

基于稀疏级联回归的快速人脸配准方法 及其在移动设备上的应用

邓健康 杨 静 孙玉宝 刘青山

(南京信息工程大学信息与控制学院 南京 210044)

摘要 如何在计算和存储能力受限的移动平台上实现高效的人脸配准是移动平台人脸应用需要解决的关键问题。主要研究了移动平台上的快速人脸配准问题,为了降低配准模型的计算与存储要求,提出了稀疏约束的级联回归模型。该模型采用稀疏性约束学习回归矩阵,不但能够筛选鲁棒的特征,而且模型的存储空间被压缩到原来的5%左右。基于稀疏级联回归模型,进一步构建了移动平台上人脸配准的快速算法。首先,在人脸检测的基础上,利用二值特征快速定位眼角、嘴角和鼻尖的关键点,估计出人脸的姿态,旋正人脸图像;然后,根据人脸的姿态,选择相应的正脸或侧脸模型,进行稀疏约束的级联回归配准,定位人脸关键点。大量实验结果表明,提出的配准方法精度高、速度快、模型小。在三星 Note3 智能手机上,每幅人脸图像的配准时间在 10ms 左右,整个 apk 文件大小仅为 4MB,非常适合移动平台的人脸应用。

关键词 移动平台,快速人脸配准,级联回归,稀疏约束

中图分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.061

Fast Face Alignment Method Based on Sparse Cascade Regression and its Application on Mobile Devices

DENG Jian-kang YANG Jing SUN Yu-bao LIU Qing-shan

(School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract Efficient face alignment is the key problem for the face applications on the mobile platform which has limited computing and storage capacity. We studied the problem of fast face alignment on the mobile platform. To reduce the computing and storage requirements for face alignment, sparse constrained cascade regression model was proposed in this paper. Sparse constraint was introduced to learn the regression matrix, which can not only select the robust features, but also compress the model size to about 5% compared to the original model. We further constructed the fast face alignment algorithm on mobile platform based on sparse cascade regression model. First, the facial landmarks on the tip of the nose, the corners of the mouth and eyes are quickly located by binary features after face detection, and face pose is estimated. Face image is rotated to frontal view according to the face pose. Then, the corresponding model (frontal model or profile model) is selected according to the face pose, and cascade regression with sparse constraint is used to face alignment. Extensive experiments show that the alignment method proposed in this paper is effective and efficient with compact model size. On the Samsung smart phone of Note3, the alignment time for each face image is about 10ms, and the size of whole apk is only 4MB, which is suitable for face applications on mobile platform.

Keywords Mobile platform, Fast face alignment, Cascade regression, Sparse constraint

1 引言

随着手机、平板电脑等移动设备性能的提升,移动设备的智能化也成了学术界与工业界的研究热点。移动智能终端上图像传感器的性能提升给移动平台上的视觉应用提供了更好的硬件条件支撑^[1]。移动终端的生物特征识别、表情分析、低频带下的视频通话、自然的人机交互等移动应用在人们的生活中已经崭露头角。这些应用都与人脸关键点的定位密切相

关,如何在计算能力和存储能力受限的移动平台上快速稳定地实现人脸关键点的定位成为解决上述问题的基础与前提。

人脸配准是在一幅人脸图像中自动地定位出人脸各个器官的准确位置以及人脸的外轮廓,为进一步处理与分析人脸图像提供重要的形状信息,是人脸识别、人脸验证、人脸表情分析、人脸动画、人脸姿态估计、人脸跟踪以及人脸图像编码等领域的关键技术之一。不精确的关键点定位往往会引起“误配准灾难”^[2]。近年来,国内外学者提出了多种人脸配准

到稿日期:2014-05-17 返修日期:2014-07-17 本文受国家自然科学基金项目(61272223, 61300162),江苏省自然科学基金项目(SBK201210296, SBK201204234),模式识别国家重点实验室开放课题基金(201204234)资助。

邓健康(1990-),男,硕士生,主要研究方向为基于移动平台的人脸图像分析, E-mail: deng_jiankang@126.com; 杨 静(1991-),女,硕士生,主要研究方向为视觉跟踪; 孙玉宝(1984-),男,博士,副教授,主要研究方向为基于稀疏和低秩的图像分析; 刘青山(1975-),男,教授,博士生导师,主要研究方向为图像分析、基于图/超图模型的视觉理解、基于事件的视频分析、医学图像分析、多光谱图像分析等。

