

基于纹理分布和变形模板的面部特征提取*

山世光¹, 高文^{1,2}, 陈熙霖²

¹(中国科学院 计算技术研究所, 北京 100080);

²(哈尔滨工业大学 计算机科学与工程系, 黑龙江 哈尔滨 150001)

E-mail: sgshan@ict.ac.cn

http://mmi.ict.ac.cn

摘要: 面部特征提取是面部感知的重要内容, 同时也是特定人的 3D 人脸动画应用中所必须的前期工作. 在一个多级人脸检测模块检测到人脸大致区域和尺寸的基础上, 提出并实现了一种基于面部图像纹理分布特性和可变形模板的由粗到细的面部特征提取策略, 旨在解决可变形模板对参数初值依赖性强和计算时间长的问题. 该策略首先利用眼睛区域的谷特性和频率特性定位两个虹膜中心点位置, 然后用积分投影确定唇部和鼻子区域的位置, 在此基础上进行关键特征点的检测, 从而可以得到预定义特征模板参数的良好初值, 最后基于贪心算法的多阶段轮换优化算法来搜索一个极小点. 实验表明, 该方案在时间和精度方面具有良好的性能.

关键词: 人脸检测; 面部特征提取; 人脸识别; 变形模板; 贪心算法

中国法分类号: TP391

文献标识码: A

人脸检测和面部特征提取是人脸识别、表情识别、唇读、视点跟踪等面部感知应用系统的前期工作, 具有重要的研究价值. 同时, 在特定人 3D 人脸模型的自动生成、人脸动画以及基于分析/合成的人脸图像编码中也有重要应用. 人脸检测和面部分割通常是面部特征提取的前提, 人脸检测技术已趋成熟^[1,2], 主要方法有模板匹配、Eigenface^[3]、肤色模型和人工神经网络等. 面部特征提取算法正日趋成熟, 但尚无令人满意的系统, 面部特征提取的技术主要有模板匹配、变形模板、Hough 变换、主动轮廓、Gabor 小波变换、曲线拟合等.

在面部特征提取技术中, 可变形模板得到了较为广泛的应用. 其主要思想是根据人脸特征的先验形状信息, 定义一个用若干参数描述的形状模型, 这些参数反映了对应特征形状的可变部分, 如位置、大小、角度等, 它们最终通过模板与图像的边缘、峰、谷和灰度分布特性动态地交互适应得以修正^[4]. 由于模板变形利用了特征区域的全局信息, 因此可以较好地检测出相应的特征形状. 但是, 由于变形模板要采用优化算法在参数空间内进行能量函数极小化, 因此, 算法有两个主要缺点: 一个是对参数初值的依赖程度高, 容易陷入局部最小; 另一个是计算时间较长.

针对可变形模板上述两方面的问题, 本文提出了一种由粗到细的面部特征提取策略. 首先要进行人脸检测. 本文继承了实验室已有的基于肤色模型、模板匹配和特征子脸的二级人脸检测模型. 然后在检测到人脸大致区域的基础上, 先利用人脸器官构造的先验知识、面部图像灰度分布的峰谷和频率特性粗略地估计眼睛、鼻子、嘴、下巴的大致区域和关键的特征点, 再借助于它们给出模板的

* 收稿日期: 1999-11-23; 修改日期: 2000-01-13

基金项目: 国家自然科学基金资助项目(69789301); 国家 863 高科技发展计划资助项目(863-306-ZT93-01-2); 中国科学院百人计划资助项目

作者简介: 山世光(1975-), 男, 山东日照人, 博士生, 主要研究领域为人脸识别, 面部动画, 图像编码; 高文(1956-), 男, 辽宁大连人, 教授, 博士生导师, 主要研究领域为多媒体数据压缩, 图像处理, 计算机视觉, 多模式接口, 人工智能, 虚拟现实; 陈熙霖(1965-), 男, 安徽芜湖人, 博士, 教授, 主要研究领域为图像理解, 计算机视觉, 多模式接口, 虚拟现实.

较好的初始参数,从而可以大幅度地提高算法的速度和精度.实验表明,该策略在精度和时间方面都具有良好的性能.

