文章编号: 1000−8152(2011)10−1461−06

自适应视听信息融合用于抗噪语音识别

梁 冰[1], 陈德运[2], 程 慧[3]

(1. 大连理工大学 创新实验学院, 辽宁 大连 116024; 2. 哈尔滨理工大学 计算机科学与技术学院, 黑龙江 哈尔滨 150080;
3. 哈尔滨工程大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘要: 为了提高噪音环境中语音识别的准确性和鲁棒性, 提出了基于自适应视听信息融合的抗噪语音识别方法, 视听信息在识别过程中具有变化的权重, 动态的自适应于环境输入的信噪比. 根据信噪比和反馈的识别性能, 通过学习自动机计算视觉信息的最优权重; 根据视听信息的特征向量, 利用隐马尔科夫模型进行视听信息的模式匹配, 并根据最优权重组合视觉和声音隐马尔科夫模型的决策, 获得最终的识别结果. 实验结果表明, 在各种噪音水平下, 自适应权重比不变权重的视听信息融合的语音识别性能更优.

关键词: 视听信息融合; 语音识别; 自适应权重; 学习自动机; 隐马尔科夫模型

中图分类号: TP273 **文献标识码**: A

Adaptive fusion of acoustic and visual information in noise-robust speech recognition

LIANG Bing[1], CHEN De-yun[2], CHENG Hui[3]

(1. School of Innovation Experiment, Dalian University of Technology, Dalian Liaoning 116024, China;
2. College of Computer Science and Technology, Harbin University of Science and Technology, Harbin Heilongjiang 150080, China;
3. College of Computer Science and Technology, Harbin Engineering University, Harbin Heilongjiang 150001, China)

Abstract: We propose the adaptive fusion of acoustic and visual information for improving the accuracy and the robustness in the speech recognition. The acoustic and visual information is involved in the recognition process with different weights, which are adaptively determined according to the signal-to-noise ratio(SNR) between the environment inputs during the process of recognition. Based on the SNR and the performance feedback, a learning automata is used for computing the adaptive weights for the visual information. A hidden Markov model is used to match the patterns of the acoustic information and the visual information. The hidden Markov model decides the final recognition results by combining the acoustic information and the visual information with optimal weights. Experiments under various noise-level conditions are performed; results show that the speech recognition based on adaptive weights surpasses the speech recognition based on fixed weights.

Key words: audio-visual information fusion; speech recognition; adaptive weights; learning automata(LA); hidden Markov model

1 引言(Introduction)

近年来, 基于多模态信息的语音识别系统逐渐成为研究热点[1,2]. 单纯依赖单模的声音信息的语音识别在无噪音环境的条件下性能较好, 然而, 当存在噪声或频率干扰时, 其识别性能将大大降低[3]. 感知实验表明, 多模态信息可以提高对语音的感知和理解, 应用视觉信息(发音时的唇形)可以有效对抗环境中噪音的干扰. 文献[4]提出基于交互多模型(interacting multiple model, IMM)的自适应多模信息融合方法, 通过滤波器计算识别目标状态, 其中每个滤波器均应用先验模型条件估计的不同组合, 因此当识别目标数量变大时, 滤波器的数量变大, 计算量也成倍增长. 文献[5]提出基于当前统计模型和自适应滤波(current statistical model and adaptive filtering, CSMAF)的自适应多模信息融合方法, CSMAF应用当前模型的优势实现了只用一个滤波器自适应识别目标, 计算量较小, 但在噪音环境中, CSMAF融合结果的精度较低. 文献[6]中的信息融合技术运用了神经网络(neural network, NN)学习率自适应调整的融合算法, 由于神经网络抗噪音干扰的能力较强, 提高了识别精度, 但同时调整权重和参数, 算法计算量大, 学习周期长.

收稿日期: 2010−05−02; 收修改稿日期: 2010−11−15.
基金项目: 国家自然科学基金资助项目(60572153); 黑龙江省博士后基金资助项目(LBH–Z09102); 哈尔滨理工大学青年科学研究基金资助项目(2009YF015); 中央高校基本科研业务费专项资金资助项目(DUT11RC(3)54).

