

基于 SRN/HMM 的非特定人连续手语识别系统*

方高林¹, 高文^{1,2}, 陈熙霖¹, 王春立³, 马继勇²

¹(哈尔滨工业大学 计算机科学与工程系,黑龙江 哈尔滨 150001);

²(中国科学院 计算技术研究所,北京 100080);

³(大连理工大学 计算机科学与工程系,辽宁 大连 116023)

E-mail: {fgl,wgao,xlchen,chlwang,jyma}@ict.ac.cn

http://mmi.ict.ac.cn

摘要: 手语识别是通过计算机提供一种有效而准确的机制将手语翻译成文本或语音.目前最新发展水平的手语识别系统在实际应用中应解决非特定人连续手语问题.提出一种将连续手语识别分解成各孤立词识别的分治方法,用于非特定人连续手语识别.把精简循环网(simple recurrent network,简称 SRN)作为连续手语的段边界检测器,把 SRN 分段结果作为隐马可夫模型(hidden Markov models,简称 HMM)框架中的状态输入,在 HMM 框架里使用网格 Viterbi 算法搜索出一条最佳手语词路径.实验结果表明,该方法的识别效果比单纯使用 HMM 要好.

关键词: 神经网络;精简循环网络;隐马可夫模型;连续手语识别;非特定人手语识别

中图法分类号: TP391 **文献标识码:** A

手语是聋人交际的一种最常用的工具.手语识别的目标就是通过计算机提供一种有效而准确的机制,将手语翻译成文本或语音,使得聋人和听力正常人之间的交流变得更方便、更快捷.它已经成为人机接口领域的一项最重要研究内容之一.此外,手语识别还有许多其他应用,例如,在虚拟环境通过识别人的手势去控制虚拟人的运动,在虚拟现实作为多模式用户接口等等.

最初进行自动手语识别的尝试始于 20 世纪 90 年代,但是大部分手语识别研究者主要集中在孤立手语词识别研究上,对于连续手语识别的研究则很少.Starner^[1]使用固定在桌面上的彩色相机作为输入设备,采用 HMM(hidden Markov models)方法识别由 40 个手语词组成的句子,识别正确率达 92%;如果将相机固定在用户的帽子上,对句子进行严格的语法限制,在实时情况下识别率能达到 98%.Liang 和 Ouhyoung^[2]使用数据手套作为输入设备,HMM 作为识别技术,能够识别手语者打出由 250 个台湾手语词组成的连续句子,但是它要求打手语的速度比正常的速度慢.Vogler 和 Metaxas^[3]使用计算机视觉方法提取手语者手的三维运动变量,将视觉的方法与 HMM 结合识别由 53 个手语词组成的连续句子.为了克服“运动插入”影响,使用上下文相关模型,识别率达 89.9%.此外,他们以音素为基元识别连续句子^[4],使用 22 个手语词组成的句子做实验,得到与词作为基元相似的识别结果.我们前面的工作在文献[5]中已经进行了描述,使用数据手套作为输入设备,采用 HMM 模型识别技术,在特定人手语识别方面,识别 5 177 个孤立手语词,正确率达 94.8%;可以识别由该词汇集中词组成的 200 个句子,正确率达 91.4%.

从前期研究来看,对连续手语的研究都集中在特定人方面,而对非特定人连续手语识别的研究在文献中还

* 收稿日期: 2001-04-12; 修改日期: 2001-07-13

基金项目: 国家自然科学基金资助项目(69789301);国家 863 高科技发展计划资助项目(863-306-ZD03-01-2);中国科学院百人计划资助项目

作者简介: 方高林(1975 -),男,安徽庐江人,博士生,主要研究领域为模式识别,统计语言模型,人工智能;高文(1956 -),男,博士,教授,博士生导师,主要研究领域为多媒体数据压缩,图像处理,计算机视觉,多模式接口,人工智能,虚拟现实;陈熙霖(1965 -),男,博士,教授,博士生导师,主要研究领域为计算机视觉,多模式接口,虚拟现实;王春立(1972 -),女,河北保定人,博士生,主要研究领域为模式识别;马继勇(1964 -),男,黑龙江哈尔滨人,博士,主要研究领域为模式识别.

