

# 中国手语音韵标记的建立、实现及其有效性验证

倪训博 赵德斌 姜 峰 程丹松

(哈尔滨工业大学计算机学院 哈尔滨 150001)

**摘 要** 手语是一门语言,手语语言学家从语言学的角度,对手语做了类似“语法”的总结而形成了音韵标注理论.针对目前中国手语无符号标记体系的现况,首先,根据中国手语的特点并结合汉语的语言习惯,文中提出了中国手语音韵标记体系.其次,针对目前音韵标注只给出一些形式化的描述或简单的手工标注的现况,文中借助基于数据手套的中国手语识别系统,提出了以手形数据模板与拉班舞谱相结合的方法,将上述的标记体系在计算机上以自动标注的方式加以实现.最后,针对数据手套的中国手语识别系统缺少先验信息的现况,应用上述音韵标记体系构建具有先验语义信息的易混词集,在构建的过程中说明并分析了音韵标记体系具备语义识别能力.实验表明,在基于统计识别的易混词集与基于先验语义信息的易混词集的识别效果相比较,在不改变统计模型的条件下,可以不同程度地辅助 HMM,SOFM/HMM, Effort 提高识别率.

**关键词** 音韵标注;拉班舞谱;语义信息;易混词集;HMM;SOFM/HMM; Effort

**中图法分类号** TP391 **DOI号**: 10.3724/SP.J.1016.2009.02438

## Establishment, Realization and Effectiveness Verification for Phonological Representation of Chinese Sign Language

NI Xun-Bo ZHAO De-Bin JIANG Feng CHENG Dan-Song

(School of Computer Science, Harbin Institute of Technology, Harbin 150001)

**Abstract** Sign language is a language. Linguists summarized grammar for sign language from the view of linguistics, and the theory of phonological representation was formed. Not having a system of sign digit in Chinese sign language, this paper firstly proposes a phonological representation according to the characteristic of Chinese sign language and language habit. Secondly, a method which combined data template of hand shape and Latin spectrum is proposed according to the characteristic of recognition system of Chinese sign language which is based on data glove, and the recognition system above is realized by the computer by means of automatic label. Finally, recognition system of Chinese sign language of data glove is lack of transcendental information, thus a confused word assemble with transcendental semantic information is established through the system of phonological representation. During the process of setting up the confused word assemble with transcendental semantic information, the semantic recognition function of the system of phonological representation is illustrated and analyzed. The confused word assemble based on statistical identification is compared with the one based on the prior sign-language semantic information. Without any change of statistic model, the experiment shows that the recognition function of phonological representation will to some extent improve the recognition rate which used the HMM,SOFM/HMM and Effort.

收稿日期:2007-12-10;最终修改稿收到日期:2009-06-06. 倪训博,男,1979年生,博士研究生,主要研究方向为模式识别、机器学习等. E-mail: nixunbo@hit.edu.cn. 赵德斌,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、多模式接口和人工智能. 姜峰,男,1978年生,讲师,主要研究方向为模式识别、机器学习、图像处理等. 程丹松,男,1972年生,博士研究生,讲师,主要研究方向为模式识别、数字图像处理.

**Keywords** phonology noting; labannotation; semantic meaning information; confusable vocabulary set; HMM; SOFM/HMM; effort

## 1 引言

### 1.1 手语音韵标记的研究状况

手语作为一种高度结构化的手势语言,是聋人进行信息交流最自然的方式和交际工具。目前,对于手语的研究主要包括理论及计算两个线索,即理论模型和计算模型。理论模型根据实验、假设,提出手势问题的基本理论,进而为计算模型提供有价值的指导;这些研究为计算手势模型的建立提供了良好的材料,许多计算模型以这些理论模型为基础。在手语语言学方面,Stokoe 首次将语言学应用在手语研究上,描述手语固有的语言特性。Stokoe 把手语构成的研究称为手形学(cherology),他认为手语有表达位置(place of articulation)、手形(handshape)及动作(movement)三种音系参数构成。他对手语作出的这种结构性的分析,为后来的手语研究打下了基础。目前世界上大多数的手语语言学者常用的是经典的 stokoe<sup>[1]</sup>表示法来表示手语,其它的一些表示方法还包括诸如英国的 kyle 和瑞典的 Bergman 的工作。以 prillwitz 为首的汉堡大学的手语研究组基于 stokoe 的工作,在标记方法中有意识地加入了国际音标(IPA),提出了“Hamnosys”标记法,并成功地将其应用于手语电子辞典的设计。传统的音韵理论赋予音素以记号,手语中的手形、位置、动作被认为是音素,而认为手势运动是复杂的,是不可标记的。Liddell 和 Johnson 打破古典手语音韵理论的框架<sup>[2]</sup>,提出了基于音节理论的手语标记法,以可以显示手语词差别的特征来进行手语标记。他们的工作定义了新的音韵单位——分节(segment),分节由手势(the posture of the hand)和动作(activity)组成。具体的如手形、指尖的方向、手掌的方向以及手的动作定义为一个模块,运动被规定发生在身体的周围;这个思路可以将复杂的手势运动简单化、模型化(Move-Hold model)。Sandler<sup>[3]</sup>于 1989 年基于音节理论和 Liddell 与 Johnson 的工作提出了 TH 模型并用于手语标记。岛越隆士<sup>[4]</sup>对这两种模型作了分析和介绍。在日本的手语标记的研究上,本名、神田等人在手语语言学领域上做了细致的研究<sup>[5]</sup>。神田<sup>[6]</sup>在 1986 年在本名他<sup>[7]</sup>的手形标记法的基础上,提出了日本手语的音韵标记,包括基于数字、英文字