针对以上问题, 考虑到噪音水平的变化和处理过程的性能反馈, 本文提出基于学习自动机的自适应视听信息融合方法(adaptive information fusion based on learning automata, AIFLA), AIFLA方法能够在有限的权重集合内通过学习选择最优的自适应参数, 因此能够获得较高的融合精度, 且算法计算量较小, 计算效率高. 首先应用隐马尔科夫模型(hidden Markov models, HMM)实现视听信息融合的语音识别方法, 并在此基础上提出基于学习自动机自适应视听信息融合方法. 学习自动机通过反馈环与环境交互, 自动机在一定概率上受到环境的奖励或惩罚[7,8]. 基于学习自动机能够与环境交互的特点, 可以应用学习自动机实现信息融合的自适应性. 自适应信息融合能够与环境交互, 实时反馈环境的噪音水平, 确定视觉和声音信息在识别过程中权重的变化, 首先通过频率倒谱提取声音特征, 通过离散余弦变量将说话者唇部区域转换为低维视觉特征, 并分别训练视觉和声音特征的HMM; 然后利用学习自动机的概率更新策略获得视觉信息的最优权重, HMM根据最优权重组合视听HMM, 概率最高的假设为最终的识别结果; 最后通过实验验证在噪音环境中基于自适应权重视听信息融合语音识别的性能.

2 基于自适应视听信息融合的语音识别方法(Adaptive acoustic and visual information fusion method for speech recognition)

基于自适应视听信息融合的语音识别方法利用HMM进行视听信息的模式匹配. 如果测试数据库和训练匹配的较好, 在无噪音的环境下, 利用HMM实现视听信息融合的语音识别在理论上最优. 但HMM方法的基本假设是在训练时生成的模型能够匹配在测试时识别的语音数据. 在有噪音干扰的情况下, 假设失效, 模式匹配时会产生错误的概率估计. 自适应信息融合技术能够实时反馈环境的噪音水平, 并决定声音和视觉传感器信息源收集到的信息在融合过程中所占的比重, 应用视觉权重降低噪声的干扰. 本文提出基于自适应视听信息融合的语音识别方法分为以下几个步骤(见图1):

1) 对声音$a_{\rm t}$和视觉信号$v_{\rm t}$进行特征提取, 获得声音和视觉特征向量;

2) 通过描述视觉和声音信号的特征训练HMM, 获取视觉和声音HMM;

3) 根据信噪比SNR与反馈的识别性能, 利用学习自动机选择最优视觉权重$\gamma_{\rm v}$;

4) 根据最优视觉权重$\gamma_{\rm v}$组合声音HMM和视觉HMM, 获得视听HMM, 概率最高的假设即是识别结果.

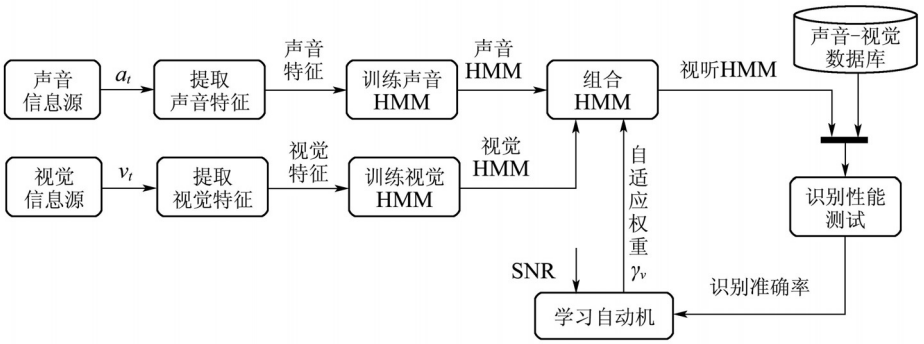

图 1 基于自适应视听信息融合的语音识别模型

Fig. 1 Speech recognition model based on adaptive acoustic and visual information fusion

2.1 声音和视觉特征提取(Acoustic and visual feature extraction)

用于HMM模式匹配的声音数据特征包括Mel频率倒谱(MFCC)以及一阶和二阶差分. 每10 ms计算25 ms汉宁窗语音样本的MFCC. 在话语层面所有声音信号通过一阶预加重滤波器$p_n = s_n - ks_n$, 其中s_n是第n个采样信号. 预加重k为0.97, 语音识别器通过HMM训练并测试.

