

基于颜色和特征匹配的视频图像人脸检测实现技术^{*})

牛德蛟 詹永照 宋顺林

(江苏大学计算机科学与通信工程学院 江苏镇江212013)

Face Detection Using Skin-Color and Feature Matching

NIU De-Jiao ZHAN Yong-Zhao SONG Shun-Lin

(School of Computer Science and Communications Engineering, Jiangsu University, Jiangsu Zhenjiang, 212013)

Abstract A face detection method using statistical skin-color model and facial feature matching is presented in this paper. According to skin-color distribution in YUV color space, we develop a statistical skin-color model through interactive sample training and learning. Using this method we convert the color image to binary image and then segment face-candidate regions in the video images. In order to improve the quality of binary image and remove unwanted noises, filtering and mathematical morphology are employed. After these two processing, we use facial feature matching for further detection. The presence or absence of a face in each region is verified by means of mouth detector based on a template matching method. The experimental results show the proposed method has the features of high speed and high efficiency, but also robust to face variation to some extent. So it is suitable to be applied to real-time face detection and tracking in video sequences.

Keywords Video image, Face detection, Skin-color model, Feature matching

人脸是人类视觉交互中最重要的模式,人脸检测与识别是计算机视觉和人机交互研究领域的一个重要方向,它在个人辨识、访问控制、多功能感知、视频会议等领域有重要的应用。传统的人脸检测技术多用单一方法,一般要求背景简单,人脸姿态变化幅度不大。主要的方法有:模板匹配法^[1],特征脸法^[2],基于统计特征的方法^[3]及人工神经网络^[4]。其中,模板匹配法主要利用人脸模板在图像中搜索匹配,该方法运算量较大,检测效率低,一般不能满足实时性要求^[5];特征脸方法主要利用统计 K-L 变换将人脸模式空间降维,利用空间投影及特征距离进行判断。目前该方法主要对正面垂直姿态且固定大小的人脸检测取得了较好的结果^[2]。基于特征的方法是比从图像中抽取的一定特征的相似性,用匹配的方法进行检测,但是特征的提取有一定困难。而人工神经网络虽然检测效果较好,但训练需要大量的样本,计算复杂度较大。随着远程协作及人机交互技术的研究与开发,对人脸检测及跟踪的实时性要求越来越高。考虑到人的皮肤颜色分布与其他物体的颜色分布有较明显的差异,而且与物体的大小、伸缩及姿态基本无关,因此,采用人的皮肤颜色模型对人脸位置进行粗定位,就成为目前人脸检测与跟踪研究的主要方法^[5~8]。

本文主要研究视频图像中对特定人脸的检测与实时跟踪。我们根据这一特定要求,对人的皮肤颜色特征进行分析,利用人的皮肤颜色在归一化 YUV 空间分布的特点,提出了一种适合多种肤色的物体颜色特征建模方法。之后,运用模型匹配的原理,采用了一种快速、简单有效的在复杂背景下人脸位置的定位方法,为视频运动图像的人脸实时检测跟踪奠定基础。

1. 建立人脸肤色模型

研究表明,人的皮肤颜色具有以下三个特点:①人的皮肤颜色分布在颜色空间中一个小范围内。②人的皮肤颜色的不同主要是由颜色的强度引起,肤色区别主要受亮度影响,受色

度影响较小。③在一定的光照条件下,人的皮肤颜色在归一化 RGB 颜色空间符合高斯分布,即归一化后的三种颜色分量分别为 $a=R/I, b=G/I, c=B/I$, 且 $I=R+G+B$, a, b 分别为一高斯分量^[6]。如图1,2,3所示。

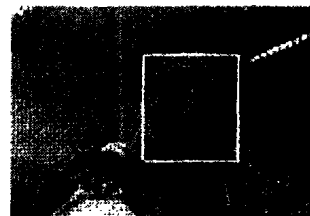


图1 采集到的彩色图像



图2 肤色区域在 RGB 空间的分布



图3 肤色区域在归一化 RGB 空间的分布

^{*})本文受到教育部科学技术研究重点项目和江苏省教育厅自然科学基金项目的资助。牛德蛟 硕士研究生,研究方向:图像处理、CSCW。詹永照 博士,教授,主要研究方向:CSCW、计算机图形学。宋顺林 教授,博导,主要研究方向:CSCW、软件工程。

另外,RGB 彩色空间不适于表征物体的颜色特征,因为空间的[R,G,B]值不仅代表色彩同时也表示亮度,RGB 三者间存在很大的相关性^[8],因此我们彩色图像颜色特征采用了 YUV 空间来表示。

