

复杂背景下人脸检测的数学形态学运算方法¹

游亚平 袁保宗

(北京交通大学信息科学研究所 北京 100044)

摘要: 皮肤的颜色特性被广泛地应用于人脸检测和跟踪中, 本文采用 HSV 空间与归一化 RGB 空间相结合的肤色模型, 提取原始图像中的类肤色区域作为候选人脸目标; 采用数学形态学算子, 分别用来完成去除噪声和干扰, 分离候选目标, 提取眼睛嘴巴‘孔洞’, 以及标示人脸轮廓等处理过程。大量实验表明, 该方法能有效地剔除臂和手等非人脸肤色目标; 能克服人脸遮挡, 姿势和方向变化, 大小差异, 光照变化等造成的困难; 该方法的计算开销较小, 易于实现, 适合快速人脸检测。

关键词: 人脸检测, 肤色模型, 形态学算子

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1009-5896(2004)12-1863-08

Mathematical Morphology Approach to Face Detection in Complex Background

You Ya-ping Yuan Bao-zong

(Institute of Info. Science, Beijing Jiaotong University, Beijing 100044, China)

Abstract Skin color character is widely used for face detection and face tracking. This paper construct a mixed model in both normalized RGB and HSV color space, which can be used to acquire possible skin regions in original image. It designs a sequential flexible morphology operators which can respectively eliminate noises and disturbances, separate face candidates, acquire eye and mouth “hole”, and gain contours of faces. The experimental results show that the proposed approach can not only effectively eliminate human hands and arms non-face objects in image, but also smooth away difficulties of face occlusion, pose and orientation, size difference and light variety. Furthermore, this approach is easy to implement, and its computational spending is relatively smaller. So it is suitable to fast face detection.

Key words Face detection, Skin color model, Morphology operator

1 引言

从一幅图像中判断是否存在人脸, 找出人脸所在位置及其占据区域, 是人脸检测的任务。人脸检测是一项基础性工作, 广泛应用于 HCI(人机交互), 人脸识别与跟踪, 图像增强和检索, 基于内容的视频编码等领域^[1-3]。

人脸检测受到诸多因素的影响, 首先是成像环境: 复杂背景, 多变的光照, 给从背景中分割出人脸造成困难; 其次是摄像机质量: 热噪声会使人脸结构等信息丢失, 几何畸变会使图像失真; 最后是图像中人脸本身的多变性: 人脸大小, 姿势, 方向, 遮拦, 面部表情等使图像中的人脸千差万别; 如果处理的是视频图像, 还要考虑计算复杂性。以上诸多原因, 使人脸检测成为一个很有挑战性的课题, 目前的方法大致有以下几大类。

(1) **模板匹配** 采用某个函数来预先定义或参数化表示一类标准人脸模式, 对于输入图像中的候选人脸, 计算它与标准模式之间的相关性, 若相关性值超过了预先设定好的阈值, 就判断该候选人脸为人脸。该方法操作简单, 但使用固定的模板来描述千差万别的非刚性的人脸是很

¹ 2003-07-24 收到, 2004-02-24 改回

困难的。为此人们提出了多尺寸模板 (Multiscale template), 子模板 (subtemplate), 变形模板 (Deformable template) 等解决办法, 在一定程度上克服了固定模板的缺陷, 但是这些方法仍只能检测到正面 (或允许少许偏向) 的人脸, 对于噪声和遮挡也很敏感^[4,5]。

(2) **基于统计和学习** 这类方法需要给出一些样本作为训练对象, 通过对这些样本的统计分析或机器学习来发现人脸和非人脸图像的相关特性, 并形成分布模型或者判别函数, 用来检测图像中的人脸。归纳起来, 这类方法包括: 特征脸法 (Eigenfaces), 基于分布方法 (Distribution-based methods), 神经网络 (Neural networks), 支持向量机 (Support vector machine), 简单贝叶斯分类器 (Naive Bayes classifier), 遗传算法 (Genetic algorithm), 隐马尔可夫模型 (Hidden Markov model), 信息理论方法 (Information theoretical approach), 归纳学习 (Inductive learning) 等^[6]。基于统计的方法有很高的检测率, 但需要事先训练检测系统, 训练时间较长, 如何选择合适的训练样本和训练策略也是一个难题。