应用灰度水平信息获得原始的图像描述, 唇部区域的水平样本为121像素宽, 采样超过3个相邻垂直线, 纵向样本为31像素高, 包括来自11个相邻列中的数据(见图2). 每个数据获得一个水平和垂直的特征向量. 唇部区域的中心通过匹配滤波器有从图像中计算得到. 视觉信号的采样速率是60帧/s. 通过离散余弦变量将唇部区域转换为低维视觉特征, 视觉特征描述基于图像转换的嘴部形状和外观, 视觉特征向量限制在32维, 视觉识别器通过HMM训练并测试.

2.2 自适应视听信息融合(Adaptive acoustic and visual information fusion)

声音和视觉信息有时是异步的, 有些字符发音唇部没有变化, 如词/six/, 包括4个音素/sIks/, 只

表现为单一的唇部运动, 但唇部形状主要由元音/I/决定. 引入HMM可以解决声音和视觉信息融合的异步性.

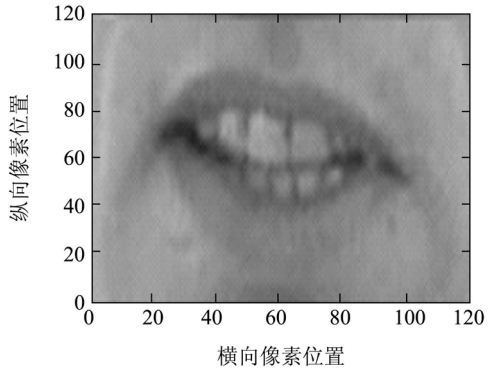

图 2 唇部区域的水平和垂直位置采样

Fig. 2 The lips horizontal and vertical position samples

基于HMM的视听信息融合方法中假定视听特征是条件独立的, 根据视听观测给定一个可能语音的假设, 则

$$\mathrm{P}(a_{\mathrm{t}}, v_{\mathrm{t}}|H_i) = \mathrm{P}(a_{\mathrm{t}}|H_i)\mathrm{P}(v_{\mathrm{t}}|H_i), \tag{1}$$

其中: $a_{\mathrm{t}}, v_{\mathrm{t}}$分别是视觉传感器、声音传感器在时刻$t$获得的观测, H_i是语音识别结果的多个假设集合. 在给定观测下, 应用Bayes规则给出假设的概率, 计算得到

$$\mathrm{P}(H_i|a_{\mathrm{t}}, v_{\mathrm{t}}) = \frac{\mathrm{P}(H_i|a_{\mathrm{t}})\mathrm{P}(H_i|v_{\mathrm{t}})}{\mathrm{P}(H_i)} \times \frac{\mathrm{P}(a_{\mathrm{t}})\mathrm{P}(v_{\mathrm{t}})}{\mathrm{P}(a_{\mathrm{t}}, v_{\mathrm{t}})}. \tag{2}$$

未加权的贝叶斯乘积可以表示为概率估计$\widehat{\mathrm{P}}$, 见式(3):

$$\widehat{\mathrm{P}}(H_i|a_{\mathrm{t}}, v_{\mathrm{t}}) = \frac{\widehat{\mathrm{P}}(H_i|a_{\mathrm{t}})\widehat{\mathrm{P}}(H_i|v_{\mathrm{t}})}{\sum\limits_{j=1}^{N} \widehat{\mathrm{P}}(H_j|a_{\mathrm{t}})\widehat{\mathrm{P}}(H_j|v_{\mathrm{t}})}. \tag{3}$$

式(3)表示N个假设的集合.

