

文章编号 1004-924X(2006)03-0473-04

采用非对称唇形轮廓模型提高汉语唇形识别效果

李 刚,王蒙军,林 凌

(天津大学 精密仪器与光电子工程学院,天津 300072)

摘要:通过分析人脸正面和侧面图像的特点,建立起一种新的唇形轮廓描述模型,能够提取出嘴唇的突出度信息,同时通过计算部分参数对时间的导数,来获得唇形轮廓的动态信息。通过对小词库试验,对比单纯采用正面图像轮廓的描述方法,该方法能够将识别效果平均提高 25% 以上,克服了以往对唇形轮廓提取时,通常都采用人脸的正面图像,不能反映出嘴唇的突出度信息的局限。

关键词:唇形轮廓;对称模型;非对称模型;差分特征

中图分类号: TP391.43 **文献标识码:** A

Improving Chinese lip-reading recognizing rate by unsymmetrical lip contour model

LI Gang, WANG Meng-jun, LIN Ling

(School of Precision Instrument and Optoelectronics Engineering,
Tianjin University, Tianjin 300072, China)

Abstract: Based on analyzing the side-face image and full-face image, a new model was presented to extract the degree of pouting from a lip contour. At the same time, the differential coefficient of some parameters to describe dynamic characteristic of the lip contour were calculated. Experimental results based on a small database of Chinese words show that the parameters from unsymmetrical lip contour model improve the recognizing rate in more than 25%, which is superior to those of the traditional symmetrical lip contour model.

Key words: lip contour; symmetrical model; unsymmetrical model; difference parameters

1 引 言

由于人的面部信息在安全、认证、人机交互等方面的所具有的广泛应用前景而越来越成为计算机模式识别领域的研究重点^[1],特别是近十几年来,唇读技术已经成为人机交互领域中讨论研究

的热点,该技术是将人讲话时嘴唇区域唇动的视觉信息作为补充,来增强自动语音识别效果。大量的实验研究表明,无论是在有无噪声的情况下,采用唇动信息(甚至是面部信息),都可以明显提高和改善语音识别的效果。可见,通过唇形识别技术是可以理解或是部分理解讲话者讲话内容的^[2-5]。研究唇读技术对基于唇动信息的人体生

收稿日期:2005-12-20;修订日期:2006-02-14.

物特征识别、视频音频传输压缩编码、汉语听觉视觉双模态数据库的建立,以及语音噪声敏感音素的辨识等各项工作都具有积极意义^[2,4]。唇形与其动态特征在人类对语言的理解过程中有着重要的作用,通过对唇形序列进行分析来识别语言的音节,完全可以应用于聋哑人与正常人交流辅助工具的研究。为将唇读技术应用于康复助残工程,本文讨论建立一个唇形识别的交互式系统,该系统能够根据说话人的唇动过程,判断说话内容,再依靠语音合成技术,还原成声音,从而能够部分还原那些后天致残但仍具有正确唇动特征残疾人的语言表达能力。

在研究唇读技术时,主要有两方面的内容:

(1)唇形轮廓特征的描述和提取;(2)唇形轮廓特征的理解和识别。这里唇形轮廓特征的描述和提取是唇形轮廓特征理解和识别的前提条件^[3-4]。

以往的研究者在提取唇形轮廓的时候,大部分采用的是人脸的正面图像,它是一个对称式的模型。通过分析人讲话时唇形的动作过程,本文认为:人在讲话过程中,不但人脸正面图像中反映出来的嘴唇的内外轮廓会发生动态变化,而且人脸侧面图像所反映出来的嘴唇突出度,下腭突出度也会发生动态变化,参考文献[6]中也阐述了这一观点。针对以上特点,采用了一种新的非对称式的嘴唇轮廓描述方法,使得该轮廓中既有正面图像中嘴唇的内外轮廓动态变化信息,又包含人脸侧面图像所反映出来的嘴唇突出度和下腭突出度动态变化信息。

2 对称唇形轮廓模型分析

2.1 正面唇形轮廓模型分析

以往的研究者在描述唇形轮廓的时候,大都采用了人脸的正面图像,其基本的特征点选择如图 1。

其中 a_1 、 a_2 为嘴唇的两个嘴角点,线段 $|a_1 a_2|$ 反映了嘴唇的宽度; b_1 、 b_2 、 b_3 、 b_4 分别为上嘴唇上轮廓中点、上嘴唇下轮廓中点、下嘴唇上轮廓中点、下嘴唇下轮廓中点,线段 $|b_1 b_4|$ 反映了外唇线轮廓的高度,线段 $|b_2 b_3|$ 反映了内唇线轮廓的高度,曲线 $a_1 b_1 a_2$ 、 $a_1 b_2 a_2$ 、 $a_1 b_3 a_2$ 、 $a_1 b_4 a_2$ 分别描述了上嘴唇上轮廓线、上嘴唇下轮廓线、下嘴唇上轮廓线、下嘴唇下轮廓线, d 为下腭点^[8-10]。

