

# 一种基于 LDAO 的唇读特征提取算法\*

何俊, 李赣平

(南昌大学信息工程学院, 南昌 330031)

**摘要:** 在语音与唇读识别应用中,传统的 LDA(linear discriminant analysis)算法一般以音节、半音节、HMM 状态等基元为类别进行数据分段,经线性判别分析后获得的特征投影方向与识别率不直接相关,影响了识别率。提出了一种新的基于 LDAO(linear discriminant analysis based on object)的唇读特征提取算法,该算法以待识别对象为类别进行线性判别分析,在理论上保证了唇读特征矢量向最具判别能力的方向投影。基于唇读数据库的实验证明,该算法明显优于现有各种唇读特征提取算法,比 DCT + LDA 算法识别率提高了 3%。

**关键词:** 唇读; 特征提取; 线性判别分析算法; 基于对象的线性判别分析算法

**中图分类号:** TP391.1      **文献标志码:** A      **文章编号:** 1001-3695(2013)01-0288-03

doi:10.3969/j.issn.1001-3695.2013.01.074

## Feature extraction based on LDAO algorithm in lipreading

HE Jun, LI Gan-ping

(School of Information Engineering, Nanchang University, Nanchang 330031, China)

**Abstract:** In speech and lipreading recognition application, LDA(linear discriminant analysis) algorithm is usually based on syllable, semi-syllable, HMM state or other class units. But the extracted features based on traditional LDA have no direct relation to recognition accuracy. This paper proposed linear discriminant analysis based on object (LDAO) algorithm on recognizing isolated words in lipreading. It selected objects to be recognized as class to LDA, which ensured feature extracting followed the most discriminant directions among objects in theory. Experiments on bimodal database show that this algorithm is superior to any other feature extracting algorithms in lipreading. Specifically, the recognition accuracy is better than DCT + LDA algorithm about 3%.

**Key words:** lipreading; feature extracting; LDA algorithm; LDAO algorithm

### 0 引言

唇读,即利用视觉通道获取视觉语言信息,主要用来提高噪声环境下的语音识别率。唇读特征提取技术是唇读识别的关键环节,一般包括三种方法<sup>[1]</sup>:a)基于形状的方法,主要提取嘴唇的几何形状、模型参数等高级特征<sup>[2-5]</sup>;b)基于像素的方法,主要是利用图像线性变换的方法提取嘴唇图像纹理等低级特征<sup>[6-11]</sup>;c)混合特征提取法,其原理是结合前两种方法的优点,提取的特征同时包括形状特征与灰度特征。形状法的优点是特征维数低,语义明确;缺点是主观选择的这些高级特征丢失了过多的语言信息,总体识别率较低。混合法的优点是结合了形状法与像素法的优点,如 AAM<sup>[12]</sup>(active area model);其缺点是训练过程过于复杂,而且匹配过程太耗时,不适用于实时人机交互场合。像素法的优点是利用了 ROI(region of interesting)全部像素,语言信息完整;缺点是特征矢量维数高、冗余度高,且对旋转、照度、尺度变化敏感。

在约翰霍普金斯大学暑期论坛上,对各种唇读特征提取技术进行了对比,在相同数据库上的实验表明像素法的唇读识别率最高,误识率为 58.14%,包含 30 维形状与灰度的 AAM 法误识率为 65.66%<sup>[13]</sup>。如图 1 所示。Potamianos 等人提出的基于像素的唇读特征提取过程如图 1 所示,共包含了三次线性

