

彩色图象中主要人脸特征位置的全自动标定

张欣 徐彦君 杜利民

(中国科学院声学所, 北京 100080)

摘要 人脸特征的定位是计算机人脸识别中必需的一步. 该文提出了一种在彩色图象中进行脸部特征定位的新方法. 该方法只取彩色图象中的红色成分作为研究对象, 以简化处理及提高信噪比; 首先利用对原图的垂直灰度投影曲线确定脸的左右边界, 并分离出此区域的高低灰度图, 以便分图定位脸部特征; 接着利用对经边缘检测的高灰度图的水平灰度投影确定下颌、嘴、鼻、眼的垂直位置, 然后在相应位置附近, 对鼻、眼、嘴依此进行准确定位. 算法最后的输出为人眼的位置、鼻端的位置及鼻宽、嘴的位置及嘴宽和下颌的垂直位置. 大量实验表明, 该算法具有很高的精度和很强的鲁棒性.

关键词 边缘检测 垂直灰度投影 水平灰度投影 有效波峰 霍夫变换

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2000)02-0138-06

Automatic Localizing Main Facial Features in Color Images

ZHANG Xin, XU Yan-jun, DU Li-min

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080)

Abstract Facial features localization is very important in face recognition by machine. In this paper a novel method which performs main facial features localization in color information is put forward. It utilizes only red component of the color image to simplify processing cost and improve SNR, segments high gray-level region and low gray-level region in order to localize facial features in segmented regions, first localizes the vertical position of chin, mouth, nose, and eyes with gray-level projection, then localizes the horizontal position of them in associated vertical positions. In this method, nose is first localized to aid other features' localization. The output of this method is positions of eyes, position of nose tip and width of nose, position of mouth and width of mouth, and position of chin. Computer simulation shows that this algorithm has high accuracy and good robustness.

Keywords Edge detection, Vertical gray-level projection, Horizontal gray-level projection, Valid peak, Hough transform

0 引言

计算机人脸识别是一个非常活跃的研究领域. 尤其进入 90 年代以来, 关于在给定的一幅图象中进行人脸的分割和标定, 以及进行眼、嘴等人脸部特征的提取的研究越来越多. 由于计算机人脸识别技术在经济、法律等领域有着很好的应用前景(如在许多情况下的身份认证), 所以它在近年来新兴的听觉视觉双模态语音识别的唇读识别方面更加得到重视, 故视觉特征的提取是其研究重点之一. 其中脸部

特征的定位则是人脸识别技术中必不可少的一步.

从参考文献中已见到几种脸部特征标定的方法. 如 Kwang Nam Choi 等^[1] 直接利用经典的模板匹配方法, 通过训练已有的图象数据提取了一套可以刻画脸部各特征的滤波器模板. Ariel Tankus 等^[2] 将脸作为一种三维结构来处理, 眼在这里被看做一个抛物面. 他们的算法虽有较高的精确度, 但过于复杂. Dieckmann^[3] 和 Karin Sobottka^[4] 使用的灰度投影是一种简单有效的方法, 但由于他们都是直接对原图做水平灰度投影, 因而其包含的干扰信息太多, 水平投影曲线也很不规则, 个体差异较大. 这

就给脸部特征的定位带来困难. Benn^[5]使用的经改进的霍夫算法效果不错,但他将眼球看作圆形不太合理,因为眼球必受眼睛开闭程度的影响.这种对眼的描述尤其不适合于眼睛较小的人.

本文在研究分析了前人研究工作的基础上,提出了一套新颖且简单有效的方法,用来进行人脸部特征(眼、鼻、嘴、下颌)的定位.其主要步骤如下:

(1) 首先利用原图的垂直灰度投影曲线确定脸的左右边界,并根据大津定律在此区域内确定分割高低灰度图的阈值,以提取高低灰度图;

(2) 然后对边缘检测后的高灰度图做水平灰度投影,并对得到的水平灰度投影曲线利用人脸各几何尺寸间的相互关系对其进行修正,使之由下至上最终只剩下4个指示下颌、嘴、鼻、眼垂直位置的有效波峰;

(3) 最后以这四个波峰所在的位置为基准,对鼻、眼、嘴进行准确定位;

(4) 算法最后的输出为人眼的位置、鼻端的位置及鼻宽、嘴的位置及嘴宽和下颌的垂直位置.