未见报导.连续手语识别的主要问题是“运动插入”影响,即两个手势词之间存在过渡动作,它随着手语词的上下文变化而改变,类似于语音识别中的“协同发音”.对于非特定人识别,由于不同人的手的大小、身高长短、打手语的幅度、习惯以及打手语句子节奏快慢等都不一样,这些都增加了识别的难度,所以它的识别比特定人连续手语识别更具有挑战性.我们针对非特定人连续手语识别研究进行了尝试,实验表明连续手语具有段特性.根据该特性,我们提出了分治策略来实现连续手语识别,将连续手语识别问题化解成各孤立词的识别问题求解;同时,对 SRN(simple recurrent network)进行改进,充分利用上下文,改进后的 SRN 作为段边界检测器;最后使用网格 Viterbi 算法,在 SRN 输出的分段里搜索出一条最佳的手语词路径.

本文首先介绍 SRN 并提出改进,然后给出基于 SRN 连续手语的边界检测方法;其次分析了在 HMM 框架下的网格 Viterbi 算法和段内概率的计算问题,并给出实验结果及其与单纯使用 HMM 方法的比较,最后得出结论.

1 基于 SRN 的边界检测

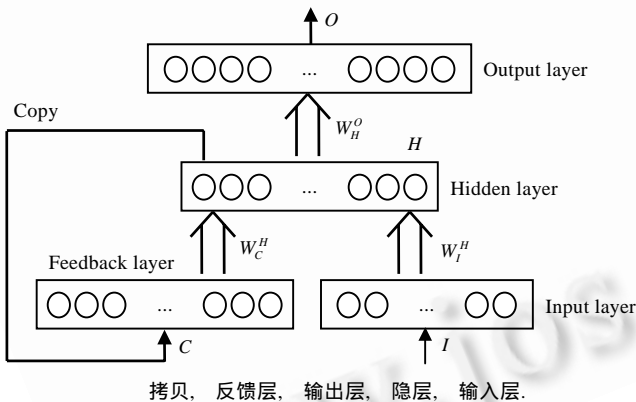
1.1 精简循环网络及其改进

Elman 在循环网基础上进行了改进,提出了精简循环网(SRN)^[6].循环网通过反馈层的引入而使网络具备了记忆和利用上文的能力,并已经成功地应用到语音识别^[7,8]、手写体识别^[9]和孤立手语词识别^[10]等方面.典型的精简循环网络(如图 1 所示)共有 4 层神经元.令网络在接收第 t 时刻输入向量 I_t 时相应各层的输出为:隐层 H_t , 反馈层 C_t , 输出层 O_t . W_C^H, W_I^H, W_H^O 分别表示反馈层到隐层、输入层到隐层、隐层到输出层的权值矩阵.

反馈层节点是隐层节点的拷贝,并加入了一个单位的延迟.网络的输入层和反馈层组成联合输入层.若 Φ, Ψ 分别为隐藏层神经元和输出层神经元的阈值向量,则有

$$H_t = f(C_t \cdot W_C^H + I_t \cdot W_I^H - \Phi), \tag{1}$$

$$O_t = f(H_t \cdot W_H^O - \Psi). \tag{2}$$



拷贝, 反馈层, 输出层, 隐层, 输入层.
Fig.1 Simple recurrent network
图 1 精简循环网络

神经元的活跃函数 $f(\cdot)$ 一般取为 Sigmoid 函数.精简循环网络采用误差反向传播(back-propagation)算法^[11]来进行网络参数训练.

反馈层的引入,使得网络输出不但与当前的网络输入有关,而且与前一时刻的网络状态有关,而前一时刻的网络状态又是前面所有输入计算的结果.因此,精简循环网络可以记忆并使用较大范围的上文^[6].