(3) **基于特征不变性** 单特征: 可作为区别于其它物体的人脸特征有: 人脸结构 (眼, 鼻, 嘴, 眉, 发及相互关系等)^[7,8], 人脸纹理^[9], 皮肤颜色^[10], 人脸轮廓形状等。

多特征: 近年来, 大量综合多种人脸特征的方法被用来检测和定位人脸。Yachida 采用模糊理检测彩色图像中的人脸^[4]。Chai 和 Ngan 采用 YCrCb 空间肤色模型提取侯选人脸目标, 通过一系列正则化操作来确认人脸^[2]。Menser 和 Bruning 采用和 Raja 一样的方法提取人脸目标, 采用形态学连接算子来简化目标, 最后分析目标的灰度分布来确认人脸^[11,12]; 该方法适合处理正面人脸的图像。Sobott 和 Pitas 的模型在 HSV 空间里, 他们精简目标的方法是椭圆匹配 (Ellipse fitting), 通过分析灰度图像来找出一些 ‘谷点’ (这些谷点被认为是眼眉鼻嘴等 ‘孔洞’ 位置), 最后判断 ‘孔洞’ 是否在椭圆区域内来确认人脸^[13], 该方法有很高的精确度但是椭圆匹配计算量很大, 而且由于需要 ‘谷点’ 信息保持完全, 因此只适合于检测正面且无旋转的人脸。

本文采用的方法属于基于特征不变性中的多特征检测方法。考虑到人眼的视觉特性, 采用 HSV 空间肤色模型是合适的, 同时为使该模型对光照有更强的适应性, 我们结合归一化 RGB 空间肤色模型, 构造了一个二者结合的混合肤色模型, 用它来从图像中获取侯选人脸目标。对每一个目标, 我们灵活地设计出一系列不同大小和形状的形态学结构元, 其中包括一个与目标尺寸大小成正比的结构元 (我们把它称为关联结构元), 以这些结构元设计算子对侯选人脸进行一系列形态学运算; 对于真正的人脸目标, 运算后其内部由眼睛嘴巴等形成的重要 ‘孔洞’ 得以保留, 噪声和其它干扰得以最大限度地去除。此方法最大的优点是可以检测出有大面积遮挡, 几乎任意大小, 侧面偏转可达 75 度角的人脸; 处理完成后我们输出的是人脸的轮廓 (对于正面人脸也可以确定出眼睛或嘴巴的位置), 可用于后续的跟踪或者编码。此外考虑到侧面人脸轮廓与椭圆相差较大, 我们的方法没有椭圆匹配过程, 计算量大大减少, 非常适合于快速人脸检测。

2 基于 HSV 与归一化 RGB 空间混合肤色模型的肤色区域获取

在杂乱无章的背景中, 人类皮肤的颜色可作为区别于其它物体的有效特征, 大量统计表明, 在同一人种内, 尽管不同的人有不同的肤色, 但是这种区别主要表现在亮度上而非色度上^[14]。为描述皮肤颜色, 人们提出了很多颜色模型: RGB 模型, HSV 模型, YIQ 模型, YcrCb 模型, YUV 和 LUV 颜色模型, 这些模型针对不同的应用场合有不同的效果。

HSV 颜色空间中亮度 (V) 和色彩信息 (H, S) 之间相互独立, 考虑到人眼对颜色比对亮度更敏感的特点, 采用 HSV 空间颜色模型更符合人眼的感知特性。在 HSV 空间中, 皮肤像素在 $H-S$ 平面上表现出明显的聚集特性, 图 1 是一张典型黄种人的人脸像素统计图, 可以看出绝大部分像素聚集在 $0 < H < 0.1, 0.3 < S < 0.7$ (归一化) 区域内。许多统计也表明, 在 HSV 空间中肤色象素满足高斯分布 (或混合高斯分布)^[10]。

类似于 Wang 和 Yuan 的方法^[15], 我们通过对 300 张典型条件下的人脸图像进行统计分析, 获得了 HSV 肤色模型的参数. 参考归一化 RGB 空间肤色模型, 构造出的混合空间肤色模型如下:

$$\begin{aligned} 0.005 < H < 0.10 & \quad 0.36 < r < 0.5 \\ 0.20 < S < 0.65 & \quad 0.28 < g < 0.363 \\ 0.60 < V <= 1 & \end{aligned}$$

其中 H, S, V 分别是 HSV 空间的 H 分量, S 分量, V 分量, 它们都已归一化. r, g 是归一化 RGB 空间的 r 分量和 g 分量.