由 RGB 空间向 YUV 空间转换的公式如下:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

其中,Y 分量表示颜色的亮度信息,两个色差信号 U 和 V 正交,成为色度信号,在二维平面空间中形成二维矢量,其模值 C_k 代表彩色的色饱和度,体现 R、G、B 的大小;而相位角 Φ 代表色调,体现 R、G、B 三者的比例。其中, C_k, Φ 定义如下:

$$C_k = \sqrt{U^2 + V^2} \quad \Phi = \tan^{-1}(\frac{V}{U}) \quad (2)$$

为了适合任何肤色的人脸检测,要对待检测的人脸图像进行学习,选取不同角度视频捕捉的人脸图像作为学习样本,交互标注出人脸范围,从而建立起肤色模型。

由于皮肤颜色具有空间聚类性,相对于整个空间而言仅是一个小的区域,利用颜色直方图的方法,根据高斯分布特点,可知二维直方图呈明显单峰,据此可建立人脸肤色模型。算法如下:统计采集到的彩色图像各像素点的(R,G,B)值,利用 RGB 彩色空间与 YUV 彩色空间的转换关系,得到像素点在 YUV 空间的对应像素值(Y,U,V),在平面二维 UV 坐标系中统计所得的数据,去除边缘不连续区域,保留 U、V 分布概率在最大值的 0.9~1.1 范围内的坐标点,就可以得到肤色在二维 UV 坐标系中的聚类区域。然后利用式(2),经过计算,将聚类区域的 U、V 值转换为相位角 Φ 值,建立人脸肤色的 Φ 模型。

为了验证该算法的正确性,我们选用标准人脸库中的图片进行实验,得到的肤色相位角 Φ 的分布保持在 104~150 之间,其结果与[6,8]等人的肤色模型完全相符。

需要注意的是:由于光照条件和采集设备的原因,肤色相位角 Φ 的分布会随光线而改变,例如在实验室灯光下,采集到一组图像的肤色 Φ 模型为: $\Phi \in [120^\circ, 165^\circ]$ 。它更适合于后继的人脸检测,因为在我们研究的实时视频人脸检测中,肤色建模与检测是两个连续过程,两者是在同一环境下进行的,因此采用我们的实时建立肤色模型并进行人脸检测的方法的鲁棒性好。

2. 基于颜色模型的人脸区域粗检测

根据肤色模型得到在一定光照下的肤色阈值,逐一对待检测彩色图像二值化,即用 $g_1(x,y)$ 表示图像(x,y)点处的二值化像素值,若 $g_1(x,y)=255$,表示白色像素点,若 $g_1(x,y)=0$,表示黑色像素点,则 $g_1(x,y)$ 计算式如下:

$$g_1(x,y) = \begin{cases} 255, & \Phi \in [120^\circ, 165^\circ] \\ 0, & \text{else} \end{cases} \quad (3)$$

得到的二值图像边缘较模糊,可能有两种误差,一种是由于噪声而引起的而被误判为肤色的孤立像素点或小区域;另一种则是由于人脸局部区域如眼睛、嘴等非肤色区域所导致整个人脸的非完全填充。首先应去除孤立像素点,可利用图像密度 $\rho(x,y)$ 在目标像素的领域内计算^[7]。若考虑 8 邻域,则 $\rho(x,y)$ 计算式如下:

$$\rho(x,y) = \frac{1}{8} \sum_{i=-1}^1 \sum_{j=-1}^1 g_1(x+i,y+j) - g_1(x,y) \quad (4)$$

若考虑 4 邻域,则 $\rho(x,y)$ 计算式为:

$$\rho(x,y) = g_1(x-1,y) + g_1(x+1,y) + g_1(x,y-1) + g_1(x,y+1) \quad (5)$$

利用此值对二值图进行平滑去噪处理,得到新的二值图像为

$$g_2(x,y) = \begin{cases} 255, & \rho(x,y) \geq t \\ 0, & \rho(x,y) < t \end{cases} \quad (6)$$

其中,t 的取法分别依测试目标像素的 8 邻域和 4 邻域而定,若使用 8 邻域,t 取值为 4×255 ;若是 4 邻域,t 取值为 2×255 。

去除噪声后的二值图像中大块区域的边缘仍较模糊,还需考虑区域合并问题。为了后继区域合并算法能得到正确的区域范围,再对图像进行数学形态学处理。其中腐蚀操作用于增强边缘,去除毛刺点。膨胀操作用于填充脸部非肤色区域,使整个脸部非填充区域尽可能变小。下式中 $\ominus$ 为腐蚀操作算子, $\oplus$ 为膨胀操作算子,结构元素 S 自定义为 3×3 数组。

$$\begin{cases} g_3(x,y) = g_2(x,y) \ominus S \\ g_4(x,y) = g_3(x,y) \oplus S \end{cases}, \text{其中 } S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