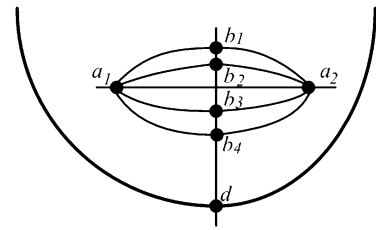


图 1 对称唇形轮廓模型

Fig. 1 Symmetrical lip contour model

从以上分析可以看出,正面图像虽然完整的刻画了整个嘴唇的大部分信息,但是不能反映出嘴唇突出度,下腭突出度信息。

2.2 侧面嘴唇轮廓模型分析

参考文献[6]考虑了采用人脸的侧面图像作为信息的来源,即在拍摄获取发音者的口型图像序列时,通过在紧靠面部侧面放置一面镜子的方法,同步记录发音时的正面和侧面图像。但是在实际应用中放置镜子的方法有时带来不便。如果采用两个摄像装置采集图像,图像序列间的同步会带来很大的问题。故从中得到启示,设计了下面的模型。

3 非对称唇形轮廓描述模型

通过分析发音者嘴唇轮廓的特点,正常人完整的唇形轮廓基本上是左右对称的。在汉语发音过程中,口形运动主要在上下、左右方向上,嘴唇轮廓左半部分和右半部分的运动基本上相互对应,不会出现扭曲和不对称的情况。基于以上的条件,综合设计了下面的唇形轮廓描述模型,如图 2。

该模型建立的方法是:在获取人脸图像时,摄像头从人脸正面转过一个角度,这样既能保证一半正面嘴唇轮廓的图像,又能反映出嘴唇突出度和下腭突出度的投影信息。

图中 A_1 、 A_2 为嘴唇的两个嘴角点,线段 $|A_1 A_2|$ 反映了嘴唇的宽度的投影; B_1 、 B_2 、 B_3 、 B_4 分别为上嘴唇上轮廓中点、上嘴唇下轮廓中点、下嘴唇上轮廓中点、下嘴唇下轮廓中点,线段 $|B_1 B_4|$ 反映了外唇线轮廓的高度,线段 $|B_2 B_3|$ 反映了内唇线轮廓的高度,曲线 $A_1 B_1$ 、 $A_1 B_2$ 、 $A_1 B_3$ 、 $A_1 B_4$ 分别描述了上嘴唇上轮廓线、上嘴唇下轮廓

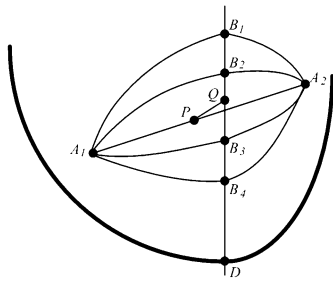


图2 非对称唇形轮廓模型

Fig. 2 Unsymmetrical lip contour model

线、下嘴唇上轮廓线、下嘴唇下轮廓线的半幅信息;D为下腭点。特别注意的是,以 $|A_1A_2|$ 的中点P和 $|B_2B_3|$ 的中点Q的线段 $|PQ|$ 是嘴唇突出度空间距离的投影,可以用它代替嘴唇突出度的信息。另外通过 $|PD|$ 能够反映出下腭突出度信息。

4 轮廓特征选择

根据以上的基本思想,从上面的模型中选择几种参数来描述唇形轮廓的特征。包括有:线段 $|A_1A_2|$ 反映的嘴唇的宽度的投影信息 W ;线段 $|B_1B_4|$ 反映的外唇线轮廓的高度信息 H_1 ;线段 $|B_2B_3|$ 反映的内唇线轮廓的高度信息 H_2 ;曲线 A_1B_1 、 A_1B_2 、 A_1B_3 、 A_1B_4 分别描述了上嘴唇上轮廓线、上嘴唇下轮廓线、下嘴唇上轮廓线、下嘴唇下轮廓线的半幅信息,该类信息利用边界傅里叶变换,用傅里叶描述子来刻画;以及嘴唇突出度空间距离 $|PQ|$ 的投影反映的嘴唇突出度信息 F_1 和线段 $|PD|$ 反映出下腭突出度信息 F_2 。

