

基于层进模型的快速人脸配准方法及其在智能移动设备上的应用

邓健康 王灿田 刘青山

(南京信息工程大学信息与控制学院 南京 210044)

摘要 主要研究了移动智能手机上人脸关键点的快速定位问题。在活动形状模型的基础上,提出了一种基于层进模型的快速人脸配准方法:首先,在人脸检测的结果上,采用二值特征快速定位眼角、嘴角等关键点,并对其在校准修正;然后,通过眼角和嘴角的关键点,并结合边缘约束,对眼睛、嘴巴和人脸外轮廓进行局部配准;最后,对整个人脸形状进行基于加权投影的形状配准。实验结果表明,提出的方法在8~10次迭代后即可收敛,在三星 I9300 智能手机上,每幅人脸图像的配准时间在40ms以下,满足实时性要求。

关键词 移动平台,人脸关键点,快速定位

中图分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.12.064

Fast Face Alignment Method Based on Hierarchical Model and its Application on Mobile Device

DENG Jian-kang WANG Can-tian LIU Qing-shan

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

Abstract This paper studied the problem of fast localization of facial landmark on a mobile smart phone. Fast face alignment method based on hierarchical model was proposed, which is derived from active shape model. Firstly, the landmarks on the canthi and angulus oris are quickly located using binary feature based on the result of face detection. After that, calibration and correction are performed on these landmarks. Secondly, combined with edge constraint on the eyes, mouth and face outline, local alignments are performed independently based on the location of landmarks on the canthi and angulus oris. Finally, the shape of the whole face is aligned by weighted projection. Experimental results show that the proposed method will converge after 8 to 10 iterations. It takes 40ms or less to complete face alignment on a single face image on a smartphone(the Samsung I9300), which satisfies the request of real time.

Keywords Mobile platform, Facial landmark, Real-time localization

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研究领域与工程领域的热点。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多的机会^[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,如何在计算能力有限和存储能力有限的移动平台上快速稳定地实现人脸关键点的定位成了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,为进一步的人脸图像的处理与分析提供重要的形状信息,是人脸识别、人脸验证、人脸动画、人脸姿态表情分析、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(Curse of Mis-alignment)”^[2]。近年来,国内外学者提出

了许多种人脸配准的方法,代表性的工作有主动形状模型(Active Shape Model, ASM)^[3,4]、贝叶斯切空间形状模型(Bayesian Tangent Shape Model, BTSM)^[5]、主动外观模型(Active Appearance Model, AAM)^[6,7]、受约束的局部模型(Constrained Local Model, CLM)^[8]、可变形部件模型(Deformable Part Models, DPM)^[9]等。

在上述方法中,T. F. Coots等人提出的ASM模型^[3,4]和AAM模型^[6,7],由于算法定位精度较高,备受研究者的关注。ASM和AAM都是基于点分布模型(Point Distribution Model, PDM)的算法。ASM的主要思想是通过若干关键点的坐标来表示人脸的形状信息,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二者融合的统计约束,在考虑形状变化的同时,采用PCA对纹理变化进行建模。ASM算法搜索定位的速度较快,单个特征点的定位精度较高,但是忽略了全局纹理信息。AAM算法由于同时考虑了形状和纹理信息,通常情况下,视

到稿日期:2013-06-25 返修日期:2013-08-16 本文受国家自然科学基金项目(61272223),江苏省杰出青年基金(SBK201210296),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、人脸配准,E-mail:deng_jiankang@126.com;王灿田(1987-),男,硕士生,主要研究方向为基于图/超图模型的图像分析;刘青山(1975-),男,教授,博士生导师,主要研究方向为人脸图像分析、基于图/超图模型的视觉理解、医学图像分析、基于事件的视频分析、多光谱图像分析等。