母、片假名、日文汉字的日文手形的音韵标记法;神田于 1991 年提出了比较完善的日本手语标记体系<sup>[8]</sup>,并于 1999 年提出了基于句子的标记体系<sup>[9]</sup>,神田又于 2007 年发表了手语认知学的研究<sup>[10]</sup>。

### 1.2 手语识别领域的研究状况

计算模型涉及基于机器学习、计算机视觉、模式接口领域等方面的工作,这些领域的研究关注的是对手语的操作,大部分工作都面向手语识别系统。特定人手语识别的尝试始于 20 世纪 90 年代。Marble<sup>[11]</sup>使用图像处理方法来理解美国手语中 31 个孤立手势词,该方法能正确识别其中的 27 个。Hinton<sup>[12]</sup>使用 VPL 数据手套和 Polhemus 位置跟踪器作为输入设备,神经网络作为手势分类器,根据手的运动轨迹、手形、手的运动方向、手的偏移量以及手的运动速度为特征形成 5 个功能网来进行识别。Kadous<sup>[13]</sup>采用 PowerGlove 作为手势输入设备,利用基于实例的学习和决策树判别等方法,实现了 Auslan 中 95 个手势词的识别,正确率达到 80%。Assan<sup>[14]</sup>系统采用 HMM 模型识别 262 个孤立手语词,正确率达 91.3%,该系统需要手语者带上有色手套,然后通过视频提取二维特征。Sagawa<sup>[15]</sup>使用基于语素的方法进行日本手语的识别,该方法识别 60 个日本手语词,识别率为 80.1%。Kim<sup>[16]</sup>对南韩手语进行了研究。系统利用一副数据手套作为感知手及手指运动的设备,利用模糊最大最小神经网络技术对 131 个手势词进行在线识别,识别率为 94.3%。Deng<sup>[17]</sup>使用基于并行的 HMM 模型去识别 192 个美国手语词,识别率为 93.3%。Sutherland<sup>[18]</sup>对爱尔兰手语进行了研究,他们提出分层的手语识别算法,在纯色手套的帮助下识别 100 个手语,识别率达到 89.6%。

### 1.3 非特定人识别领域的研究状况

非特定人手语识别可以减少单独一个用户训练所需的样本,是推动手语识别系统实用化所必需解决的问题。从现有的一些非特定人手语识别结果来看,非特定人手语识别与特定人手语识别相比仍有较大差距。非特定人手语识别的部分研究工作我们总结在表 1 当中。目前大部分的非特定人手语识别在训练集和测试集仅仅包括 2~10 人的数据。非特定人手语识别中所用最大的训练集规模是 20 人数据<sup>[19-21]</sup>,这远远少于训练一个友好的语音识别系统

所需人数. 当一个用于训练的数据中只含有较少人数的数据时, 非注册集的结果会很差. Kadous<sup>[22]</sup>使用 4 个人训练, 非注册集只达到平均 15%~80% 的准确率. Assan<sup>[23]</sup>使用一个人的数据训练, 注册集和非注册集的结果分别为 94% 和 51%. 当用更多人的数据训练时, 效果会有所改善. Vamplew<sup>[24]</sup>使用 7 个人的数据, 得到注册集 94.2%, 非注册集 85.3% 的识别结果. 当仅仅考虑手形数据时, 非注册集和注册集的识别率会相对接近, 这也许是因为与其他成分相比, 不同人手形上的变化相对较小. 例如, 文献<sup>[25-26]</sup>报告了使用手形分类器达到了注册集 93%~96%, 非注册集 85%~91% 的准确率. 有趣的是, Kong<sup>[27]</sup> 对非特定人 3D 轨迹进行分类, 报告了类似的好结果: 注册集

99.7%, 非注册集 91.2% 的准确率. 我们前期设计的 SOFM/HMM 模型, 是将数据转换成一个紧凑、重要的低维表示形式, 使用隐含的方法进行特征提取<sup>[28]</sup>. 这种提取是在缺少对手语运动理解的情况下进行的, 因此不可避免地会造成有意义特征的丧失. 在语音识别领域, 人们采用说话人自适应算法克服特定人和非特定人系统各自的缺点, 利用少量的训练语音调整系统的参数, 使得系统性能有所提高. Ong<sup>[29]</sup>将 MAP 用于基于贝叶斯网络的成分分类器的估计, 达到了 88.5% 的识别率. 在大词汇量的识别系统, MAP 算法对适应训练语音中出现的语音模型做更新识别效果理想, 而对未出现过的语音模型自适应速度缓慢且识别率低.

表 1 非特定人手语识别的方法与系统

Signer-independent SLR approaches							
Researchers	D/V	Recognition methods	Sign Vocab.	Subjects		Rec. rate/%	
				Train	Test	Reg.	Unreg.
Kadous	D	Instance based learning and decision tree learning	95 Auslan	5	5	80	
				4	1		12-15
Vamplew & Adams	D	Multilayer perceptron NN & Nearest neighbor lookup	52 Auslan	7	4	94.2	
				7	3		85.3
Imagawa	V	PCA+clustering	33 JSL	6	6		72-94
Fang	D	HMM/SOFM	208 CSL	5	2	92.1	
				5	1		85

D/V: Data Glove based approach or Vision based approach

以上的这些方法虽然研究了问题的不同层面, 但却没有回答一些基础的问题. 如何设计中国手语的音韵标记方法? 如何实现对手语的自动标记? 如何将先验信息引入传统的统计识别模型当中去? 本文将对这些问题展开深入的讨论与研究.



