1 多级人脸检测模型

人脸检测是面部特征提取的第一步工作.本文继承了文献[1]的工作,采用了一种基于肤色模型、模板匹配和特征子脸的三级人脸检测模型.肤色模型具有检测速度快的优点,但其缺点是可靠性差,易受噪声影响;模板匹配虽实现简单,但精度不高;特征子脸检测算法具有较高的精度,但其缺点是速度较慢.这3种技术在检测性能上各有其优缺点,合理地将它们组合在一起,可以取得更好的检测性能.在分析各种检测模型性能的基础上,本文建立了一个逐步求精的多级人脸检测模型,如图1所示.该模型较好地兼顾了检测精度与检测速度的矛盾,具有检测精度高、鲁棒性强等特点,较好地解决了复杂背景下人脸检测的问题.

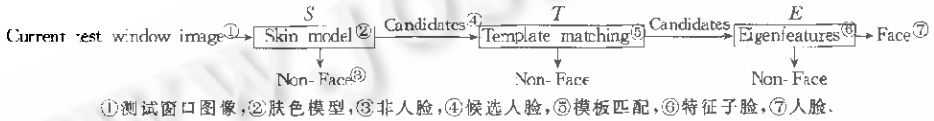


Fig. 1 Multi Level face detecting model
图1 多级人脸检测模型

2 面部特征粗定位

上述人脸检测模块提供了输入图像中人脸区域的位置和尺寸的粗略信息,在此基础上,本文根据虹膜在人脸图像中是最明显的特征这一观察结果,首先定位虹膜的位置,然后根据两个虹膜中心点位置、面部器官结构关系的统计先验知识和面部灰度分布特性来定位其他器官的位置.

2.1 面部器官结构关系统计

尽管不同人的脸形不同,但是可以合理地假设这种构形满足正态分布,这些知识可以为特征定位提供有利的支持.因此,本文采用手工测量的方法,对人脸器官的位置关系进行了统计.人脸样本数目为100人,统计结果见表1.由此可见,不同个体的面部构形具有较大的相似性,可以用来指导面部特征区域的粗略定位.

Table 1 Facial feature conformation statistics

表1 面部特征结构关系统计

Ratio to distance of the two eyes ^①	Eye width ^②	Eye to nose tip ^③	Eye to lip ^④	Eye to chin tip ^⑤	Nose width ^⑥	Mouth width ^⑦	Mouth height ^⑧	Eye to eyebrow ^⑨
Mean ^⑩	0.42	0.51	0.88	1.59	0.57	0.70	0.27	0.35
Variance ^⑪	0.03	0.05	0.04	0.09	0.05	0.06	0.02	0.05

①与双目距之比,②眼睛宽,③眼到鼻尖,④眼到唇上,⑤眼到下巴尖,⑥鼻宽度,⑦嘴宽,⑧嘴高,⑨眼到眉心,⑩均值,⑪方差.

2.2 基于面部纹理分布特性的特征检测

在本文的特征检测算法中,两个虹膜中心点的定位是其他器官定位的基础,这是因为从视觉效果上看,虹膜在人脸图像中往往最为突出,因此它的检测精度和可靠性可以得到保障.在此基础上,定位其他特征就相对容易了,因此,本文首先实现了一个定位两个虹膜中心点的算法.

2.2.1 虹膜中心点检测

对于由人脸检测框定的人脸区域,首先进行区域的扩充,以使之包含全部的头部区域A(包括大部分或者全部头发以及整个面部);然后对区域A用合适的阈值进行二值化,在得到的二值图像

B 中,用区域标号算法对不同的“黑”连通区进行标号化,用预置的面积阈值 θ 滤掉由于噪声造成的小的孤立连通区,可以认为,位于区域 A 最上方的最大的连通区 H 即为头发区域.当头发区域 H 确定后,在剩下的连通区域中搜索满足下列条件的连通区对 L 和 R :

- (1) L 和 R 的中心要在头发区域 H 的下方;
- (2) L 和 R 下方的一定距离内不能有其他的连通区域;
- (3) L 和 R 的中心横坐标差值不能过小;
- (4) L 和 R 的中心纵坐标差值不能太大;
- (5) L 和 R 在原图像对应区域内的频率要足够大.