的方法,大体上可以分为基于参数化模型的方法和基于回归的方法。

基于参数化模型的方法中,代表性的工作有活动形状模型(Active Shape Model, ASM)^[3,4]、活动外观模型(Active Appearance Model, AAM)^[5,6]、受约束的局部模型(Constrained Local Model, CLM)^[7],以及基于树^[9]和图^[10]的结构模型。该类方法利用关键点之间的位置关系,以迭代的方式进行关键点定位,运用参数化模型约束迭代过程中的形状,如Coots等人提出的ASM模型^[3,4]和AAM模型^[5,6]。ASM和AAM都是基于点分布模型(Point Distribution Model, PDM)的算法,ASM采用主元分析(Principal Component Analysis, PCA)方法对形状变化进行建模;AAM则采用了形状和纹理相融合的统一约束,在考虑形状变化的同时,采用PCA对纹理变化进行建模。PCA能够保留形状变化的主要模式,但细微的形状变化往往被当作噪声抑制掉。参数化模型的表达能力有限,很难应对自然场景下复杂的人脸变化,尤其在夸张表情和复杂姿态情况下很难精确配准。

基于回归的方法中,代表性的工作有基于深度卷积神经网络回归的方法^[10,11]和基于级联回归的方法^[12-17]。基于深度卷积神经网络的方法直接从人脸图像定位出关键点坐标,由于深度网络模型具有层级抽象的能力,能够学习全局的上下文信息来进行关键点定位,因此它取得了很好的定位效果,但是,训练深度网络模型需要大量的标注样本,同时训练过程中的参数调节也比较费时费力。基于级联回归的方法主要依赖局部描述子的稳健性能,通过级联的弱回归器拟合复杂的非线性映射,能够有效定位人脸关键点坐标。文献^[14]采用快速SIFT特征,通过最小二乘回归实现关键点快速配准。文献^[17]通过利用二值特征,进一步提高了配准速度。文献^[16]通过显式的回归遮挡信息,增强了级联回归模型在局部遮挡情况下的鲁棒性。基于级联回归模型的方法精度高、定位速度快,但是模型占用的存储空间较大。

传统的人脸配准算法需要一定计算复杂度和存储空间的支持,所以对计算能力和存储能力有限的移动设备来说,很难直接将其移植应用。为了能在移动智能设备上实现实时人脸配准,如何设计计算与存储资源要求更低的匹配算法是需要解决的难题。本文提出了稀疏约束的级联回归配准模型,该模型对计算和存储资源要求较低,具有精度高、速度快、模型小的特点。在此模型的基础上,构建了移动平台上人脸关键点的快速定位算法。该算法在人脸检测的基础上,利用二值特征快速定位眼角、嘴角和鼻尖的关键点,估计出人脸的姿态,旋正人脸图像;然后,根据人脸的侧角,选择相应的正脸($<30^\circ$)或侧脸($>30^\circ$)模型,进行稀疏约束的级联回归配准,定位人脸关键点。在LFPW、Helen、COFW等自然场景数据库上验证了本文方法的配准效果。在三星Note3智能手机上,每幅人脸图像的配准时间在10ms左右,整个apk文件大小仅为4MB,非常适合移动平台的人脸应用。

2 基于稀疏约束的级联回归模型

为了在移动平台上实现人脸关键点快速、准确的定位,本文采用了基于稀疏约束的级联回归模型(Sparse Cascade Regression, SCR),方法流程如图1所示。(1)在人脸检测的基础上快速定位眼角、嘴角、鼻尖的7个关键点,并估计出人脸

姿态。实验研究表明,在人脸关键点的定位过程中,眼角、嘴角、鼻尖比其他关键点的定位更鲁棒、更精确,所以对这7个关键点单独提取特征,训练局部模型,进行特征的匹配与关键点的定位。在移动平台上的实现过程中,我们巧妙地利用了特征的二值特性,将点积浮点运算变成了二进制位运算,加快了计算速度^[18]。在眼角、嘴角和鼻尖快速定位之后进行简单的校验修正,然后估计出人脸的姿态,通过平面内旋转角度将人脸图像旋正,之前图像的尺度已经缩放至人脸检测框宽度,即200个像素,用人脸的平面外旋转角度来选择级联回归的模型,±30°之内的使用正脸模型,±30°以上的使用侧脸模型。(2)通过姿态选择对应的级联回归模型,进行基于稀疏约束级联回归的人脸配准。对于每个关键点,选取64×64的局部区域提取SIFT特征,对于每幅配准图像,仅计算一次梯度,在五次级联回归配准的过程中直接统计不同位置的梯度方向直方图。在训练回归器的过程中增加了稀疏约束,用Lasso^[19]取代最小二乘进行求解,获得稀疏的回归矩阵。在实际运用过程中,这种稀疏性能够筛选鲁棒的特征,同时有效地压缩模型的存储空间。