## 2 中国手语音韵标记体系的建立

同一般的人体运动范畴相比, 手语有着高度的结构性. 这种结构性主要来自于手语本身形式化的定义, 并可以理解为对待识别手语丰富的先验信息. 对手语的标注是对结构性有效、直观的体现, 从而可以使我们能够更好地描述、理解并识别手语. 手语识别研究中, 如何利用手语本身的结构性, 亦即先验信息, 以提高分类器的性能, 是一个迫切需要讨论与研究的问题. 手语的音韵标记包括对手形、方向、位置、运动的标记以及手语节、手语各组成部分的时序关系的标记规则. 而这些就是我们手语识别系统所需要的先验信息.

### 2.1 手形标记记号

关键手形是指手语字典中所规定的手形形式, 在打手语的过程中表现为保持阶段所对应的手形. 一个手语词中关键手形的个数是由手语字典中规定的情况确定的. 我们以数字和汉语拼音为主要的键手形标记. 并对手语中出现的键手形, 按照与主要的键手形标记的相似程度, 我们对键手形作了如表 2 的整理和归类.

表 2 数字与拼音的关键手形标记记号

数字标记符号							
	4		5				
拼音标记符号							
	a		b		c		d
	e,3		f,v,2		g,i,1		h
	j,9		k		l <sub>1</sub>		m
	n		o,0		p		q,7

(续 表)

拼音标记符号			
	r, l, 2, 8		s
	t		u
	w		x
	y, 6		z
	zh		ch
	sh		ng

来源于手语关键手形,根据手形与汉语拼音的相似程度进行分类,为了便于说明我们将此类手形称为“组合手形”,如表 3 所示。

表 3 组合手形的关键手形标记记号

a 类相似组合手形		a <sub>1</sub>		a <sub>2</sub>		a <sub>3</sub>
c 类相似组合手形		c <sub>1</sub>		c <sub>2</sub>		c <sub>3</sub>
		bc		4c		4c <sub>1</sub>
e 类相似组合手形		e <sub>1</sub>		e <sub>2</sub>		e <sub>3</sub>
h 类相似组合手形		h <sub>1</sub>		h <sub>2</sub>		h <sub>3</sub>
		j <sub>1</sub>		j <sub>2</sub>		j <sub>3</sub>
		fj		gj		oj
j 类相似组合手形		rj				
		k <sub>1</sub>		k <sub>3</sub>		
k 类相似组合手形						
n 类相似组合手形		n <sub>1</sub>				
o 类相似组合手形		o <sub>1</sub>		o <sub>2</sub>		
q 类相似组合手形		q <sub>1</sub>		q <sub>2</sub>		
r 类相似组合手形		r <sub>1</sub>		r <sub>3</sub>		
t 类相似组合手形		t <sub>1</sub>				
x 类相似组合手形		x <sub>1</sub>		x <sub>2</sub>		x <sub>3</sub>
z, zh 类相似组合手形		z <sub>1</sub>		zh <sub>1</sub>		
ch, sh 类相似组合手形		ch <sub>1</sub>		ch <sub>2</sub>		ch <sub>3</sub>
		sh <sub>1</sub>				
ng <sub>1</sub> 类相似组合手形		ng <sub>1</sub>		ng <sub>2</sub>		

## 2.2 方向标记记号

手形的方向包括手指的方向和手掌的方向,由于手指的情况比较复杂,根据数据手套的特点,用手腕的方向来代替手指的方向.此方向数据通过腕部的传感器获得,用表征手与参照间的方向的三维数据刻画手掌的方向.采用手腕方向代替手指的方向,可以避免手指伸开或闭合时带来的方向的不确定性.这里手掌的用 P 表示,手腕用 W 表示.具体方向的标记见表 4.

表 4 方向的标记记号

方向	右	左	中	上	下	前	后	左前	左后	右前	右后
标记	R	L	M	U	D	F	B	LF	LB	RF	RB

手形和方向的具体标记的例子,如图 1 和图 2 所示.图 1 可分别标记为 fPUWB,其中,f 表示手形所标记的汉语拼音 f,PU 表示手掌向上,WB 表示手腕方向向后.图 2 可分别标记为 gPLWB,g 表示手形所标记的汉语拼音 g,PU 表示手掌向上,WB 表示手腕方向向后.

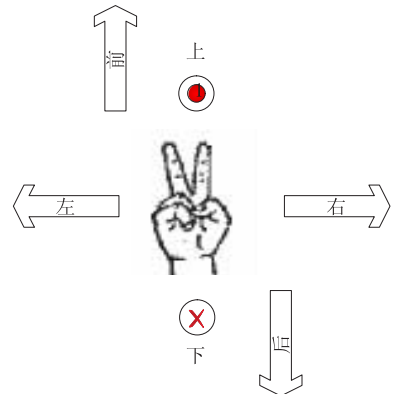


图 1 “f”手形和方向的具体标记示例

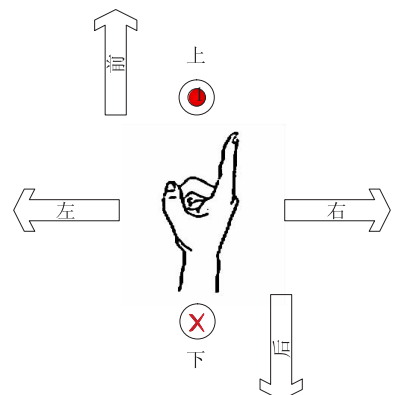


图 2 “g”手形和方向的具体标记示例

## 2.3 位置标记记号

基于语言学手语标记的位置记号包括身体的各

个部位的记号、手与身体各个部位的距离记号以及两手位置的关系的记号(见表5)。其中手与身体各个部位的距离记号表示根据身体的部位细化剩余空间位置、关键手形与身体各个部位的远近。在双手手语词中,两手的位置关系就是双手打手语词的姿势。在经典的手语音韵理论中,将双手为相同的手形称为对称条件;双手手形不一致时,副次手(通常指左