变换:a)每帧图像进行离散余弦变换(discrete cosine transformation,DCT)并保留 24 维较大的 DCT 系数;b)均值减归一化后以该帧为中心连接前后 7 帧共 15 帧图像共同作为新的特征矢量,再以 HMM 状态为类进行 LDA 变换;c)利用最大似然线性回归(maximum likelihood linear regression,MLLR)对特征矢量作自适应处理。在唇读特征提取中 Potamianos 借鉴了成功应用于语音识别的 LDA 算法。是 Haeb-Umbach 等人<sup>[14]</sup>最早将 LDA 算法应用到语音识别,其基本思想是以德语的亚音素为类别,将前后若干帧矢量连接形成超帧,语音识别率有了显著提高。根据语言种类不同陆续有文献将音素、音节、半音节、HMM 状态等基元作为 LDA 的类别<sup>[15,16]</sup>,由于 Viterbi 算法能自动搜索 HMM 的最佳状态转移序列,在传统 LDA 算法中以 HMM 状态为类别的 LDA 效果最好。

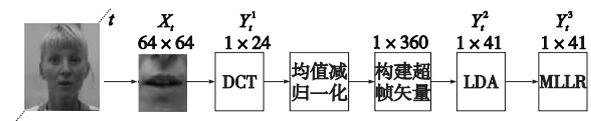


图1 基于DCT+LDA的唇读特征提取示意图

按以上各种类别进行 LDA 获得的投影主轴并不与识别率直接相关,即原始特征矢量向这些主轴投影后,只能保证各个观察特征矢量类间距离最大、同时类内距离最小,但这并不意味着最好的识别率。因此,本文提出了一种新的 LDAO 算法,其基本

收稿日期: 2012-05-02; 修回日期: 2012-06-14      基金项目: 国家自然科学基金资助项目(61064004)

作者简介:何俊(1969-),男,江西东乡人,副教授,博士,主要研究方向为模式识别(boxhejun@tom.com);李赣平(1972-),男,讲师,博士,主要研究方向为过程控制、模式识别。

思想是当识别对象有限时(如命令词),以所有待识别对象为类别进行 LDA,这样获得的投影主轴代表了各个待识别对象具有最好分类效果的方向,从而在理论上保证了最优的识别率,在特定人唇读数据库上的实验证明了该算法的优异性能。

### 1 LDAO 算法原理

LDA 依据式(1)所示的 fisher 准则函数,目的是找到最具判别能力的降维投影主轴方向  $W$ ,原始特征矢量在这个主轴方向进行投影降维。对于多类分类,  $S_b$ 、 $S_w$  分别表示训练样本的总体类间散布矩阵和总体类内散布矩阵,按以上各种类别进行 LDA 获得的投影主轴并不与识别率直接相关,即原始特征矢量向这些主轴投影后,只能保证各个观察特征矢量类间距离最大、同时类内距离最小。

$$J(W) = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|} \quad (1)$$

传统 LDA 类别的选择方向与 LDAO 类别选择方向示意图如图 2 所示。

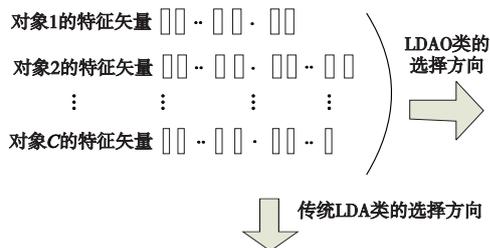


图2 传统LDA类别的选择方向与LDAO类别选择方向示意图

假设共有  $C$  个待识别对象,即共  $C$  类。从语音或者唇读识别角度看,语音与唇读识别实际是个分类问题,若按传统 LDA 类别选择方向,无论以音素、亚音素还是半音节等基元为类别进行 LDA,获得的投影方向  $W$  能使属于这些基元的特征矢量之间具有最好的分类性能,但并不意味在对象识别上最优;其次,根据最大后验概率原理,最优的分类准则应为

$$k^* = \arg \max_{1 \leq k \leq C} P(M_k | X) = \arg \max_{1 \leq k \leq C} \frac{P(X|M_k)P(M_k)}{P(X)} \quad (2)$$

其中: $C$  为识别对象总数;若用 HMM 作为识别模型,  $M_k$  是第  $k$  个对象的 HMM;  $X$  是特征矢量序列。由于 HMM 的 Baum 迭代训练算法是次优的,即训练出来的模型参数只能保证输出该特征序列的概率最大,并不能保证各识别对象的分类效果最好,故传统 LDA 算法并不能在理论上保证最优的分类能力。