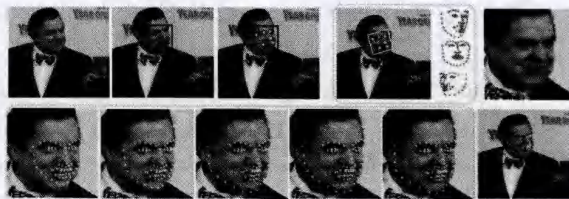


图1 关键点定位算法示意图

2.1 人脸姿态估计

对于一张输入的人脸图像,首先进行人脸检测,然后缩放图像,使得人脸检测框的宽度为200个像素,这样主要是为了方便在同一尺度上提取局部特征。接着,在人脸检测结果的窗口上快速定位7个关键点(眼角、鼻尖、嘴角),并估计出人脸姿态。我们选择了四层LBP金字塔特征^[20]来描述这7个关键点,统计了训练集中这些点在人脸检测窗口中出现的位置,得到其外接矩形区域(该区域也是测试时关键点的搜索区域),以标定点为正样本,其余视为负样本,利用支持向量机(Support Vector Machine, SVM)进行训练,如图2所示。

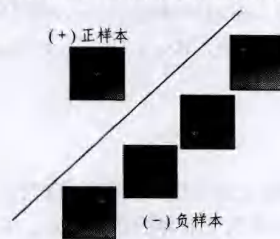


图2 SVM训练右眼角特征

为了进一步降低计算复杂度,考虑到移动平台不擅长浮点计算的实际情况,并且结合LBP特征的二值特性,我们对每个 w_j ($j=1,2,\dots,6$)做了近似^[18],将其表示成一组基向量的组合:

$$w_j \approx \sum_{i=1}^{N_b} \beta_i b_i$$

其中, $b_i \in \{-1,1\}^D$, D 即为特征描述子的维数。对于 w_j 的二值近似,是将其投影到与之同向的基向量 $b_i = \text{sign}(w_j)$,得到投影系数 $\beta_i = \langle b_i, w_j \rangle / \|b_i\|^2$ 。为了减少信息的损失,计

算投影之后的残差 $w_j = w_j - \beta_j b_j$, 重复上述过程 N_b 次, 最终得到 w_j 的二值近似值。

为了能够进一步加速计算, 我们还考虑了 $b_i = b_i^+ - \overline{b_i^+}$, $b_i^+ \in \{0, 1\}^D$, 其中 b_i^+ 与 $\overline{b_i^+}$ 互补, 所以

$$\begin{aligned} \langle w_j, x_j \rangle &\approx \sum_{i=1}^{N_b} \beta_i (\langle b_i^+, x_j \rangle - \langle \overline{b_i^+}, x_j \rangle) \\ &= \sum_{i=1}^{N_b} \beta_i (2 \langle b_i^+, x_j \rangle - |x_j|) \end{aligned}$$

这样就点积浮点运算变成了二进制位运算, 可以通过“与”门快速实现。通过大量的实验测试, 当 $N_b = 4$ 时 w_j 二值近似结果对最终的匹配精度几乎没影响。

在实际的定位过程中往往存在局部遮挡, 因此引入校验修正的机制增强定位的鲁棒性。先对训练数据集人脸图像的眼角、嘴角、鼻尖 7 个关键点位置进行普氏分析 (Procrustes Analysis), 然后进行 PCA 分析, 得出形状的主要变化模式, 此处采用 PCA 主要是因为正交空间投影计算比较方便。当通过局部特征定位了 7 个关键点的位置后, 通过相似变换消除其与训练得到的平均形状之间的刚性差异, 然后将其投影到 PCA 空间, 得到平滑后的位置, 分别统计各个点的偏差, 相应地进行校验修正。

$$\Delta p_i = \frac{|p_i - p_i^*|}{\sum_{i=1}^6 |p_i - p_i^*|} * 100\%$$

其中, p_i 是投影前的形状, p_i^* 是经 PCA 空间平滑后的形状。当误差集中在一两个关键点时, 修正的策略是直接投影到 PCA 空间中形状的对点替换掉误差很大的那一两个点, 其他点保持原状, 然后通过相似逆变换把校正后的形状还原。与 ASM 形状约束不同的是, 个别点的匹配误差不会平均到其他所有点, 并且局部误配准可以通过其他点加以修正。当然, 训练每个关键点特征的判别分类器也可以用来判断关键点遮挡与否, 从而根据精确定位的人脸关键点子集进行姿态估计。

