

脸部特征定位方法

林维训 潘纲 吴朝晖 潘云鹤

(浙江大学计算机系, 杭州 310027)

摘要 脸部特征定位是人脸分析技术的一个重要组成部分,其目标是在图象或图象序列中的指定区域内搜索人脸特征(如眼、鼻、嘴、耳等)的位置。它可广泛应用于人脸检测和定位、人脸识别、姿态识别、表情识别、头部像压缩及重构、脸部动画等领域。近年来,该领域的研究有了较大的发展,为了让相关领域内的理论研究和开发人员对目前的进展有一个全面的了解,将近年来提出的脸部特征定位方法根据其所依据的基本信息类型,分为基于先验知识、几何形状、色彩、外观和关联信息等5类,分别作了介绍,同时,对各类方法的性能作了一些比较和讨论,对未来的发展作了展望。

关键词 模式识别(520·2040) 脸部特征定位 脸部特征提取

中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2003)08-0849-11

A Survey on Facial Features Localization

LIN Wei-xun, PAN Gang, WU Zhao-hui, PAN Yun-he

(College of Computer Science, Zhejiang University, Hangzhou 310027)

Abstract Facial features localization, an important technique in human face analyzing, is specialized in searching for facial features (eyes, nose, mouth, ears, etc) within a given region in an image or image sequence. It finds applications in various areas, such as face detection, face recognition, gesture recognition, expression recognition, face image compression and reconstruction, face cartoon, etc. This paper attempts to give an overview of the latest development in this field by classifying the newly proposed methods into five categories, namely the ones based on knowledge, geometry information, color, appearance and relative location. All these methods appeared in related papers or works published on international, as well as Chinese, journals and conferences in recent years. Then their performances in accurate rate, robustness and computational expense are roughly estimated and compared while some discussions about the criteria on which the estimation and comparison bases are provided. At the end of this paper, a prospect of the future development in this field is presented based on the observation of how an effective face feature location system should be build.

Keywords Pattern recognition, Facial features localization, Facial features extraction

0 引言

在模式识别领域中,与人脸相关的技术,如人脸识别、检测、跟踪、特征定位等,因其难度大、应用面广,一直是近年来的研究热点。脸部特征定位作为其中的一个重要部分,也正受到越来越多的重视。脸部特征定位,即在图象或图象序列的给定区域内搜索部分或所有人脸特征(如眼、鼻、嘴、耳等)的位置、关

键点或轮廓线。这3种信息都有十分广泛的应用,例如,脸部特征的位置信息既可以在人脸检测中用于定位人脸^[1~3]、验证人脸检测的结果^[4]以及精确指明人脸位置,也可在人脸识别中用于人脸对齐,或作为识别的依据之一^[5],在姿态识别中,它更是识别的重要依据之一;关键点信息可用于头部(Talking Head)图象的压缩和重构、脸部动画等领域,轮廓线信息则可用于图象的分割等。

评价一个算法的首要标准是其定位准确率,即

基金项目:国家自然科学基金(60273059);国家863计划(2001AA4180);浙江省自然科学基金青年科技人才培养专项基金(RC01058)

收稿日期:2002-11-12;改回日期:2003-03-28

被正确定位的特征数与待定位的特征总数之比;其次,算法应该有较强的鲁棒性,能够最大限度地适应人脸的各种变化,以提高算法的鲁棒性,这些变化包括:各种肤色和种族的差异,各个方向上的转动,尺寸、表情和光照条件上的变化,遮挡,某些特征如胡须、眉毛、眼镜等的出现或缺失等;此外,算法的计算量大小也是一个重要的评价标准,它影响到算法的可应用范围。因此,本文将主要从定位准确率、鲁棒性和计算量3个方面对各种定位算法进行考察。

关于人脸检测近期已有两篇很好的综述^[6,7],但尚未见脸部特征定位方面的综述。

1 脸部特征定位方法

近年来,国内外学者们已提出了许多种脸部特征定位方法。根据定位所依据的基本信息的类型,可以将现有方法分为基于先验规则、基于几何形状信息、基于色彩信息、基于外观信息和基于关联信息等5类:

(1) 基于先验规则 根据脸部特征的一般特点总结出一些经验规则,搜索前,先对输入图象作变换

使目标特征得到强化,而后根据上述规则从图中筛选出候选点或区域。

(2) 基于几何形状信息 根据脸部特征的形状特点构造一个带可变参数的几何模型,并设定一个评价函数度量被检测区域与模型的匹配度。搜索时不断调整参数使评价函数最小化,从而使模型收敛于待定位的脸部特征。

(3) 基于色彩信息 使用统计方法建立起脸部特征的色彩模型,搜索时遍历候选区域,根据被测点的色彩与模型的匹配度筛选出候选点。

(4) 基于外观信息 将脸部特征附近一定区域(窗口)内的子图象作为一个整体,映射为高维空间中的一个点,这样,同类脸部特征就可以用高维空间中的点集来描述,并可以使用统计方法得到其分布模型。在搜索中,通过计算待测区域与模型的匹配度即可判定其是否包含目标脸部特征。

(5) 基于关联信息 在局部信息的基础上,引入脸部特征之间的相对位置信息,以缩小候选点范围。

表1从运算量、准确率与鲁棒性(包括图象质量需求和姿态表情光照等影响)3个方面对各类方法的性能作了粗略的比较。

表1 脸部特征定位方法分类

算法分类	运算量	准确率	鲁棒性		
			图象质量需求	姿态表情光照等影响	
先验规则	镶嵌图	大			
	几何投影	小	较低		
	二值化定位	小		高	
	广义对称	大	较高	大	
几何信息	Snakes	大	高	高	
	可变形模板	大	高	大	
	ASM	较大	较高	较高	
色彩信息		小	低	较高	
外观信息	神经网络				
	PCA	大	高	低	
	SVM				较小
关联信息	概率网络				小
	DLA	较大	高	较低	
	GWN				较小

1.1 先验规则

先验规则是关于脸部特征一般特点的经验描述。人脸图象有一些明显的基本特征,如脸部区域通常包括双眼、鼻和嘴等脸部特征,其亮度一般低于周边区域;双眼大致对称,鼻、嘴分布在对称轴上等。为了利用这些基本特征进行脸部特征定位,一般要先对输入图象作变换,使所要使用的特征得到强化,而