精简循环网络通过隐藏层,利用所有前导输入的信息作出当前判断,但是无法利用下文,为此,这里进行改进.策略 1,将下文矢量作为当前输入的一部分,从而使上下文的微

结构信息同时得到有效利用,这样,新的输入为 $I_t = [I_t, I_{t+1}]$,其他计算同 SRN.策略 2,训练时,结构不变,将训练样本正向输入网络,训练出正向 SRN;将训练样本反向输入网络,训练出反向 SRN.识别时,数据分别以正向输送到正向 SRN,得出正向输出结果,再将数据逆向输送到反向 SRN,得到逆向结果.然后将两种结果进行综合,通过两个网络充分利用了上下文信息.本文针对这两种策略进行手语边界检测实验,由于策略 2 的正、反向网络输出有时会产生冲突,很难用统一的标准进行判断,所以实验结果不如策略 1 好,这里,我们采用策略 1 作为改进的 SRN.

1.2 连续手语的边界检测

采用数据手套和 3 个位置跟踪器作为数据采集设备.数据手套用于采集手形变化信息,左、右手各 18 维数

据,共 36 维;位置跟踪器用于采集方向、位置、运动轨迹特征,共 12 维输入数据.这样,整个输入为 48 维数据.由于不同人的手形大小、体形、打手语的习惯以及幅度不一样,我们让打手语者做几个固定的姿势,然后对每个人的数据进行标定处理;另外,由于 48 维数据的动态范围不一样,需要对每维数据进行归一化处理,使得动态范围为 0~1.实验表明,上述的数据预处理还不能满足 SRN/HMM 的要求.实验是采用 48 维数据作为 SRN 的输入,选用 30 个隐含节点,3 个输出节点,在 PIII450(192M 内存)微机训练网络 28 个小时.但是这种形式 SRN 分段的效果不是十分理想,分段召回率仅为 87%,在可扩展性方面和识别率方面都是远远不够的.使用分段召回率的概念是因为这里仅考虑将实际所有段检测出来,对于一个词可能被检测成几段的情况,在 HMM 框架下通过网格 Viterbi 算法搜索去解决,这里不作考虑.

$$\text{分段召回率} = \frac{\text{作出正确分段的个数}}{\text{实际所有分段的个数}} \times 100\%, \quad (3)$$

因而在前述的数据处理基础上用自组织特征映射(SOFM)作为特征提取网,提取出来的输出表现出很强的段特性(如图 2 所示,横轴为帧数,纵轴为 SOFM 的输出),SOFM 的输出经过编码输入到 SRN.SOFM 的输入为 48 个节点,经过实验,SOFM 的输出选取 256 个节点;SRN 选用 16 个输入节点,15 个隐含节点,3 个输出节点,在 PIII450(192M 内存)微机上训练 40 分钟,分段召回率达到 98.8%.下面分别讨论 SRN 的输入、输出、训练以及识别.

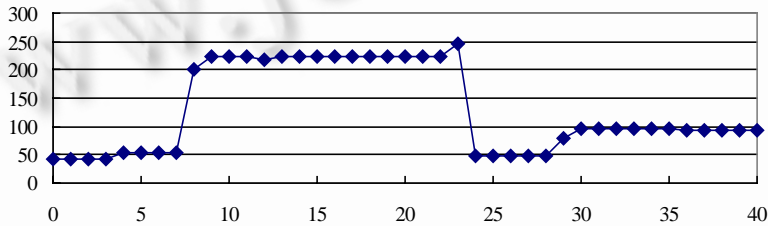


Fig.2 The segments property of sign language 'when will we leave'
图 2 手语“我们什么时候走”段特性图

输入:对 SOFM 的 256 个输出节点进行二进制编码,需 8 位,同时引进下文单元 8 位作为输入,这样共选用 16 个输入节点,节点的输入值为 $I_i \in \{0,1\}, i=1,2,\dots,16$.

输出:使用 3 个输出节点表示,分别定义段的左边界 1,对应节点 o_i^1 ,段的右边界 2,对应节点 o_i^2 ,段内部 3,对应节点 o_i^3 .

$$O_i = \begin{cases} [1 & 0 & 0], & \text{输出为左边界} \\ [0 & 1 & 0], & \text{输出为右边界} \\ [0 & 0 & 1], & \text{输出为段内部} \end{cases} \quad (4)$$

训练:由于连续手语中间没有停顿,不能确定其边界,这里采用自动分割的方法.设训练集中的样本句子 $W = w_1 w_2 \dots w_k$,对应的帧序列为 $T = t_1 t_2 \dots t_l$,对于任何一帧 t_i 判断它是属于词 w_m 还是词 w_{m+1} ,如果 $t_i \in w_m$, $t_{i+1} \in w_{m+1}$,那么帧 t_i 为段的右边界,帧 t_{i+1} 为另一段的左边界.利用孤立词模型参数(见第 3 节)按照孤立词识别方法计算 t_i 属于词 $w_m (m=1,\dots,k)$ 状态的概率,然后通过受限的动态规划算法搜索出一条最佳的词分割序列.受限搜索只能按照 $w_1 w_2 \dots w_k$ 这样的词序列进行搜索.我们以此搜索结果作为训练的目标边界输出.