在快速定位 7 个人脸关键点之后, 我们对人脸的三维姿态进行了估计。通过人脸的平面内旋转角度, 将人脸图像旋正, 使得双眼保持水平。通常 SIFT 特征需要旋转到主方向以实现特征的旋转不变性, 然而, 这个步骤比较耗时, 本文用人脸图像旋正取代旋转每个关键点位置的 SIFT 特征。另外, 由于侧脸会引起人脸的自遮挡, 侧脸形状与正脸形状存在明显差异, 因此训练了对应的正脸和侧脸的 3 个模型, 模型分类的界限是旋转角度。在实际配准过程中, 我们通过 7 个关键点位置估计人脸侧角, 选择相应的级联回归模型。

2.2 稀疏级联回归模型

级联回归模型通过级联简单的回归器不断拟合配准残差以完成人脸配准。级联回归方法的成功得益于局部描述子的稳健性能, 以及级联的弱回归器拟合复杂的非线性映射的能力。我们通过级联 T 个回归器 (R_1, R_2, \dots, R_T) 不断拟合配准残差。

$$R_i = \arg \min_{R_i} \sum_{i=1}^N \| (X_i^i - X_i^i) - R_i \Phi(I^i, X_i^i) \|^2$$

其中, I^i 是人脸图像, N 是训练集人脸图像的数目, X_i^i 是标定的人脸形状, X_i^i 是迭代过程中的人脸形状, R_i 是每步迭代的回归矩阵, $\Phi(I^i, X_i^i)$ 是图像 I^i 在形状 X_i^i 位置下的 SIFT 特征描述, 每一步回归的目标都是减少残差, 该优化问题可以通过最小二乘求解, 存在闭式解。迭代过程中的形状更新按照

累加的形式完成。

$$X_{i+1}^i = X_i^i + R_i \Phi(I^i, X_i^i), t=1, \dots, T$$

在实际的应用过程中, 迭代 4 次或 5 次即可收敛。由训练的目标函数可知, 迭代过程中的形状实际上是在人脸形状的线性子空间中, 因此, 引入了隐性的人脸形状约束, 这种约束相对于参数化模型灵活性较高, 在姿态变化和夸张表情下表现较好。

由迭代过程中的形状增量 $R_i \Phi(I^i, X_i^i)$ 可知, 每个关键点迭代过程中位置的变化都与所有关键点的特征相关, 参数维度较高, 容易过拟合。对此, 我们对回归矩阵 R_i 引入了稀疏约束, 即假设某个关键点的位置变化仅仅与一部分关键点的特征相关。

$$\arg \min_{R_i} \sum_{i=1}^N \| \Delta X_i^i - R_i \Phi(I^i, X_i^i) \|^2 + \lambda \| R_i \|_1$$

其中, $\Delta X_i^i = X_i^i - X_i^i$, 该优化问题可由 Lasso^[19] 求解, λ 是正则系数, 控制回归矩阵 R_i 的稀疏度, 我们通过交叉验证确定 $\lambda = 0.1$, 此时, R_i 中每行的非零元素约为 300 左右, 远小于 8704 (68 * 128), R_i 可以获得 5% 左右的压缩率, 大大缩小了模型的存储空间。同时, 稀疏的回归矩阵对应着稀疏的特征选择, 相对于最小二乘的求解方法, 稀疏约束抑制了过拟合倾向, 筛选了鲁棒的特征。在实际的模型训练过程中, 我们根据人脸的姿态差异, 分别训练了正脸和侧脸的 3 个模型。

在测试的过程中, 关键点的位置未知, 稀疏级联回归模型以迭代的方式从当前位置提取特征 $\Phi(I^i, X_i^i)$, 然后乘以稀疏回归矩阵 R_i , 获得形状更新的增量 $R_i \Phi(I^i, X_i^i)$, 往往经过 4、5 次迭代便可精确定位人脸关键点。