后根据规则从图中筛选出候选点或区域。此类方法的难点在于,如何将人们的直观印象精确地表述为可应用的代码化规则,以及如何处理规则的精确性与普适性之间的矛盾。

1.1.1 镶嵌图

文献[8]中提出了镶嵌图(Mosaic Image,又称马赛克图)方法,他们先将图象划分成一组相同大小

的方格,每个方格的灰度取格中各像素灰度的均值,根据一组规则确定可能为人脸的区域;然后将方格的边长减半,重新构建镶嵌图,并根据一组规则分别定位出眼、鼻、嘴等脸部特征;最后,将前两步所得到的脸部区域二值化,并使用边缘跟踪最终确定脸部特征的位置。为了实现尺寸无关性,对所有可能的方格尺寸都作了测试,因此算法的运算量很大,准确率也不高。但是由粗到精的搜索策略对后来的研究有很大的影响。

使用长宽比可变的矩形单元代替方格^[9],以便更好地拟合人脸。为了设定矩形单元的长宽比,利用灰度值在 X 和 Y 方向上的投影确定眼、鼻、嘴和脸颊边界的位置,从而确定出脸部的大致范围,据此设定矩形单元的长宽比。这样缩小了搜索的范围,也避免了循环测试,不但使运算量大幅减少,也使准确率大大提高(在M2VTS数据库^[20]上,对眼、鼻、嘴等的定位准确率均在94.6%以上),但是他们的算法无法处理复杂背景或多人的情况,对目标对象姿态的变化较敏感。

1.1.2 几何投影

几何投影是脸部特征领域的一类经典算法,它利用脸部特征灰度值较皮肤低的特点,先统计出 X 或 Y 方向上的灰度值和或灰度函数值和(投影),找出特定的变化点,然后根据先验知识将不同方向上的变化点位置相结合,即得到脸部特征的位置。Kanade等首先将这一方法应用于人脸识别^[11]。Brunelli等在投影中使用 X 、 Y 两个方向上的梯度信息,以取代原始灰度值^[12],减小了对光照条件的敏感度。

Feng和Yuen等提出使用VPF(Variance Projection Function 方差投影函数)函数来定位脸部特征^[13],该函数的作用是计算指定方向上的所有像素的灰度值方差,与简单求和相比,它对复杂背景的适应力较强。他们在人眼上定义了6个特征点,分别对应于VPF函数值在 X 、 Y 方向上的特定突变点,同时假定瞳孔为圆形,上下眼眶为抛物线。搜索时,先依据人体测量标准^[14]在人脸区域上划出搜索范围,然后根据其中VPF函数值的变化情况求出特征点位置,最后根据眼形假设拟合出瞳孔形状和眼眶线。与可变形模板法相比,该方法计算量较小,对初始搜索位置要求不高,但在眼睛状态变化的情况下容易失效。

姜军等提出以“或”运算代替求和运算的广义投影运算^[15],他们先将图象二值化,而后在一定高度范围内

作水平方向投影得到一系列短线,最后根据先验规则对短线两两配对,从而筛选出最终结果。该算法能应对多人的情况,运行速度很快,但错误接受率较高。

几何投影法计算量较低,但在大的姿态变化或复杂背景下容易失效。

1.1.3 二值化定位

Liming Zhang等先用直方图阈值法将图象二值化,然后根据其中黑色区域的面积、形状和相对位置等几何特征确定出瞳孔的位置,最后通过边缘追踪依次找到上眼眶、眼角和下眼眶^[16]。范宏深等提出了一种与其类似的方法^[17],即将图象二值化后,计算出黑色区域的有效面积、质心、外接矩形、圆性质和添实度(即黑斑的有效面积与黑斑外接矩形面积之比)等几何特征,然后依据经验标准筛去一部分黑色区域,并将余下的黑色区域作配对筛选,从而得到双眼位置。此类方法运算量很小,但易受光照条件和图象质量的影响,在追踪边界时容易发生断线问题,稳定性无法保证。

1.1.4 广义对称

对称性是识别物体可利用的基本性质之一,它包括点对称性(也叫中心对称性)和轴对称性。绝大多数自然物体或人造物体都存在着这两种对称性,对于人脸图象来说,人眼、眉毛、嘴等都具有很强的点对称性。为了描述物体的点对称性,定义了广义对称变换GST^[18,19](Generalized Symmetry Transform),它只考察人眼中心点的强对称性和脸部特征的几何分布,对人脸偏转、表情变化、光照变化等条件不敏感,因而具有很好的鲁棒性。但计算量很大,且只利用了各点的对称性,易产生大量的候选点,使定位精度受到影响。

在此基础上提出的方向对称变换DST^[20](Directional Symmetry Transform),利用对称性在不同方向上的分布情况反映物体的基本形状特征,从而提高了定位的精确度,此外,因还事先利用了一些先验规则估算尺度因子和人眼的可能区域,减小了运算量。该算法在40多幅MIT的实验图象和100多幅自拍的人脸图象上作过测试,准确率为93.5%,与使用快速算法前相比,运算速度提高了20倍以上。

离散对称变换DST(Discrete Symmetry Transform)^[21],在计算对称性之前加入一个对图象灰度不均匀区域的检测过程,仅对灰度不均匀区域执行广义对称变换以加快系统的运行速度。在50幅自搜集图象上测得的准确率为92%。

此类方法的共同缺点是对闭着的眼睛无效。

基于先验规则的方法构造简单,仅使用一些简单的低层图象信息。由于这类算法强烈依赖于之前设定的先验规则,因此算法的普适性不高,对问题的条件通常也有较多限制,如要求图象中只能有单人、背景不能过于复杂、分辨率不能过低、姿势和表情的变化不能过大等。这些都限制了此类方法的应用。

1.2 几何形状信息

几何形状信息,即目标对象的几何形状特征,具有很好的直观性,易于理解和应用,因而很早就的模式识别领域得到广泛的使用。基于几何形状信息类方法的一般思路是:根据脸部特征的形状特点构造一个带可变参数的几何模型,并设定一个相应的评价函数以量度被检测区域与模型的匹配度,搜索时,不断调整参数使能量函数最小化,使模型逐渐收敛于待定位的脸部特征。