实验表明这种模型比单一空间的模型对光照有更强的适应性. 图 2 展示了该模型提取类皮肤区域的效果, 可以看出, 在提取脸部区域的同时, 也提取出了胳膊以及其它和皮肤颜色相近的区域, 这需要在后续的过程中把它们剔除.

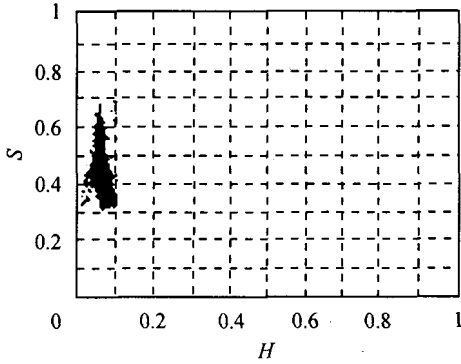


图 1 肤色像素 $H-S$ 平面分布



图 2 基于混合空间肤色模型提取类皮肤区域

3 数学形态学运算

形态学本是生物学的一个分支, 常用它来处理动物和植物的形态和结构, 后来 Euler, Steiner, Crofton, Minkowshi 等人把它当作数学研究内容, 称为数学形态学 (本文所说的形态学就是指数学形态学). 本世纪 80 年代, Serra 和 Matheron 在法国建立了枫丹白露数学形态学研究中心, 在随后的几年里, 他们逐步建立并且完善了数学形态学的理论体系, 此后, 他们又研究了基于数学形态学的图像处理系统^[16].

数学形态学定义了一些基本的数学运算, 运用这些基本运算的组合, 可以方便快速地完成提取目标边缘, 提取连接部分, 进行区域填充, 提取集合的凸壳, 细化, 粗化, 骨骼化目标等任务; 尤其重要的是, 在图像恢复中, 利用形态学方法可以有效去处噪声和其他干扰, 又可保留图像中有用的信息. $(B)_x$.

设 A, B 为 Z^2 中的集合, B 为结构元, 示意图如图 3; \emptyset 为空集, $(B)_x$ 表示将 B 中的元素平移 x 单位, 运算定义如下:

A 被 B 膨胀: $A \oplus B = \{x | [(B)_x \cap A] \neq \emptyset\}$

A 被 B 腐蚀: $A \ominus B = \{x | (B)_x \subseteq A\}$

开运算: $A \circ B = (A \ominus B) \oplus B$

闭运算: $A \bullet B = (A \oplus B) \ominus B$



图 3 结构元示意图：圆形半径为 2(像素) 和方形半径为 2(像素)

使用形态学算子处理图像的效果取决于结构元和形态运算的选取，本文根据人脸检测的需要选择了形状不同大小各异的结构元，设计出一系列形态学算子，分别用来去除噪声，分离候选目标，提取眼睛嘴巴‘孔洞’，以及提取人脸轮廓，我们分别叙述如下。

3.1 去除噪声和干扰算子

设 A 为肤色模型提取出的类皮肤区域，结构元 B_1 一般取为圆形，半径为 1 或 2(像素)，我们称它为简单结构元，做简单的闭运算和开运算：

$$(A \bullet B_1) \circ B_1 \quad (1)$$

经此处理，可去掉噪声和小块干扰，可加快检测速度，实验表明这一步骤是非常必要的。(算子的标号分别与下述处理流程中的标号相对应)

3.2 分离候选目标算子

若 O_f 是包含于类皮肤区域 A 中的某个目标，假设 O_f 中的一个点 P 已知，那么下述迭代可以得到 O_f 中所有元素：

$$X_k = (X_{k-1} \oplus B_2) \cap A, \quad k = 1, 2, \dots \quad (2)$$

其中 $X_0 = P$ ，为使收敛速度加快，结构元 B_2 可以取较大半径。如果得到 $X_k = X_{k-1}$ ，则表明算法收敛，并使 $O_f = X_k$ 。