该算法具有如下特点,即只采用三原色中的红色成分为研究对象,因而简单有效,并且对眼睛使用类椭圆轮廓来描述,其中心由与其内切的两圆圆心的中点决定;多处利用了人脸一些几何尺寸间的相互关系.



图1 取自CVASR数据库的一幅人脸图象
(原图象为彩色)

实验表明,该方法具有很高的定位精度,并且对于照明、倾斜、旋转、眼镜等变化具有很强的鲁棒性.

1 研究对象的选择

在以往人脸特征提取的研究中一般以灰度图象为研究对象,灰度图象虽有易于处理的特点,且大多数经典的图象处理方法都基于灰度图象.但彩色图象毕竟包含了比灰度图象更为丰富的信息,而且由于人脸肤色有别于其它物体,这些信息将有助于提高特征提取的精度与鲁棒性.本文便以BMP格式24位真彩图为研究对象.本法只取三原色中的红色成分进行处理.因为人脸肤色中红色是基调,它包含了绝大部分的人脸有效信息,信噪比高.而灰度图象是由三原色加权组合而得,它必会模糊一些边缘.同时只取红色还有简化处理,减少计算量的优点.基于灰度图象的处理方法也可以很方便地移植过来.当然,通过对人脸彩色信息的进一步的统计研究,有可能存在更好的利用彩色信息的方法.

我们从CVASR数据库中选取了一幅人脸图象(图1)作为研究对象,然后提取其中的红色成分(图2)对比这两幅图我们可以看到图2包含了图1中几乎全部的人脸信息,且其对比度高,有效边缘得到加强.



图2 图1色彩中的红色成分

2 人脸图象的预处理

2.1 脸左右边界的确定

我们利用垂直灰度投影来确定人脸左右边界.垂直灰度投影图的纵坐标为图象一列上所有象素点的灰度值之和,横坐标为列号,它反映了图象在水平方向上灰度的变化.设图大小为 $M \times N$,各点象素值为 $I(x, y)$,则垂直灰度投影函数为

$$P_y(x) = \sum_{y=1}^N I(x, y) \quad (1)$$

脸的平均灰度值明显不同于背景中的其它物体,而人脸内部灰度值的分布除眼眉处有变化外,但对脸平均灰度值影响不大,其余部分则比较均匀.因此投影曲线必会在脸所在区域处显得比较平坦,而在脸与背景的交界处则会有突变,即分界处的那一点会有一个大的梯度值.这样整个曲线看起来会有一个大的凹槽或凸起,可以据此找到脸的大致边界.

即在这条曲线上找到一个具有最大梯度值(为正值)的点和一个具有最小的梯度值(为负值)的点,它们所在的列即可作为脸的左右边界,图3中大致呈正态分布(白线)是经低通滤波的垂直灰度投影曲线,两竖线即为脸的左右边界.这样做一般会把人耳,甚至人脸的一部分框在脸部区域外,但决不会把眼睛框在外面,因此不但不会影响脸部特征的初定位,反而有利于净化信息.以后各步的处理皆局限在此区域内.脸内部物(例如发型的影响)和背景物在投影曲线上可能产生意外的尖峰,它的大梯度会干扰脸定位,所以必须预先对投影曲线进行低通滤波.这里要求脸是图中最大的物体,因此垂直灰度投影曲线上所有这种凸起(或凹槽),以对应于脸的那一部分最宽,在频域上对应更低的频段.这样我们就可以把低通滤波器的门限定得很低,以保证滤去属于高频段的尖峰.