1.2.1 Snakes

Snakes方法是一种边界检测和图象分割的方法^[22],也可用于脸部特征的定位^[23,24]。Snakes模型是一条由 n 个控制点组成的连续闭合曲线,它用一个能量函数作为匹配度的评价函数,定义为

$$E_{\text{snake}} = E_{\text{internal}} + E_{\text{external}}$$

其中, E_{internal} 和 E_{external} 分别为内部和外部能量函数,前者定义了模型的内部属性和内在运动趋向,后者则与被测区域的图象属性相关,使模型偏离内在运动趋向。定位前,首先将模型定位在目标对象预估位置的周围,不断迭代使能量函数最小化,当内外能量达到平衡时即得到目标对象的边界。这种方法可以避免二值化方法时易发生的边缘断线的问题,但缺点是运算量很大,易受初始位置的影响,对非凸形体检测效果不佳,而且由于模型本身不包含目标对象信息,因而当存在局部变形、噪音和遮挡等干扰时容易出错。针

对这些问题,Wu等在搜索前,先根据一系列先验规则,利用灰度和色彩信息找出脸部特征的大致位置,作为设定初始位置的基准^[24]。为解决非凸形体检测问题,他们根据各个脸部特征的形状特点,使用多个Snake模型拟合一个脸部特征,如在眼、眉、嘴的上下方和鼻的左、右、下3个方向各放置一个Snake模型并使之各自收敛,直至相交。此外,还在能量函数中加入了边界、肤色相似性和红色相似性3种信息,以减小成像条件、表情及个体差异所带来的影响。在无遮挡、图象质量较好的情况下,该算法可以取得较好的效果,且对目标对象尺寸不敏感。

1.2.2 可变模板

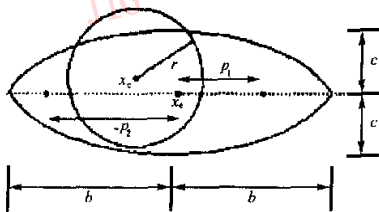
Yuille等提出使用参数化的可变模板来定位眼和嘴^[25],其中人眼模型如图1(a)所示。他们同样使用了一个能量函数作为匹配度评价函数,定义如下

$$E = E_v + E_c + E_p + E_t + E_{\text{internal}}$$

其中,外部能量由峰(E_v)、谷(E_c)、边缘(E_p)及光强(E_t)4部分信息组成, E_{internal} 的定义为

$$E_{\text{internal}} = \frac{k_1}{2}(x_c - x_c)^2 + \frac{k_2}{2}\left(p_1 + \frac{1}{2}(r+b)\right)^2 + \frac{k_2}{2}\left(p_2 + \frac{1}{2}(r+b)\right)^2 + \frac{k_3}{2}(b-2r)^2$$

其中, k_1, k_2, k_3 为权值,不同的权值组合可以使函数具有不同的功能,其他参数的含义见图1(a)所示。定位分为4个阶段,通过设定不同的权值并使能量函数最小化,可变形模板首先被“拖”到人眼附近;而后逐步完成尺寸、偏转角度以及位置和形状上的微调,整个过程如图1(b)所示。这个算法在人脸相关技术领域有很大的影响,但它的缺点是:模板的初始位置对定位结果有很大的影响;能量函数优化过程收敛缓慢,耗时长;能量函数中各种权值通过经验设定,难以保证最优。



(a) 可变形模板



(b) 处理过程

图1 可变形模板

针对以上问题,在能量函数中引入眼白信息,并增加了参数数量,使定位的准确率提高到 84%^[26]. Chow 等分两步实现人眼定位^[27],首先使用 Hough 变换求出瞳孔的位置,然后用一个仅包含眼眶的经简化的模型进行搜索,运行速度大为提高.用眼角检测来指导遍历过程^[28],即首先根据人体测量学知识在人脸区域中划定候选范围,然后使用边缘检测算法^[29]找到眼睛的 4 个边缘点(左右眼角及瞳孔与上下眼眶的交点),并以此为依据确定模板的初始参数.实验表明,该模型达到最优位置的时间为 Yuille 方法的 40%,准确率也有所提高.

1.2.3 PDM

PDM(Point Distributed Models,点分布模型)是一种参数化的形状描述模型^[30],它用一组离散的控制点来描述对象的形状,并用 PCA 方法^[31]建立起各个控制点的运动模型,对控制点的基准位置和移动模式作了约束,从而保证模型的整体形变始终在可接受的范围之内.Lanitis 等首先将这一方法应用于脸部特征定位及人脸识别,建立起了一个由 152 个控制点组成的人脸模型,并对各个控制点附近的与形状无关的局部灰度信息也分别建立了 PCA 模型^[32].搜索时,将基准模型放在目标图象附近,首先根据每个控制点当前位置附近的局部灰度信息计算出一个目标位置,然后使整个模型根据约束改变形状,使各个控制点与目标位置的距离尽可能地接近,如此反复,直至各个控制点都移动到合适的位置.与 Snakes 方法相比,此类方法在模型中引入了与目标特征有关的信息,从而降低了对噪音和局部变形的敏感度,但运算量仍然较大.

几何信息类方法在原理上大多较为简单易懂,但使用的时候面临很多困难.这是因为,作为一种低阶信息,几何信息受各种因素影响较大,鲁棒性较差.因此,为了保证提取的精度,此类方法通常都对图象质量有较高的要求,且对姿态、表情及光照条件的变化等也都有一定的限制.此外,由于脸部特征附近的对比度较低,边界搜索过程收敛缓慢,因而运算量一般都比较大.

1.3 色彩信息

色彩信息类方法的基本思想是用统计方法对目标对象的色彩建模,在搜索中根据被测点的色彩与模型的匹配度,筛选出待测特征的可能位置.色彩信息在人脸检测中有广泛的应用,而在脸部特征定位方面的应用则相对较少,这是因为,肤色在色彩空间

中具有聚合性,而脸部特征的色彩信息较为复杂,如眼与嘴,以及眼睛的瞳孔与眼白的色彩有显著区别,这给统一建模造成了一定的难度.