3.3 提取眼嘴‘孔洞’算子

在任意一幅图像中，人脸以及人脸器官的大小可能相差很大，但是人脸大小和人脸器官的大小比例却相差不大，例如脸宽 / 眼珠半径保持在 22–31.8 之间，脸宽 / 唇高保持在 12–31.8 之间^[17]。因此可以考虑设计一个大小随待处理人脸候选目标大小改变而改变的结构元，我们称之为关联结构元。

设 B_f 代表人脸宽度， R_e 代表眼半径，它一般为眼珠半径，由于成像条件影响，实际取值略大于它。定义脸宽与眼半径之比：

$$C = B_f / R_e \quad (3)$$

根据文献 [17]， $C \in (20.3, 50.2)$ ， C 的统计平均值是 29.5。

为了更有效地去除噪声和干扰，结构元选择越大越好，但是为了把它用于下一步判断的眼嘴‘孔洞’，结构元也不能过大，它的半径有一个上限。由式 (3) 得到 $R_e = B_f / C$ ， $20.3 < C < 50.2$ 关联结构元 B_3 半径应小于 R_e 的最大值，即 $B_3 < B_f / 20.3$ 。

考虑到图像的人脸可能有旋转，可对每一个分离出来的候选目标，先计算它们所占矩形区域的宽和长，再把宽和长中较小的一个作为脸宽值 B_f 。然后对目标进行形态学闭，开运算：

$$(O_f \bullet B_3) \circ B_3 \quad (4)$$

算子处理的结果中既保留了眼嘴等‘孔洞’，同时又去掉了噪声和其他相关干扰。如图 4 所示，提取出来的‘孔洞’在图 4(c) 中以白点表示，上面的人脸经处理后保留了双眼和嘴‘孔洞’，下面的人手由于没有类似于眼嘴等较大器官而不会留下‘孔洞’。

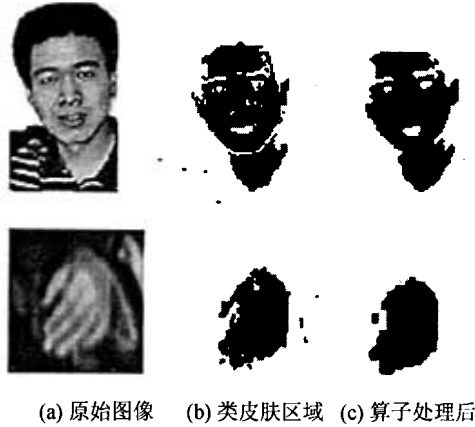


图 4 ‘孔洞’提取算子处理结果

3.4 提取人脸轮廓算子

设 O_f 是保留了眼嘴‘孔洞’的目标，若需要提取人脸的轮廓，可以先对 O_f 进行闭运算：

$$O'_f = O_f \bullet B_4, \quad E_f = O'_f - (O'_f \ominus B_5) \tag{5}$$

其中 B_4 是一个半径较大的圆形结构元，它的半径取值为 B_2 的 5 - 15 倍，这样取值可确保运算后眼嘴‘孔洞’也被剔除，只剩下整块的人脸目标 O'_f ，然后提取出 O'_f 的边缘 E_f ，它即是该目标所代表的人脸轮廓。 B_5 是半径为 1 的圆形结构元，‘-’符号代表集合运算中的差。

4 人脸检测流程

人脸检测的流程如图 5 所示。

整个流程分成 3 个大的部分：单个候选人脸目标的提取 (包括 (1), (2), (3), (4))，候选人脸目标的确认 (包括 (5), (6), (7))，真实人脸区域的定位 (包括 (8))。对于正面人脸 (也允许有稍许偏转)，若需要定位眼睛，嘴巴，可相应地在 (6), (7) 中完成，本文所处理的图像也包括偏转较大的人脸，定位步骤未给出。

(4) 候选人脸的分离还可以采用区域生长或者区域聚合等方法实现，本文采用数学形态学运算提取连接部分；(5) 根据人脸的几何尺寸关系，取脸宽 / 脸长值在 $[0.5, 1]$ 内为合法人脸目标，超出该范围的目标视为非人脸而被舍弃。(7) 采用包围盒算法判断孔洞是否在脸部区域内。经此流程处理后部分输出如图 6。