觉效果要稍好于 ASM,但是其计算复杂度明显高于 ASM。

针对经典 ASM 模型存在的一些实际问题,在全局形状模型、局部纹理模型、搜索匹配策略方面出现了很多改进的方法。针对训练集中样本数不足的问题,Cootes 等人在 ASM 模型中引入了有限元分析^[10]的思想;针对训练集中可能有错误样本的问题, Mike 和 Jim 将稳健分析应用于 ASM 模型中^[11],采用了两种稳健参数估计技术: M. estimators 和 Random Sampling;为了处理形状的非线性变化,Cootes 等人提出了混合高斯模型 ASM 算法^[12];为了适应人脸姿态变化, Pomdiani 等人提出用 Kernel PCA 方法建立基于多视角非线性的 ASM 模型^[13];为了适应光照和表情变化, Huang 等人提出了基于部件框架的人脸配准方法^[14];为了改进图像搜索匹配过程, Shuicheng Yan 提出了 TC-ASM^[15];基于小波变换建立局部纹理模型, Feng 等人提出了 W. ASM/Gabor ASM^[16], Fei 等人提出 Haar-wavelet ASM^[17];基于分类器建立局部纹理模型,研究者提出了 K-NN ASM、AdaBoosting ASM、Fisher-Boost ASM。这些改进的 ASM 算法在鲁棒性和定位精度方面都有不同程度的提升,但同时带来了计算复杂度增加的问题,需要一定内存空间和计算复杂度的支持,所以对计算能力和存储能力有限的移动设备来说,基本很难直接进行移植应用。

为了能在移动智能设备上实现实时人脸配准,本文基于 ASM 框架,提出了一种新的基于层进模型的快速人脸配准方法,其能在 Andriod 移动平台(三星 I9300)上达到实时性要求(<40ms/帧)。该方法在人脸检测结果的基础上,快速定位眼角、嘴角,接着定位眼睛、嘴巴和人脸外轮廓的关键点,这样的策略使得所有关键点的初始化相对准确,最后通过加权投影的方式进行关键点迭代定位。实验结果表明,本文方法在不失精度的同时,算法的迭代次数明显下降,效率明显提升。

2 层进定位模型

为了在移动平台上实现人脸关键点的快速定位,我们采用了层级的定位机制。实验研究表明,在使用 ASM 进行人脸关键点的定位过程中,眼角、嘴角明显比其他关键点的定位更鲁棒、更精确,所以我们对眼角、嘴角单独提取特征,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,加速了计算^[18]。在眼角和嘴角关键点快速定位之后,进行眼睛、嘴巴和人脸外轮廓的局部形状定位,初始位置通过角点信息估计,关键点搜索匹配的过程中引用边缘约束策略,这种基于局部器官的配准方式在一定程度上可以增强对局部器官形变的适应性。最后,根据已有信息初始化其他所有关键点,并进行迭代定位,形状平滑采用加权投影的方式^[19],利用迭代过程中的信息尽可能使得结果跳出局部极值,增强模型的鲁棒性。整个算法的示意图和流程如图 1、图 2 所示。

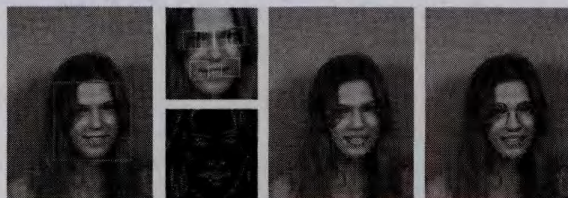


图 1 关键点层定位算法示意图

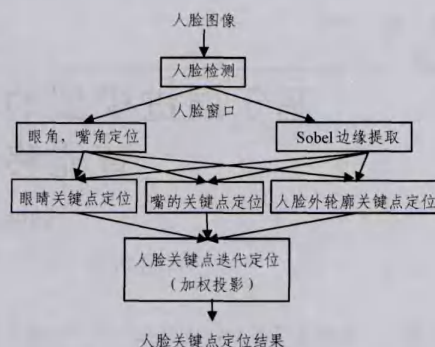


图 2 关键点层定位算法流程

2.1 关键点定位

对于一张输入的人脸图像,首先进行人脸检测^[20],并对检测返回的人脸窗口进行一定的预处理,使其成为大小固定的灰度人脸图像。在移动平台上进行快速人脸检测,可以结合肤色特征排除一部分候选人脸的区域,并且搜索窗口的移动策略是从中间向四周辐射,搜索窗口大小的初值也可以相应调整,结合具体应用决定是否在检测到一张人脸后就立刻返回检测结果。