与传统 LDA 算法的类基元选择方向不同, LDAO 算法以每个待识别对象为类,训练类间与类内散布矩阵  $S_b$ 、 $S_w$  获得投影主轴方向  $W$ ,向该方向投影降维则意味着待识别对象之间有最好的判别能力,即最好的分类效果。以下是基于 LDAO 算法的唇读特征提取过程。

设有  $C$  个待识别汉字的唇读数据,每帧 ROI 大小为  $24 \times 24$ , DCT 变换后用 Zig-Zig 法保留 55 个较大的 DCT 系数,每个汉字规整为相同长度的 30 帧。因为每个汉字  $X$  的特征维数  $d$  ( $55 \times 30$ ) 远大于样本数量  $N$ ,导致矩阵  $S_w$  奇异,即  $\text{rank}(S_w) \leq N - C$ 。因此在 LDA 之前必须对特征矢量  $X$  降维,可利用 PCA (principle component analysis) 算法,即利用样本  $X$  构造产生矩阵,得到投影方向矩阵  $W_{\text{PCA}}$ ,根据式(3)将  $d$  维特征  $X$  降低到  $d_1$  ( $d_1 \leq N - C$ ) 维特征  $Z$ ,即

$$Z_k = W_{\text{PCA}}^T X_k \quad k = 1, 2, \dots, N \quad (3)$$

在 PCA 子空间,对低维  $Z_k$  应用 LDA 算法提取特征并进一步降维。对新样本  $Z_k$  定义各类总的类内散布矩阵和类间散布矩阵分别为

$$S_w = \frac{1}{N} \sum_{i=1}^C \sum_{k=1}^{N_i} (x_{ik} - \mu_i)(x_{ik} - \mu_i)^T \quad (4)$$

$$S_b = \frac{1}{N} \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T \quad (5)$$

其中: $N$  表示总样本数,  $C$  表示类别数,  $N_i$  表示第  $i$  类样本的数量,  $\mu_i$  表示第  $i$  类样本的均值,  $\mu$  表示总体样本均值。LDA 特征提取是从  $d_1$  维空间向  $C - 1$  维空间投影为

$$Y_k = W_{\text{LDA}}^T Z_k \quad k = 1, 2, \dots, N \quad (6)$$

使式(1)中准则函数  $J(W)$  最大解出投影矩阵  $W$ 。  $J(W)$  是广义的 Rayleigh 熵,其值可以用 Lagrange 乘法求解,因  $S_w$  矩阵非奇异,求解  $W$  等效为求解式(7),也就是求出矩阵  $S_w^{-1} S_b$  的特征值。解出矩阵  $W_{\text{LDA}}$  后,可按照特征值大小顺序排列,选择前  $M$  个特征值对应的特征矢量构成维数更低的投影矩阵  $W^*$ 。

$$S_w^{-1} S_b W_{\text{LDA}} = \lambda W_{\text{LDA}} \quad (7)$$

在完成训练并获得变换矩阵  $W_{\text{PCA}}$  与  $W_{\text{LDA}}$  后,进行唇读特征提取时只要将原特征矢量代入式(8)进行两次线性变换。

$$Y_k = W_{\text{PCA}}^T W_{\text{LDA}}^T X_k \quad k = 1, 2, \dots, N \quad (8)$$