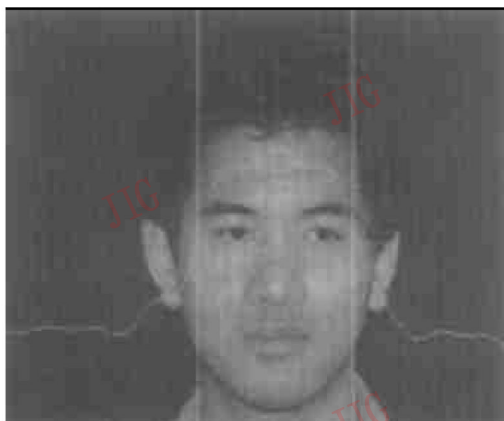


图3 人脸垂直灰度投影曲线

2.2 分离高低灰度图

我们依据大津准则计算阈值 $thresh$, 以此分离高低灰度图. 大津准则: 把直方图在某一阈值处分割成两组, 当被分成的两组间方差为最大时, 并据此决定阈值.

分离高灰度图时, 将脸部区域内灰度值小于 $thresh$ 的像素点的灰度值赋为 0, 即

$$I_H(x, y) = \begin{cases} I(x, y), & I(x, y) \geq thresh \\ 0, & I(x, y) < thresh \end{cases} \quad (2)$$

分离低灰度图 $I_L(x, y)$ 方法类同. 若是对图象序列进行处理, 且预知光照情况变化不大, 可只取一幅图象作抽样计算 $thresh$.

3 脸部各特征垂直位置的标定

首先我们用 Sobel 算子对高灰度图做一边缘检

测. 因后续分析(鼻定位)所需, 在对边缘检测器输出的图象二值化时, 所使用的阈值必须尽量避免边缘图象在一个区域内产生断裂的边缘. 所以可以取一个较低的阈值: $\frac{4}{5} thresh$.

然后对该经边缘检测后的二值图象做一水平灰度投影曲线; 该曲线的横坐标为图象一行上所有像素点灰度值之和, 纵坐标为行号, 它反映了图象在垂直方向上的灰度变化. 水平灰度投影函数为

$$P_x(y) = \sum_{x=1}^M I(x, y) \quad (3)$$

边缘检测后的图象必然在下颌、嘴、鼻、眼处出现大量的边缘点, 而且水平灰度投影曲线必在这些部位对应处形成 4 个大的波峰, 图 4 为图 3 脸部区域经边缘检测输出的二值图象. 左侧白点组成的曲线是该图的水平灰度投影曲线, 该曲线经过 5 点平均. 其中的 4 条短线所指示的为与眼、鼻、嘴、下颌相对应的 4 个有效波峰位置.

我们定义具有足够深度的波峰为有效波峰, 该深度条件要求波峰两侧都必须满足. 我们首先找出该投影曲线上所有的有效波峰. 然后利用人脸各几何尺寸的相互关系对其进行修正, 除去可能由于衣服、镜框、胡子及唇下阴影的干扰产生的波峰, 使之由下至上的 4 个有效波峰恰好对应于下颌、嘴、鼻、眼的垂直位置(见图 4 中横线所标的波峰). 其中人脸各几何尺寸的相互关系指的是嘴和下颌的距离大于鼻和下颌的距离的一半、鼻和下颌间距应大于脸宽的一半、眼和鼻的间距大于脸宽的 $1/4$ 等等.

该步输出的 4 个位置中, 确定的鼻、嘴、下颌垂直位置即为最后输出, 而眼的垂直位置有待下一步更精确的定位.



图4 人脸经边缘检测后的水平灰度投影曲线及二值图象

4 脸部各特征水平位置的标定

4.1 鼻水平位置的标定

从边缘检测后的水平灰度投影曲线我们已得到鼻端的纵坐标,即与鼻相对应的有效波峰所指示的垂直位置.对图4中鼻端纵坐标位置附近的水平带状区域做一次垂直灰度投影,可得一条垂直灰度投影曲线.由于鼻在此区域存在最大的连续边缘,因此在这条垂直灰度投影曲线上则对应有一条最长的无零点的线段,可据此确定鼻的左右边界.在取水平带状区域时,鼻端以上的部分应比鼻端以下的部分取得多些,以防误取了嘴的边缘点.另外为防止可能存在的眼镜框下沿和鼻边缘相连造成的误判,必须检查已确定的鼻宽是否小于脸宽的一半,若不是,则可认为离脸中点较远的那一个边界是被误判,因而需在该无零点的线段上重新再找一个最靠外的波谷点代替它,并再次检查鼻宽是否满足条件.