手)不动的情况,称之为优位条件。在对称条件与非对称条件下,两手位置的关系的记号表示有所不同。其中,组合(lin)是在对称条件下,除了相同(idn)、交替(alt)、交叉(cro)、接触(con)以外的其他的位置关系。手指(FI)和小臂(LA)交叉(cro)的位置关系分别标记,如图3所示。

表5 身体各部位及其距离与位置的标记记号

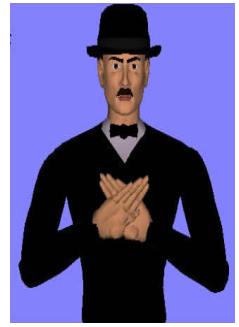
身体的各个部位的记号	HD:头,FA:脸,FH:额,EY:眼睛,NO:鼻,MO:口,EA:耳,NK:脖,TH:牙,TN:舌,CK:脸颊,CN:颞,TM:太阳穴,BR:胸,BL:腹,SH:肩,EL:肘,UA:上臂,LA:小臂,WD:手腕外侧,WV:手腕内侧,HB:手背,HP:手心,HA:手的周围,FG:指尖
手与身体各个部位的距离记号	con:接触,near:近距离,mid:中距离,far:远距离 双手手语的情况,同侧用“1”标记,不同侧用“C”标记
两手位置的关系的记号	对称条件 Idn:同,alt:交替,cro:交叉,con:接触,lin:组合 非对称条件 左右手的相互关系用“/”加以分隔;主手(通常指右手)在“/”之前,副次手(通常指左手)在“/”之后



(a) 对称条件下双手手形相同(idn)



(b) 对称条件下双手手形的交替(alt)



(c) 对称条件下双手手形的交叉(cro)



(d) 对称条件下双手手形的手指交叉(ficro)



(e) 对称条件下双手手形的接触(cro)



(f) 对称条件下双手手形的关系(lin)

图3 对称条件下双手手形

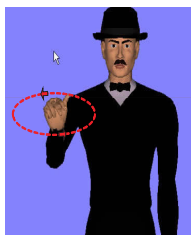
## 2.4 运动标记记号

经典的手语语言学中,至今还没有对动作有较为明确的阐释。在现代的手语音韵理论中,把动作分为轨迹运动(path movement)和局部运动(local movement)。所谓的局部运动(local movement)在这里指的就是手的运动。手的运动包括手指的运动、手腕的运动以及肘的运动。运动的种类简单地可分为弯曲、伸展、转动。根据动作的幅度可分为大幅度

的动作、短暂重复性的动作、平稳缓慢的动作、快速的动作。但是在实际的应用中,存在着很多复杂的运动,例如由各种运动组合时所产生的与拇指摩擦的运动;顺序数数时手指屈伸的动作等等。由于手语运动是复杂的,本文中去掉了一些不必要的细分类,为了保证音韵标记的完整性,并根据中国手语的特点做了如表6的整理。

表 6 运动的标记记号

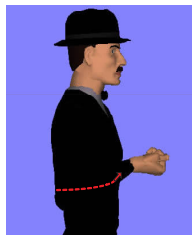
基本运动标记	旋转(ro)、弯曲(be)、伸展(ex)、开放(op)、闭合(cl)、摆动(wg) 为了限定特定的部位,可以加上动作的相应的身体部位,例如手指(FI)、手腕(WR)、肘(EL)
复杂运动标记	tw:肘的转动,co:数数时手指的顺序弯折,ru:手指摩擦、互捻,wr 书写汉字.此外,还要附加一个动作,敲击或盖章的动作用(st)进行标记.
运动轨迹	运动路线 无标记:直线,C:曲线
	运动面 HP:与地面平行的平面称为平面,VP:与平面(HP)垂直的平面称为立面,SP:与身体平行的平面称为表面;MP 与表面垂直的平面称为断面,UP:沿离开身体的方向上升的平面称为上升面.
	运动方向 仍用 2.2 节所介绍的方向标记进行表示.
运动样态	运动的样态 分别用大(BIG)、小(SML)、快(FAS)、慢(SLO)、重复(*).例如:重复 2 次用 *2 来标记,重复 3 次用 *3 来标记.不指定次数的重复场合进用 * 来标记,例如一些抖动的动作等.