### 2 LDAO 算法实现

因 LDAO 算法提取的特征维数很低而且相同,可不必用 HMM 这种复杂的分类模型,用简单的欧式距离方法即可。如式(9)所示,可获得每个测试数据的欧式距离,距离最小的即为识别结果。

$$\text{dist}(k) = W^T \times X - \mu_{\text{tran}} \quad (9)$$

其中: $X$  是待测试特征矢量,  $W$  是变换矩阵  $W_{\text{PCA}}^T W_{\text{LDA}}^T$ ,  $\mu_{\text{tran}}$  是变换后的数据均值,  $k$  是类索引。但欧式距离以均值为判据,忽略了各个类的方差,用高斯混合模型更为合理,因为高斯混合模型本质上是一种多维概率密度函数。一个具有  $M$  个混合成分的  $D$  维 GMM 的表达式为

$$P(x) = \sum_{i=1}^N w_i \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)\right] \quad (10)$$

其中: $\mu_i$ 、 $\Sigma_i$  代表第  $i$  个高斯元的均值和协方差;  $N$  是高斯混元的数量;  $w_i$  表示第  $i$  个高斯混元的权值,一个词的模型  $\lambda$  可表示为  $\lambda = (w_i, \mu_i, \Sigma_i)$ ,  $i = 1, 2, \dots, M$ 。各个高斯密度函数可表示不同人的唇读特征,实现非特定人唇读识别。模型的训练可采用 EM(expectation maximization) 迭代算法<sup>[17]</sup>,可获得 GMM 中的权值、均值与方差的迭代式为

$$\bar{w}_i = \frac{1}{T} \sum_{t=1}^T P(i|x_t, \mu_i, \Sigma_i) \quad (11)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T P(i|x_t, \mu_i, \Sigma_i) x_t}{\sum_{t=1}^T P(i|x_t, \mu_i, \Sigma_i)} \quad (12)$$

$$\bar{\Sigma}_i = \frac{\sum_{t=1}^T P(i|x_t, \mu_i, \Sigma_i) (x_t - \bar{\mu}_i)(x_t - \bar{\mu}_i)^T}{\sum_{t=1}^T P(i|x_t, \mu_i, \Sigma_i)} \quad (13)$$

模型参数训练好后,对未知样本的识别只需将其代入各个模型计算输出概率,概率最大的即为识别结果。

### 3 实验结果与分析

实验采用的数据库为自制双模态唇读数据库,该数据库包含 30 人,识别任务是 10 个命令字与数字 0~9,每个字发音 12 遍。该数据库的特色是直接拍摄 ROI,避免了 ROI 定位带来的误差,ROI 大小为  $160 \times 160$  的 24 bit 真彩图像。数据训练采用交叉训练法,即每次实验将其中 6 遍用于训练,剩下的 6 遍发音用于识别。在进行 LDAO 特征提取之前所有 ROI 图像均下采样为  $24 \times 24$ 。分别进行了两项实验:a) LDAO 算法与目前唇读识别率最好的 DCT + LDA 算法进行识别率比较;b) 检验 LDAO 算法与其他特征提取算法的降维性能。实验平台为 MATLAB 7.0 + VC 6.0,在 VC 上完成图像的采集,在 MATLAB 上实现算法。训练数据的获取是利用 Premiere 软件,根据语音波形与唇读图像的对应关系将每个字的发音图像帧切分出来。因 LDAO 算法只能处理固定维数的特征,而各个字发音的帧长度不同,故需将每个字规整为相同帧长度。按正常语速一个字的发音时间小于 1.2 s,帧率为 25 fps 时,则发音的最大总帧数小于 30 帧,在应用 LDAO 算法时,本文事先将所有字的发音帧数采用基于样条的插值法规整到 30 帧。

为可视化 LDAO 提取的唇读特征矢量,取 LDAO 投影矩阵前两个最大特征值对应的特征矢量为坐标轴,图 3 是对 5 个命令字“前、后、左、右、曲”基于 LDAO 提取的唇读特征在两个坐标轴上的投影示意图,显示出了很好的分类能力。

#### 3.1 LDAO 算法与传统 LDA 算法比较实验

为检验 LDAO 算法性能,本文分别采用了以半音节为类与以状态为类的 DCT + LDA 算法与之进行比较。ROI 图像 DCT 后,传统 LDA 的训练和识别都基于连续 HMM,每个汉字用 6 个 HMM 状态,每个状态包含 2 个混合高斯元。基于 LDAO 算法的训练与识别采用高斯混合模型。