一种简单的解决办法是直接利用脸部特征与皮肤在色彩上的差异,如在已找到的皮肤区域上搜索“空洞”,并根据空洞的面积、相对位置等判断其是否为脸部特征^[4].这种方法避开对脸部特征色彩信息的直接建模,直接利用了基于色彩信息的人脸检测的搜索结果,速度快,但精度不高.

在直接建模方面,Fu 等在 YES 色彩空间中,使用 SPDNN 型(Self-growing Probabilistic Decision Neural Network 自增长概率决策神经网络)神经网络对瞳孔的 E 和 S 分量(即色彩分量)建模^[33].通过对包含 151 人、755 幅图象的自建数据库的测试,他们声称该算法的准确率可达 91.9%以上,速度快,能适应 $\pm 22.5^\circ$ 的转动,但无法区分眉毛和瞳孔.

Rein-Lien Hsu 等对脸部特征的色彩信息作了更细致的研究^[34].他们使用了 $YCbCr$ 色彩模型.搜索前,他们用 Reference White 方法对光照条件变化所造成的影响进行补偿.对于脸部特征的色彩特点,他们指出,人眼的 C_b 值高、 C_r 值低,且瞳孔与眼白的亮度反差大.据此,他们对人脸图象的色彩信息部分(C_b 、 C_r 分量)^[34]和光强信息部分(Y 分量)分别进行重构,强化了这些特征,重构图相“与”后,即得到眼睛的位置,过程如图 2 所示.对嘴的处理与此相似,与脸部其他区域相比,嘴唇的红色素较强,而蓝色素较弱,这些特点同样在重构图中得到强化.在 227 幅自建图象库上作了测试,其准确率可达 91.63%,且对各种姿态的变化有较好的适应性.

色彩信息类方法对光照条件和图象采集设备的特性较为敏感,易受环境因素的干扰,精度难以稳定,但其优势在于系统的运行速度快,受姿态、尺寸、

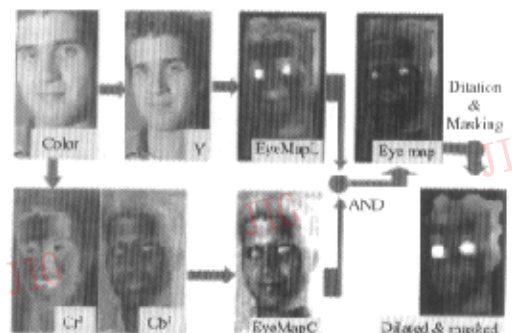


图 2 Rein-Lien Hsu 基于色彩的脸部特征定位算法

表情等变化的影响小,适合于粗定位或一些对运行时间有较高要求的应用。

1.4 外观信息

与几何信息类方法和色彩信息类方法不同的是,外观(Appearance)信息类方法不再面向像素级判断,而是将一定区域(窗口)内的图象整体视为一个随机变量,并映射为高维空间中的一个点,这样,同一类型的脸部特征就可以描述为高维空间中的一个点集,可以用统计方法对其分布规律建模,在搜索中,通过计算待测区域与模型的匹配度即可判定其是否包含目标脸部特征。这类方法对图象质量、环境条件和目标对象状态等方面的变化有较好的适应能力,近年来引起研究者的广泛关注。

1.4.1 神经网络

人工神经网络方法的研究具有悠久的历史,它很适合于解决非线性问题,因而在模式识别领域有着广泛的应用。在文献[35]中, Waite 等将这一方法应用于人眼定位。他们指出,完整的人眼图象受个体差异、眼睛状态和目标对象姿态等变化的影响较大,而子特征点(micro-features,包括左右眼角和上下眼眶顶点)附近区域相对稳定,根据这一特点,他们以各子特征点附近的灰度图象为输入,分别建立神经网络。检测时,先用各个神经网络对目标区域进行遍历搜索,再结合先验知识对搜索结果进行筛选和结合。这个算法训练过程比较简单,有较强的鲁棒性,但区分度不足。Reinders 等提出以梯度向量信息代替原始灰度值作为系统输入^[36],其优点是受光照、阴影、视角和目标物体尺寸的影响较小^[37]。同时,他们用一个统一模型对各个子特征点作同步检测,从而减少了遍历次数,他们在自建图象库(包含10人、50幅图,并有状态和一定范围的尺寸变化)上作了测试,对各个特征的定位准确率平均达到96%,且对尺寸、朝向、目标状态和光照条件都有良好的适应能力。

1.4.2 PCA

PCA(Principal Component Analysis 主元分析法)是模式识别领域的一类经典算法^[31],它通过K-L变换将表征人脸的高维向量映射到由若干个特征向量(也称 Eigenface 特征脸)张成的子空间中,从而实现人脸图象的最优解析和重构。作为分类器利用的是次元子空间(特征空间的补空间),先用 PCA 模型重构被检区域,求出重构图与原图之间的距离,当距离小于一定阈值时,即被认定为候选区域。

Turk 和 Pentland 首先将这一算法应用于人脸识别和定位^[38]。Cootes 等提出使用多 PCA 模型来辅助确定 PDM 模型初值^[1]。为了降低 PCA 分类器对目标物体的尺寸和角度偏转的敏感性,他们针对不同角度不同尺寸的脸部特征分别建立了模型,并在搜索中采用了金字塔搜索策略,因此他们的算法运算量较大。这种方法的缺点是 PCA 技术主要着眼于最优化地解析和重构图象,当作为分类器使用时会产生大量的虚警点,需要结合其他算法加以剔除。

1.4.3 SVM

SVM(Support Vector Machine 支持向量机)是20世纪90年代以来依据统计学习理论发展起来的一种通用的模式识别方法^[39],与传统学习理论力图使经验风险最小化不同,它着眼于使学习过程中的结构风险(Structural Risk)最小化,较好地解决了小样本情况下的学习问题。文献[40]~文献[42]将 SVM 方法用于脸部特征检测。其中,文献[40]使用了方形扫描窗口,将眉毛与眼睛作为一个整体作为定位的对象,从而减少了眉毛对定位的干扰。文献[41]提出了一个二级的 SVM 系统,先用线性核心的 SVM 筛选出候选点,再用多项式核心的 SVM 作最终判定。对 ORL 图象库^[43]中400幅图象的实验显示,他们的系统对眼、鼻、嘴等脸部特征的定位成功率均达到96%以上。直接应用 SVM 方法的主要困难首先是训练 SVM 需要求解二次规划问题,计算复杂度高,内存需求量大,其次是,当训练样本个数较大时,会得到大量的支持向量,使分类器计算量过高。