3 人脸配准的实验结果

3.1 实验数据与定位结果评价

本文使用 300-W 人脸配准比赛^[22] 的标注数据进行训练和测试, 其中包括 AFW(337)、iBug(135)、XM2VTS(2360)、LFPW(811+224)、Helen(2000+330)。另外, 我们从网上下载并整理了 2000 张自然场景下的人脸图像, 按照 Multi-PIE 68 个关键点的定义方式^[21] 进行标注, 关键点的定义如图 3 所示。COFW(500+507) 数据库^[16] 局部遮挡比较多, 常被用来测试模型对局部遮挡的鲁棒性, COFW 的标注是 29 个关键点, 与 68 个关键点的标注不一致, 所以我们对其所有数据都进行了重新标注。LFPW 测试集的 224 张人脸图像、Helen 测试集的 330 张人脸图像以及 COFW 测试集的 507 张图像均用来进行测试, 其余图像都用于训练。为了进一步扩充数据集, 对训练数据进行了镜像操作, 同时生成对应的标注信息。级联回归模型对形状的初始化比较敏感, 因此对初始化的平均形状进行了扰动以产生更多的训练数据。根据人脸检测初始化的平均形状和每个人脸标注形状之间存在的旋转、平移、缩放差异, 我们估计其多元高斯分布的参数, 并用以产生不同的初始化。根据平均脸与每个人脸形状归一化后的平均脸的差异, 同样估计其多元高斯分布的参数, 并用以产生不同的初始化。对于以上两种产生不同初始化的方式, 前者模拟了人脸变化的刚性形变, 后者模拟了人脸变化的非刚性形变, 但是, 这两种形变方式的期望都是人脸的平均形状。我们对每幅训练集图像产生 10 个形状扰动, 因此, 最终获得了 162860 (8143 * 2 * 10) 个训练数据。

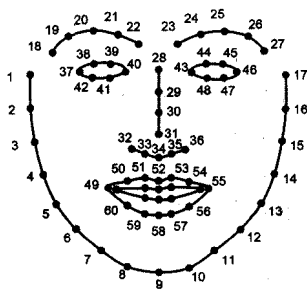


图3 关键点的标定信息

在人脸关键点定位的精度评价方面,关键点的平均定位误差定义为:

$$\bar{\epsilon} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^M |p_{ij} - g_{ij}|_2}{|l_i - r_i|_2}$$

其中, N 为测试集人脸图像的数目, M 为每幅人脸图像人脸关键点的数目, $|p_{ij} - g_{ij}|_2$ 是每幅图像每个关键点预测位置与标定位置的欧氏距离, $|l_i - r_i|_2$ 是每幅人脸图像左右眼角之间的欧氏距离,如图3所示,即第37与第46个关键点之间的距离。对于每幅测试图像而言,所有关键点的平均误差在0.1以上就表示配准失败。除了给出平均定位误差,还给出了累积误差曲线,以便更直观地反映误差的分布情况。在定位速度方面,仅仅统计了从给出人脸检测框到人脸配准结束的时间。在Android平台上,采用JAVA的JNI技术调用C++代码实现人脸配准,因此在PC平台的精度测试与实际Android平台定位精度一致,为方便起见,在PC平台上对配准精度进行测试评价。

本文设计了4组实验:Gabor、LBP、SURF、SIFT、HOG局部特征描述子性能比较;LFPW和Helen数据库定位精度比较;COFW和“情绪拼图”配准分析;Android平台人脸关键点定位速度和空间复杂度分析。前3个实验在计算机(i7-2600 CPU 3.4GHz,内存16GB)上实现,最后的实验在Android平台(三星Note3)上实现。

3.2 PC平台测试

3.2.1 局部描述子性能分析

在实验过程中,分别采用Gabor、LBP、SURF、SIFT、HOG局部特征描述子进行比较。实验训练使用Helen数据库的2000张人脸图像,测试使用Helen数据库的330张人脸图像,训练采用文献[14]的SDM方法,迭代4次,各种局部特征的性能比较如表1所列。

表1 局部描述子性能分析

局部特征描述	精度(*0.01)	速度(ms)
Gabor	7.2	32.7
LBP	6.5	1.6
SURF	5.8	2.4
SIFT	3.9	10.9
HOG	4.2	15.3

从定位精度来看,SIFT和HOG特征定位精度较高,SUFT特征稍微差一些,Gabor和LBP性能一般,因为人脸关键点分布的位置往往在人脸器官的边缘,包含很强的梯度信息,所以SIFT和HOG的梯度方向直方图特征比较适合。从定位速度来看,LBP和SURF明显较快,SIFT和HOG特征定位速度也能达到实时的水平,Gabor特征提取速度较慢。权衡精度和速度的表现,最终选择了SIFT特征描述方式。