以半音节为类的训练方法是每个汉字按 1:2 比例分为前半音节和后半音节,将图像帧分配给对应的前半音节和后半音节,再将这些帧归属到相应的半音节训练出类内与类间散布矩阵,求出投影方向矩阵  $W$ 。以 HMM 状态为类的训练方法是先将图像帧平均分配给六个状态,利用 Baum 算法训练 HMM,用 Viterbi 算法找出属于各个状态的图像帧,然后基于这些 HMM 状态为类的数据训练出投影方向矩阵  $W$ 。为克服 HMM 帧间不相关的假设,以当前图像帧为中心连接前后若干帧构成超帧特征矢量。

如表 1 所示,以 HMM 状态为类的 LDA 最高识别率为 93.8%,略高于以半音节为类时的最高识别率 93.6%,因为以 HMM 状态为类时是利用 Viterbi 算法获得最佳状态序列对应的特征帧,说明 Viterbi 算法得到的半音节特征帧比手工按比例切分出的半音节对应的特征帧更准确。当增加超帧数量时,二者识别率均略有增加。当超帧数量增加到 3 时,识别率有明显增加,之后随着超帧数量的增加识别率并无明显增加,说明相邻帧之间确实存在相关性,但相隔较远的帧之间相关性逐渐减弱。与以上两种传统的 LDA 算法相比,实验显示了 LDAO 算法的优异性能,虽然 Potamianos 在 DCT、LDA 后又采用了 MLLR,但 MLLR 的作用是实现非特定人唇读特征的自适应调整,本实验是基于特定人,故对算法没有影响。总体而言,LDAO 算法识别率比以 HMM 状态为类的 LDA 算法识别率提高了约 3%。

表 1 LDAO 算法与 DCT + LDA 算法的唇读识别率比较

算法	超帧数量			
	1	3	5	7
以半音节为类的 LDA	93.2%	93.5%	93.6%	93.6%
以 HMM 状态为类的 LDA	93.6%	93.8%	93.8%	93.8%
LDAO	96.7%			

#### 3.2 LDAO 算法降维性能实验

算法的降维性能关系到系统的实时性,图 4 是 LDAO 算法与 DCT、PCA 以及 LDA 算法降维性能的比较示意图。当采用 PCA 法提取特征时,最高识别率为 6 维静态特征加上 6 维动态特征时的 89%;当采用 DCT 法提取特征时,最好识别率为 27 维静态特征加上 27 维动态特征时的 93%,略高于 PCA 法;当基于以 HMM 状态为类且超帧数为 3 的 LDA 算法时,最好识别率是特征维数为 18 时的 93.8%。以上三种算法给出的特征维数是某个对象(汉字)其中单帧或者超帧的特征维数,若要表示对象的特征维数,以上三种算法提取的总特征维数还要乘以帧数。而基于 LDAO 算法,当对象特征维数为 4 时识别率就达到了 94.5%,当特征维数增加到 8 时识别率就达到了 96.7%,之后随着特征维数的增加识别率基本保持不变。实验证明了 LDAO 算法不仅分类性能优异,而且降维效果非常好。

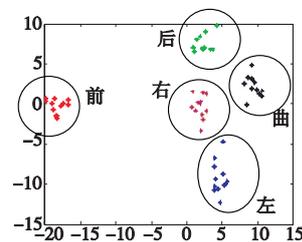


图 3 基于 LDAO 提取的唇读特征分类能力示意图

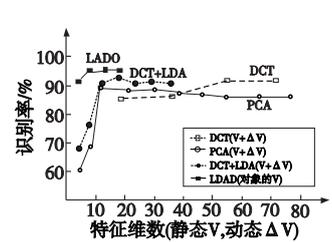


图 4 LDAO 算法与其他特征提取算法降维性能比较

在模型复杂度方面,DCT、PCA 以及传统的 LDA 均采用 HMM 进行训练与识别,而 LDAO 算法采用高斯混合模型,简化了训练与识别时间,提高了实时性。以上实验证明,LDAO 算法不仅识别率明显高于其他算法,而且特征维数低,节省了存储空间,实时性好。