自适应视听信息融合通过视觉权重γ_{v}对式(3)进行加权, 见式(4):

$$\mathrm{P}(H_i|a_{\mathrm{t}}, v_{\mathrm{t}}) = \mathrm{P}^{(1-\gamma_{\mathrm{v}})}(H_i|a_{\mathrm{t}})\mathrm{P}^{\gamma_{\mathrm{v}}}(H_i|v_{\mathrm{t}})\delta(\gamma_{\mathrm{v}}). \tag{4}$$

视觉权重是环境噪音水平的函数, 视觉权重随着信噪比SNR呈正比变化. 应用$\delta(\gamma_{\mathrm{v}})$归一化概率估计, 得到式(5):

$$\delta(\gamma_{\mathrm{v}}) = 1/\sum_{j=1}^{N} \mathrm{P}^{(1-\gamma_{\mathrm{v}})}(H_i|a_{\mathrm{t}})\mathrm{P}^{\gamma_{\mathrm{v}}}(H_i|v_{\mathrm{t}}). \tag{5}$$

如果γ_{v}的值接近于1, 则声音和视觉信息源在可能性估计上的贡献相同, 而当γ_{v}的值小于1时, 则削弱了视觉信息源的重要性.

识别的任务是获得HMM概率最高的假设, HMM表示的概率在log空间, 所执行的解决方法见式(6):

$$u = \arg \max_{i=1,2,\cdots,N} \{\gamma_{\mathrm{v}} \log \mathrm{P}(H_i|a_{\mathrm{t}}) + (1-\gamma_{\mathrm{v}}) \log \mathrm{P}(H_i|v_{\mathrm{t}})\}, \tag{6}$$

其中u是最有可能的话语.

每个识别的结果可作为一个假设. 对于有限的词汇量, 可以评估每个可能的假设, 识别器返回最高联合概率的假设.

3 基于学习自动机的自适应权重计算 (Adaptive weight computation based on learning automata)

通过学习自动机计算式(4)中的自适应视觉权重, 首先通过学习自动机的动作集合描述视听信息融合的视觉权重集合, 每个权重都有一个初始概率, 输出集合描述信息融合的性能反馈; 其次, 通过学习自动机的概率更新策略获得概率最高的权重, 即为最优视觉权重.

3.1 基于学习自动机的自适应视听信息融合模型(Adaptive acoustic and visual information fusion model based on learning automata)

学习自动机(learning automata, LA)由一组内部状态、输入动作、状态概率转移函数和强化策略构成, 并且通过反馈回路与环境连接, 见图3所示.

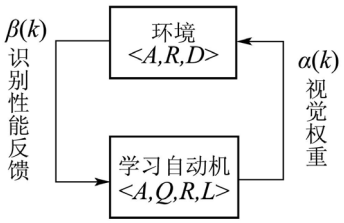

图 3 学习自动机模型

Fig. 3 The learning automata model

自适应视听信息融合模型定义为学习自动机四元组$M =< A, Q, R, L >$, 其中: 动作集合A描述视听信息融合系统的视觉权重集合, $R = \{0, 1\}$表示融合系统的性能反馈, 如果当前选择的视觉权重a_i, 可使识别准确率上升, 那么$R = 1$, 否则$R = 0$; L是更新M状态的概率更新算法. 视听信息融合所操作的环境定义为三元组$E =< A, R, D >$.

学习自动机四元组$< A, Q, R, L >$, 其中:

1) 动作集合$A = \{\alpha_1, \alpha_2, \cdots, \alpha_r\}$, 其中: $\alpha(k)$是自动机在时刻k的动作, $\alpha(k) \in A(k = 0, 1, 2, \cdots)$, r是动作的总数. A是自动机的输出集合, 同样也是环境的输入;

2) 状态集合$Q(k)=\mathrm{E}[P(k)\ \ D(k)]$, 其中: $\mathrm{P}(k)=[p_1(k)\ \ p_2(k)\ \ \cdots\ \ p_r(k)](\sum_{i=1}^{r}P_i(k)=1)$是动作概率向量; $\widehat{D}(k)=[\widehat{d_1}(k)\ \ \widehat{d_2}(k)\ \ \cdots\ \ \widehat{d_r}(k)]$是动作反馈概率估计;

3) 环境的响应域R. $\beta(k)$表示自动机在时刻k的响应, $\beta(k)\in R$. $\beta(k)$是环境的输出, 也是自动机的输入;

4) L是更新学习自动机状态的学习算法或强化策略. $Q(k+1)=L[Q(k)\ \ \alpha(k)\ \ \beta(k)]$.