接着在人脸检测结果的窗口上快速定位面部的角点。我们选择了 4 层 LBP 金字塔特征^[21]来描述眼角与嘴角,统计了训练集中角点在人脸检测窗口中出现的位置,得到其外接矩形区域(该区域也是角点匹配时的搜索区域),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 3 所示。

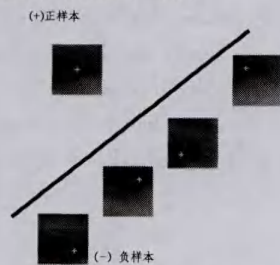


图 3 SVM 训练眼角特征

为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。

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

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 $w_j = w_j - \beta_i b_i$,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,所以有:

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

这样就把点积浮点运算变成了二进制位运算,其可以通过“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。

对于角点间的全局位置约束,我们引入校验修正的机制增强定位的鲁棒性,解决个别角点存在遮挡的问题。与ASM形状平滑不同的是,个别点的匹配误差不会平均到其他所有点上去,也就是说,宁可6个点当中有个别点偏了,也不要影响其他定位点的精度。这样的方法是合理的:1)角点定位建立在人脸检测或跟踪的基础之上,位置不会偏得太离谱;2)角点定位的后续步骤是面部器官的独立配准,也就是说角点定位的作用是使得人眼和嘴巴配准时初始化更精确,以减少迭代次数;3)人脸所有的关键点配准采用的是加权投影的方式,偏离的个别点可以得到较快的修正。

对训练数据集人脸图像的眼角、嘴角6个关键点位置进行普氏分析(Procrustes analysis),然后进行主成分分析(Principal Component Analysis, PCA),得出角点的主要变化模式,此处使用PCA主要是考虑到正交空间投影计算比较方便。当通过局部特征定位了眼角和嘴角位置后,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,然后投影到PCA空间,得到平滑后的位置,分别统计各个点偏差,相应地进行校验修正。

$$\Delta p_i = \frac{|p_i - p_i^*|}{\sum_{i=1}^6 |p_i - p_i^*|} \times 100\% \quad (3)$$

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角点的时候,修正的策略为直接用影到PCA空间中形状的对应点替换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。这与ASM的迭代搜索定位的策略是不一样的,这样不会使个别点的误差平均到所有点,由于不存在迭代的过程,处理速度也非常快。当人脸的6个角点中有一两个点被遮挡或者人的姿态变化将眼角自遮挡时,可以通过其他点的信息加以修正。

2.2 局部器官定位

在眼角、嘴角快速定位的基础之上,可以进行人脸姿态的初步分析,在人机交互的游戏中可以用人脸姿态进行游戏角色控制。眼角与嘴角定位之后,如图4所示,我们先把眼睛和嘴巴的关键点配准好,它们之间的配准是独立的,可以同时进行。其原理与ASM的类似,只不过训练的形状模型是人脸局部器官的形状模型。提前定位眼睛和嘴的主要原因是:1)移动平台上的一些应用,例如目光估计、唇动分析等可以在这个程序框架下进行扩展,得到提前处理;2)人脸的变化比较复杂,全局统计模型例如PCA、Kernel PCA或者混合高斯模型^[12]也很难给人脸精确配准,基于局部人脸器官的高斯形状分布模型可以增强对局部器官形变的适应性^[14];3)在后续的关键点迭代定位的过程中采用加权投影可以使得匹配较好的点稳定性较高,受到其他误匹配的点影响较少,所以眼睛和嘴上的关键点定位会减少后面的迭代次数。