这样处理后的二值图像边缘较前清晰,但可能存在不连续区域,因此必须找到肤色区域的包围盒。我们采用二维空间投影的方法进行区域合并。具体算法如下:

1) 将形态学处理后的二值图像作水平方向的投影,H 表示其水平投影的结果,显然 H 为一维数组,数组长度即为二值图的宽度。

$$H(x) = \sum_{y=1}^N g_4(x,y) \quad x=1, \dots, M \quad (8)$$

其中,M,N 分别表示二值图像的宽度和高度。

2) 由水平投影图可以看出,投影图中突变的地方分别对应着候选区域的边界。对 $H(x)$ 逐个检测,每当出现从 0 变到非 0 或从非 0 变到 0 (即出现不连续区域) 时,记下该水平点所对应的 x 坐标值。如果是从 0 变为非 0,则记为白色起始点,如果是从非 0 变为 0,则记为白色终止点。为了排除毛刺点和不应出现的不连续中断点,考察前一白色终止点与下一白色起始点的距离,如果该距离大于某一阈值 α ,则认为出现了新的候选区域,否则认为该不连续由于干扰所致,应将这两区域合并。这样扫描完 $H(x)$ 的所有元素后可以得到各肤色候选区域在水平方向的起始和终止坐标,及水平方向的分区个数 horiNum。实验中取 $\alpha=5$ 。

3) 二值图水平方向的各个分区逐一进行如下操作:将该区域的图像作垂直方向的投影, V_{num} 表示其垂直投影的结果 ($num=1 \dots \text{horiNum}$), V_{num} 也为一维数组,数组长度为二值图像的高度。

$$V_{num}(y) = \sum_{x=1}^M g_4(x,y) \quad y=1, \dots, N \quad num=1, \dots, \text{horiNum} \quad (9)$$

每个 $V_{num}(y)$ 也按与 2) 类似的方法进行区域的划分与合并,最后可以得到各肤色候选区域垂直方向的起始和终止坐标。

4) 根据各个肤色候选区域的水平、垂直方向起始与终止坐标,记下原彩色图像上相应的矩形框,就得到候选区域的包围盒。



图4 采集到的彩色

按照上述方法进行的粗检测得到的结果图如图4~7所示。包围盒中可能有与人脸肤色相近的其他物体区域,如人的手臂等。为此,必须对粗检测到的人脸候选区域进行特征验证,以确定人脸区域。



图5 用肤色模型分割得到的二值图



图6 滤波及形态学处理后的二值图



图7 原彩色图像中找出的出肤色区域

3. 特征匹配

为了降低特征匹配的运算量和复杂度,我们首先检查候选区的形状,比例,结构和灰度分布等,淘汰形状显然与人脸比例不符的区域。

若一个区域长度/宽度比例范围介于0.8~2.5之间,则认为该区域可能为人脸区域,否则该区域认为非人脸区域。实验表明人脸上边界的定位很准确,但裸露的脖子也被框入待选区域中,如图8所示,导致人脸待选区域的长宽比与实际不符,所以我们将比例范围放宽到0.8~2.5。

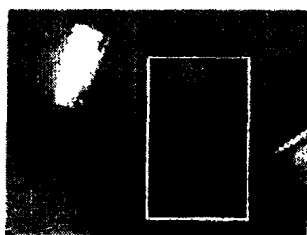


图8 裸露的脖颈也被框入

我们的人脸检测选取嘴作为一个显著特征,而不用眼睛来确定人脸区域,这样可以适用于戴眼镜的人。我们选用YCbCr彩色空间对唇色建模。由RGB空间向YCbCr空间转换的公式如下:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2990 & 0.5870 & 0.1140 \\ -0.1687 & -0.3313 & 0.5000 \\ 0.5000 & 0.4187 & -0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (10)$$

实验数据分析表明嘴唇区域的Cb值较小,而眼睛区域的Cr值较小,它更适合于特征区域的检测。例如:对某人的多幅嘴唇图像学习后,建立唇色模型为:Cb ∈ [-10, 8], & Cr