实验表明,通过上述 5 条规则的限制,总可以找到一对满足条件的连通区,而这这对连通区恰好就是由两个眼睛所形成的区域.在该算法检测得到的 L 和 R 区域的二值图像 B 和二值化的边缘图像 E 中,利用虹膜区域的谷特征和眼睛区域的高频特性,分别形成两个眼睛的区域点集,描述为

$$R_{ng} = \left\{ (x, y) \mid B(x, y) = 0, \sum_{(i, j) \in \Omega} E_{(i, j)} > \theta, \Omega \text{ 为 } (x, y) \text{ 的 } 8 \text{ 邻域} \right\}, \quad (1)$$

从而得到两个眼睛区域的粗略点集,分别计算这两个点集的重心,以该重心为中心得到一些候选的瞳孔点,对每一个候选点 P ,在边缘图像上计算下述支持函数:

$$S_p = \frac{1}{N} \sum_{(x, y) \in A} E_{(x, y)}, \quad (2)$$

其中 E 为边缘图像, A 为由半径范围 $R_1 \leq r \leq R_2$ 所决定的圆环区域, N 为该圆环中的像素总数.分别取两个候选点集中具有最大支持函数的点作为粗定位阶段的虹膜中心点:

$$P_{LeftIris} = \arg(\text{MAX}_{p \in R_{ng_{left}}} S_p), \quad P_{RightIris} = \arg(\text{MAX}_{p \in R_{ng_{right}}} S_p). \quad (3)$$

实验表明,这里提出的定位算法具有良好的性能.即使在图像质量较差的情况下也能够有效地定位出眼睛区域.表 2 给出了该算法在一个人脸库上检测虹膜的识别率.

Table 2 Results of irises localization

表 2 虹膜定位结果

Performance of the recognition ^①	Look left ^②	Look down ^③	Frontal ^④	Look up ^⑤	Look right ^⑥
Test samples ^⑦	56	56	400	56	55
Errors ^⑧	4	3	5	2	3
Ratio ^⑨ (%)	92.8	94.6	98.8	95.4	94.6

①识别情况,②左转平视,③俯视,④准正面,⑤仰视,⑥右转平视,⑦总样本数,⑧定位错误数,⑨检测率.

2.2.2 图像平面内旋转的矫正

对于人脸图像可能存在的图像平面内的旋转,可以通过定位出的两个虹膜中心点的位置用仿射变换进行矫正.进行该矫正的主要目的在于:(1)方便后续的其他特征区域的定位;(2)旋转矫正后可变形模板时就可以减小计算耗时非常大的角度参数的寻优范围,以提高模板变形速度.

2.2.3 基于积分投影的嘴部、鼻子区域检测

观察表明:鼻孔处和中唇线在人脸灰度图像中总会在双眼的下侧中垂线上形成两个谷区.这里利用了这一特性,用水平方向的积分投影曲线的分析进行嘴部和鼻子区域的检测.首先根据第 2.1 节中得到的统计参数,估计出嘴部区域在垂直方向上的大致范围,然后在此范围内取窗口进行水平方向的积分投影.对该投影曲线进行高斯平滑,计算曲线的极小点,具有最小值的极小点可以作为中唇点定位结果.同样,利用鼻孔和鼻侧翼造成的阴影区域,可以用与嘴部的检测算法相类似的方法检测鼻子区域,不同的只是搜索区域.鼻孔的检测则采用垂直方向的投影得到.在两个鼻孔点上方的区域内搜索具有最大邻域灰度均值的点可以检测到鼻尖.

3 关键特征点检测

关键点主要是指眼角点、嘴角点、下巴曲线上的主要特征点,借助于它们可以给出相应的模板参数的较好的初值,从而可以使模板更快地变形到全局最优。

眼睛关键点的检测 眼睛关键点包括左右眼角点和上下眼皮的界限点.本文采用积分投影算法来确定这些点,不同的是,积分投影过程是在边缘图像中,而不是在灰度图像中进行的.该算法的实验结果如图 2 所示.



Fig. 2 Localization of the key feature points

图 2 关键特征点定位

嘴部关键点的检测 嘴部关键点包括两个嘴角点和上唇最高点.左右嘴角点用模板匹配得到.上唇最高点采用边界跟踪的方法,跟踪的结果如图 2 所示.下唇往往因为反光而很难找到准确的外部边缘,所以对于下唇最低点此处采用估计的方法给出.当然,在后续的模板算法中,对下唇还要进行变形处理.

下巴关键点的检测 下巴区域有 7 个可选的关键点,分别是:左右嘴角的延长线与下巴的交点、过中唇点的垂线与下巴的交点、过左右嘴角点的垂线与下巴的交点、过左嘴角点往左下 45° 直线与下巴的交点、过右嘴角点往右下 45° 直线与下巴的交点.下巴区域在通常情况下都具有良好的边缘,依据统计的先验知识,可以限定这些点的搜索范围,因此,在边缘图像中可以较好地估计出这 7 个点中的全部点或者部分点.图 2 给出了实验结果.