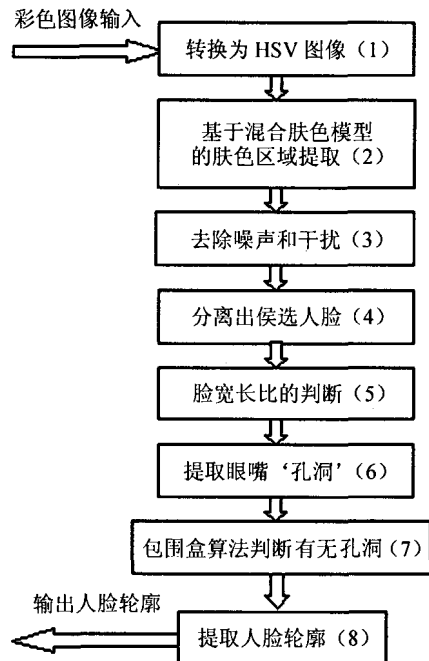


图 5 人脸检测流程图



图 6 实验结果示意

5 实验结果及分析

为构造 HSV 与归一化 RGB 空间混合肤色模型, 我们选取了 300 张不同成像环境的人脸照片, 从中选取 150 万个皮肤像素。通过分析这些皮肤像素的统计特性得到肤色模型的参数。各步骤中结构元的设计按表 1 确定。

表 1 结构元形状和大小说明

结构元	简单结构元 B_1	关联结构元 B_3	候选目标提取结构元 B_2	人脸轮廓提取结构元 B_4
形状	圆形 (或方形)	圆形 (或方形)	圆形 (或方形)	圆形
半径 (像素个数)	1 或 2	$B_f/C, C = 30$	5	$B_f/4$

我们的测试图像有 3 组, 第 1 组包括 36 幅, 主要取自 Internet 和扫描的照片。第 2 组包括 94 幅, 部分来源于欧洲 ACTS 项目支持的 M2VTS database, 部分是从 VCD 视频中抓拍的, 还有一部分取自 Internet, 绝大部分是室外自然背景下的照片, 大小从 200×160 到 600×450 像素之间不等。第 3 组是本实验室用 USB 摄像头采集的人脸库, 包括 40 个人的单人脸图像, 它们包括正常和喜怒哀乐等 5 种表情, 分别向上下左右偏转 30° 和 60° 等八种姿态, 每人 13 张, 一共是 520 幅。表 2 为用上述方法测试效果统计, 其中检测率是指: 成功检测出的人脸数和图像中出现的总人脸数的比值; 误检率是指: 被误检的臂手目标数与图像中出现的臂手目标总数的比值。

表 2 实验结果统计

内容	幅数	人脸数	检出数	检测率	臂手目标数	误检臂手数	误检率
第 1 组	36	51	47	92.1%	22	0	0
第 2 组	94	132	128	97.0%	110	4	3.6%
第 3 组	520	520	507	97.5%	---	---	---

部分检测结果如图 7。

实验表明, 尽管图像有着复杂的背景和不同的光照, 我们的方法检测率也较高, 3 组实验分别达到 92.1%, 97.0% 和 97.5%。误检率很低, 分别只有 0 和 3.6%(第 3 组中没有臂或手, 故不统计误检率); 如图 7 中 A 类图像所示, 被检测的人脸也不要求是正面的, 允许人脸在平面内做任意角度的旋转, 允许在深度方向有较大角度的偏转, 更多的实验结果显示, 允许的深度方向偏转角度可达 75° 。如 B 类所示, 由于有关联结构元构成的算子处理, 使得大小不一的人脸都能一一检测出来, 而对于图像中出现的臂和手等其他非人脸皮肤区域, 绝大部分都在处理过程中被有效地排除掉。如图 7 中 C 类图像, 尽管是有头发, 眼镜和衣服等遮挡住的人脸, 只要双眼和嘴巴 3 个器官中有一个完整地漏出, 该人脸就能被检测出来。