∈ [15, 32]。

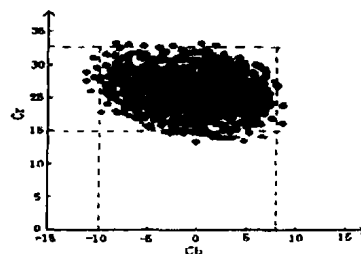


图9 唇色在 YCbCr 空间的聚类分布

为了加快检测速度,在待选区域内特征的检测范围可进一步缩小。我们知道嘴在人脸区域的下半部分,所以我们也只在待选区域的下半部分进行搜索检测。但如果待选区域中包含脖颈部分则可能导致嘴部位于待选区域的上半部分,无法检测到嘴部。由于一般人脸的长宽比约为1.3:1,因此可将待选区域由上到下的长度定为宽度的1.3倍,截去部分脖颈区域,从而缩小了特征的检测范围。

由于视频图像中人脸会有平移与转动,特征的大小会随之改变,传统的方法采用变形模板匹配,分别取不同比例尺寸的模板逐一匹配。我们则根据修正后待选区域的大小自适应调整模板的大小,选出最佳模板进行匹配,这样大大降低了时间和空间复杂度。初始模板在建立唇色模型时手工框出确定。我们在待检测区域内滑动与模板等大的窗口,若窗口内符合唇色模型的像素点数在一定阈值范围内,则认为检测到了嘴,匹配成功,否则,待选区域被认为非脸部区域。实验结果表明该方法能有效地排除人体其它部位或与人脸颜色相近的背景而形成的人脸待选区域,如手臂等。加特征匹配后,人脸检测结果如图10所示。

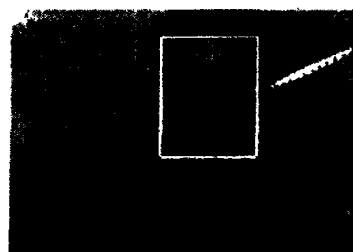
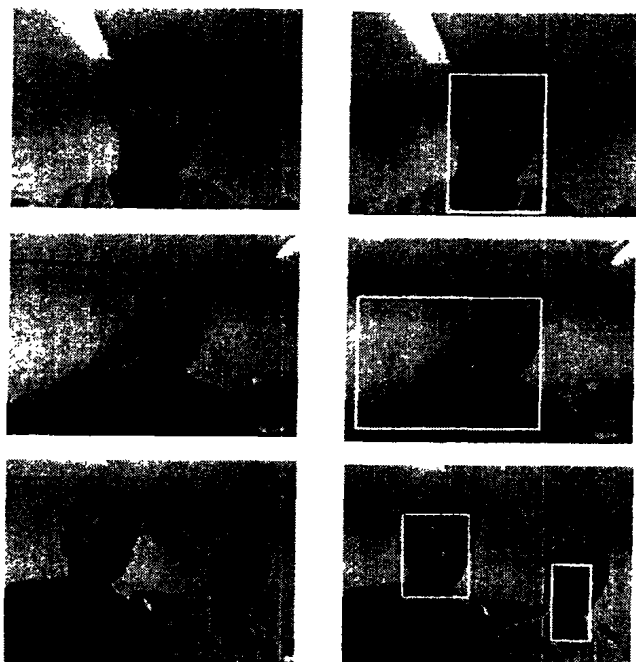


图10 特征检测后去除手臂

4. 试验结果

用本文介绍的方法对视频彩色图像进行了人脸检测,结果较满意。视频采集到的图像为BMP格式,大小为320×240像素/帧。在Pentium III PC机上使用Visual C++ 6.0编程实现,训练图像15幅。实验结果如图11所示,图11(a)为采集到的各帧彩色图像,然后分别用肤色模型分割、滤波和形态学处理后,去除大部分噪声,图11(b)为特征匹配后的各帧人脸区域。可以看到此方法对人脸的平移和转动鲁棒性较好。但我们 also 发现当人脸被手臂或其他物体遮挡后,导致检测到的人脸区域范围过大或检测不到人脸。当图像中有多个脸时,该方法仍然有效。

结论 本文提出了一种基于颜色和特征匹配的视频图像人脸检测方法。该方法根据人脸肤色分布特征建立肤色模型,利用肤色模型进行人脸区域的粗检测,然后针对待选区域利



(a)采集到的彩色图像

(b)检测到的人脸区域

图11 部分试验结果

用嘴唇这一特征进行特征匹配,找到最后的人脸区域。实验表

明,此方法执行速度快,鲁棒性好,检测结果正确率高,且对人的头部运动适应性好。此方法适合于视频实时图像中的人脸检测与跟踪,该方法的进一步改进还可用于一般物体的实时检测与跟踪。