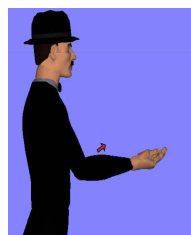
(a) 旋转(ro)运动



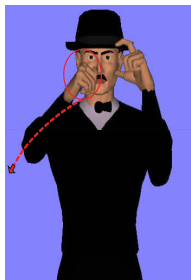
(b) 弯曲(be)运动



(c) 伸展(ex)运动的初始位置



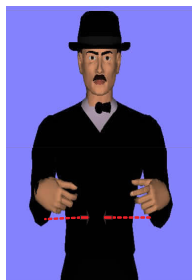
(d) 伸展(ex)运动的结束位置



(e) 开放(op)运动的初始位置



(f) 开放(op)运动的结束位置

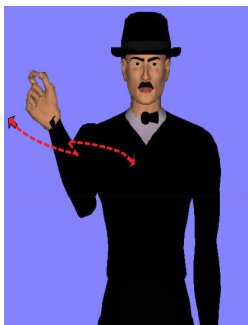


(g) 闭合(cl)运动的初始位置

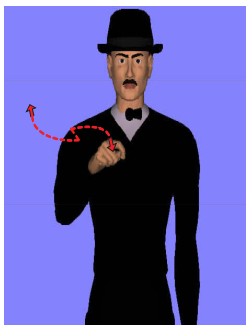


(h) 闭合(cl)运动的结束位置

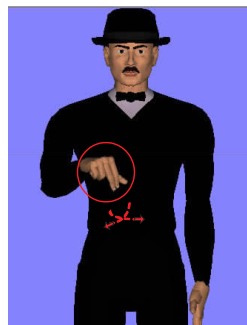
图 4 基本运动示例



(a) 摆动(wg)运动重复一次的初始位置



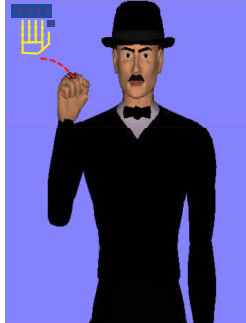
(b) 摆动(wg)运动重复一次的初始位置



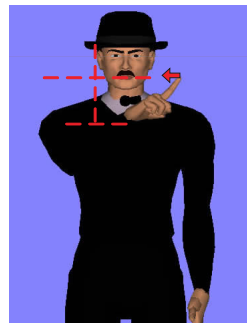
(c) 手指摩擦,互捻可设为(ru)



(d) 数数时手指的顺序弯折(co)的初始位置



(e) 按照图中手指的数字顺序到达结束位置



(f) 书写汉字(wr)运动

图 5 复杂运动示例

## 2.5 音韵标记的其它一些规定

- ① 顺序发生:用“·”符号标记.
- ② 同时发生:用“+”符号标记.
- ③ 保持:⋈(保持的手形(方向)、空间位置等).
- ④ 交叉:以{手语节 1}+{手语节 2}·{手语节 1}·{手语节 2}的形式加以标记.

## 3 中国手语音韵标注基于数据手套的手语识别系统的实现

目前我们的手语识别系统所使用的数据手套,主要是以手形、方向、运动位置为主要特征的,因而中国手语音韵标注的实现也是围绕上述特征完成的,而上述特征也是我们实现中国手语音韵标注的客观条件.

### 3.1 数据手套的识别系统的介绍

#### 3.1.1 CyberGlove 型号数据手套

我们目前研究的手语识别系统,是以美国 Virtual Technologies 公司生产的具有 18 个传感器的 CyberGlove 型号数据手套及其配套设备 Polhemus FASTRAK 3-D 位置跟踪器作为手势输入设备.该型号的数据手套代表了当今最先进的设备手套,由该设备采集的手语数据精确度较高,其测量角度示意图如图 6 所示.

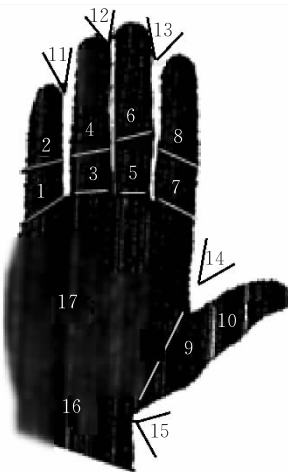


图 6 Cyberglove 测量角度示意图

利用具有 18 个传感器的 CyberGlove 型号数据手套,将得到手骨及其各关节的原始传感数据,即小指 MPJ(Pinkie MPJ)、小指 PIJ(Pinkie PIJ)、无名指 MPJ(Ring MPJ)、无名指 PIJ(Ring PIJ)、中指 MPJ(Middle MPJ)、中指 PIJ(Middle PIJ)、食指 MPJ(Index MPJ)、食指 PIJ(Index PIJ)、拇指 MPJ(Thumb MPJ)、拇指 IJ(Thumb IJ)、小指无名指间

外展(Pinkie-ring Abduction)、无名中指间外展(Ring-middle Abduction)、中食指间外展(Middle-index Abduction)、拇指外展(Thumb Abduction)、拇指旋转度(Thumb Rotation/TMJ)、腕的外展(Wrist Yaw)、腕的曲度(Wrist Pitch)、掌的弧度(Palm Arch),人手各种骨及关节的解剖图(见图 7).这些角度体现了手指的弯曲与伸展、手掌的曲率以及手腕的旋转,输出大小与相应传感器所在位置的弯曲程度成线性关系,范围为 0~255. MPJ、PIJ、DIJ 及腕的关节曲度随着 A/D 值的增加而增加,外展传感器输出的 A/D 值随着外展关节的扩张而减少,腕关节向小指一侧的弯曲对应着 A/D 值的增加.