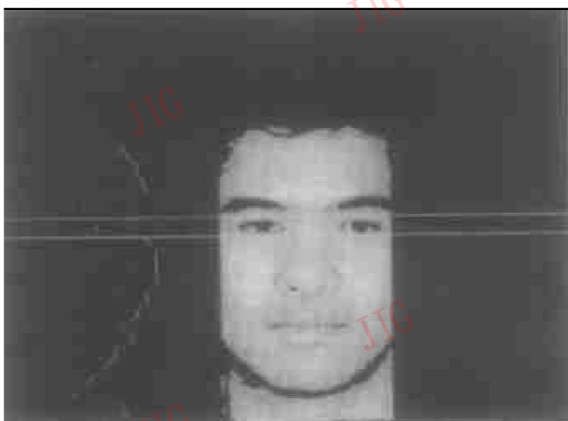


图5 图2高灰度图的水平灰度投影曲线

对戴镜者而言,其镜框的下沿在确定眼所在的垂直区域带时已被除去,上边沿有可能只被去掉一部分,因此需进一步除去镜框鼻架及两侧边沿的影响.这时我们将借助于已获得的鼻定位信息.因为鼻在两眼中间,所占的水平区域基本不和眼重合,所以首先可以将已确定的眼垂直区域带中与鼻所占的水平区域相重合的那一部分除去,以除去鼻架的影响.若脸倾斜稍大,这样做会去掉眼的一部分,但丢失的这一部分可在后续分析得知脸的倾斜度后找回来.灰度投影.垂直区域带内若存在镜框的上边沿则其在边缘检测图象上呈 \cap 状,并将在垂直灰度投影曲线上出现一个波谷;若不存在镜框的上边沿,则镜框的侧边沿和眼球间应有一段零点,从而可在垂直灰度投影曲线上找到这个波谷点或零点,作为眼所在区域的左右边界.

4.2 眼的标定

眼的定位以标在眼球内或附近,且不出眼眶为准.在实际操作中,为提高精确度,则以标在眼球中心为目标.首先将眼的可能区域框至最小.从人脸高灰度图及其水平灰度投影曲线可以看到眼部附近的灰度性状非常稳定,眼所对应的波谷的上下边界,几乎就是眼框的上下边界(图5、图6).但水平灰度投影曲线则在此处易受眼镜框上沿的影响,有时会使眉毛引起的波峰和眼镜框上沿引起的波峰相连,使眼和眉毛间的波谷消失(图6),导致眼区域确定的失误.故本法首先根据投影曲线来确定眼所占的垂直区域带.标定眼的位置只需在与眉、鼻对应的第3、第5个有效波峰间的区域进行,并在眼垂直位置附近,找到该曲线上和眼对应的波谷,然后找出和该波谷相邻且靠上的那个波峰作为眼的上边界.该波谷靠下的那个波峰峰形常不明显,多显平坦状(图5).所以我们在波谷往下,幅值增长的区域内找一个梯度绝对值最大的点作为眼的下边界.



图6 带眼镜人脸高灰度图的水平灰度投影曲线

现在我们可以以上确定的一个很小的区域内来寻找眼球.因眼球灰度非常低,为了避免眼皮的影响,我们对边缘检测后的低灰度图进行霍夫变换用来寻找眼球(图7).由于眼皮的影响,眼不应看作圆形,除非眼睁得很大,以至几乎露出整个眼球,否则将眼看成圆形是不妥当的.图8为我们对眼模型的描述,其中大圆代表眼球,大圆的虚线部分代表眼球被眼皮遮盖的部分,上下两条曲线代表上下眼皮形成的边缘,它和大圆的实线部分组成了我们在图7中看到的眼的形状.此形状可由图8中两个和它相切的小圆来描述,这两个圆圆心连线的中点便可认为是眼球的中心.