1.4.4 相关讨论

外观信息类方法研究的一个重要问题是如何确定扫描窗口的长宽比和中心在窗口中的相对位置。在文献[42]中,Heisele 等提出了一种自增长算法来解决这一问题。对于每个脸部特征,他们从一个较小的初始窗口开始,使之相对于中心向上下左右4个方向扩展,从而得到一系列窗口,而后为每个窗口训练出一个分类器,并将其中预期错误率最小的一个作为相应脸部特征的扫描窗口模板。他们所求得的各个脸部特征的扫描窗口长宽比例如表2所示。

此外,外观信息类方法对目标对象尺寸大多较为敏感,为了实现尺寸无关性,一种常用的办法是采用金字塔搜索策略^[44],即保持扫描窗口大小不变,而对目标图象做多级采样,得到一系列不同分辨率的图象,然后在这些图象上分别作遍历搜索。

外观信息类方法大多基于一定的统计模型,具有

表 2 文献[42]得到的各脸部特征的扫描窗口比例

脸部特征	宽度	高度
眉	19	15
眼	17	17
两眼间距	18	16
鼻	15	20
鼻孔	22	12
脸颊	21	20
嘴	31	15
嘴唇	13	16
嘴角	18	11

良好的可扩展性,可以通过“自举”(bootstrap)策略^[45]来使系统的精确性和兼容性得到提高,因而近年来一直是研究的一个热点.但是此类方法的缺点是:运算量较大;使用的多为抽象的高维统计信息,难以与直观的物理特征相联系;仅使用局部灰度信息难以有效定位脸部特征^[46];在大型图象库上性能不佳.

1.5 关联信息

上述方法大多从局部着眼,试图依据像素本身(色彩信息类方法)或邻近像素之间的相互关系(几何、外观类方法)来做出判断,但局部信息的区分度有限,在不受限制的条件下往往存在着大量的相似点,这就增大了后期处理的难度.为了解决这个问题,研究者们引入了关联信息,试图利用脸部特征之间存在着较稳定的相对位置关系来缩小候选点范围.与前面提到的先验知识不同的是,这些方法所使用的关联信息都是通过统计学习得到的.

1.5.1 概率网络

Kin 和 Cipolla 在文献[3]中使用一个 3 层的贝叶斯概率网络来对脸型建模,其网络结构如图 3 所示.他们在搜索中采用了自底向上的搜索策略,首先,结合使用高斯滤波器和边缘检测算法找出双眉、双眼、鼻和嘴的候选点(对应于网络中的第 1 层),然后根据邻近候选点之间的相对关系两两配对为水平或垂直组合(对应于网络中的第 2 层),并进一步归入脸的上、下、左、右 4 个区域(对应于网络中的第 3 层).

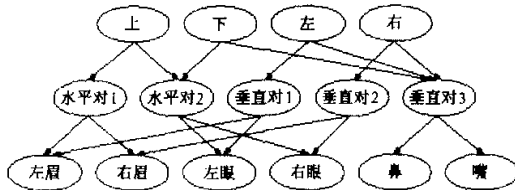


图 3 Kin 和 Cipolla 的概率网络

层),从而筛去虚警点.他们的算法在包含 100 幅图象的自建图象库上测得的准确率达到 92%,对残缺脸像及各种视角(包括侧面)、各种尺寸的脸像都有很好的适应能力.

1.5.2 DLA

DLA(Dynamic Link Architecture,动态链接模型)是一种神经网络模型^[47],它既保留了神经元的独立性,又能够很好地表述神经元之间的相关语义信息,因而对目标对象的局部突变有很好的适应能力.Lades 等将其应用于模式识别领域^[48],他们的动态链接模型是一个神经元矩阵,每个节点由 40 个不同方向、不同频率的 Gabor 小波函数组成,其返回值组成一个向量,称为 jet,用以表征局部信息.

Wiskott 等在此基础上作了三方面改进^[2]:首先,他们将节点的分布由阵列式改为散布式,每个节点分别对应一个脸部特征点,这使得模型更加直观和灵活;其次,他们引入了集束(Bunch)的概念,一个集束由若干个 jet 组成,分别对应于特征点的不同状态(如眼睛的开闭),检测时,以集束中的各个 jet 与被检测点的匹配度的最大值作为相似度估计值,从而增强了模型对不同状态的适应能力;第三,他们在匹配函数中引入了相位信息,提高了匹配函数的区分能力,使定位更为准确.他们的算法对目标对象在姿势和表情方面的变化有很好的适应能力,但运算量很大.

Rui 等指出,对同源派生的小波函数而言,由于中心频率与频带宽之积为常数,因此,中心频率越高,空间域上的有效带宽越窄,所表征的区域就越小^[49].与低频率分量相比,高频率分量对位置变化更敏感,可靠性差.基于以上考虑,他们只采用了后 16 个小波函数的返回值(对应于最低的两个中心频率)组成 jet 参与匹配计算,这使得运算量减少为文献[48]中方法运算量的 1/3.