另外,由于讲话过程中人的嘴唇是连续变化的过程,并且在整个汉字发音的过程中,变化的快慢程度不一致。在汉语拼音中,一个汉字就是一个音节,该音节由声母和韵母组成,在发音时,声母持续的时间很短,迅速转到韵母的发音过程,整个过程中口型变化与时间有很大的关系^[11]。为了在特征中反映出这一时间变化特点,选取了能够反映此类变化的特征,考察它们对时间的导数关系,这其中包括 dW/dt 、 dH_1/dt 、 dH_2/dt 、 dF_1/dt 、 dF_2/dt 等差分特征参数。

在实验时通过手工标定的方法来跟踪各点的

位置,没有对唇形轮廓进行追踪。

5 实验结果分析

根据以上的模型,采用自己建立的基于常见汉字的小字库,对小字库中的30个常见汉字发音过程的图像序列进行分析处理,所有图像分辨率为 320×240 ,30帧/s拍摄,每个字由实验室的10名同学分别发音10次,其中的8次作为样本集合来对隐马尔可夫模型(HMM)进行训练,2次用来验证。

这里采用隐马尔可夫模型(HMM)来识别唇动过程是由于隐马尔可夫模型(HMM)是一个双重的随机过程,这与人的语言唇动过程是相吻合的。唇动信息本身就是一个可观察的序列,它是头脑中(不可观察的)根据言语需要和语法知识(状态选择)所发出的音素(词句)的参数流。隐马尔可夫模型(HMM)的基本原理在参考文献[12]中有表述。一个隐马尔可夫模型(HMM)是一个双随机序列,包括状态序列和可观察值序列,其中(1) N :模型中隐含的马尔可夫链状态数目。记 N 个状态为 $1, 2, \dots, N$, t 时刻马尔可夫链所处的状态为 q_t ,其中 $q_t \in (1, 2, \dots, N)$ 。(2) M :每个状态对应的可能的观测值数目。记 M 个观察值为 v_1, v_2, \dots, v_m , t 时刻观测到的观察值为 O_t ,其中 $O_t \in (v_1, v_2, \dots, v_m)$ 。

通过实验发现,采用非对称唇形轮廓描述模型后,利用嘴唇突出度信息 F_1 和下腭突出度的信息 F_2 ,以及 dW/dt 、 dH_1/dt 、 dH_2/dt 、 dF_1/dt 、 dF_2/dt 等差分特征参数之后,识别的效果得到极大地提高。对于隐马尔可夫模型(HMM)状态数 N 和可能的观测值数 M 选取不同的值,选取不同的唇动特征组合后,识别效果的对比结果如表1所示。

通过表格中数据的对比,可以看到,引入嘴唇突出度信息 F_1 和下腭突出度信息 F_2 ,以及 dW/dt 、 dH_1/dt 、 dH_2/dt 、 dF_1/dt 、 dF_2/dt 等差分特征之后识别的效果较单纯采用嘴唇的宽度的信息 W ,外唇线轮廓的高度信息 H_1 ,内唇线轮廓的高度信息 H_2 的识别效果可以提高25%以上。

表 1 HMM 模型下不同特征参数识别结果(%)

Tab. 1 Recognizing rate in different parameters based on HMM

HMM 模型 参数选 取	马尔夫可夫链状 态数目 N	可能的观测值 数目 M					
		4	4	4	5	5	
唇动特 征参数选取	不同的 HMM 模型特征参数组合所对应的不同的识别结果(%)	WH_1	38.2	38.7	39.5	38.8	39.3
		WH_2	38.2	38.5	39.3	39.0	39.2
		$WH_1 F_1$	50.2	51.5	53.3	52.3	51.4
		$WH_1 F_2$	49.5	50.1	51.1	50.9	50.5
		$WH_1 F_1$	55.2	56.4	58.9	56.1	58.3
		$dW/dt dH_1/dt$	54.8	56.2	58.5	56.4	57.9
		$WH_2 F_2$	64.3	65.4	66.4	64.5	65.8
		$dW/dt dH_1/dt dF_1/dt$	64.2	65.9	66.2	64.9	65.3
		$WH_2 F_2$	65.3	68.5	69.3	65.7	68.1
		$dW/dt dH_2/dt dF_1/dt$	65.1	68.9	69.2	65.2	67.9
		$WH_1 F_1$					
		$dW/dt dH_1/dt dF_2/dt$					
		$WH_2 F_2$					
		$dW/dt dH_2/dt dF_2/dt$					