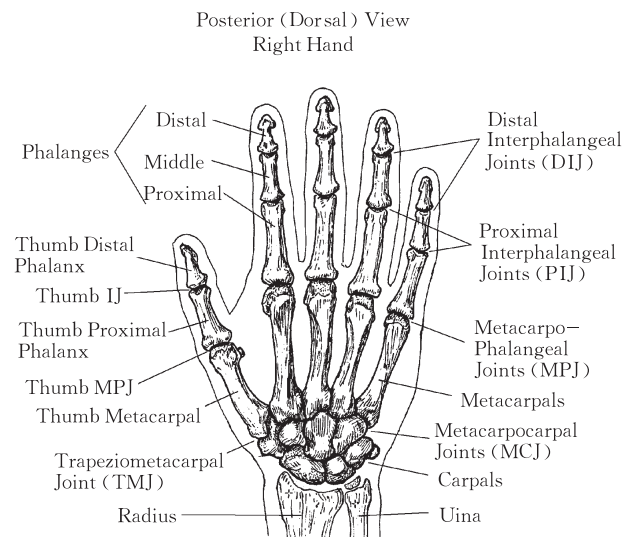


图 7 人手解剖图

#### 3.1.2 FASTRAK 位置跟踪器

上述数据手套都需要再配备三个位置跟踪器才能成为一套完整的手语输入设备,利用带有两个置于双手腕部接收器的 Pohelmus FASTRAK 3-D 位置跟踪器将得到相对于发射器的手的位置及手的朝向等信息. FASTRAK 跟踪系统利用电磁场来确定远处物体的位置及方向.该系统基于两种技术:(1)利用 3 个并置、静止天线组成的单个装置(称为发射器)产生邻近域、低频电磁场向量技术;(2)利用 3 个并置、遥感天线组成的单个装置(称为接收器)检测磁场向量的技术.3 个接收器分别位于左右手的腕部和身体的背部,如图 8 所示.

通过数学算法,被感知的信号被计算为相对于发射器的接收器的位置及方向.一个样本是一个帧序列.帧的数目从 10~20.一帧由 48 维数据组成,包括手套数据和位置传感器的数据,我们采用基于几何分析的算法提取与打手语者位置无关的特征,

每维数据都归一化到 0~1 之间. 每只手用 24 维数据描述, 18 维表示手形, 3 维表示手的位置, 3 维表示手的方向. 接收器相对发射器的几何关系如图 9 所示.



图 8 数据手套和三维跟踪器(3个接收器分别固定在双手和背部)

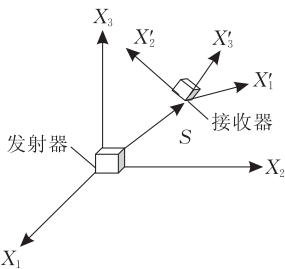


图 9 接收器相对发射器的几何关系

接收器的位置坐标及方向角度允许值的范围如图 10 所示.  $x_{\max}, y_{\max}, z_{\max}, x_{\min}, y_{\min}, z_{\min}$  的值分别为: 200, 200, 200, -200, -200, -200.  $a_{\max}, e_{\max}, r_{\max}, a_{\min}, e_{\min}, r_{\min}$  的值分别为 180, 90, 180, -180, -90, -180.

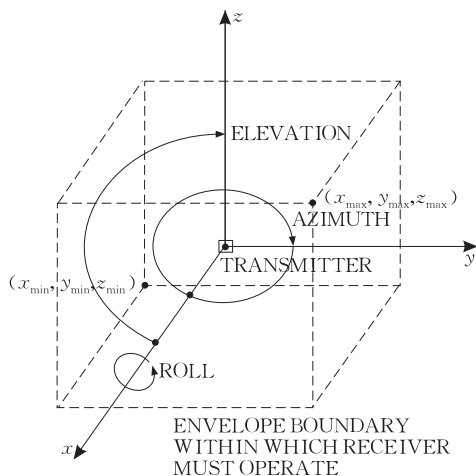


图 10 FASTRAK 3-D 位置跟踪器的接收范围

### 3.1.3 CyberGlove 型号数据手套及 FASTRAK 跟踪系统采集的数据说明

我们目前所使用的手语数据总共是 51 维. 具体如下:

① 右手相对参照(即颈部的接收器)的 3 维方向数据和 3 维位置数据;

② 左手相对参照(即颈部的接收器)的 3 维方向数据和 3 维位置数据;

③ 左右手之间的欧式距离, 1 维数据表示;

④ 左手与参照之间的欧式距离, 1 维数据表示;

⑤ 右手与参照之间的欧式距离, 1 维数据表示;

⑥ 左手手形 18 维数据, 右手手形 18 维数据.

### 3.2 手形音韵标记的实现

左右手共有 36 维数据存放左右手的手形, 其中左手 18 维, 右手 18 维, 即 3.1.3 节数据说明中的⑤和⑥. 我们首先对手势词中表征 36 维的手形数据, 分别对左、右手形数据帧进行分段, 主要采用沿时间轴的贪心聚类算法 (GLATA, Greedy Clustering Algorithm along the Time Axis), 使之能对手形数据帧进行分段. 来获得用于标记汉语拼音和数字的 28 个手形数据, 以及用于标记组合手形的 46 个的手形数据(即 2.1 节中所介绍手形标记). 最后, 我们应用 GLATA 的训练算法 (TAGLATA, Training Algorithm based on GLATA), 将手形数据训练成用于标记的手形模板. 我们使用滑动窗技术可以减少观察值短时间变化所带来的不良影响, 这些不良影响是由错误和特征提取不完善造成的. 详细的手形音韵标记的介绍和算法请参考文献[30].

### 3.3 方向和运动音韵标记的实现

由于手语运动是非常复杂的, 在手语语言学上仅是对一些简单的运动做了概念上的定义, 仅用模板匹配的方法对运动做标记是困难的. 中国手语音韵标记体系对手语运动的描述是“静态的”, 仅能描述动作出现的关键点, 并不能描述手语运动的“细节”. 由于手语有肢体运动的特性, 拉班舞谱是通过每个人体运动关节方向的细微变化来描述人体动作, 为此引入了拉班舞谱来对方向和运动音韵标记进行“动态”的实现.