1.5.3 GWN

GWN(Gabor Wavelet Network,Gabor 小波网络)是一种基于 Gabor 小波的神经网络^[50],Krüger 等将其引入图象处理领域^[61].他们使用一组同源派生的 Gabor 小波函数取代 RBF 神经网络的基函数,通过训练,可以将目标图象分解为若干个小波函数的线性组合.与传统的 Gabor 小波分解不同的是,GWN 的训练中同时对相关权值和小波函数本身的参数作优化,这使得 GWN 模型可以用很小数量的小波函数实现对目标对象的解析和重构.Feris 使用

两层的 GWN 树模型来定位脸部特征^[52], 两层 GWN 分别用于表征全脸和各个脸部特征。在训练中, 他们为每幅训练图建立一个 GWN 树模型, 并标定出各脸部特征的位置, 存入人脸库中。实际搜索的时候, 他们首先通过全脸比对从库中找出与目标图象最接近的一个模型, 然后以该模型的标定位置为搜索起点, 在小范围内, 通过与该模型中相应的脸部特征信息的比对求出脸部特征的精确位置。通过对 Yale 图象库^[53]和 FERET 图象库^[54]的测试, 该算法对眼角、鼻侧的识别率在 94% 以上, 对嘴角的识别在 87% 左右。

可以看出, 随着相对位置信息的引入, 脸部特征定位与人脸识别、定位等相关领域之间的界线渐趋模糊, 如在文献[2]中, 在脸部特征的定位与匹配度计算同步完成, 而文献[52]则利用相似人脸的经验信息来确定搜索的初始位置, 这可以说是目前脸部

特征定位研究的一个重要趋势。随着使用信息类型的增多, 定位的准确率随之提高, 对眼睛状态、目标对象姿态上的变化有一定的适应能力, 但同时也增大了算法的运算量。

2 常用人脸图象数据集及算法评价标准

绝大多数脸部特征定位算法在设计过程中都需要有一定数量的训练图作为设定参数的依据或统计学习的样本, 此外, 还需要有一定数量的测试图库用于评测和改进性能。换言之, 这些算法所建立的脸部特征模型实际上都是建立在有限图库基础上的经验模型, 因此所选用的图库对各种情况的覆盖面, 将在很大程度上影响到模型的普适性和鲁棒性。表 3 列出了一些常用的人脸库, 并对它们的特点作了简单介绍。

表 3 常用人脸图象数据集

数据库	链接	特点
M2VTS 人脸库 ^[10]	http://poseidon.csd.auth.gr/M2VTS/index.html	包含图象序列
MIT 人脸库 ^[28]	ftp://whitechapel.media.mit.edu/pub/images	16 人, 每人 27 幅, 有光照、尺度和头部偏转变化
ORL 人脸库 ^[45]	http://www.uk.research.att.com	40 人, 每人 10 幅
Yale 人脸库 ^[53]	http://cvc.yale.edu	有表情、光照变化及眼镜
FERET 人脸库 ^[54]	http://www.nist.gov/humanid/feret	大量的男女脸像, 有表情变化
UMIST 人脸库 ^[55]	http://images.ee.umist.ac.uk/danny/databased.html	20 人 564 幅图象, 包含从正面到侧面的各个角度
Harvard 人脸库 ^[56]	ftp://ftp.hrl.harvard.edu/pub/faces	各种光照条件, 脸像经裁剪和掩盖
Purdue AR 人脸库 ^[57]	http://rv11.ecn.purdue.edu/~aleix/aleix-face-DB.html	3 276 张脸像, 有表情、光照变化和遮挡
Kodak 人脸库 ^[58]	Eastman Kodak Corporation	彩色图象, 有尺寸、姿态和光照变化
CMU PIE 图象库 ^[59]	http://www.ri.cmu.edu/projects/project_418.html	68 人共超过 40 000 幅脸像, 每人 13 种姿势、43 种光照条件和 4 种表情
Bern 大学人脸库	ftp://iamftp.unibe.ch/pub/Images/FaceImages/	30 人, 每人 10 幅正面像和 5 幅侧面像

在评价一个算法的时候, 除了要求在相同的数据集上测试之外, 还应考虑许多其他因素, 如:

(1) 搜索范围 脸部特征定位的搜索范围可以大致分为全图搜索和给定区域搜索两类, 前者的难度显然要高于后者。

(2) 定位结果评价标准 研究者们对“定位准确”有着不同的评价标准, 如 Rowley 等在文献[60]中使用的标准为: 与手工标定位置比较, 误差在 4 个像素以内, 求得的尺寸误差不超过 20%。

(3) 应用范围 不同的应用对算法性能要求的侧重点各不相同, 如姿态识别对特征定位的准确率有较高要求, 而实时应用则允许牺牲准确率以减小运算量; 在用于确认人脸检测结果的时候, 必须降低错误接收率, 而自底向上的人脸检测算法则要求有

较低的错误拒绝率等。

(4) 其他因素 本文列出的实际上都是系统的整体性能, 它并不完全等价于算法的性能, 因为一个完整的系统往往包含有前期处理、粗定位和微调等阶段, 核心算法和其他因素对整体性能的贡献度难以区分。

上述因素对评价结果的影响应引起人们的注意。

表 4 列出了部分算法的性能比较。由于测试平台不尽相同, 无法直接比较各算法的运算量和速度, 因此本表中没有列出速度性能。此外, 大部分研究者的结果报告都是在自建图象库上测试得到的, 且对结果准确的定义也不尽相同, 因此这里所列出的准确率也仅供参考。

表 4 部分算法性能比较

方法	代表	测试图库	准确率(%)	鲁棒性	定位对象
镶嵌图	文献[9]	M2VTS库 ^[10]	94.6	—	眼、鼻、嘴和眉间点
广义对称	文献[20]	40幅MIT库及100多幅自拍图象	93.5	有一定光照变化,背景中略有干扰,可戴眼镜	眼
可变换板	文献[26]等	60幅自建图象库	84	包括戴眼镜	眼
色彩	文献[34]	227幅自建图象库	91.63	有光照、肤色、位置、尺寸、方向、3D姿态和表情变化,复杂背景	眼和嘴
神经网络	文献[36]等	50幅自建图象库	96	有尺寸、朝向、目标状态和光照等变化	眼
SVM	文献[40]	ORL库 ^[48]	96	有尺寸和姿态变化	眼、鼻和嘴
概率网络	文献[3]	100幅自建图象库	92	复杂背景,可戴眼镜,有尺寸和姿态变化(包括侧面像)	眼、鼻和嘴
GWN	文献[52]	FERET库 ^[54] 和Yale库 ^[53]	87~94	有表情、光照变化,可戴眼镜	眼、鼻和嘴

3 结 论

脸部特征定位经常被作为人脸检测算法的一部分^[3,4,36],或与人脸检测同时实现(如PDM^[30,32]、DLA^[2]).与人脸检测相比,脸部特征定位的特点在于特征的图象区域小,本身荷载的信息量较少,区分度有限,更易受目标对象自身状态和外部因素的干扰.另一方面,脸部特征之间存在着一些相对固定的位置、朝向关系等,可在定位过程中加以利用.