参考文献

- 1 Yang G. Human face detection in a complex background. *Pattern Recognition*, 1994, 27(10): 53~63
- 2 Brunelli R, Poggio T. Face recognition: feature versus template [J]. *IEEE Trans PAMI*, 1993, 15(10): 1042~1052
- 3 Sung K, Poggio T. Example-based learning for view-based human face detection. *IEEE Trans Pattern Analysis and Machine Intelligence*, 1998, 18(1): 39~51
- 4 Rowley H A, Baluja S, Kanade T. Neural network-based face detection. *IEEE Trans Pattern Analysis and Machine Intelligence*, 1998, 20(1): 23~28
- 5 艾海舟, 梁路宏, 徐光佑, 张钹. 基于肤色和模板的人脸检测. *软件学报*, 2001, 12(12): 1784~1790
- 6 康学雷, 邵凌, 张立明. 一种基于肤色和模板的人脸检测方法. *红外与毫米波学报*, 2000, 6: 209~214
- 7 Chai D, Ngan K N. Face segmentation using skin-color in video-phone application. *IEEE Trans. CSVT*, 1999, 9(4): 550~566
- 8 姚鸿勋, 刘明宝, 高文. 基于彩色图像的色系坐标变换的面部定位与跟踪法. *计算机学报*, 2000, 2(2): 158~165

(上接第144页)

- 16 Chen C-Y, Klette R. Image Stitching-Comparisons and New Techniques. *CAIP*, 1999. 615~622
- 17 Kang S B, Weiss R. Characterization of errors in compositing panoramic images. In: *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'97)*, San Juan, Puerto Rico, 1997. 103~109
- 18 Gumustekin S. A Introduction to Image Mosaicing. <http://likeya.iyte.edu.tr/>, July, 1999
- 19 Sawhney H S, Kumar R. True multi-image alignment and its application to mosaicing and lens distortion correction. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 1997. 450~456
- 20 Peleg S, Herman J. Panoramic mosaics by manifold projection. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 1997
- 21 Rousso B, Peleg S, Finci I. Mosaicing with generalized strips. In: *DARPA Image Understanding Workshop*, May 1997
- 22 Russo B, Peleg S, Finci I, Rav-Acha A. Universal mosaicing using pipe projection. In: *Proc. of IEEE Intl. Conf. on Computer Vision*, 1998
- 23 Rademacher P, Bishop G. Multiple-center-of-projection images. In: *Proc. of ACM SIGGRAPH*, 1998
- 24 Reddy B S, Chatterji B N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Processing*, 1996, 5(8): 1266~1271
- 25 Xie Hongjie, et al. Automatic image registration based on FFT algorithm and IDL/ENVI. In: *Intl. Conf. on Remote Sensing and GIS/GPS-ICORG*, 2001
- 26 Badra F, Qumsieh A, Dudek G. Rotation and Zooming in Image Mosaicing. In: *IEEE WACV*, 1998. 50~55
- 27 Richmond J. Image Mosaicing <http://www.cs.ubc.ca/~jrichmo/>, 1999
- 28 Blaszkia T, Rachid Deriche Recovering and Characterizing Image Features Using An Efficient Model Based Approach: [Technical Report 2422]. INRIA, Nov. 1994
- 29 Zoghalmi I, Faugeras O, Deriche R. Using Geometric Corners to Build a 2D Mosaic from a Set of Images. In: *Proc. of the Intl. Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, IEEE, 1997
- 30 Nielsen F. Adaptive Randomized Algorithms for Mosaicing Systems. *IEICE TRANS. INF. & SYST.*, VOL. E83 CD, 2000(7)
- 31 Bao P, Xu D. Complex Wavelet-based Image Mosaics using Edge-Preserving Visual Perception Modeling. *Computers and Graphics*, 1999, 23(3): 309~321
- 32 Kang E, Cohen I, Medioni G. A graph-based global registration for 2D mosaics. In: *Proc. of the 15th Intl. Conf. on Pattern Recognition*, Barcelona, Spain, 2000
- 33 Sawhney H S, Kumar R G, Bergen J, Dixon D, Paragano V. Videobrush™: Experiences with consumer Video Mosaicing. In: *IEEE WACV*, 1998. 56~62
- 34 Shum H-Y, Szeliski R. Panoramic Image Mosaics, Microsoft Research: [Technical Report MSR-TR-97-23]
- 35 WONG, Chun Kit Wilbur YANG, Ruiduo ZHANG, Yan. Image Mosaics under Simple Planar Scenes and its Application, A Final Report Submitted to The Course COMP641F "Introduction to Image Processing" at The Hong Kong University of Science and Technology, 2001