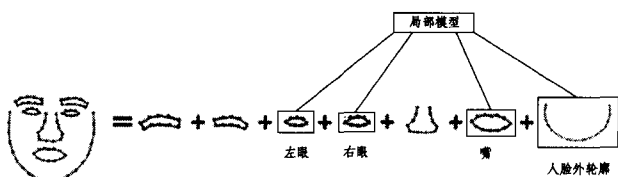


图4 局部器官形状配准

另外,考虑到人脸图像中许多关键点处于边缘信息较强的位置上,如人眼、嘴巴以及人脸外轮廓上的关键点,可以在

这些关键点搜索过程中融入边缘信息。首先对人脸窗口进行索贝尔边缘提取,然后利用边缘值对最佳匹配的距离进行改进。

$$d = (325 - k)(g_i' - \bar{g}_i)^T C_i^{-1} (g_i' - \bar{g}_i) \quad (4)$$

其中, k 为索贝尔算子得到的边缘值。

2.3 基于加权投影的人脸配准

经过上面的人脸角点快速定位以及眼睛、嘴巴与人脸外轮廓上关键点的局部配准的操作后,人脸的关键点便有了相对精确的初始化。关键点的局部搜索仍然采用垂直于轮廓的剖面信息,主要是考虑了这种局部模型的实现速度。对于迭代定位过程中的形状平滑,采用了加权的投影方式,融入了搜索过程中的信息。

在ASM中,形状子空间模型表示为:

$$x' = \bar{x} + \phi_i b_i \quad (5)$$

对于任意形状 x , 在ASM形状子空间中找到一个与 x 相似的 x' 。因为形状空间是正交分解的,所以 x' 是点 x 在形状子空间的正交投影,则点 x' 在形状子空间中的参数 b_i 为:

$$b_i = \phi_i^T (x - \bar{x}) \quad (6)$$

ASM用正交投影对搜索中的中间形状进行校正,对形状中的所有点一视同仁,于是个别点的误差会平均到所有点上,从而增加了迭代次数,降低了收敛速度。其实,搜索过程中的信息如某点与目标点的相似程度,或者某点在迭代过程中稳定性描述等信息,可以帮助我们解决上述问题,对于正交投影,通过引入搜索过程中的评价信息形成加权投影。

ASM中的形状 x' 是形状 x 在解空间中欧氏距离最近的形状,也就是最小化

$$distance = (x - \bar{x})^T (x - \bar{x}) \quad (7)$$

我们采用加权投影改进ASM的正交投影,最小化 $distance = (x - \bar{x})^T w (x - \bar{x})$, 其中

$$w = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{2N} \end{bmatrix} \quad (8)$$

$w_{2i} = w_{2i+1} = \frac{1}{f_{\min}^i(t) + \alpha d_{f_i}^i + \beta d_{p_i}^i}$ 为当前点和目标点的相似程度的度量,与局部模型有关。

$d_{f_i}^i = |f_{\min}^i(t) - f_{\min}^i(t+1)|$ 为相继两次评价值的稳定性度量。

$d_{p_i}^i = |p_i - p_{i+1}|$ 为相继两次位置的稳定性度量。

α, β 分别表示 $d_{f_i}^i$ 和 $d_{p_i}^i$ 的影响程度。

为了尽量避免搜索中的局部极值,定义了权重的敏感度,设置权重的上限。

$$1 = \max \left(\frac{|w_{2i} - w_{2(i-1)}|}{w_{2i}}, \frac{|w_{2i} - w_{2(i+1)}|}{w_{2i}} \right) \quad (9)$$

如果计算得到的权重过大,则将相邻两点的权重均值赋予给它,以免权重过大带来噪声。眼睛、嘴、鼻子、外轮廓空间分布是不连续的,所以对于不同器官需要进行不同处理。

接下来,在解空间中求使得 $distance = (x - \bar{x})^T w (x - \bar{x})$ 最小的点。

将 $x' = \bar{x} + \phi_i b_i$ 代入得:

$$distance = (x - \bar{x})^T w (x - \bar{x}) = \sum_{i=1}^{2N} w_i (x_i - \bar{x}_i - \sum_{j=1}^i \phi_{ij} b_j)^2 \quad (10)$$

对 b 求导,并令其等于零,得到:

$$\frac{\partial distance}{\partial b_k} = \frac{\sum_{i=1}^{2N} \frac{\partial [w_i(x_i - \bar{x}_i - \sum_{j=1}^L \phi_{ij} b_j)^2]}{\partial b_k}}{\sum_{i=1}^{2N} w_i(x_i - \bar{x}_i - \sum_{j=1}^L \phi_{ij} b_j) \phi_{ij}} = 0 \quad (11)$$

