

基于 OpenCV 的口唇检测研究

房昭菊¹, 郭朋²

(1. 电子科技大学电子工程学院, 成都 611731; 2. 四川理工学院自动化与电子信息学院, 四川 自贡 643000)

摘要:唇读技术在提高噪音环境中语音识别系统的识别率和帮助有听力障碍的人更好的理解语言方面起到了一定的辅助作用。由于说话过程中口型变化较大,直接提取口型区域较困难;而在发音过程中鼻子形状不会发生较大变化,且 OpenCV 自带的人脸检测器可以很好的检测出人脸。提出了一种利用嘴唇与人脸和鼻子的相对位置关系进行检测嘴唇区域的方法。实验证明,该方法可以比较快速,精确地提取出嘴唇区域,从而利于精确的进行口唇分割。

关键词:OpenCV; 口唇跟踪; 唇读; adaboost 算法; haar 特征

中图分类号:TP391

文献标识码:A

引言

考虑到语音识别系统在噪音比较大的环境中识别率较低,或对于听力方面有障碍的人来说,依靠观察口型可有助于理解语言,故唇读技术对于提高语音识别和聋哑人的语言理解能力起到了一定的辅助作用。准确的提取出人脸和嘴唇,对获取能较好的区分各类口型的特征点起到了重要作用,从而对提高语音识别率或提高聋哑人的语言理解能力起到了至关重要的作用。

在文献[1]中提出了排红法,针对唇色和肤色相似,但红色分量占有很大比例的特点,可用 G 和 B 分量进行阈值分割来提取出口唇形状,然而由于视频获取到的不仅仅是肤色和唇色,所以需要提取出嘴唇的周围区域以便尽量减少非唇色的比例。本文采用 Intel 公司开发的开源的计算机视觉库(OpenCV)进行口唇定位和跟踪。

1 Adaboost 算法

OpenCV^[2]提供了基于 adaboost 级联的 Haar 分类器。Adaboost 算法^[3]提供了一个高效快速的特征提取机制,它是一个迭代过程,每一层分类过程都使用一个弱分类器^[4] $h_t()$,最小化

$$Z_t = \sum_i D_t(i) \exp(-y_i h_t(x_i)) \quad (1)$$

其中, $D_t(i)$ 是第 t 层的第 i 个样本的权重, $y_i \in \{-1, 1\}$ 是样本的目标标签, x_i 是样本, $h_t()$ 是弱分类器。

每一层分类之后权重 $D_t(i)$ 都被更新为:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-y_i h_t(x_i))}{Z_t} \quad (2)$$

分类器 $h_t()$ 有两个可能值 $h_+ = \log \sqrt{\frac{W_{++}}{W_{+-}}}$ 和 $h_- = \log \sqrt{\frac{W_{--}}{W_{-+}}}$, 此处, W_{pq} 是实际为 q 类而给出 p 类的样本的权重,这个预测确保下一层分类时分类界限两边的正样本和负样本是均衡的。

最小化 Z_t 即最小化第 t 层迭代的加权指数。而最小化每一层分类中的 Z_t 也是最小化强分类器的训练错误率的上界 $\prod Z_t$ 的一种贪婪算法。因此 adaboost 算法的主要优点是学习速度快,如果把能与其它非目标物体区分开的主要特征形成的结构作为首要的分类器即可去除许多的假样本,从而达到提高训练速度的目的。

Adaboost 算法示意图如图 1 所示:

2 Haar 特征

由上述所示,adaboost 算法仅仅提供了分类器的级联方式,而如果需要进行口唇检测还需要选取能区分其

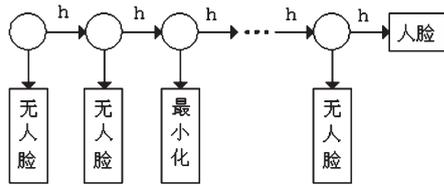


图 1 Adaboost 算法示意图

它非样本的特征。OpenCV 中选择使用类 Haar 特征^[5]。在假设可以快速计算窗口中任意竖直或 45 度旋转角方向矩形中的像素和的情况下,类 Haar 特征可表示为:

$$feature_i = \sum_{i \in I = (1, \dots, N)} \omega_i \cdot RecSum(r_i) \quad (3)$$

类 Haar 特征模型共有 14 种,分别为 4 种边缘特征,如图 2 所示;8 种线特征,如图 3 所示;和 2 个中心特征,如图 4 所示:



图 2 边缘特征

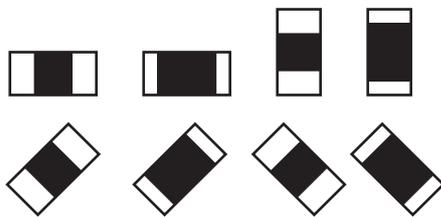


图 3 线特征



图 4 中心特征

3 编程实现口唇定位及分割

由于在发音过程中口型变化较大,如 a 音和 i 音口型差别明显较大,不能根据嘴唇在脸部的比例进行很好的定位,而以往方法是仅取脸部的三分之一作为嘴唇区域进行分割,此时的三分之一区域或许会包含鼻孔等非肤色和非唇色区域,对较好的嘴唇定位产生影响。该实验提出使用 OpenCV 中的 adaboost 级联的 Haar 分类器进行人脸检测和鼻子检测,由于鼻子在口型变化时相对比较稳定,故可采用检测出的鼻子和人脸对口唇的位置进行估计,使用检测到的脸和鼻子的区域位置以及实验得到的嘴唇相对人脸和鼻子的位置比例关系可以比较精确地得到嘴唇的区域位置。