我们通过霍夫变换来找到这两个圆的圆心.由于搜索区域很小,霍夫变换受到的噪声干扰也很少.

在这我们采用文献[6]中改进的霍夫变换方法,对圆心的纵坐标和横坐标分别进行投标,左右眼的投标情况统计是独立分区进行的(以鼻中心为分界点),经边缘检测后的眼图象的霍夫变换结果见图9.



图7 边缘检测后的眼 图8 对眼的模型描述

从眼的边缘检测图象的霍夫变换图(图9)上可以看到横坐标上灰度最高点对应的并不是眼的正中心,眼的正中心处在最高峰和次高峰之间,说明眼的正中心在两个圆的圆心连线的中点上.这证明了前面我们对眼描述的正确性.在该变换结果上两眼的纵坐标首先被确定,并可由此得知脸的倾斜方向和大致倾斜度,以补回在前面去镜框鼻架时因脸倾斜而丢失的在鼻所占区域内眼的一部分,供对眼横坐标进行投标时使用.

闭眼时,以上标定过程在高灰度图中进行.眼的开闭可由眼所在区域内的边缘点数判定.将该区域左右均分两半,若其中有一个的边缘点数不足,就认为处闭眼状态.



图9 图7的霍夫变换结果

4.3 嘴的标定

嘴左右边界的标定通过对边缘检测后的高灰度图内嘴的纵坐标附近的水平带状区域再做垂直灰度投影来实现.利用嘴比鼻宽这一特点,我们可以不必对整个水平带状区域做垂直灰度投影,而只需对鼻以外一半鼻宽的两个区间作投影.这样做,需考虑脸倾斜,对这两个区间做一修正.然后只需在这两个区间的垂直灰度投影曲线上各找一个梯度最大值的点作为嘴边界即可.

由于嘴附近的噪声较多,嘴的标定易受干扰,所以我们对嘴的标定需作进一步修正.在眼的标定中

脸的倾斜角 α ,嘴中心的横坐标 X_{mouth} 和鼻中心的横坐标 X_{nose} 近似有:

$$X_{\text{mouth}} = X_{\text{nose}} + (Y_{\text{nose}} - Y_{\text{mouth}}) \times \tan\alpha \quad (4)$$

若发现两者确定的嘴中点不重合,则认为受了噪声的干扰.这种情况下,则选取和式(4)确定的中心相距较近的一个,并将它关于该中心对称的另一端点作为嘴的一新边界.

5 实验及结果分析

我们采用C语言实现了算法,并使用中科院声学所交互信息实验室汉语双模态语音识别工作组于1998年初创建的汉语视觉听觉双模态数据库CVASR中的视觉的人脸图象数据进行了大量的实验.该数据库的20名实验者(男12人,女8人,其中戴眼镜者6人)是随机抽取的,选取时对他们的发型、穿着、眼镜大小式样均未作出限制.其图象是由Hi8的摄像机在钨光灯照明下拍摄,经MPEG转换为MPG格式的文件保存.因此其拍摄条件和实际生活中的各种条件相近,具有很高的可信度.本实验对CAVSR中20位实验者每人各取出了50幅图,共1000幅人脸图象,利用该算法进行了人脸特征位置标定,取得了很高的正确率,表明该算法具有很高的鲁棒性.

图10、图11是两幅代表性人脸特征位置的最后标定结果,两图中眼、鼻、嘴附近白线即为标定的眼鼻嘴位置.表1为对20位实验者每人50幅图进行主要脸部特征定位的结果.

该算法具有如下的优点:

- (1) 对不同的人具有很高的鲁棒性;
- (2) 阈值的计算使之对光照变化不敏感,完全可以在实际光线下应用;
- (3) 对背景及被标定者的发型、着装、脸部表情皆无严格限制,仅要求脸在图象中是最大的物体;
- (4) 能够很好地消除眼镜的影响,有较强的抗眼镜反光的能力;
- (5) 有一定的抗脸倾斜和侧向的能力,在脸倾斜度不至于使之存在一条可以从两眼中间穿过的水平线及脸侧向不丢失眼睛的情况下,则不影响效果;
- (6) 在人闭眼时也能将眼位置准确的检测出来.