### 4 结束语

针对唇读技术中视觉语言特征提取,本文提出了一种新颖的适用于孤立词识别的 LDAO 唇读特征提取算法。不同于传统的以音节、状态等基元为类别的 LDA 算法,LDAO 以各个待识别对象为类进行 LDA,在理论上保证了 LDA 获得的方向是最优的分类方向,克服了传统 LDA 算法特征分类方向与对象识别不相关的缺陷。在特定人数据库上 LDAO 算法分别与 DCT、PCA 以及传统 LDA 算法进行了唇读识别实验比较,结果表明 LDAO 算法不仅特征维数低,而且识别率明显优于现有各种唇读特征提取算法。

#### 参考文献:

- [1] POTAMIANOS G, NETI C, GRAVIER G, et al. Recent advances in the automatic recognition of audio-visual speech[J]. Proceedings of the IEEE, 2003, 91(9): 1306-1326.
- [2] CHAN M T, ZHANG Y, HUANG T S. Real-time lip tracking and bimodal continuous speech recognition[C]//Proc of the 2nd IEEE Workshop Multimedia Signal Processing. 1998: 65-70.

(下转第 294 页)

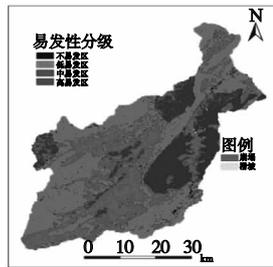


图11 研究区地质灾害易预测分布图

### 4 结束语

本文以汶川地震数据为背景,利用遥感与 GIS 等技术,首先对地质灾害信息进行了提取,并结合研究区域的高程、坡度、坡向、地层岩性、地震烈度、水系等因子对研究区地质灾害空间分布特征进行了分析,最后采用层次分析法对研究区地质灾害易发性进行了预测。实验结果表明,本文方法对该地区的震后安置工作具有重要指导意义,不仅适用于汶川的地质区域,通过适当的改善也可以运用到其他的地质区域。

#### 参考文献:

[1] 王涛,马寅生,龙长兴,等. 四川汶川地震断裂活动和次生地质灾害浅析[J]. 地质通报,2008,27(11):1913-1922.  
 [2] 崔鹏,韦方强,何思明,等. 5·12 汶川地震诱发的山地灾害及减灾措施[J]. 山地学报,2008,26(3):280-282.  
 [3] 杨武年,廖崇高,濮国梁,等. 数字区调新技术新方法——遥感图

像地质解译三维可视化及影像动态分析[J]. 地质通报,2003,22(1):60-64.  
 [4] 刘传正,李铁锋,温铭生,等. 三峡库区地质灾害空间评价预警研究[J]. 水文地质工程地质,2004,31(4):9-19.  
 [5] 刘传正,李铁锋,温铭生,等. 长江三峡库区地质灾害成因与评价研究[M]. 北京:地质出版社,2007.  
 [6] WANG H B, SASSA K, XU W Y. Analysis of a spatial distribution of landslides triggered by the 2004 Chuetsu earthquakes of Niigata Prefecture, Japan[J]. *Natural Hazards*,2007,41(1):43-60.  
 [7] SATO H P, HARP E L. Interpretation of earthquake-induced landslides triggered by the 12 May 2008, M7.9 Wenchuan earthquake in the Beichuan area, Sichuan province, China using satellite imagery and Google Earth [J]. *Landslides*,2009,6(2):153-159.  
 [8] 许强,李为乐. 汶川地震诱发滑坡方向效应研究[J]. 四川大学学报:工程科学版,2010,42(5):7-14.  
 [9] 程思,易加强. 四川省汶川县地质灾害的成因及防治对策[J]. 地质灾害与环境保护,2007,18(4):1-6.  
 [10] 黄润秋. 汶川地震地质灾害研究[M]. 北京:科学出版社,2009.  
 [11] LEE C T, HUANG C C, LEE J F, et al. Statistical approach to earthquake-induced landslide susceptibility[J]. *Engineering Geology*,2008,100(1/2):43-58.  
 [12] KAMP U, GROWLEY B J, KHATTAK G A, et al. GIS-based landslide susceptibility mapping for the 2005 Kashmir earthquake region [J]. *Geomorphology*,2008,101(4):631-642.