外唇线轮廓的高度信息 H_1 , 内唇线轮廓的高度信息 H_2 对识别结果的影响大致相同。嘴唇突出度信息 F_1 , 下腭突出度信息 F_2 对识别结果有影响, 但差别不是很明显, 在多汉字大字库下的影响还需进一步验证。

6 结 论

根据人讲话时唇动信息的特点, 提出了一种新型的非对称唇形轮廓描述模型, 使得该轮廓模型中既有人脸正面图像中嘴唇的内外轮廓动态变化信息, 又包含人脸侧面图像反映出来的嘴唇突出度和下腭突出度动态变化信息。通过实验证明, 该模型能够获取更多的唇动信息, 提高唇形识别效果。将来要在大字库下, 综合考虑其它唇形区域信息, 如唇形轮廓曲线的傅里叶描述子, 唇间灰度等信息的基础上进一步验证该模型, 提高汉语唇读识别效果。

参考文献:

- [1] 梁毅雄, 龚卫国, 潘英俊, 等. 基于奇异值分解的人脸识别方法[J]. 光学 精密工程, 2004, 12(5): 543-549.
LIANG Y X, GONG W G, PAN Y J, et al. Singular value decomposition-based approach for face recognition[J]. *Optics and Precision Engineering*, 2004, 12(5): 543-549. (in Chinese)
- [2] WANG R, YAO H X, GAO W. Recognition of sequence lip images and its application[C]. *IEEE Fourth International Conference on Signal Processing*, 1998, (): 849-854.
- [3] MATTHEWS I, COOTES T F, BANGHAM J A, et al. Extraction of visual features for lip reading[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(2): 198-213.
- [4] SCANLON P, REILL Y R. Feature analysis for automatic speech reading[C]. *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, Page(s): 625-630.
- [5] ZHANG X, MERSEREAU R M, CLEMENTS M, et al. Visual speech feature extraction for improved speech recognition[C]. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002, 2: 1993-1996.
- [6] 王志明, 蔡莲红. 汉语语音视位的研究[J]. 应用声学, 2002, 21(3): 29-34.
WANG Z M, CAI L H. Study of Chinese viseme[J]. *Applied Acoustics*, 2002, 21(3): 29-34. (in Chinese)
- [7] SEGUIER R, CLADEL N. Multiobjectives genetic snakes: application on audio-visual speech recognition[C]. *Fourth EURASIP Conference focused on Video/Image Processing and Multimedia Communications*, 2003: 625-630.
- [8] 姚鸿勋, 高文, 李静梅, 等. 用于口形识别的实时唇定位方法[J]. 软件学报, 2000, 11(8): 1126-1132.
YAO H X, GAO W, LI J M, et al. Real-time lip locating method for lip-movement recognition[J]. *Journal of Software*, 2000, 11(8): 1126-1132. (in Chinese)
- [9] CHANDRAMOHAN D, SILSBEE P L. A multiple deformable template approach for visual speech recognition[C]. *Fourth International Conference on Spoken Language*, 1996, 1: 50-53.
- [10] 王磊, 莫玉龙, 戚飞虎. 基于弹性模板的嘴巴轮廓提取[J]. 上海大学学报(自然科学版), 1998, 4(5): 579-585.

- WANG L, MO Y L, QI F H. Mouth feature extraction based on deformable template[J]. *Journal of Shanghai University (Natural Science)*, 1998, 4(5): 579-585. (in Chinese)
- [11] 晏洁. 文本驱动的唇动合成系统[J]. *计算机工程与设计*, 1998, 19(1): 31-34.
YAN J. Text-driven lip motion synthesis system[J]. *Computer Engineering and Design*, 1998, 19(1): 31-34. (in Chinese)
- [12] 王波波, 来忠信, 黄廉卿, 等. 基于多尺度隐马尔可夫模型的 CR 影像降噪方法研究[J]. *光学精密工程*, 2000, 10(2): 188-193.
WANG B B, LAI Z X, HUANG L Q, et al. MHMM-based computed radiography image denoising method[J]. *Optics and Precision Engineering*, 2002, 10(2): 188-193. (in Chinese)

作者简介:李 刚(1959 -),男,博士,天津大学精密仪器与光电子工程学院教授,主要研究领域为生物医学工程,信息检测与处理,嵌入式计算机及其应用,主持和参与了 30 多项国际合作、国家自然科学基金、省部委以及企业合作项目。