在进行人脸检测^[6-7]和鼻子检测时,可以自己收集大量的人脸和鼻子数据进行训练,但由于 OpenCV 训练出的人脸特征和鼻子特征已经达到较好的效果,故本实验中采取 OpenCV 中已训练好的 haarcascade_frontalface_alt2.xml 和 haarcascade_mcs_nose.xml 分别检测人脸和鼻子,部分代码如下:

```

cascade =
    (CvHaarClassifierCascade * ) cvLoad( " haarcascade_
    frontalface_alt2.xml" ); //加载分类器
faces =
    cvHaarDetectObjects ( small_image, cascade, storage,
    1, 2, 2, CV_HAAR_DO_CANNY_PRUNING ); //检测人脸
    
```

若检测出的人脸区域为 CvRect 类型的 face_rect 值,而鼻子的矩形区域为 nose_rect 值,则实验证明,根据以上两值可以较精确的跟踪出嘴唇的矩形区域 lip_rect 为:

$$lip_rect = cvRect (nose_rect.x - nose_rect.width/13, nose_rect.y + nose_rect.height * 9/11, nose_rect.width * 19/14, face_rect.y + face_rect.height * 10/11 - nose_rect.y - nose_rect.height * 2/5);$$

由此,根据以上方法对口唇进行定位,可以尽可能的去掉唇色和肤色之外的其他颜色,而且矩形框可以较好的包围在口唇周围。此时再利用基于颜色的图像分割方法即可将口唇精确的提取出来,在文献[8]中提出利用 fisher 准则,即类内离散度比较低而类外离散度比较大的特点对口唇进行分割,其步骤如下:

- (1) 计算检测出的矩形区域的均值 mean;
- (2) 根据公式:

$$cor = \frac{\sqrt{\sum_{i=1}^{width} \sum_{j=1}^{height} k(i,j) - mean}}{(width * height - 1)} \quad (4)$$

计算检测出的矩形区域的协方差,其中 width 和 height 分别为口唇区域的宽和高, $k(i,j)$ 为口唇区域第 i 行第 j 列的像素值,mean 为(1)中的均值;

- (3) 利用计算出的均值和方差之和作为域值进行域值分割,可以很好的提取出嘴唇。

4 检测结果及分析

使用 Adaboost 级联的 Haar 分类器对下面四幅图像进行口唇检测,其检测的矩形区域如图 5 所示:

检测出的矩形区域如图 6 所示:

检测出的嘴唇区域如图 7 所示:

利用上述检测出的嘴唇区域对其进行口唇分割,其

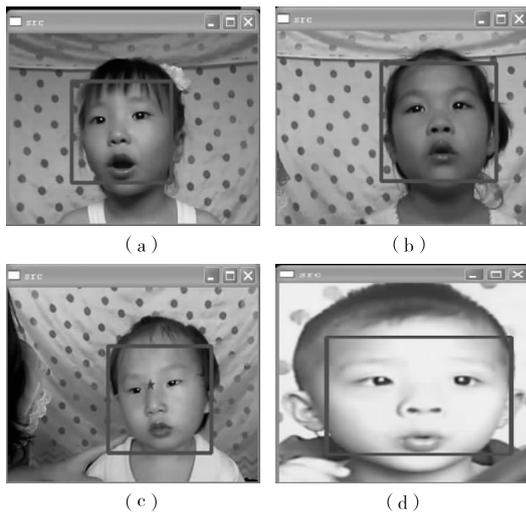


图5 原图像中的检测区域

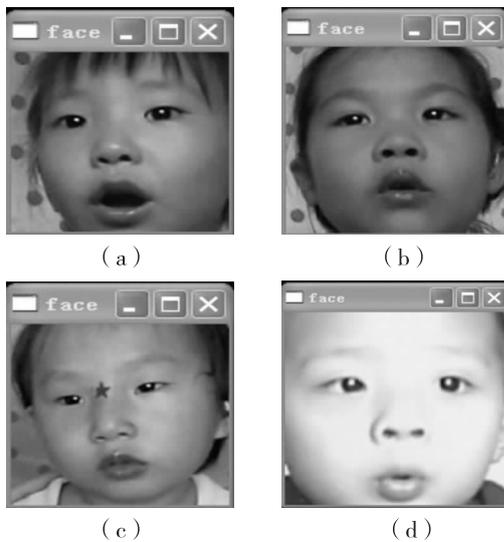


图6 人脸区域

效果如图8所示:

