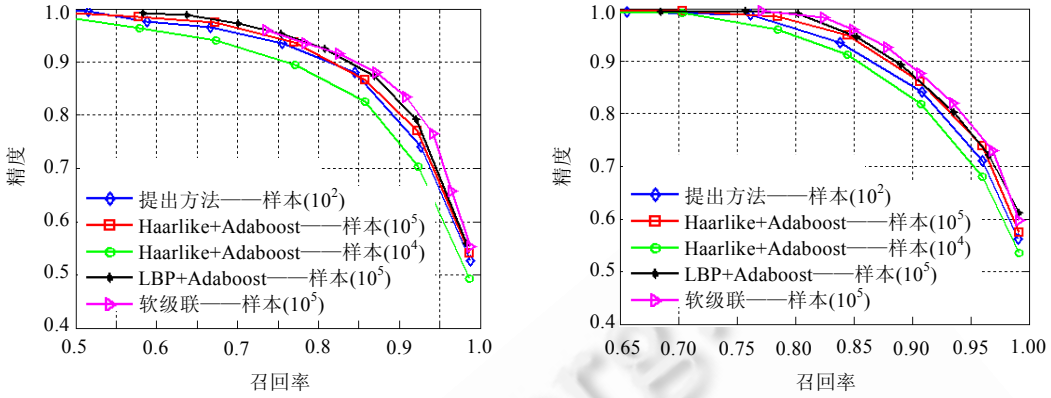
实验中,利用检测窗与真实目标窗的相对重叠面积来衡量检测精度,即

$$S(C,C_{gt})=\frac{C\cap C_{gt}}{C\cup C_{gt}} \tag{11}$$

其中,C表示检测窗,C_{gt}表示真实目标窗.实验规定当$S(C,C_{gt})>0.5$时,窗口为正确检测.另外,如果多个窗口重复检测一个目标,则只有 1 个正确,其余窗口都为误检.因此,实验中采用非极大值抑制方法[12]对检测窗进行融合,以防止重复检测.

图 7 给出了在 Caltech 和 BioID 数据库上的精度-召回率曲线.由图 7 可知:在近似的检测性能情况下,本文的方法仅采用很少的训练样本(数量级 10^2),但其他实时人脸检测方法(包括 Haarlike-Adaboost、LBP-Adaboost、软级联)则需要采用大量训练样本(数量级 10^5);Haarlike-Adaboost 方法检测人脸的效果易受训练样本的影响,当训练样本数量减少为 10^4 时,本文的方法优于 Haarlike-Adaboost 方法.表 1 给出了在 BioID 数据库中,训练样本

在不同数量级下,Haarlike-Adaboost 和本文方法平均检测精度的比较.可以清晰地看出:本文方法在小样本($O(10^2)$)训练下,接近 Haarlike-Adaboost 方法在大量训练样本($O(10^5)$)下的检测结果;当训练样本不足时,本文方法要明显优于 Haarlike-Adaboost 方法.



(a) Caltech 数据库 (b) BioID 数据库

Fig.7 Precision recall curves in the standard face databases

图 7 标准人脸数据库精度召回率曲线

Table 1 Average detection precision comparison of Haarlike-Adaboost and the proposed method

表 1 Haarlike-Adaboost 和提出方法平均检测精度的比较

方法	样本数量级			
	$O(10^2)$	$O(10^3)$	$O(10^4)$	$O(10^5)$
本文方法	0.950	–	–	–
Haarlike-Adaboost	–	0.874	0.942	0.958

2.4 训练样本对检测性能的影响

为了验证文中模型,同时为了尽可能小地受到训练样本的影响,采用 Yale 和 Face94 两个不同数据库进行训练,采用 Caltech 数据库做测试,实验结果见表 2.可以看出:尽管两个数据库样本的数量和采集方式差别很大,但检测性能基本不受影响.

Table 2 Influence of the different training samples on detection performance

表 2 不同训练样本对检测性能的影响

训练数据库	测试数据库	人脸数	误检数	召回率
Yale 数据库	Caltech	450	61	0.842
Face94 库	Caltech	450	60	0.849

2.5 可视化检测结果与分析

为了更直观地分析本文方法的性能,实验分别给出了部分遮挡、复杂表情、人脸偏转和光照变化这 4 种情况下,本文方法和 LBP-Adaboost 方法的可视化检测结果.由图 8(其中,C1,C3 为本文方法的结果,C2,C4 为级联 adaboost 方法的结果)可知:相对于 LBP-Adaboost 方法,本文提出的方法在前 3 种情况下有更好的检测效果.原因在于:尽管目标整体已经产生了较大的变化,但目标某些局部区域并未发生变化,所以利用局部区域来进行人脸检测受干扰较小;反之,LBP-Adaboost 方法利用全局特征来表征人脸,当人脸整体产生较大变化时,就很容易漏检.图 8(d)显示出,在光照条件欠佳的情况下,本文方法不如 LBP-Adaboost 方法效果好.究其原因主要是:文中提取特征的方法过于简单,所以光照强度的变化会导致特征不稳定,从而影响检测效果;而 LBP-Adaboost 方法的特征具有一定程度的光照不变性,可以更好地去除光照干扰.