在时刻k, 学习自动机从动作集合A中选择动作$\alpha(k)$, 选择依赖于目前的动作概率向量$\mathrm{P}(k)$; 选择的动作$\alpha(k)$成为环境的输入, 环境反馈随机响应$\beta(k)$成为学习自动机的输入; 如果$\alpha(k)=\alpha_i$, 其预期值为$d_i(k)$, 学习自动机利用强化策略计算$Q(k+1)$.

环境定义为三元组$<A,R,D>$, A和R的定义如上. $D=\{d_1,d_2,\cdots,d_r\}$是反馈概率, 其中$d_i(k)=\mathrm{E}[\beta(k)|\alpha(k)=\alpha_i]$.

3.2 计算视听信息的自适应权重(Adaptive weight computation of acoustic and visual information)

根据目前的噪音水平, 通过学习自动机选择视听信息融合系统的视觉权重γ_{v}. 建立γ_{v}值的集合

$$A=\{0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\},$$

$d_i(k)=\mathrm{E}[\beta(k)|\alpha(k)=\alpha_i]$表示与识别结果相对应的视觉权重的输出概率.

基于学习自动机的视觉权重γ_{v}计算算法描述如下:

学习自动机的一次循环包括两个操作:

1) 概率更新. 根据噪音水平对选择视觉权重的响应, 学习自动机修改视觉权重集合的概率分布.

2) 视觉权重的选择. 基于新的概率分布, 学习自动机选择一个新的视觉权重.

执行以上操作的算法需要实现概率更新策略, 计算时间依赖于动作集的成员数量r, 计算复杂度为$\mathrm{O}(r)$. 本文提出的概率更新策略可将学习自动机的概率集合离散化, 从而降低计算复杂度, 以下是概率更新策略:

$$\begin{cases} d_i(k+1)=d_i(k)+1/c, \\ \text{如果 } a(k)=a_i,\ \beta(k)=1,\ d_i(k)<1, \end{cases} \tag{7}$$

$$\begin{cases} d_i(k+1)=d_i(k)-1/c, \\ \text{如果 } a(k)=a_i,\ \beta(k)=0,\ d_i(k)>0. \end{cases} \tag{8}$$

初始时, $d_i(k)=0,\ \forall i, c(c=100)$为实数, 取值大小根据问题的精度要求. 根据以上概率更新策略, 动作的基本概率从有限的集合中取值, $d_i(k)\in\{0,1/c,2/c,\cdots,1\}$; 那么, $\sum_{j=1}^{r}d_j(k)$的值取自有限集合, 动作概率$\mathrm{P}_i(k)=d_i(k)/\sum_{j=1}^{r}d_j(k)(i=1,\cdots,r)$同样取值于有限集合.

算法继续执行, 根据集合A的概率分布, 选择$\alpha_{\mathrm{m}}\in A$, 与输出概率的最大期望值$d_{\mathrm{m}}=\max\{d_i\}(i=0,1,2,\cdots,r)$相对应, 当时间$k$趋向无穷, 期望得到视觉权重$\alpha_{\mathrm{m}}$趋向一致.

因此, 在识别过程中, 学习自动机能够根据当前的噪音水平和性能反馈从集合中选择语音识别准确率最高的自适应视觉权重$\gamma_{\mathrm{v}}(\gamma_{\mathrm{v}}=\alpha_{\mathrm{m}})$.

4 实验结果与分析(Experimental results and analysis)

应用XM2VT声音–视觉数据库. XM2VT包括295例数据, 其中2/3用于训练HMM, 1/3用于测试HMM. 声音信号人为加噪, 信噪比分别为15, 10, 5, 0, –5, –10, –15分贝. 识别准确率ACC(percentage accuracy)评估语音识别系统的性能, 见式(9):

$$\mathrm{ACC}=\frac{N-D-S-I}{N}\times 100\%. \tag{9}$$

式中: N为参照句子中的词数, D为删除错误词数, S为替代错误词数, I为插入错误词数.