引入自适应学习速率、附加动量项的反向传播算法作为 SRN 基本学习算法,将训练集中的样本经过 SOFM 网的变换,然后进行编码,与下文矢量一起输入到 SRN 网中,计算出相应各时刻的输出,将该输出与理想输出比较,计算出误差并反向传播,根据误差调整 SRN 的网络权值.在学习开始时,将权值矩阵、隐层神经元阈值和输出层神经元的阈值都赋予 $(-1,+1)$ 区间内的随机值,反馈单元初始化为 0.5.学习速率初始值为 0.01,附加动量项的动量因子为 0.95.期望误差最小值为 0.01.

识别:对于测试集中的连续手语,首先输入到 SOFM 中经过特征提取,形成量化的输出,将量化的输出经过编码,与下文矢量一起输送到 SRN,经过 SRN 的边界检测,在 t 时刻的检测输出结果为 $i^* = \arg \max_i (o_i^i)$.在检测时,

根据左右边界相邻特性给出限制.

2 HMM 框架

SRN 的分段结果输送到 HMM 框架中,一段或若干段(2~4 段)组合在一起形成 HMM 的状态.这是由于手语词的长短不一,一个手语词可能由几个段组成,需要以段为单位,在段中搜索出一条最佳词路径.这里的最佳指两方面,一方面是指重新组合的段序列最佳,另一方面是指在重新组合的最佳段序列中所选的词序列最佳.为了适应这种网格搜索,标准的 Viterbi 算法被改进,这里称之为网格 Viterbi(lattice Viterbi)算法.

为了区分,我们定义两种 Viterbi 算法.孤立词 Viterbi 算法:对于输入数据帧序列,预先知道它是由一个词组成,在识别时,Viterbi 算法每走一帧时只搜索该词的所有状态,最后选择最佳状态序列.连续语句 Viterbi 算法:对于帧序列,预先不知道它是由几个词组成,在识别时,Viterbi 算法每走一帧时,不但搜索该词的所有状态,而且还要搜索识别空间中其他所有词的状态,最后选择最佳状态序列.对于同一段手语数据帧序列,使用孤立词 Viterbi 算法识别比使用连续语句 Viterbi 算法识别精度高,识别时间短.

孤立词 Viterbi 算法和连续语句 Viterbi 算法都是一帧一帧地搜索,本质上是一样的,都属于标准的 Viterbi 算法.然而网格 Viterbi 算法与它们的区别是可以跨段搜索.下面详细讨论该算法.

2.1 网格 Viterbi 算法

定义三元组 $\langle t, t', q \rangle$ 为网格图中的重组后的段,以第 t 段开始,以第 t' 段结束,代表第 q 个词, $0 \leq t < T, t < t' \leq T$;所有三元组 $\langle t, t', q \rangle$ 形成集合 L ,它是由段重组和手语词的段特性共同决定.定义 $\delta(t, t', q)$ 为包含三元组 $\langle t, t', q \rangle$ 的局部最优累加器.为了能回溯出最佳路径,定义辅助变量 $\psi(t, t', q)$ 指示当前 $\delta(t, t', q)$ 局部最大时前一个三元组变量.定义 $b(t, t', q)$ 为从第 t 段到第 t' 段是词 q 的发射概率,具体计算将在第 2.2 节讨论. $P(q|q')$ 为从词 q' 到词 q 的转移概率,我们以 1995 年、1996 年《人民日报》作为语料统计出转移概率.网格 Viterbi 算法如下:

初始化:

$$\delta(0, t, q) = b(0, t, q), \tag{5}$$

$$\psi(0, t, q) = \text{NULL}. \tag{6}$$

递归:

$$\delta(t, t', q) = \max_{\langle t'', t', q' \rangle \in L} \delta(t'', t', q') P(q|q') b(t, t', q), \tag{7}$$

$$\psi(t, t', q) = \arg \max_{\langle t'', t', q' \rangle \in L} \delta(t'', t', q') P(q|q'). \tag{8}$$

终止:

$$P^* = \max_{\langle t, T, q \rangle \in L} \delta(t, T, q), \tag{9}$$

$$\langle t_1^*, T, q_1^* \rangle = \arg \max_{\langle t, T, q \rangle \in L} \delta(t, T, q). \tag{10}$$

路径回溯:令 $T = t_0^*$,递归 $\langle t_{i+1}^*, t_i^*, q_{i+1}^* \rangle = \psi(t_i^*, t_{i-1}^*, q_i^*)$ 直到 $\langle t_{k+1}^*, t_k^*, q_{k+1}^* \rangle = \text{NULL}$, 这样 $q_k^* \dots q_1^*$ 为搜寻的最佳词序列.

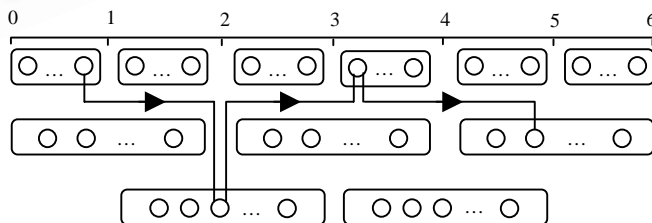


Fig.3 The sketch map of the lattice Viterbi algorithm

图 3 网格 Viterbi 算法示意图

图 3 以一个含有 6 段连续手语句子来表示网格 Viterbi 算法搜索过程(为了表达清晰,第 3 段和第 4 段组合在图中没有画出)。例如,在图 3 中,当 $t=4, t'=6$ 时, $t''=3$ 或 $t''=2$, 经过公式(8)的计算, $t''=3$ 是使得 $\delta(4,6,q)$ 段局部最大, 这样 $\psi(4,6,q) = \langle 3,4,q' \rangle$ 。在该图中最终的搜索结果是 $\langle 0 \ 1 \rangle \langle 1 \ 3 \rangle \langle 3 \ 4 \rangle \langle 4 \ 6 \rangle$ 4 段, 4 段中相应的词(用圆圈表示)构成了最佳词序列。

2.2 段内概率计算

段内概率计算相当于一个孤立词识别的计算过程,但是它保存候选词的概率值,作为 HMM 框架中的发射概率 $b(t, t', q)$, 最后通过网格 Viterbi 算法去选择最佳路径。孤立词识别时我们采用 SOFM/HMM 方法计算,该方法是对基于 HMM 手语识别方法的改进,使用自组织特征映射(SOFM)的输出构造 HMM 的状态概率,将 SOFM 很强的自组织能力和 HMM 良好处理时间序列功能结合起来,在识别过程中使用后验概率对状态概率进行修正,增强了模型的模式辨识能力,提高了手语识别的正确率。

根据段内概率计算时所采用的模型不一样,我们分为嵌入式训练识别和无嵌入式训练识别。将训练集中的句子经过第 1.2 节自动分割方法分割成词及相应的帧序列,我们把分割后的词作为训练集,训练出一套模型参数,称为嵌入式模型参数。将嵌入式模型参数和孤立词模型参数合并在一起作为段内概率计算的候选模型进行识别,称为嵌入式训练识别。只是将孤立词模型参数作为段内概率计算的候选模型进行识别,称为无嵌入式训练识别。由于嵌入式训练识别通过参数训练将部分的上下文相关模型包含在参数里,所以它的识别率要高于无嵌入式训练识别,具体见实验结果。