即得到方程组:

$$\phi_i^T W \phi_i b_i = \phi_i^T W(x - \bar{x}) \quad (12)$$

所以

$$b_i = (\phi_i^T W \phi_i)^{-1} \phi_i^T W(x - \bar{x}) \quad (13)$$

3 实验结果

3.1 实验数据与定位结果评价

本文使用了 MUCT^[22] 人脸数据库进行模型的训练和测试。MUCT 数据集的人脸图像包括 276 人的 3755 张图像, 其中女性占 51%, 佩戴眼镜的人数占 18%, Viola Jones 人脸检测失败率为 1.8%。每张人脸图像有 76 个标定定点, 而我们仅采用了 61 个标定定点, 舍去了嘴唇内的 7 个点, 仅使用了嘴唇外轮廓的关键点, 同时眼睛轮廓上的关键点也仅保留了 8 个。图 5 显示了人脸的标定结果。我们对数据集进行了划分, 将 MUCT 人脸库中 60% 作为训练集, 40% 作为测试集, 如表 1 所列。



(a) MUCT 人脸数据集的标定定点 (b) 实验采用的标定定点

图 5

表 1 数据集划分

数据集	训练	测试
百分比	60%	40%
图像数	2253	1502

我们设计了 3 组实验: 关键点特征定位性能分析, 层进迭代算法与 ASM 算法在精度一致的情况下迭代次数与迭代时间的比较, Android 平台人脸关键点定位速度统计。前两个实验在计算机上实现 (i7-2600 CPU 3.4GHz, 内存 16G), 后面的实验在 Android 平台 (三星 I9300) 上实现。

在人脸关键点定位的精度方面, 我们统计了人脸关键点定位结果与实际手工标定结果的偏差, 关键点定位的平均偏差定义为:

$$\epsilon = \frac{1}{M} \sum_{i=1}^M \sqrt{(p_{ix} - p_{ix}^*)^2 + (p_{iy} - p_{iy}^*)^2} \quad (14)$$

对于人脸关键点定位结果的评价, 研究者们对此有着不同的评价标准, 我们采用了 Rowley 等使用的标准: 与手工标定位置比较, 误差在 4 个像素以内^[23]。在定位速度方面, 我们的算法未加入跟踪的方法, 没有利用相邻帧之间的运动连续性统计从人脸检测给出人脸窗口开始到关键点迭代定位收敛所花的时间和迭代次数。在 Android 平台上, 我们采用 JAVA 的 JNI 技术调用 C++ 代码实现人脸关键点定位方法, 定位的过程也没有引入跟踪的方法。另外, 代码没有做专门的优化, 这意味在某些应用中算法的效率还可以进一步提升。

3.2 PC 平台测试

3.2.1 关键点特征定位性能分析

在实验的过程中, 分别采用 4 层 LBP 的金字塔特征、Brief 特征^[24] 进行实验, 并权衡速度和精度表现选择角点最终的特征提取方式, 模型均采用了二值近似, 将点积浮点运算变成了二进制位运算, 局部定位结束后进行校验修正。另外, 使用 Brief 特征需要对图像做平滑, 对人脸检测返回的结果进行高斯滤波, 在图像纵横方向采用的高斯核 $K=(1/4 \ 1/2 \ 1/4)$, 由于都是 2 的幂次, 可以采用移位的方式快速计算。从表 2、表 3 列出的实验结果中可以看到, Brief 特征虽然速度很快, 这种二值描述子从理论上而言非常适合在移动平台上运行, 但是定位精度不及 4 层 LBP 金字塔特征。所以, 最终选用了 4 层的 LBP 金字塔特征。表 4 统计所有测试图像 6 个角点定位偏差的范围。在计算机上将点积浮点运算变成二进制位运算, 提速不是很明显, 原因是 PC 机本身非常擅长浮点运算, 但对于不擅长浮点计算的移动平台而言, “与”门运算明显比点积浮点计算要快很多。图 6 是在计算机上测试角点定位的实例图, 实验结果表明, 我们的方法对姿态变化、表情变化、轻度遮挡、佩戴眼镜等情况都有一定的适应性。