实验分为两个部分, 首先验证基于学习自动机的自适应信息融合与不变权重的信息融合的识别性能测试, 然后比较基于学习自动机的自适应信息融合与基于IMM, CSMAF, NN模型的自适应信息融合方法的语音识别性能. 仿真实验条件为P4, CPU3.00 GHz, 512 MB内存, WinXP.

4.1 自适应视听信息融合识别准确率测试(Percentage accuracy test of adaptive acoustic and visual information fusion)

在各种噪音水平下, 首先验证单模系统即只根据声音信息或视觉信息的语音识别性能, 然后验证视听信息融合系统在不变权重和自适应视觉权重下的语音识别性能, 实验结果如表1所示.

由表1可见, 在噪声逐渐增大的情形下, 只根据声音信息的单模系统识别性能急剧下降, 而基于视觉信号的识别性能并没有变化.

在视听信息融合方面, 基于不变权重的视听信息融合能够提高语音识别性能, 但在15 dB, 10 dB, 5 dB和0 dB条件下, 与基于声音信息的识别性能相

比, 分别提高1%, 1.2%, 2.7%和1.4%, 但提高并不显著; 随着噪音的增大, 在–5 dB, –10 dB, –15 dB条件时, 不变权重的视听信息融合方法与基于声音信息的识别性能相比分别提高17.2%, 19.7%和25.4%.

表 1 语音识别准确率

Table 1 Percentage accuracy of speech recognition

SNR/dB	单模系统/%		视听信息融合系统/%	
	声音	视觉	不变权重	自适应权重
15	90.3	31.4	91.3	95.8
10	87.4	31.4	88.6	93.0
5	73.5	31.4	76.2	90.3
0	69.2	31.4	70.6	82.9
–5	34.5	31.4	51.7	65.2
–10	19.7	31.4	39.4	52.5
–15	10.2	31.4	35.6	41.8

最后, 将基于自适应权重的视听信息融合识别方法与不变权重的方法进行比较, 自适应权重识别性能想比不变权重识别性能, 平均提高了9.7个百分点, 说明基于自适应权重的视听信息融合能够进一步提高语音识别性能.

总之, 自适应信息融合技术能够实时反馈环境的噪音水平, 并决定声音和视觉传感器信息源收集到的信息在融合过程中的权重, 实验结果表明, 尤其在噪音环境中的语音识别, 基于自适应视觉权重的语音识别准确率更高.

4.2 自适应视听信息融合方法性能比较测试 (Performance comparisons of adaptive acoustic and visual information fusion method)

将本文基于学习自动机的自适应信息融合方法(adaptive information fusion based on learning automata, AIFLA)与IMM, CSMAF, NN以及不变权重的信息融合(invariable weight information fusion, IWIF)方法进行比较, 并从语音识别的时间和准确率两方面进行测试, 实验结果如表2, 3所示.

根据表2对比不同自适应融合方法的测试结果, 随着信噪比的降低, 识别结果的误差也增大了. 在不同信噪比条件下, AIFLA和NN方法的识别准确均高于IMM和CSMAF方法, 而AIFLA和NN方法的获得了相差无几的识别准率, 但从表3的识别时间可见, AIFLA方法识别的速率几乎是NN方法的1倍, CSMAF方法的5倍, IMM方法的10倍. AIFLA方法由于要根据信噪比计算权重, 因此识别时间比不变权重的信息融合方法(IWIF)要多出数秒, 但随着信噪比的降低, 时间的增加是呈线性增长的, 识别准确率比IWIF方法平均提高了9.7%(不变权重的信息融合方法识别准确率见表1).