#### 3.3.1 拉班舞谱的简介

拉班舞谱<sup>[31]</sup>是记录和分析人体运动的标记体系. Rudolf Laban 于 1928 年首次发表了“Kinetographie”标记记号, 这些记号被继承并发展, 并称为“拉班舞谱”, 见图 11~13. 拉班舞谱是一种既精密又合理的描述每一个人体运动细节的方法, 用一些文本性质的“结构化的记号”来描述人体运动关节的方向变化. 拉班舞谱起初是用来记人体舞蹈动作的标记, 后来被用于人体有韵律特性动作的检测.

根据拉班舞谱的定义, 一个动作可以分解为若

干个拉班运动节拍. 一个拉班节拍可以由几个拉班记号组成. 拉班记号对应着人类的关节. 拉班记号用分割空间的子空间来描述关节和与其相反的父关节. 这些分割由 9 个水平分割和 3 个垂直分割所组成. 拉班舞谱记号的不同的形状和阴影及经组合后可以区分不同的子空间. 用记录的前、后、左、右的方向来定义躯干动作流的方向. 因此每一个拉班运动节拍, 都定性地描述了在常数时间间隔内身体的姿

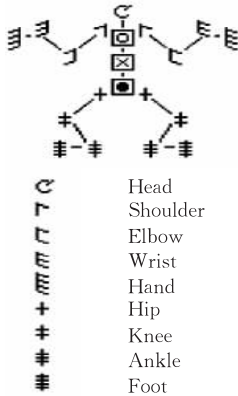


图 11 拉班运动的关节符号

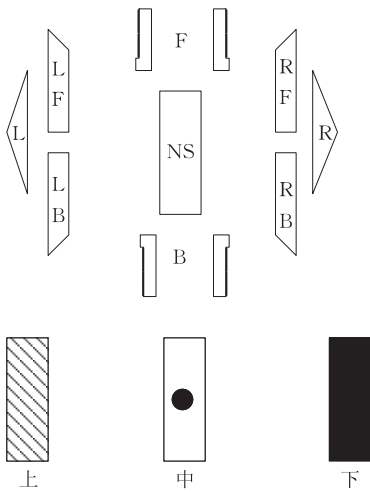


图 12 拉班运动的方向符号

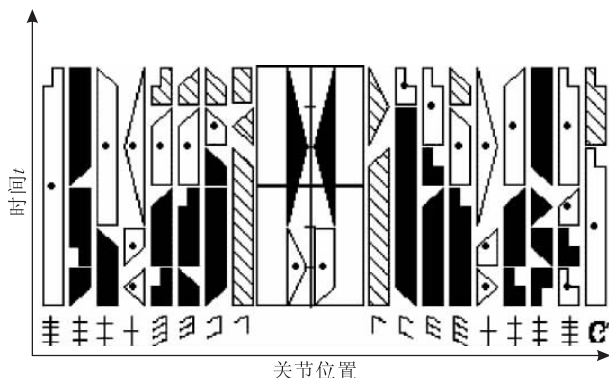


图 13 拉班运动序列的标记

势, 一个粗略的填满拉班节拍的序列, 描述了一个动作的表达形式. 在人体运动的检测中, 拉班舞谱由 17 个关节组成. 由 19 个记号所组成的拉班节拍, 追加了两个特殊的拉班运动符号来描述根方向和位置. 在手语动作里, 相关动作发生的关节均出现在上半身.

### 3.3.2 拉班舞谱对方向和运动音韵标记的实现

我们所用到的数据用于表征方向的仅有右手相对参照(即颈部的接收器)的 3 维方向和左手相对参照(即颈部的接收器)的 3 维方向, 即数据手套传感器数据①②的 3 维方向数据. 所能用到的关节仅有腕部(W)根据我们的具体问题对拉班舞谱做了改造, 即描述手势运动的拉班舞谱由 2 个关节和 12 个方向所组成. 上述图中的 NS 均代表颈部传感器的位置.

对拉班舞谱规定的 12 个方向(包括中立位置 NS)设定检测阈值  $\theta_F$ , 先用 LW 或 RW 标记左右手的腕关节, 然后按照时序标记相应的方向变化做拉班运动标记. 11 个方向对应的音韵标记: 前(F)、后(B)、左(L)、右(R)、左前(LF)、左后(LB)、右前(RF)、右后(RB)、上(U)、下(D)、中间(M). 手势词的序列为  $O = \{o_1, o_i, \dots, o_T\}$ , 对于其中的每一帧, 分别获得右手与左手方向标记, 并将标记加入队列 RWQ 与 LWQ.

每一个拉班记号可以唯一地映射到相对参照(即颈部的接收器)的 3 维方向特征向量. 这个方向特征向量能够从整个空间的均匀采样中选择对应于子空间的位置. 对 3 维方向特征向量赋予符号化的度量. 两个手势之间的差异可以被两个拉班节拍  $B_i$  和  $B_j$  所计算, 如下所示

$$D(B_i, B_j) = \sum_{k=1}^n \omega_k \|v_{i,k} - v_{j,k}\|^2 \quad (1)$$

$v_{i,k}$  和  $v_{j,k}$  是 3 维符号化的度量所构成的向量, 分别由腕部(W)关节的节拍  $B_i$  和  $B_j$  所映射.  $\omega_k$  是在拉班节拍内腕部(W)关节的权值. 每个关节的权值是按照对动作姿态影响的程度来设定的. 考虑到每个动作中与时间相关的速度和加速度. 我们在拉班序列中每一格节拍都构建一个居中的小窗. 设窗口的大小为  $2n+1$ , 这个滑动窗内的拉班节拍的距离度量可以由下面的式子获得:

$$LabanDist(B_i, B_j) = \sum_{l=-n}^n \omega_l D(B_{i+l}, B_{j+l}) \quad (2)$$

$B_{i+1}$  和  $B_{j+1}$  是在两个窗口内分别与  $B_i$  和  $B_j$  相邻的节拍. 而拉班运动序列是由上述的若干个关节节拍

所组成,我们目前所使用的手语数据仅有两个腕关节的数据,因而也就只有两个腕关节节拍.关于拉班舞谱对手语运动的标记算法、拉班舞谱方向标记的符号化度量以及拉班舞谱标记的运动序列的相似性判别,为了节省篇幅在这里不再做过多的叙述,详细描述请参考文献[30].