表 2 关键点定位速度 (不包括人脸检测的时间)

局部特征提取方式	时间 (ms)
4 layer LBP	3.4
Brief	2.1

表 3 关键点定位精度对比

偏差	<3 像素	<6 像素	<9 像素	<12 像素
4 layer LBP Pyramid	35.50%	63.7%	83.40%	95.30%
Brief	20.30%	42.6%	73.20%	89.40%

表 4 关键点定位精度统计

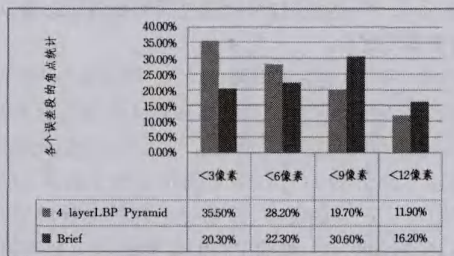


图 6 关键点定位示意图

3.2.2 人脸配准性能分析

基于眼角与嘴角定位的结果, 并结合边缘约束, 来对眼睛、嘴巴和人脸外轮廓上的关键点进行定位, 然后在此基础上初始化整个人脸形状, 进行基于加权投影的迭代搜索直到

收敛。算法实现过程如图 7 所示。



图 7 基于层进定位的人脸配准结果

为了评价所提方法的有效性,将其与 ASM 进行精度和速度的比较。本文方法称为层次加权活动形状模型(Hierarchical Weighted Active Shape Model, HWASM), ASM 的实现算法来自文献[25],它们均在相同的人脸检测结果上进行人脸配准,实验设置以及算法迭代停止的条件都是相同的,均设迭代上限为 30 次。

表 5 统计了测试集上各个算法的平均定位时间与收敛的迭代次数,可以看出,HWASM 在速度上有明显的优势。对于 HWASM 和 ASM 迭代次数的比较,更合理的评价应该是包括眼睛、嘴巴和人脸外轮廓迭代定位的次数,由于三者独立并行定位,迭代次数选取三者中最大的来表示。实验表明,迭代定位次数最大的部位往往是嘴巴,为了不影响后面程序的执行,嘴唇上关键点迭代次数上限定为 4,在这样的情况下,人脸局部器官的迭代定位平均次数是 3.7。表 5 统计的迭代次数仅仅是人脸形状初始化后的迭代次数。

表 5 ASM 与 HWASM 人脸关键点定位方法的性能对比

关键点定位方法	时间(ms)	平均迭代次数
ASM	51.4	21.2
HWASM	23.8	6.4

表 6 比较了测试集上 HWASM 与 ASM 的定位精度,结果表明 HWASM 比 ASM 的精度略高一些。按照 Rowley 等使用的标准^[23],ASM 算法在测试集上平均误差在 4 个像素以内的占 69.5%,而 HWASM 的平均误差在 4 个像素以内的为 73.4%。这主要得益于我们将边缘信息引入局部器官的配准过程当中,眼睛、嘴巴以及人脸外轮廓上关键点定位非常准确。平均定位误差大于 8 像素的测试图像往往就是最终没能收敛的情况,搜索陷入了局部极值,相对 ASM 而言,HWASM 中的形状平滑采用了加权投影,收敛效果好一些。

表 6 ASM 与 HWASM 人脸关键点定位方法的精度比较

定位方法	平均误差			
	<2 像素	<4 像素	<6 像素	<8 像素
ASM	29.7%	69.5%	79.7%	89.5%
HWASM	33.8%	73.4%	85.9%	92.4%

3.3 Andriod 平台测试

在 Andriod 平台上,我们也对算法进行了实验测试,包括对单帧图像、摄像头实时数据流和拍摄好的视频流进行测试,详见图 8。测试的硬件为三星 I9300 智能手机。实验结果显示,在三星 I9300 上,每帧的人脸关键点定位时间都小于 40ms,完全达到了实时的目标。作为对比实验,我们同时也在移动平台上实现了 ASM 算法。为了避免搜索陷入局部极值并且加快搜索速度,我们对人脸窗口建立了 3 层图像金字塔来进行多分辨率分析。HWASM 由于采用了层进定位,迭代的次数明显减少,整体的处理时间比经典 ASM 算法少很多,比较结果见表 7。