表 2 IMM, CSMAF, NN和AIFLA方法的识别准确率

Table 2 Percentage accuracy of recognition of IMM, CSMAF, NN and AIFLA method

自适应信息融合方法	SNR/dB						
	15	10	5	0	–5	–10	–15
IMM/%	90.7	89.2	84.5	78.3	55.8	45.6	37.3
CSMAF/%	87.5	84.6	82.5	76.9	54.7	42.1	36.2
NN/%	95.3	92.8	91.0	81.6	66.3	50.7	40.2
AIFLA/%	95.8	93.0	90.3	82.9	65.2	52.5	41.8

表 3 IMM, CSMAF, NN, IWIF和AIFLA方法的识别时间

Table 3 Time of recognition of IMM, CSMAF, NN and AIFLA method

自适应信息融合方法	SNR/dB						
	15	10	5	0	–5	–10	-15
IMM/ s	89.3	94.5	112.2	124.3	146.8	198.3	203.5
CSMAF/ s	35.2	45.0	56.4	66.5	78.9	89.2	112.2
NN/ s	12.8	15.3	17.4	24.6	45.8	55.7	74.3
AIFLA/ s	6.4	8.2	10.5	14.5	25.7	30.6	39.5
IWIF/ s	5.5	6.9	7.3	10.8	14.9	22.6	29.3

综上, 由于NN方法需要同时调整权重和参数, 识别时间是AIFLA方法的2倍. 而AIFLA方法获得的自适应权重的集合是有限的, 计算效率高. 因此, AIFLA方法在识别时间和识别准确率的整体性能上优于IMM, CSMAF, NN和IWIF方法.

5 结论(Conclusion)

本文通过自适应融合视觉和听觉信息实现语音识别, 其中不同模态信息的权重根据噪音水平动态调整. 首先通过散余弦变量和Mel频率倒谱离提取视听信息特征, 提出基于HMM的视听信息融合方法, 引入学习自动机实现与噪音环境的交互, 并通过学习自动机的概率更新策略获得最优的视觉信息权重. 实验结果表明基于学习自动机的自适应视听信息融合语音识别方法, 能够获得在不同噪音水平下比不变权重更好的识别性能, 能够有效提高语音识别的准确率和时间效率. 进一步的工作将针对多个说话者的语音识别任务, 如何根据视觉注意选择特定说话者的方式确定视觉权重, 提高基于自适应视听信息融合的语音识别的检测能力、响应速度和准确率.

参考文献(References):

[1] FARAJ M, BIGUN J. Audio-visual person authentication using lip-motion from orientation maps[J]. *Pattern Recognition Letters*, 2007, 28(11): 1368 – 1382.

[2] STIEFELHAGEN R, BERNARDIN K, EKENEL H. Audio-visual perception of a lecturer in a smart seminar room[J]. *Signal Processing*, 2006, 86(12): 3518 – 3533.

[3] MARTIN A, MAUUARY L. Robust speech/non-speech detection based on LDA-derived parameter and voicing parameter for speech recognition in noisy environments[J]. *Speech Communication*, 2006, 48(2): 191 – 206.

[4] ZHANG J, BALASURIYA A, CHALLA S. Vision based data fusion for autonomous vehicles target tracking using interacting multiple dynamic models[J]. *Computer Vision and Image Understanding*, 2008, 109(1): 1 – 21.

[5] HU H, JING Z. Unscented fuzzy-controlled current statistic model and adaptive filtering for tracking maneuvering targets[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2006, 11(8): 961 – 972.

[6] ZHANG J. Neural network-based state fusion and adaptive tracking for maneuvering targets[J]. *Communications in Nonlinear Science and Numerical Simulation*, 2005, 10(4): 395 – 410.

[7] ZAHIRI S. Learning automata based classifier[J]. *Pattern Recognition Letters*, 2008, 29(1): 40 – 48.

[8] TAZFESTAS G, RIGATORS S. Stability analysis of an adaptive fuzzy control system using Petri nets and learning automata[J]. *Mathematics and Computers in Simulation*, 2000, 51(3/4): 315 – 339.

作者简介:

梁 冰 (1981—), 女, 工程师, 博士, 研究领域为数据融合、数据挖掘、模式识别和形式语言与自动机, E-mail: newpek@163.com;

陈德运 (1962—), 男, 教授, 博士生导师, 研究领域为探测与成像技术、基于Web的信息处理技术、激光扫描图像测量技术, E-mail: Chendeyun@hrbust.edu.cn;

程 慧 (1980—), 女, 博士研究生, 研究领域为数据库和数据挖掘, E-mail: chenghui@hrbeu.edu.cn.