4 基于变形模板的特征形状提取

从图 2 可以看出,上节中给出的关键点定位结果大多数情况是不错的,但是也显然存在一定的误差和一些不很可靠的关键点.究其原因,主要是因为这些定位算法是基于局部特征的,因此容易受噪声等干扰的影响,可变形模板是一种在一定程度上强调利用全局信息的形状检测算法,因此可以用来进行面部区域形状的精确定提取.本节以眼睛为例来介绍用变形模板提取面部特征形状的过程.

4.1 眼睛区域特征形状检测

4.1.1 眼睛模板

眼睛用如图 3 所示的一个圆和两条抛物线来描述,共有 9 个参数: $a, b, c, r, X_T, Y_T, X_C, Y_C, \theta$, 用这些参数对抛物线和圆所作的数学描述分别为:

上眼皮边缘对应的抛物线

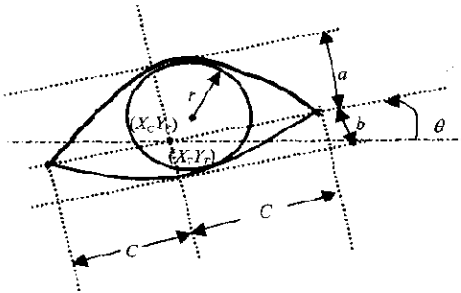


Fig. 3 Eye template and its parameters

图3 眼睛模板及其参数定义

可以控制得很窄,从而可以提高模板的变形速度。

4.1.2 边缘图计算

为了进行模板变形,需要计算边缘图像.本文采用了一种多边缘检测响应值加权累加的算法,首先计算 Canny 边缘和 Sobel 边缘,然后将它们加权求和得到边缘图 E ,形式描述为式(7).这样做的意义在于,既可以尽可能地保留所有的边缘,又可以突出对模板变形有用的边缘。

$$e(x, y) = \alpha C(x, y) + (1 - \alpha)S(x, y), \quad (7)$$

其中 C 和 S 分别为 Canny 和 Sobel 边缘的响应输出.同时,为了将模板从最初的可能离实际边缘较远的区域引导到目标边缘去,需要对边缘图像进行能量的分散.本文用二维高斯卷积实现这一目的。

4.1.3 能量函数

变形模板技术的另一个关键性的内容是能量函数的定义.本文采用了下述能量项,其中下列公式中的 $e(x, y)$ 表示由式(7)得到的边缘图像, n, m, s 分别表示相应边缘上的点数。

上眼皮能量项

$$E_u = -\sqrt{\frac{1}{n} \sum_{(x,y) \in L_{upperPan}} e^2(x, y)}, \quad (8)$$

下眼皮能量项

$$E_l = -\sqrt{\frac{1}{m} \sum_{(x,y) \in L_{lowerPan}} e^2(x, y)}, \quad (9)$$

虹膜边缘能量项

$$E_c = -\sqrt{\frac{1}{s} \sum_{(x,y) \in L_{circle}} e^2(x, y)}. \quad (10)$$

为了充分利用人眼构形的先验知识,采用了内部能量项,定义为

$$E_i = (2r - (a - b))^2 + [(X_r - X_c)^2 + (Y_r - Y_c)^2] + (c - 2a)^2. \quad (11)$$

另外,为了阻止模板朝显然不合理的方向变形,引入了一个参数合法性判断的能量项,定义为

$$E_p = \begin{cases} \text{MAX} & \text{if } (b > a \text{ or } a > c \text{ or } b > c \text{ or } r > c \text{ or } d > a \text{ or } d > c) \\ 0.0 & \text{otherwise} \end{cases} \quad (12)$$

最后得到下述能量函数:

$$E = \omega_p E_p + \omega_i E_i + \omega_u E_u + \omega_l E_l + \omega_c E_c, \quad (13)$$

其中 $\omega_p, \omega_i, \omega_u, \omega_l, \omega_c$ 为经验权系数。

4.1.4 参数初值估计

如前所述,可变形模板能否最终搜索到全局最优,除了优化算法的影响以外,还对参数的初值有较强的依赖性.本文的初值估计在上述关键点检测的基础上进行.事实上,本文在大多数情况下给出的参数的初值都较好地反映了眼睛的实际形状,从而提高了模板的变形速度和性能.