基于实用角度考虑,脸部特征定位算法需有较快的运行速度,且能在各种条件下保持稳定,如:不同人种、肤色;有大幅度转动、侧像;有表情变化;有部分遮挡;图象质量差、分辨率低.

速度与稳定性的要求在一定程度上是矛盾的,要保证算法具有足够的推广能力,就必须尽可能多地从图象中挖掘并利用信息,避免使用一些易受干扰的简单特征.但复杂的模型必然导致运算量的增大和运算速度的降低.

为了解决这个矛盾,一种思路是利用目标点远远少于非目标点的特点,将搜索过程划分为多个阶段,采用由粗到精的搜索策略,先使用一些运算量小、错误拒绝率低的算法尽快将大部分非目标点排除;而后用准确率较高的算法筛选出最终的结果,这样不同算法的优势得到互补,使系统的整体运行速度和准确率都得以保证.

另一种可行的思路是综合使用各类信息.在单个脸部特征定位的基础上,引入上下文关系,将统计信息与先验知识相结合,可以有效地排除虚警点,提高识别的正确率.但上下文关系往往易受姿态变化和遮挡的干扰,应注意避免由此引入的对特定条件

的依赖.

此外,根据特定应用背景的要求,还可以对问题条件作一定限制,从而降低算法设计的难度.例如在实时应用中,由于对运算量和运行速度的要求远高于对准确率的要求,因此可以选用一些模型简单、运算量小的算法.

参 考 文 献

- 1 Cootes T F, Taylor C J. Locating faces using statistical feature detectors[A]. In: Proceedings of Automatic Face and Gesture Recognition[C], Killington, VT USA, 1996:204~209.
- 2 Wiskott L, Fellous J M, Krüger N *et al.* Face recognition by elastic bunch graph matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997,19(7):775~779.
- 3 Kin C Y, Cipolla R. A probabilistic framework for perceptual grouping of features for human face detection [A]. In: Proceedings of Automatic Face and Gesture Recognition [C], Killington, VT USA, 1996:16~21.
- 4 Son Lam Phung, Douglas Chai, Bouzerdoum. Skin color based face detection [A]. In: Proceedings of Intelligent Information Systems Conference [C], Australian and New Zealand, 2001: 171~176.
- 5 Heisele B, Ho P, Poggio T. Face recognition with support vector machines: global versus component-based approach [A]. In: Proceedings of International Conference on Computer Vision [C], Vancouver BC, Canada, 2001,2:688~694.
- 6 Yang M H, Kriegman D J, Ahuja N. Detecting faces in images: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(1):34~58.
- 7 Hjelmas E. Face detection: A survey [J]. Computer Vision and Image Understanding, 2001,83(3):236~274.
- 8 Yang G, Huang T S. Human face detection in a complex background [J]. Pattern Recognition, 1994,27(1):53~63.
- 9 Kotropoulos C, Pitas I. Rule based face detection in frontal views [A]. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing [C], Munich,

- Germany, 1997, 4: 2537~2540.
- 10 Pigeon S, Vandendorpe L. The M2VTS multimodal face database [A]. In: Lecture Notes in Computer Science [C], Crans-Montana, Switzerland, 1997, 1206: 403~409.
- 11 Kanade T. Picture processing by computer complex and recognition of human faces [R]. Department Information Science, Kyoto University, Kyoto, Japan, 1978.
- 12 Brunelli R, Poggio T. Face recognition: features versus templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(10): 1042~1052.
- 13 Feng G C, Yuen P C. Variance projection function and its application to eye detection for human face recognition [J]. Pattern Recognition Letters, 1998, 19(9): 899~906.
- 14 Verjak M, Stephancic M. An anthropological model for automatic recognition of the male human face [J]. Annals Human Biology, 1994, 21(4): 363~380.
- 15 姜军, 张桂林. 一种基于知识的快速人脸检测方法 [J]. 中国图象图形学报, 2002, 7A(1): 6~10.
- 16 Zhang L M, Lenders P. Knowledge-based eye detection for human face recognition [A]. In: Proceedings of Knowledge-Based Intelligent Engineering Systems and Allied Technologies [C], Brighton, UK, 2000, 1: 117~120.
- 17 范宏深, 倪国强, 申会莹. 利用几何特性及神经网络进行人脸探测技术的研究 [J]. 光学技术, 2002, 28(2): 105~107.
- 18 Reisfeld D, Yeshurun Y. Robust detection of facial features by generalized symmetry [A]. In: Proceedings of International Conference on Pattern Recognition [C], The Hague Nethzrland, 1992: 117~120.
- 19 Reisfeld D, Wolfson H, Yeshurun Y. Context-free attentional operators: the generalized symmetry transform [J]. International Journal of Computer Vision, 1995, 14(3): 119~130.
- 20 周杰, 卢春雨, 张长水等. 基于方向对称变换的人脸定位方法 [J]. 电子学报, 1999, 27(8): 12~15.
- 21 刘文子, 潘峰. 离散对称变换在人脸图像眼睛定位中的应用 [J]. 红外与毫米波学报, 2001, 20(5): 375~380.
- 22 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models [A]. In: Proceedings of International Conference on Computer Vision [C], London, UK, 1987, 1: 259~268.
- 23 Huang C L, Chen C W. Human facial feature extraction for face interpretation and recognition [J]. Pattern Recognition, 1992, 25(12): 1435~1444.
- 24 Wu H, Yokoyama T, Pramadihanto D *et al.* Face and facial feature extraction from color image [A]. In: Proceedings Automatic Face and Gesture Recognition [C], Killington VT, USA, 1996: 345~350.
- 25 Yuille A L, Cohen D S, Halliman P W. Feature extraction from faces using deformable templates [A]. In: Proceedings of Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 1989: 104~109.
- 26 Shackleton A, Welsh W J. Classification of facial features for recognition [A]. In: Proceedings of International Conference on Computer Vision and Pattern Recognition [C], Mani, HI USA, 1991: 373~579.
- 27 Chow G, Li X. Towards a system for automatic facial feature detection [J]. Pattern Recognition, 1993, 26(12): 1739~1755.
- 28 Lam K M, Yan H. Locating and extracting the eye in human face images [J]. Pattern Recognition, 1996, 29(5): 771~779.
- 29 Xie X, Sudhakar R, Zhuang H. Corner detection by a cost minimization approach [J]. Pattern Recognition, 1993, 26(8): 1235~1243.
- 30 Cootes T F, Taylor C J. Active shape models—"Smart Snakes" [A]. In: Proceedings of British Machine Vision Conference [C], Leeds, UK, 1992: 266~275.
- 31 Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces [J]. Journal Optical Society, 1987, 4(3): 519~524.
- 32 Lanitis A, Taylor C J, Cootes T F. A unified approach to coding and interpreting face images [A]. In: Proceedings of International Conference on Computer Vision [C], Cambridge MA, USA, 1995: 368~373.
- 33 Fu H C, Lai P S, Lou R S *et al.* Face detection and eye localization by neural network based color segmentation [A]. In: Proceedings of Neural Networks for Signal Processing [C], Sydney, Australia, 2000, 2: 507~516.
- 34 Rein-Lien Hsu, Abdel-Mottaleb M, Jain A K. Face detection in color images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 696~706.
- 35 Waite J, Vincent J. A probabilistic framework for neural network facial feature location [J]. British Telecom Technology Journal, 1992, 10(3): 20~29.
- 36 Reinders M J T, Koch R W C, Gerbrands J J. Locating facial features in image sequences using neural networks [A]. In: Proceedings of Automatic Face and Gesture Recognition [C], Killington, VT USA, 1996: 230~235.
- 37 Bichsel M. Strategies of robust object recognition for the automatic identification of human faces [D]. Zurich Switzerland; ETH Zurich, 1991.
- 38 Turk M, Pentland A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71~86.
- 39 Vapnik V N. Statistical learning theory [M]. New York: John Wiley & Sons, 1998.
- 40 Pan G, Liu W X, Wu Z H *et al.* An eye detection system based on SVM filter [A]. In: Proceedings of SPIE, Electronic Imaging and Multimedia Technology [C], 2002, 4925: 326~331.
- 41 Li Dihua, Podolak I T, Lee S W. Facial component extraction and face recognition with support vector machines [A]. In: Proceedings of Automatic Face and Gesture Recognition [C], Washington, DC USA, 2002: 76~81.
- 42 Heisele B, Serre T, Pontil M *et al.* Component-based face detection [A]. In: Proceedings of Computer Vision and Pattern Recognition [C], Kauai HI, USA, 2001, 1: 657~662.
- 43 Samaria F S. Face recognition using hidden markov models [D].