3 实验比较与结果分析

在实验中,我们采用的数据库按如下方法构造。词汇表为 208 个孤立词,请 7 位手语老师分别将 208 个孤立手语词每个采集 3 遍,选取 5 位,每位 2 组(共 10 组)作为孤立词训练集,利用 SOFM/HMM 模型进行训练,训练出孤立词模型参数。从孤立词训练集中的 5 个选取 2 人(用 A, B 表示),从不在孤立词训练集中的 2 人选取 1 人(用 C 表示),共 3 人。由这 3 个人按照正常的手语速度把 208 个词组成的 100 个句子分别打 2 遍,共 6 组数据分别标记为 $A_1, A_2, B_1, B_2, C_1, C_2$ 。我们选取 A_1, B_1 作为训练集,该训练集是作为 SOFM, SRN 和嵌入式训练的训练集。从 A_2, B_2 中任选一个作为已注册测试集(Reg.), C_1, C_2 任选一作为未注册测试集(UnReg.)。我们分别将 SRN/HMM 和单纯使用的 HMM 在非特定人连续手语识别中作比较,结果见表 1~表 3。

Table 1 Un-Embedded training recognition, test set is Unreg.
表 1 无嵌入式训练识别,测试集为未注册测试集

Method	Recognition accuracy (%)	Recognition time (s/word)
HMM	68.1 (S=58, I=40, D=19)	0.503
SRN/HMM	73.6 (S=57, I=8, D=32)	0.241

识别方法, 识别率, 识别时间。

Table 2 Embedded training recognition, test set is Unreg.
表 2 嵌入式训练识别,测试集为未注册测试集

Method	Recognition accuracy (%)	Recognition time (s/word)
HMM	81.2 (S=35, I=25, D=9)	1.016
SRN/HMM	85.0 (S=33, I=6, D=16)	0.479

识别方法, 识别率, 识别时间。

Table 3 Embedded training recognition, test set is Reg.
表 3 嵌入式训练识别,测试集为已注册测试集

Method	Recognition accuracy (%)	Recognition time (s/word)
HMM	90.7 (S=13, I=18, D=3)	1.025
SRN/HMM	92.1 (S=12, I=5, D=12)	0.485

识别方法, 识别率, 识别时间。

上述实验均使用 Bi-gram 语言模型,在 PIII450(192M 内存)微机运行,S(substitution)表示替代错误, I(insertion)表示插入错误,D(deletion)表示删除错误,在测试集中总的手语词数为 367。结果表明,SRN/HMM 比单

纯使用 HMM 具有更高的识别率和更快的识别速度. 形成这样的原因可能有以下几个方面:(1) 由于 HMM 模型在识别过程中采用连续语句 Viterbi 算法, 很容易受“运动插入”的影响. SRN/HMM 通过识别出连续语句的词边界, 在边界前后根据 SOFM 输出结果去过渡帧, 这样可减少“运动插入”的影响, 提高了识别率;(2) SRN/HMM 寻求的是最佳的词序列, 而不像 HMM 模型寻求的是最佳的状态序列;(3) 采用孤立词 Viterbi 算法识别单个词正确率高于采用连续语句 Viterbi 算法识别;(4) 由于 SRN/HMM 采用的网格 Viterbi 算法是跨段搜索, 只需在 SRN 输出的分段处检测是否为词分断点, 而 HMM 在识别时将每一帧都视为潜在的词分断点进行检测, 所以 SRN/HMM 比 HMM 识别速度快.

但是, SRN/HMM 由于采用分治的策略, 可能造成错误累加, 为了避免这种情况, 我们在 SRN 分段时采用软分段而不是确定性的分段的措施, 最后在网格 Viterbi 算法中才确定该句子中每个词的边界, 以保证算法具有较好的识别效果. 在实时识别时, 需要在数据采集集中缓存一段数据, 但是这并不影响实时识别的效果.

4 结 论

本文针对非特定人连续手语识别进行研究, 提出了一种基于分治策略、使用显式分割过程的方法. 首先通过改进的 SRN 进行分段, 分段的结果重新组合作为 HMM 的状态输入, 然后通过网格 Viterbi 算法从中搜寻一条最佳的手语词路径. 它可以减少“运动插入”带来的影响. 实验结果表明, 它的识别速度比单纯使用 HMM 更好的识别结果和更快. 此外, 改进的 SRN 算法、网格 Viterbi 以及实验得出的连续手语段特性不仅可以用于该算法之中, 也可以作为进一步研究其他算法的基础. 我们下一步的工作将是在其基础上进行大规模手语词汇的非特定连续手语的研究.