表 7 移动平台上 ASM 与 HWASM 在测试图像上的处理速度对比

关键点定位的方法	时间(ms)
ASM	110.6
HWASM	39.7



图 8 Android 平台人脸配准结果

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,提出了一种基于层级的快速定位机制。首先在人脸检测的基础上快速定位眼角和嘴角,并进行简单校验修正,接着配准眼睛、嘴巴以及人脸外轮廓上的关键点,在此基础上迭代定位所有关键点。其中,眼角和嘴角的定位引入了二值近似的方法,将点积浮点运算变成了二进制位运算,加快了计算速度。局部器官配准时引入了边缘约束,提高了搜索的精度。关键点的形状平滑采用加权投影的方式,利用搜索过程中的信息,尽可能地跳出局部极值。在三星 I9300 手机上的测试结果表明,每幅人脸图像的处理时间在 40ms 以下,满足实时性要求。

目前,我们的算法对于侧脸、严重遮挡等情况的适应性还有待提高。在今后的工作中,我们将尝试在线更新模型的状态基,通过不断融入用户的人脸形状,从普适的模型逐步过渡到用户特定的模型,从而减少形状参数,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 S, Chang Y, Gao W, et al. Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution [C]//Sixth IEEE International Conference on Automatic Face and Gesture Recognition. Proceedings. IEEE, 2004:314-320
- [3] Cootes T F, Taylor C J. Active shape models—'smart snakes' [M]//BMVC92. Springer London, 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] Zhou Y, Gu L, Zhang H J. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. IEEE, vol. 1, 2003:109-116
- [6] Cootes T F, Edwards G J, Taylor C J. Active appearance models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6):681-685
- [7] Matthews I, Baker S. Active appearance models revisited[J]. International Journal of Computer Vision, 2004, 60(2):135-164

- [8] Cristinacce D, Cootes T. Automatic feature localisation with constrained local models[J]. *Pattern Recognition*, 2008, 41(10): 3054-3067
- [9] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition[J]. *International Journal of Computer Vision*, 2005, 61(1): 55-79
- [10] Cootes T F, Taylor C J. Combining point distribution models with shape models based on finite element analysis[J]. *Image and Vision Computing*, 1995, 13(5): 403-409
- [11] Rogers M, Graham J. Robust active shape model search[M]// *Computer Vision—ECCV 2002*. Springer Berlin Heidelberg, 2006: 517-530
- [12] Cootes T F, Taylor C J. A mixture model for representing shape variation[J]. *Image and Vision Computing*, 1999, 17(8): 567-573
- [13] Romdhani S, Gong S, Psarrour A. A Multi-View Nonlinear Active Shape Model Using Kernel PCA[C]// *BMVC*. 1999: 483-492
- [14] Huang Y, Liu Q, Metaxas D N. A component-based framework for generalized face alignment[J]. *Systems, Man, and Cybernetics, Part B: Cybernetics*, *IEEE Transactions on Biometrics Compendium*, *IEEE*, 2011, 41(1): 287-298
- [15] Yan S, Liu C, Li S Z, et al. Face alignment using texture-constrained active shape models[J]. *Image and Vision Computing*, 2003, 21(1): 69-75
- [16] Jiao F, Li S, Shum H Y, et al. Face alignment using statistical models and wavelet features[C]// *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. *IEEE*, vol. 1, 2003: 321-327
- [17] Zuo F. Fast facial feature extraction using a deformable shape model with haar-wavelet based local texture attributes[C]// *2004 International Conference on Image Processing (ICIP'04)*. *IEEE*, 2004, 3: 1425-1428
- [18] Hare S, Saffari A, Torr P H S. Efficient online structured output learning for keypoint-based object tracking [C]// *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. *IEEE*, 2012: 1894-1901
- [19] Zhao M, Li S Z, Chen C, et al. Shape evaluation for weighted active shape models[C]// *Proc. of the Asian Conference on Computer Vision*. 2004: 1074-1079
- [20] Viola P, Jones M J. Robust real-time face detection[J]. *International journal of computer vision*, 2004, 57(2): 137-154
- [21] Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with local binary patterns[J]. *Pattern recognition*, 2009, 42(3): 425-436
- [22] Milborrow S, Morkel J, Nicolls F. The MUCT landmarked face database[J]. *Pattern Recognition Association of South Africa*, 2010, 201
- [23] Rowley H A, Baluja S, Kanade T. Neural network-based face detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(1): 23-38
- [24] Calonder M, Lepetit V, Strecha C, et al. Brief, Binary robust independent elementary features[M]// *Computer Vision-ECCV 2010*. Springer Berlin Heidelberg, 2010: 778-792
- [25] Milborrow S, Nicolls F. Locating facial features with an extended active shape model[M]// *Computer Vision-ECCV 2008*. Springer Berlin Heidelberg, 2008: 504-513
- [26] 韩玉峰, 王小林. 一种基于改进的 ASM 的人脸特征点定位方法[J]. *计算机科学*, 2013, 40(4): 271-274