图 7 部分人脸检测结果

为测试该方法的检测速度, 我们使用了一台 P4, 主频 1.7G 的 PC 机编写 C++ 程序来统计检测时间, 采用的图像一共有 30 张, 它们是从以上 3 组图像中抽取的, 大小在 240×180 – 240×240 像素之间。测试的结果是图像的平均检测速度为 191.3ms/幅, 较之传统的基于区域分割的方法 3.67s/幅有了很大的提高, 较之 Wang 和 Yuan 的基于 Agent 的方法 413ms/幅也有很大提高 [15]。

漏检的人脸有 3 种情况, 一是人脸太暗或者有阴影, 二是人脸反光, 三是霓虹灯干扰; 误检的主要是一部分双手或双臂交叠的目标, 交叠区的阴影在目标区域产生了较大的‘孔洞’, 如果该区域的宽长比又恰好满足人脸宽长比, 就会出现误判。

实验中我们还注意到, 在该方法可以准确处理的图像中, 人脸颜色信息要求保持较完整, 这可能是我们的肤色模型还不够完善的原因, 应该构造更为精确的能适用于更加广泛的光照条件的模型, 这也是下一步研究的内容。

参 考 文 献

- [1] Tsapatsoulis N, Avrithis Y, Kollias S. Facial image indexing in multimedia databases. *Pattern Analysis & Applications*, 2001, 4(2/3): 234–241.
- [2] Chai D, Ngan K N. Face segmentation using skin-color map in videophone applications. *IEEE Trans. on Circuits and Systems for Video Technology*, 1999, 9(4): 551–564.
- [3] Chua Tat-Seng, Zhao Yunlong, Kankanhalli S. Detection of human faces in a compressed domain for video stratification. *The Visual Computer*, 2002, 18(2): 774–782.
- [4] Mao J, Yin B. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. *Pattern Recognition*, 1999, 32(7): 1237–1248.
- [5] Edwards G J, Taylor C J, Cootes T. Learning to identify and track faces in image sequences. Proc. Fifth IEEE Int'l Conf. Computer Vision, Chicago, USA, 1998: 317–322.
- [6] Yang Ming-Hsuan, Krieman J, Narendra Ahuja. Detecting faces in images: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, PAMI-24(1): 34–58.
- [7] Amit Y, Geman D, Jedynak B. Efficient focusing and face detection. Face recognition: From Theory to Applications, NATO ASI Series F, Berlin: Springer-Verlag, 1998, 163: 124–156.

- [8] Leung T K, Burl M C, Perona P. Probabilistic affine invariants for recognition. Proc. IEEE Conf. Computer Vision and Pattern Recognition, Chicago, USA, 1998: 678-684.
- [9] Dai Y, Nakano Y. Face-texture model based on SGLD and its application in face detection in a color scene. *Pattern Recognition*, 1996, 29(6): 1007-1017.
- [10] McKenna S, Raja Y, Gong S. Tracking color objects using adaptive mixture models. *Image and Vision Computing*, 1999, 17(3/4): 223-229.
- [11] McKenna S, Raja Y, Gong S. Object tracking using adaptive colour mixture models. In ACCV, HongKong, China, 1998, 1: 615-622.
- [12] Mensor B, Bruning M. Segmentation of Human Faces in Color Images. In Proc. Seventh IEEE Intl. Conf. on Image Processing and its Applications IPA'99, Kobe, Japan, 1999, 3: 652-636.
- [13] Sobottka K, Pitas I. Face location and feature extraction based on shape and color information. IEEE Int'l Conf. Image Processing, Lanssanne, 1996: 483-486.
- [14] Yang J, Waibel A. A real-time face tracker. Third Workshop Applications of Computer Vision, Sarasota, FL, USA, 1996: 142-147.
- [15] Wang Y J, Yuan B Z. A novel approach for human faces detection from color images under complex background. *Pattern Recognition*, 2001, 34(10): 1983-1992.
- [16] 阮秋琦. 数字图像处理学. 北京: 电子工业出版社, 2001年1月第一版, 第九章.
- [17] 阿尔文. R. 蒂利, 著, 朱涛, 译. 人体工程学图解. 北京: 中国建筑工业出版社, 1998: 30-25.

游亚平: 男, 1978年生, 硕士生, 主要研究方向是图像处理和模式识别.

袁保宗: 男, 1932年生, 北京交通大学信息科学研究所名誉所长, 教授, 博士生导师, 主要研究领域包括语音处理、计算机视觉、AVR、多模态信息处理.