3.2.2 LFPW和Helen上的实验结果

LFPW和Helen都是自然场景下的人脸数据库,其中姿态、光照、表情变化比较大,并且存在一定程度的遮挡。本文方法主要与SDM^[14]进行比较,我们实现了4层级联回归的SDM模型,它与文献[14]训练的模型相比,性能上略差。

(1)LFPW

LFPW的测试结果如图4所示,提出的方法比SDM略好一些,主要原因是图像的旋正使得复杂姿态问题简化,分模型训练使得侧脸定位效果更好,另外稀疏的约束提取了鲁棒的特征,对局部遮挡适应性更强。SDM的平均误差为0.0393,原文的结果为0.0349。图5显示了每个关键点的定位误差,每个关键点上圆的半径代表平均定位误差的大小,由图可知,眼睛上的关键点定位比较准确,而人脸外轮廓和眉毛上的关键点定位误差较大。表2给出了各种级联回归方法在LFPW上的平均定位误差,提出的方法仅次于LBF,其他方法的平均定位误差来自文献[17]。

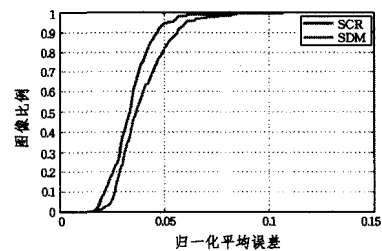


图4 LFPW累积误差曲线

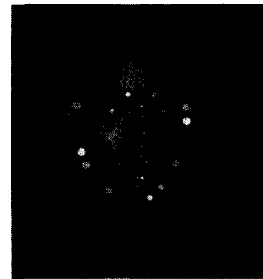


图5 LFPW测试集的归一化配准误差

表2 LFPW测试集归一化平均误差(*0.01)

ESR ^[13]	SDM ^[14]	RCPR ^[16]	LBF ^[17]	SCR
3.47	3.93	3.50	3.35	3.43

(2)Helen

我们在Helen数据库上做了同样的测试,具体比较结果见图6,每个关键点的平均定位误差见图7,表3给出了各种级联回归方法在Helen上的平均定位误差。结果表明提出的方法误差最小,其他方法的平均定位误差来自文献[17]。对比LFPW和Helen的定位结果可以发现,提出的配准方法定位误差比较稳定,而其他方法在Helen数据库上的结果明显比在LFPW上差些,这主要是因为Helen数据库的姿态变化更加复杂,遮挡也相对严重一些。由此可见,本文提出的方法由于分姿态选择模型和引入了稀疏约束,可以有效地缓解姿态问题和遮挡问题。提出的算法通过一次训练,在多个数据库上进行测试,定位结果能够保持稳定。由此可见,该方法在数据库间可以很好地迁移,相对于最小二乘的回归方式,本文提出的方法有效抑制了在特定数据库上的过拟合倾向。

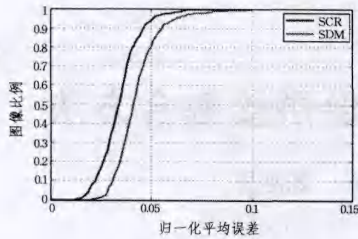


图6 Helen测试集累积误差曲线



图7 Helen测试集的归一化配准误差

表3 Helen测试集归一化平均误差(*0.01)

ESR ^[13]	SDM ^[14]	RCPR ^[16]	LBF ^[17]	SCR
5.7	4.27	6.5	5.41	3.41

3.2.3 COFW和“情绪拼图”上的实验结果

COFW^[16]是遮挡数据库,标注了29个关键点,RCPR^[16]的平均定位误差为0.0580,SCR的平均定位误差为0.0543,SDM的平均定位误差为0.1176(统计了与68个关键点定义一致的21个关键点的平均定位误差),具体结果见图8。

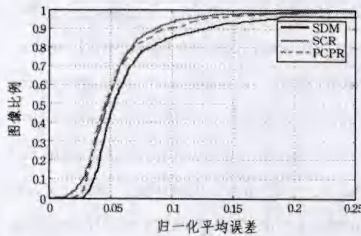


图8 COFW测试集累积误差曲线



图9 COFW测试集上的配准结果对比

图9展示了SCR、ESR^[13]、SDM^[14]、RCPR^[16]、LBF-fast^[17]、LBF^[17]在示例图片上的定位结果,从结果来看,SCR能够很好地应对遮挡和姿态变化,说明稀疏约束能够筛选对遮挡鲁棒的特征,独立的姿态模型能够更好地适应姿态变化。