(上接第 290 页)

[3] WANG S L, LIEW A W C, LAU W H, et al. An automatic lipreading system for spoken digits with limited training data [J]. *IEEE Trans on Circuits and Systems for Video Technology*,2008,18(12):1760-1765.  
 [4] GURBUZ S, TUFECKI Z, PATTERSON E, et al. Application of affine-invariant Fourier descriptors to lipreading for audio-visual speech recognition [C]//Proc of International Conference on Acoustics, Speech, and Signal Processing. 2001:177-180.  
 [5] KASS M, WITKIN A, TERZOPOULOS D. Snakes: active contour model [J]. *International Journal of Computer Vision*, 1988, 1(4):321-331.  
 [6] HE Jun, ZHANG Hua. Research on visual automatic speech recognition [J]. *Journal of Information and Computational Science*, 2008,5(4):1567-1575.  
 [7] WU Guan-yong, ZHU Jie. An extension 2DPCA based visual feature extraction method for audio-visual speech recognition [C]//Proc of the 8th Annual Conference International Speech Communication Association. 2007:714-714.  
 [8] LV Guo-yun, ZHAO Rong-chun, JIANG Dong-mei, et al. A robust visual feature extraction based BTSM-LDA for audio-visual speech recognition [C]//Proc of 2nd International Conference on Communications and Networking in China. 2007:637-641.  
 [9] ALIZADEH S, BOOSTANI R, ASADPOUR V. Lip feature extraction and reduction for HMM-based visual speech recognition systems [C]//Proc of the 9th International Conference on Signal Proceedings. 2008:561-564.

[10] DUPONT S, LUETTIN J. Audio-visual speech modeling for continuous speech recognition [J]. *IEEE Trans on Multimedia*, 2000, 2(3):141-151.  
 [11] NEFIAN A V, LIANG Lu-hong, PI Xiao-bo, et al. Dynamic Bayesian networks for audio-visual speech recognition [J]. *EURASIP Journal on Applied Signal Processing*,2002(1):1274-1288.  
 [12] MATTHEWS I, COOTES T F, BANGHAM J A, et al. Extraction of visual features for lipreading [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*,2002,24(2):198-213.  
 [13] NETI C, POTAMIANOS G, LUETTIN I J, et al. Audio-visual speech recognition [EB/OL]. (2000-10-12). [http://www.research.ibm.com/AVSTG/ws\\_00\\_avsr.pdf](http://www.research.ibm.com/AVSTG/ws_00_avsr.pdf).  
 [14] HAEB-UMBACH R, NEY H. Linear discriminant analysis for improved large vocabulary continuous speech recognition [C]//Proc of IEEE International Conference on Acoustics, speech, and Signal Processing. 1992:13-16.  
 [15] SAKAI M, KITAOKA N, NAKAGAWA S. Generalization of linear discriminant analysis used in segmental unit input HMM for speech recognition [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2007:333-336.  
 [16] POTAMIANOS G, GRAF H P. Linear discriminant analysis for speechreading [C]//Proc of the 2nd IEEE Workshop on Multimedia Signal Processing. 1998:221-226.  
 [17] LAIRD N M, LANGE N, STRAM D. Maximum likelihood computations with repeated measures: application of the EM algorithm [J]. *Journal of the American Statistical Association*,1982,82(397):97-105.