References:

- [1] Starner, T., Weaver, J., Pentland, A. Real-Time American sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(12):1371~1375.
- [2] Liang, R.H., Ouhyoung, M. A real-time continuous gesture recognition system for sign language. In: Yachida, M., ed. *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition*. New York: IEEE, 1998. 558~565.
- [3] Vogler, C., Metaxas, D. ASL recognition based on a coupling between HMMs and 3D motion analysis. In: Davis, L., *et al.*, eds. *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE, 1998. 363~369. <http://citeseer.nj.nec.com/vogler98asl.html>.
- [4] Vogler, C., Metaxas, D. Toward scalability in ASL recognition: breaking down signs into phonemes. In: Braffort, A., Gherbi, R., Gibet, S., *et al.* eds. *Gesture-Based Communication in Human-Computer Interaction*. Berlin: Springer-Verlag, 1999. 400~404.
- [5] Gao, Wen, Ma, Ji-yong, Shan, Shi-guan, *et al.* HandTalker: a multimodal dialog system using sign language and 3-D virtual human. In: Tan, Tie-niu, Shi, Yuan-chun, Gao, Wen, eds. *Advances in Multimodal Interfaces-ICMI*. Berlin: Springer-Verlag, 2000. 564~571.
- [6] Elman, J.L. Finding structure in time. *Cognitive Science*, 1990, 14(2):179~211.
- [7] Robinson, T. An application of recurrent nets to phone probability estimation. *IEEE Transactions on Neural Networks*, 1994, 5(2): 298~305.
- [8] Kershaw, D.J., Hochberg, M., Robinson, A.J. Context-Dependent classes in a hybrid recurrent network-HMM speech recognition system. *Advances in Neural Information Processing Systems*, 1996, 8:750~756.
- [9] Senior, A., Robinson, A.J. Forward-Backward retraining of recurrent neural networks. In: *Advances in Neural Information Processing Systems*, Vol 8. 1996. 743~749.
- [10] Murakami, K., Taguchi, H. Gesture recognition using recurrent neural networks. In: *Proceedings of the CHI'91 Human Factors in Computing Systems*. New York: ACM Press, 1991. 237~242. <http://portal.acm.org/citation.cfm?id=108900&coll=portal&dl=ACM&CFID=662409&CFTOKEN=62522583>.
- [11] Rumelhart, D.E., Hinton, G.E., Williams, R.J. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L., eds. *Parallel Distributed Processing*, Vol 1. Cambridge, MA: MIT Press, 1986. 318~362.

A Signer-Independent Continuous Sign Language Recognition System Based on SRN/HMM*

FANG Gao-lin¹, GAO Wen^{1,2}, CHEN Xi-lin¹, WANG Chun-li³, MA Ji-yong²

¹(Department of Computer Science and Engineering, Harbin Institute of Technology, Harbin 150001, China);

²(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China);

³(Department of Computer Science and Engineering, Dalian University of Technology, Dalian 116000, China)

E-mail: {fgl,wgao,xlchen,chlwang,jyma}@ict.ac.cn

<http://mmi.ict.ac.cn>

Abstract: Sign language recognition is to provide an efficient and accurate mechanism to transcribe sign language into text or speech. State-of-the-Art sign language recognition should be able to solve the signer-independent continuous problem for practical applications. In this paper, a divide-and-conquer approach, which takes the problem of continuous CSL (Chinese sign language) recognition as subproblems of isolated CSL recognition, is presented for signer-independent continuous CSL recognition. In the proposed approach, the SRN (simple recurrent network) is used to segment the continuous CSL. The outputs of SRN are regarded as the states of HMM (hidden Markov models) in which the lattice Viterbi algorithm is employed for searching the best word sequence. Experimental results show that SRN/HMM approach has better performance than the standard HMM.

Key words: neural network; simple recurrent network; hidden Markov model; continuous sign language recognition; signer-independent sign language recognition

* Received April 12, 2001; accepted July 13, 2001

Supported by the National Natural Science Foundation of China under Grant No.69789301; the National High Technology Development 863 Program of China under Grant No.863-306-ZD03-01-2; the 100 Outstanding Scientist Foundation of the Chinese Academy of Sciences of China