(a) 部分遮挡

(b) 复杂表情

(c) 人脸偏转

(d) 光照变化

Fig.8 Examples of the visual detection results

图 8 可视化检测结果示例

3 结论与展望

本文提出了一种基于局部区域稀疏编码的人脸检测方法.该方法的主要优点包括:1) 不需要负样本即可训练得到模型,且训练过程简单、快速;2) 人脸检测的鲁棒性更强,在人脸遮挡、复杂表情和人脸偏转情况下具有很好的检测效果;3) 本文方法仅利用少量的正样本就可以达到与几种经典级联方法接近的检测精度和召回率,且人脸检测效果受样本变化的影响较小.未来的研究工作包括:1) 研究引入光照不变的特征,以增强方法在光照变化情况下人脸检测的鲁棒性;2) 研究利用各种偏转人脸的局部区域训练字典,从而改善对偏转人脸检测的效果;3) 把本文的方法推广到其他目标类的检测.

References:

[1] Zhang C, Zhang ZY. A survey of recent advances in face detection. Technical Report, MSR-TR-2010-66, Washington: Microsoft Research, 2010.

[2] Sirohey SA. Human face segmentation and identification. Technical Report, CS-TR-3176, Maryland: University of Maryland, 1993.

[3] Augusteijn MF, Skujca TL. Identification of human faces through texture-based feature recognition and neural network technology. In: Proc. of the IEEE Conf. on Neural Networks. IEEE, 1993. 392−398. [doi: 10.1109/ICNN.1993.298589]

[4] Yang G, Huang TS. Human face detection in complex background. Pattern Recognition, 1994,27(1):53−63. [doi: 10.1016/0031-3203(94)90017-5]

[5] Heiseleu B, Serret T, Pontils M, Poggiot T. Component-Based face detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2001. 657−662. [doi: 10.1109/CVPR.2001.990537]

[6] Lanitis A, Taylor CJ, Cootes TF. An automatic face identification system using flexible appearance models. Image and Vision Computing, 1995,13(5):393−401. [doi: 10.1016/0262-8856(95)99726-H]

[7] Rowley H, Baluja S, Kanade T. Neural network-based face detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998,20(1):23−38. [doi: 10.1109/34.655647]

[8] Osuna E, Freund R, Girosi, F. Training support vector machines: An application to face detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 1997. 130−136. [doi: 10.1109/CVPR.1997.609310]

[9] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2001. I-511−I-518. [doi: 10.1109/CVPR.2001.990517]

[10] Bourdev L, Brandt J. Robust object detection via soft cascade. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2005. 236−243. [doi: 10.1109/CVPR.2005.310]

[11] Liao SC, Zhu XX, Lei Z, Zhang L, Li SZ. Learning multi-scale block local binary patterns for face recognition. In: Proc. of the Int'l Conf. on Biometrics. IEEE, 2007. 828−837. [doi: 10.1007/978-3-540-74549-5_87]

[12] Felzenszwalb PF, Girshick RB, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(9):1627−1645. [doi: 10.1109/TPAMI.2009.167]

[13] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. on Image Processing, 2006,15(12):3736−3745. [doi: 10.1109/TIP.2006.881969]

[14] Wagner A, Wright J, Ganesh A, Zhou ZH, Mobahi H, Ma Y. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(2):372−386. [doi: 10.1109/TPAMI.2011.112]

[15] Ortiz EG, Wright A, Shah M. Face recognition in movie trailers via mean sequence sparse representation-based classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2013. 3531−3538. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6619297&queryText%3DFace+recognition+in+movie+trailers+via+mean+sequence+sparse+representation-based+classification

[16] Pham D, Venkatesh S. Joint learning and dictionary construction for pattern recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2008. 1−8. [doi: 10.1109/CVPR.2008.4587408]

[17] Mairal J, Leordeanu M, Bach F, Hebert M, Ponce J. Discriminative sparse image models for class-specific edge detection and image interpretation. In: Proc. of the Computer Vision–ECCV 2008. Springer-Verlag, 2008. 43−56. [doi: 10.1007/978-3-540-88690-7_4]

[18] Song HO, Zickler S, Althoff T, Girshick R, Fritz M, Geyer C, Felzenszwalb P, Darrell T. Sparselet models for efficient multiclass object detection. In: Proc. of the Computer Vision–ECCV 2012. Springer-Verlag, 2012. 802−815. [doi: 10.1007/978-3-642-33709-3_57]

[19] Zhu YL, Chen SC. Sub-Image method based on feature sampling and feature fusion for face recognition. Ruan Jian Xue Bao/Journal of Software, 2012,23(12):3209−3220 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4199.htm [doi: 10.3724/SP.J.1001.2012.04199]

[20] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. on Signal Processing, 2006,54(11):4311−4322. [doi: 10.1109/TSP.2006.881199]

[21] Tropp JA. Greed is good: Algorithmic results for sparse approximation. IEEE Trans. on Inform. Theory, 2004,50(10):2231−2242. [doi: 10.1109/TIT.2004.834793]

[22] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Process, 1993,41(12):3397−3415. [doi: 10.1109/78.258082]

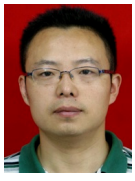
[23] Jiang ZL, Lin Z, Davis LS. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2011. 1697−1704. [doi: 10.1109%2fCVPR.2011.5995354]

附中文参考文献:

[19] 朱玉莲,陈松灿.特征采样和特征融合的子图像人脸识别方法.软件学报,2012,23(12):1−12. http://www.jos.org.cn/1000-9825/4199.htm [doi: 10.3724/SP.J.1001.2012.04199]

张抒(1985－),男,四川绵阳人,博士生,主要研究领域为目标检测,机器学习.
E-mail: jlu_zhangshu@163.com

解梅(1955－),女,博士,教授,博士生导师,主要研究领域为数字图像处理,模式识别.
E-mail: mxie@uestc.edu.cn



蔡勇(1979－),男,博士生,讲师,主要研究领域为人体活动分析,机器学习.
E-mail: caiiong@163.com