- University of Cambridge, Cambridge, UK, 1994.
- 44 Burt P J. The pyramid as a structure for efficient computation [A]. In: Rosenfeld A eds. Multi-Resolution Image Processing and Analysis[M]. Berlin: Springer-Verlag, 1984: 6~37.
- 45 Sung K K, Poggio T. Example-based learning for view-based human face detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 39~51.
- 46 Burl M C, Leung T K, Perona P. Face localization via shape statistics [A]. In: Proceedings of Automatic Face and Gesture Recognition[C], Zurich, Switzerland, 1995: 154~159.
- 47 von der Malsburg C. The correlation theory of brain function [R]. Department of Neurobiology, Max-Planck-Institut für Biophysikalische Chemie, Göttingen, Germany, 1981.
- 48 Lades M, Vorbrüggen J C, Buhmann J *et al.* Distortion invariant object recognition in the dynamic link architecture[J]. Computers, 1993, 42(3): 300~311.
- 49 Liao R, Stan Z Li. Face recognition based on multiple facial features [A]. In: Proceedings of Automatic Face and Gesture Recognition[C], Grenoble, France, 2000: 239~244.
- 50 Zhang Q, Benveniste A. Wavelet networks [J]. Neural Networks, 1992(3): 889~898.
- 51 Krüger V, Sommer G. Wavelet networks for face processing [J]. Journal of Optical Society of America, 2002, 19(6): 1112~1119.
- 52 Feris R S, Gemmell J, Toyama K *et al.* Hierarchical wavelet networks for facial feature localization[A]. In: Proceedings of Automatic Face and Gesture Recognition[C], Washington DC, USA, 2002: 118~123.
- 53 Bellhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces; Recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720.
- 54 Phillips P, Moon H, Rizvi S *et al.* The FERET evaluation methodology for face-recognition algorithm [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1090~1034.
- 55 Graham D B, Allinson N M. Characterizing virtual eigensignatures for general purpose face recognition [A]. In: Face Recognition: From Theory to Applications [M], Springer-Verlag, 1998, 163: 446~456.
- 56 Hallinan P. A deformable model for face recognition under arbitrary lighting conditions [D]. Harvard University, MA, USA, 1995.
- 57 Martinez A, Benavente R. The AR face database[R]. Computer Vision and Control Technology Report 24, Purdue University, Indiana, U. S., 1989.
- 58 Loui A C, Judice C N, Liu S. An image database for benchmarking of automatic face detection and recognition algorithms [A]. In: Proceedings of International Conference Image Processing[C], Chicago, Illinois, USA, 1998: 146~150.
- 59 Sim T, Baker S, Bsat M. The CMU Pose, Illumination, and Expression (PIE) database [A]. In: Proceedings of Automatic Face and Gesture Recognition [C], Washington, DC, USA, 2002: 53~58.
- 60 Rowley H, Baluja S, Kanade T. Neural network-based face detection [A]. In: Proceedings Computer Vision and Pattern Recognition[C], Santa Barbara, CA, USA, 1996: 203~208.



林维训 1978年生, 2000年获天津大学计算机科学与技术系学士学位, 现于浙江大学计算机学院攻读硕士学位, 研究方向为模式识别和图象处理。



潘纳 1976年生, 1998年获浙江大学计算机科学与工程系学士学位, 现于该系攻读博士学位, 主要研究方向为人脸识别、图象处理、模式识别和计算机视觉等。



吴朝晖 1966年生, 中德联合培养博士, 浙江大学教授、博士生导师。主要研究方向为生物认证、模式识别与网络计算等。



潘云鹗 1946年生, 教授, 中国工程院院士。主要研究方向为人工智能、多媒体技术和智能CAD等。