进一步的研究工作,可以在该算法标定结果的基础上,使用统计学的方法,建立精细的人脸特征的模型,以此进一步提高人脸特征定位的精确度和鲁棒性.



图 10 标定结果示例 1

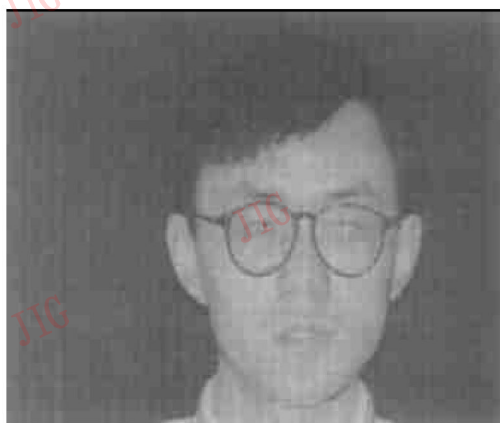


图 11 标定结果示例 2

表 1 对每人 50 幅图进行的主要脸部特征定位结果

单位: %

	1*	2	3 女	4	5* 女	6	7* 女	8	9*	10	11*	12	13 女	14 女	15	16 女	17 女	18* 女	19	20
眼	100	100	98	100	96	100	94	100	90	98	100	98	100	100	100	100	98	98	98	100
鼻	100	100	98	100	96	100	94	100	90	98	100	98	100	100	100	100	98	98	98	100
嘴	100	100	98	100	98	100	94	100	100	98	100	100	100	100	98	100	98	98	98	100
下颌	100	100	98	100	98	100	100	100	100	98	100	100	100	100	100	100	100	98	100	100

注: 表中, 眼以标在眼球内或附近, 且不出眼眶为准, 下颌的标定允许一定的偏差;

* 为戴镜者; 其中实验者 9 的眼镜反光极为严重, 故准确率稍低.

参 考 文 献

- 1 Kwang Nam Choi, Andrew D J Cross, Edwin R Hancock. Localising facial features with matched filters. In: AVBPA '97, Department of Computer Science, University of York, York, YO1 5DD, UK.
- 2 Ariel Tankus. Hezy Yeshurun and Nathan Intrator. Face Detection by Direct Convexity Estimation. In: AVBPA '97, Department of Computer Science Tel-Aviv University, Ramat-Aviv 69978, Israel.
- 3 Dieckmann U, Plankensteiner P, Schamburger R et al. SESAM: A biometric person identification system using sensor fusion. In: Proceedings of First International Conference of Audio and Video-Based Biometric Person Authentication, Crans-Montana, Switzerland, March 1997.

- 4 Karin Sobottka, Ioannis Pitas. A fully automatic approach to facial feature detection and tracking. In: AVBPA '97 Department of Informatics, University of Thessaloniki 540 06, Greece.
- 5 David E Benn, Mark S Nixon, John N Carter. Robust eye centre extraction using the hough transform. In: AVBPA '97 Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO171BJ, England.



徐彦君 1989年毕业于山东大学电子工程系, 获理学学士学位, 1992年于山东大学电子工程系获理学硕士学位, 现为中国科学院声学研究所博士生. 主要研究领域为三维视觉理论、图象理解与听觉视觉双模态语音识别.



张 欣 1998年毕业于北京大学电子学系, 获理学学士学位, 现为中国科学院声学研究所博士生.



杜利民 1983年、1987年、1991年分别于北京大学、中国科技大学研究生院、中国科学院声学研究所获理学学士、工学硕士、理学博士学位, 1996年美国麻省理工学院(MIT)高级访问科学家, 现为中国科学院声学研究所副所长, 研究员, 博士生导师, IEEE高级会员, 中国电子学会理事.