“情绪拼图”是摄影师捕捉的人脸夸张表情序列,在此定性测试本文算法对于夸张表情的鲁棒性,实验结果如图10所示。



图10 “情绪拼图”上的人脸配准结果

3.3 Android平台测试

在Android平台上对SCR算法进行了测试,测试的硬件为三星Note3智能手机。实验结果如图11所示,在人脸检测的基础上,每帧的人脸关键点定位时间大约是10ms,完全达到了实时的目标,在人脸跟踪的情况下,上一帧的定位结果作为下一帧的初始化,速度可以达到25帧/秒。如图12所示,4层SDM模型大小约为8.45MB,而4层SCR模型大小仅仅只有0.56MB。SCR的整个apk文件仅为4MB,其中还包括1MB左右的人脸检测模型。

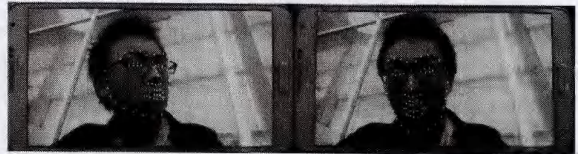


图11 Android平台人脸配准结果

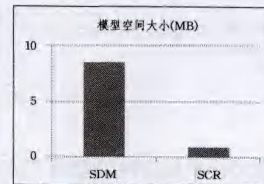


图12 回归矩阵的存储空间比较

结束语 本文主要研究了基于移动平台的快速人脸配准问题,提出了一种基于稀疏约束的级联回归模型。首先在人脸检测的基础上利用二值特征快速定位眼角、嘴角和鼻尖,并进行简单校验修正;接着估计人脸的姿态,旋正并缩放人脸图像,选择对应的侧脸或正脸模型进行基于稀疏约束的级联回归配准。其中,眼角、嘴角和鼻尖的定位引入了二值近似的方法,将点积浮点运算巧妙地转变成二进制位运算,加快了计算速度。级联回归的模型分姿态进行训练提高了模型对姿态变化的鲁棒性。回归器的训练通过引入稀疏约束,筛选了鲁棒

(下转第310页)

- [11] 孙华燕, 周道炳, 李生良. 一种序列图像的拼接方法[J]. 光学精密工程, 2000, 8(1): 35-39
Sun Hua-yan, Zhou Dao-bing, Li Sheng-liang. Method for sequence image coalescence[J]. Optics and Precision Engineering, 2000, 8(1): 35-39
- [12] 李洋洋, 赵刚, 刘利. 多摄像机图像拼接视觉归一化技术研究[J]. 计算机工程与应用, 2013, 49(20): 188-193
Li Yang-yang, Zhao Gang, Liu Li. Research on visual normalization of multi-camera image stitching[J]. Computer Engineering and Applications, 2013, 49(20): 188-193
- [13] Papadimitriou C H, Steiglitz K. Combinatorial Optimization; Al-

- [14] Yazdi S M S, Savari S A. A Max-Flow/Min-Cut Algorithm for Linear Deterministic Relay Networks[J]. IEEE Transactions on Information Theory, 2011, 57(5): 3005-3015
- [15] 刘坤, 郭雷, 陈敬松. 基于区域分割的序列红外图像融合算法[J]. 红外与激光工程, 2009, 38(3): 553-558
Liu Kun, Guo Lei, Chen Jing-song. Sequence infrared image fusion algorithm using region segmentation[J]. Infrared and Laser Engineering, 2009, 38(3): 553-558
- [16] Fang X Y, Pan Z G, Luo B, et al. Robust image mosaic with RANSAC and bundle adjustment[J]. Journal of Computational Information Systems, 2008, 4(4): 1613-1619

(上接第 305 页)

的特征, 增强了模型对人脸局部遮挡的鲁棒性, 同时可以将模型的大小压缩到原来的 5% 左右。本文提出的模型在 LF-PW、Helen 和 COFW 等自然场景下的人脸数据库上取得了很好的配准结果, 具有速度快、精度高、模型小的特点。在三星 Note3 智能手机上的测试结果表明, 每幅人脸图像的配准时间大约为 10ms, 满足实时性的要求。