4.1.5 优化过程控制

至此,可变形模板算法变成了一个多参数非线性函数的优化问题,这里采用了基于贪心算法的分阶段轮换算法进行能量函数的极小化.具体地说,就是将优化过程分成虹膜圆边缘搜索、上眼皮抛物线搜索、下眼皮抛物线搜索3个阶段进行迭代,在每一个阶段内对有关的参数用贪心算法择优,直到每个阶段内的每个参数的变化都不再改变能量值为止,最终模板参数即为提取的特征.

4.2 嘴部和下巴形状检测

嘴部模板和下巴的初值估计与优化变形过程采用与眼睛相类似的策略.因此,我们不再描述相关内容,仅简单介绍唇部和下巴模板情况.

如图4所示,嘴部模板主要由抛物线和4次曲线两种曲线构成,共12个参数: $x_c, y_c, h_1, h_2, h_3, h_4, a, w_0, w_1, \theta, q_l, q_r, q_b$.其中,位置和方向由中心坐标 (x_c, y_c) 和转角 θ 确定,上唇的外边缘4次曲线中心离坐标原点的偏移量为 a ,高为 h_1 ,参数 q_l 表示左侧4次曲线偏离抛物线的距离, q_r 表示右侧4次曲线偏离抛物线的距离;下唇的外边缘4次曲线高为 h_4 ,参数 q_b 表示该4次曲线偏离抛物线的距离;上、下外唇的宽度为 w_0 ;内唇上、下抛物线的高分别为 h_2 和 h_3 ,其宽度均为 w_1 .

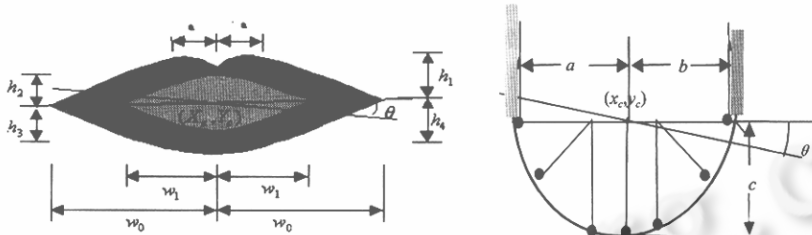


Fig. 1 Templates for mouth and chin

图4 嘴部和下巴模板

下巴模板由左、右两条抛物线组成,如图4所示.它有 $x_c, y_c, a, b, c, \theta$ 共6个参数,分别表示中心坐标、两条抛物线的横轴截距、高度和旋转角度.

5 实验结果

图5给出了本文系统的一些检测结果,其中原始输入图像的大小为 256×256 ,256级灰度,其中人脸的典型尺度为 130×140 .在Pentium III 450MHz的机器上,从检测人脸到完成上述全部特征的检测平均需要2秒.文献[4]在SUN4上定位眼睛和嘴唇,需要5~10分钟;文献[5]在SUN4上眼睛模板的平均时间为11.545秒,嘴唇模块运行时间为3.23秒;而文献[6]在33MHz的486上定位眼睛区域窗口需要3~4秒,完成眼睛的模板变形需要8~20秒,却没有报道其他特征的检测情况.