(上接第 296 页)

- [2] Zhao Feng, Fan Jiu-lun, Liu Han-qiang. Optimal-selection-based suppressed fuzzy c-means clustering algorithm with self-tuning non local spatial information for image segmentation[J]. *Expert Systems with Applications*, 2014, 41(9): 4083-4093
- [3] Wu Peng-fei, Liu Yi-guang, Li Yong-zhong, et al. TRUS image segmentation with non-parametric kernel density estimation shape prior[J]. *Biomedical Signal Processing and Control*, 2013, 8(6): 764-771
- [4] Nguyen T M, Wu Q M J, Mukherjee D, et al. A finite mixture model for detail-preserving image segmentation[J]. *Signal Processing*, 2013, 93(11): 3171-3181
- [5] Nguyen M T, Wu Q M J, Mukherjee D, et al. A finite mixture model for detail-preserving image segmentation[J]. *Signal Processing*, 2013, 93(11): 3171-3181
- [6] Portela N M, Cavalcanti G D C, Ren T I. Semi-supervised clustering for MR brain image segmentation[J]. *Expert Systems with Applications*, 2014, 41(4): 1492-1497
- [7] Wang X, Fang L, Li M. Image segmentation based on adaptive mixture model[J]. *Journal of Optics*, 2013, 15(3): 035407
- [8] Khayati R, Vafadust M, Towhidkhah F, et al. Fully automatic segmentation of multiple sclerosis lesions in brain MR FLAIR images using adaptive mixtures method and Markov random field model[J]. *Computers in Biology and Medicine*, 2008, 38(3): 379-390
- [9] Xie Cong-hua, Song Yu-qing, Liu Zhe. Density-based Clustering Algorithm Using Kernel Density Estimation and Hill-down Strategy[J]. *Journal of Information & Computational Science*, 2010, 7(1): 135-142
- [10] 刘哲, 宋余庆, 陈健美, 等. 基于二类切比雪夫正交多项式非参数混合模型的图像分割[J]. *计算机研究与发展*, 2011, 11(48): 2008-2014
- [11] Zribi M, Ghorbel F. An unsupervised and non-parametric Bayesian classifier[J]. *Pattern Recognition Letters*, 2003(24): 97-112
- [12] Melnykov V, Melnykov I. Initializing the EM algorithm in Gaussian mixture models with an unknown number of components[J]. *Computational Statistics & Data Analysis*, 2012, 56(6): 1381-1395
- [13] BrainWeb; Simulated Brain Database [OL]. <http://www.bic.mni.mcgill.ca/brainweb>
- [14] Zhang Hui, Jason E, Fritts B, et al. GoldmanImage segmentation evaluation: A survey of unsupervised methods[J]. *Computer Vision and Image Understanding*, 2008, 110(2): 260-280
- [15] Nikou C, Galatsanos N, Likas A. A class-adaptive spatially variant mixture model for image segmentation[J]. *IEEE Transactions on Image Processing*, 2007, 16(4): 1121-1130
- [16] 刘哲, 谭振江, 王洪君. 基于规范化的 B 样条密度模型的聚类算法[J]. *吉林大学学报: 信息科学版*, 2014, 31(5): 522-527