由图8我们可以清楚的看出,该实验取得了比较好的分割效果。此处使用了 OpenCV 训练好的数据,如若想得到更好的效果,可以采集大量的人脸图像和鼻子图像对其进行训练,从而可提高嘴唇定位的精确度。

5 结束语

本文提出了利用 OpenCV 中 adaboost 级联的 Haar 分类器检测出的人脸区域和鼻子区域进行较好的定位口唇区域的方法,并讲述了该方法中使用的 adaboost 算法的基本原理和粗略介绍了利用类内离散度低、类外离散度高的特点进行较好的口唇分割方法,同样由此分割出的口唇形状可以较精确的对其进行轮廓提取作为可变形模板^[9]的初始值进行较精确的跟踪动态视频中的

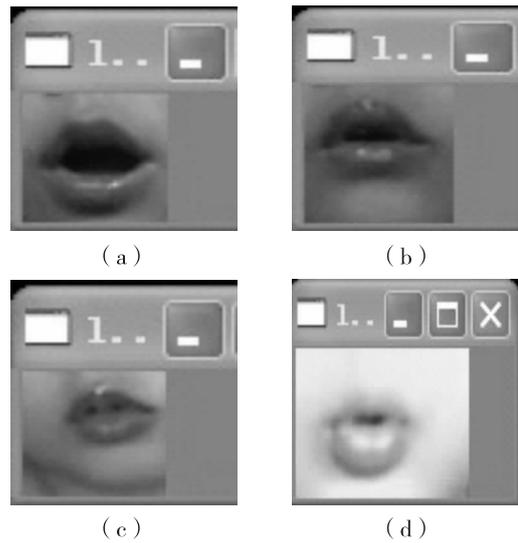


图7 嘴唇区域

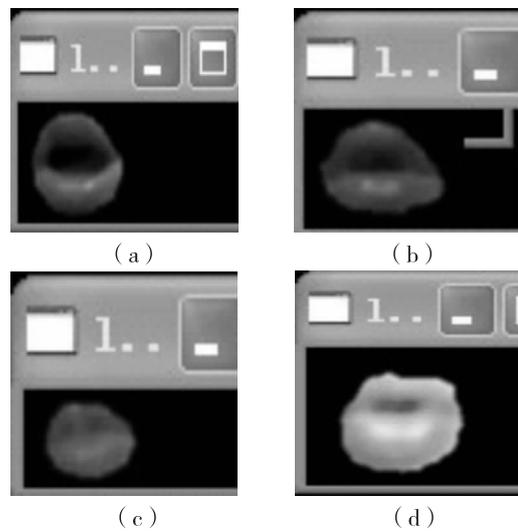


图8 嘴唇分割的效果图

口唇形状。实验结果证明,本方法的口唇检测分割效果比较好。

参考文献:

- [1] Trent W Lewis, David M W. Powers. Lip Feature Extraction Using Red Exclusion[J]. Pan-Sydney Workshop on Visual Information Processing, 2000, 12: 61-67.
- [2] Gary Bradski, Adrian Kaehler. Learning OpenCV[M]. O'Reilly Media, Inc. 2008.
- [3] 赵楠. 基于 Adaboost 算法的人脸检测[D]. 北京大学, 2005.
- [4] Wu Jian Xin, Brubaker S C, Mullin D M, et al. Fast asymmetric learning for cascade face detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelli-

- gence,2008,30(3):369-382.
- [5] Rainer Lienhart,Jochen Maydt.An Extended Set of Haar-like Features for Rapid Object Detection[J].International Conference on Image Processing,2002,1:900-903.
- [6] 邓亚峰,苏光大,傅博.一种基于 Adaboost 的快速动态人脸检测算法[J].计算机工程,2006,32(11):222-224.
- [7] 潘志庚,邹鹏程,梁荣华,等.基于特征人脸和肤色统计的人脸检测[J].系统仿真学报,2004,16(6):1346-1349.
- [8] 黄永慧,潘保昌,梁坚,等.一种自适应唇区检测及定位方法[J].计算机工程与应用,2010,46(21):17-20.
- [9] Huang Yong-Hui,Pan Bao-Chang,Zheng Sheng-Lin,et al. Lip-reading detection and localization based on two stage ellipse fitting[J]. Wavelet Analysis and Pattern Recognition,2008:168-171.

Research on Lip Detection Based on OpenCV

FANG Zhao-ju¹, GUO Peng²

(1. School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China;

2. School of Automation and Electronics Information, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: Lip-reading systems play an important role to improve the recognition rate of the speech recognition systems on the noise environment and to help the disabled in hearing communicate with others. The shape of lip has obvious changes when somebody is pronouncing, so it is difficult to directly detect the lip region; on the contrary, the nose shape has not obvious changes, and the face detection in OpenCV software has achieved considerable effect on detection face, so a method of detection lip is presented which adopts the relative position of lip against to face and nose to detect the lip region. The experiments show that this method can quickly and efficiently extract the lip region, and help to improve the accuracy of lip segmentation.

Key words: OpenCV; lip tracking; lip-reading; adaboost algorithm; haar features