目前, 本文算法对于极度夸张的表情、严重遮挡等情况的适应性还有待提高。在今后的工作中, 我们将尝试融合人脸检测和关键点定位的算法, 以进一步提高定位速度和精度, 增强模型的鲁棒性。

参 考 文 献

- [1] Hua G, Fu Y, Turk M, et al. Introduction to the special issue on mobile vision[J]. International Journal of Computer Vision, 2012, 96(3): 277-279
- [2] Shan Shi-guang, Chang Yi-zheng, Gao Wen, et al. Curse of misalignment in face recognition: problem and a novel mis-alignment learning solution[C]//Proceedings of the Automatic Face and Gesture Recognition, 2004. Seoul, Korea, 2004: 314-320
- [3] Cootes T F, Taylor C J. Active shape models—‘smart snakes’[C]//Proceedings of the British Machine Vision Conference, 1992. London, England, 1992: 266-275
- [4] Cootes T F, Taylor C J, Cooper D H, et al. Active shape models—their training and application[J]. Computer vision and image understanding, 1995, 61(1): 38-59
- [5] Cootes T F, Edwards G J, et al. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685
- [6] Matthews I, Baker S. Active appearance models revisited[J]. International Journal of Computer Vision, 2004, 60(2): 135-164
- [7] Cristinacce D, Cootes T. Automatic feature localisation with constrained local models[J]. Pattern Recognition, 2008, 41(10): 3054-3067
- [8] Zhu Xiang-xin, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]//Proceedings of the Computer Vision and Pattern Recognition, 2012. Rhode, America, 2012: 2879-2886
- [9] Zhou Feng, Brandt J, Zhe Lin. Exemplar-based graph matching for robust facial landmark localization[C]//Proceedings of the International Conference on Computer Vision, 2013. Sydney, Australia, 2013: 1025-1032
- [10] Yi Sun, Wang Xiao-gang, Tang Xiao-ou. Deep convolutional net-

work cascade for facial point detection[C]//Proceedings of the Computer Vision and Pattern Recognition, 2013. Portland, America, 2013: 3476-3483

- [11] Zhou Er-jin, Fan Hao-qiang, Cao Zhi-min, et al. Extensive Facial Landmark Localization with Coarse-to-Fine Convolutional Network Cascade[C]//Workshops of the International Conference on Computer Vision, 2013. Sydney, Australia, 2013: 386-391
- [12] Piotr D, Welinder P, Perona P. Cascaded pose regression[C]//Proceedings of the Computer Vision and Pattern Recognition, 2010. San Francisco, America, 2010: 1078-1085
- [13] Cao Xu-dong, Wei Yi-chen, Wen Fang, et al. Face alignment by explicit shape regression[C]//Proceedings of the Computer Vision and Pattern Recognition, 2012. Rhode, America, 2012: 2887-2894
- [14] Xiong Xue-han, De T F. Supervised descent method and its applications to face alignment[C]//Proceedings of the Computer Vision and Pattern Recognition, 2013. Portland, America, 2013: 532-539
- [15] Yan Jun-jie, Zhen Lei, Dong Yi, et al. Learn to Combine Multiple Hypotheses for Accurate Face Alignment[C]//Workshops of the International Conference on Computer Vision, 2013. Sydney, Australia, 2013: 392-396
- [16] Burgos A, Xavier P, Perona P, et al. Robust face landmark estimation under occlusion[C]//Proceedings of the International Conference on Computer Vision. Sydney, Australia, 2013: 1513-1530
- [17] Ren Shao-qing, Cao Xu-dong, Wei Yi-chen, et al. Face Alignment at 3000 FPS via Regressing Local Binary Features[C]//Proceedings of the Computer Vision and Pattern Recognition. Columbus, America, 2014: 1685-1692
- [18] Hare S, Saffari A, Torr P H S. Efficient online structured output learning for keypoint-based object tracking[C]//Proceedings of the Computer Vision and Pattern Recognition. Rhode, America, 2012: 2887-2894
- [19] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society, 1994, 58(1): 267-288
- [20] Heikkila M, Pietikainen M, Schmid. Description of interest regions with local binary patterns[J]. Pattern Recognition, 2009, 42(3): 425-436
- [21] Gross R, Matthews I, Cohn J, et al. Multi-pie[J]. Image and Vision Computing, 2010, 28(5): 807-813
- [22] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge[C]//Workshops of the International Conference on Computer Vision. Sydney, Australia, 2013: 397-403