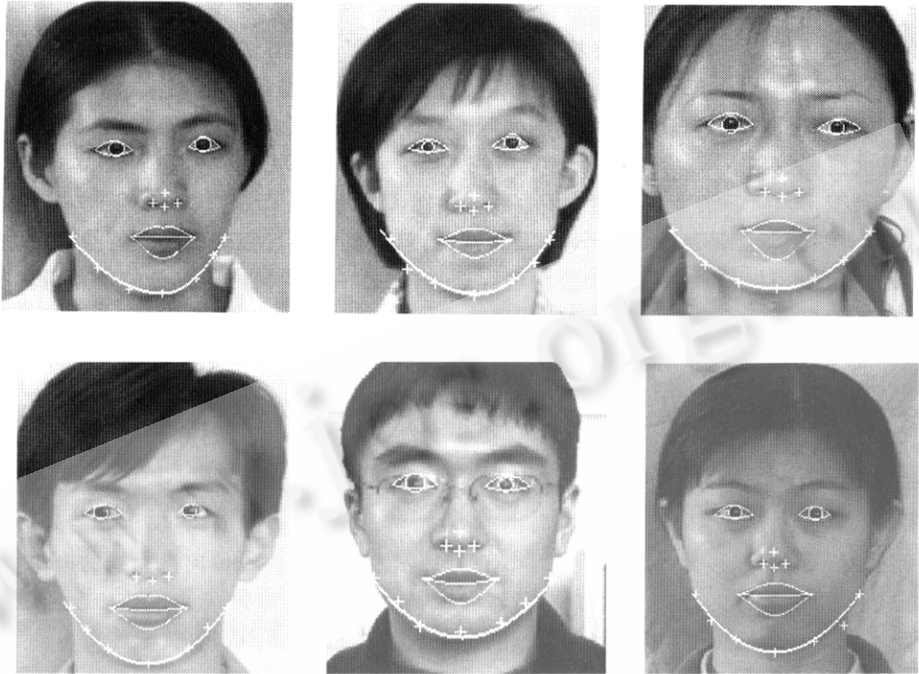


Fig. 5 Results of facial feature extraction
图 5 面部特征提取结果

6 结 论

可变形模板是进行面部特征形状提取的有效算法,但模板参数的增多使得模板变形的速度很慢,同时,优化过程能否收敛到全局最优点,往往在很大程度上依赖于参数的初值.本文针对这些问题提出的逐步求精的特征提取策略首先使用一些低层次的图像处理算法进行一些关键特征点的粗略定位,以便给出模板接近全局最优的初始值,从而极大地减少了寻优时间,提高了收敛到全局最优点的可能性.与文献中类似算法运行时间和效率的比较表明,本文提出的策略在时间和精度方面都有较好的性能.

本文的工作在特定人真实感 3D 自动人脸合成中得到了应用,所提取的面部特征用来作为修改一般人脸模型的依据以生成特定人的 3D 人脸模型,在纹理映射阶段则可以提供所需的面部纹理.同时,提取的特征还可以用于基于面部几何特征的人脸识别系统中.

References:

- [1] Rama, Chellappa, Wilson, C. L., Saad, Sirohey. Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 1995,83(5):705~740.
- [2] Liu, Ming-bao. Study on human face detection and tracking [Ph. D. Thesis]. Harbin: Harbin Institute of Technology, 1997 (in Chinese).
- [3] Turk, M., Pentland, A. Eigen-Faces for recognition. *Journal of Cognitive Neuroscience*, 1991,3(1):71~86.
- [4] Yuille, A. L., Hallinan, P. W., Cohen, D.S. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 1992,8(2):99~111.
- [5] Gloria, Chow, Li Xiao-bo. Towards a system for automatic facial feature detection. *Pattern Recognition*, 1993,26(12):

1739~1755.

- [6] Xie, X., Sudhakar, R., Zhuang, H. On improving eye feature-extraction using deformable templates. *Pattern Recognition*, 1994, 27(6): 791~799.

附中文参考文献:

- [2] 刘明宝. 人脸检测与跟踪的研究[博士学位论文]. 哈尔滨, 哈尔滨工业大学, 1997.

Facial Feature Extraction Based on Facial Texture Distribution and Deformable Template*

SHAN Shi-guang¹, GAO Wen^{1,2}, CHEN Xi-lin¹

¹*(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China);*

²*(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China)*

E-mail: sgshan@ict.ac.cn

http://mmi.ict.ac.cn

Abstract: Facial feature extraction is an important aspect in facial image perception system. And it is also a prerequisite in animation system for generating a given person's 3D-face image. In this paper, a coarse-to-fine facial feature extraction strategy is presented based on facial texture distribution and deformable template, using the pre-result of a multi-level face detection, which aims at solving such problems as the searching highly depending on the initial parameters and time-consuming that deformable template algorithm often suffers from. In proposed strategy, firstly, the center of the two irises is localized making use of the valley and frequency characteristics in the two eye regions. Then integral projection is used to localize the coarse position of the mouth and the nose. Secondly, some key feature points about these organs are estimated. Finally, according to these feature points, good initial parameters for the pre-defined templates are given and an optimal algorithm based on greedy algorithm and multi-epoch cycle is used to search for the minimum solution. Experiments indicate that the implementation of the proposed strategy is with good performance in both speed and accuracy.

Key words: face detection; facial feature extraction; face recognition; deformable template; greedy algorithm

* Received November 23, 1999; accepted January 13, 2000

Supported by the National Natural Science Foundation of China under Grant No. 69789301; the National High Technology Development Program of China under Grant No. 863-305-2T03-01-2; the Hundred Talents Foundation of the Chinese Academy of Sciences