## 4 中国手语音韵标记及其匹配算法有效性的验证

所谓的易混词集,是在使用传统 HMM 对大词汇量非特定人手语进行统计识别的过程中,将识别错误的手语词以及导致这种识别错误的其它手语词所构成的手语词集.由于这种识别错误是由于统计模型区分能力所造成的,从严格意义上来讲上述手语词集应当称为基于统计识别所构建的易混词集,这里我们简称为易混词集.

根据手语节的定义,手语节= $\{\text{手形}; \text{空间位置}; \text{运动}\}$ .我们对手形和运动建立基于语义形式的数据模板.在实现的过程中我们将手语节简化成:手语节= $\{\text{关键手形}; \text{运动}\}$ 的形式.所以一个手语词标记= $\{\text{若干关键手形}; \text{拉班手势运动}\}$ 的形式,因而,我们在构建易混词集之前把所有手语词按上述形式进行标记,使得待测的易混词集具备了语义信息,这样就可以把先验知识引入到基于 HMM 的识别系统当中.在每个易混词集当中做相应的语义标记匹配,并构建出基于语义信息的易混词集,在该词集上用不同的统计模型进行识别,在不对统计模型做任何改动的情况下,减少了统计模型的识别错误.

### 4.1 基于统计识别的易混词集的构建

传统构建易混集的方法有 N-BEST 方法等<sup>[32-33]</sup>,这些方法在已有的 MLE 模型基础上,需要对模型进行全局区分性训练,并且选择  $N$  个打分最高的模型构建易混集.姜峰、叶剑君等<sup>[34]</sup>在基于支持向量机的二层 HMM 模型上对易混集的构造进行了相应的研究,对本文易混集构造有一定借鉴意义,本文在文献<sup>[35]</sup>的基础上构建了如下基于统计识别的易混词集.

易混词集的构建算法:

1.  $N$ =训练手语词总数, $M$ =识别错误的手语词的总数.
2. 为每个识别错误的手语词  $O_i$  建立一个易混词表  $Table(1, i)$ ,初始化  $Table(1, i)$  为空,其中  $i=1, 2, \dots, M$ .
3. 如果手语词  $O_j$  被错误识别成手语词  $O_i$ ,那么根据识别概率的大小将  $O_j$  放入  $O_i$  的易混词表  $Table(1, i)$  中,直到

没有识别概率比  $O_i$  再大的手语词出现为止.

4. 对上述的易混集进行合并,被错误识别的手语词  $O_i$  和  $O_j$  相互被误识,则  $Table(1, i)$  与  $Table(1, j)$  合并为  $Table(r, i), r \leftarrow 2 \quad Table(r, j)$  中前  $r(r=2)$  个元素分别存储识别错误的手语词  $O_i$  和  $O_j$ ,其余元素按照  $i, j$  的顺序依识别概率大小的顺序保存到  $Table(r, j)$ ;  $Table(r, j)$  中第  $N$  个和第  $N-1$  个元素以倒序的方式分别保存  $Table(1, i)$  与  $Table(1, j)$  中元素的个数;  $Table(r, j)$  表内重复出现的元素加以合并,其中  $r=1, 2, \dots, M$ .

5. 返回易混词表  $Table(r, i), r, i=1, 2, \dots, M$ .

从上述算法的描述来看,我们所构建的易混词集使得易混词集内部元素之间(手语词)相对于 HMM 是依概率不可分的,而易混词集之间对于识别错误的手语词而言是依概率可以区分的.本文所说的“易混词集”非简单意义上的几个手语易混.按照本文所说的易混词集  $Table(B)$  真正意义, A 误识为 B 和 B 误识为 A 并不等价,因此两事件并不具有对称性.易混词集的得到,是通过大量实验测试出每个手语词会被误识别成哪些词,这些词组成该词的易混词集.例如:假设在进行测试的时候,手语词 A 被误识为手语词 B;手语词 C 也被误识为手语词 B;手语词 D 也被误识为手语词 B.那么 B 的“易混词集” $Table(B)$  中含有 A、C、D(这里的  $Table(B)$  相当于  $Table(1, j)$  的第一个元素为 B).基于统计识别的易混词集具体的构建过程如图 14 所示,图中的 HMM 识别过程实质是一个过程,为了便于说明在图 14 中分开表示.我们根据基于统计识别的易混词集的构建过程提出如下的结论:

(1) 如果一个统计模型的识别能力比 HMM 强(在所有手语词所构成的词集  $S$  上),那么在词集  $C$  上的识别率要比 HMM 高,在词集  $R$  上的识别率至少要与 HMM 相当;如果统计模型在词集  $R$  上的识别能力呈现出弱化的趋势,那么它在词集  $C$  上的手语词正确识别的个数应远远大于 HMM 在词集  $R$  上的手语词正确识别的个数.

(2) 如果在词集  $S$  上一个统计模型的识别能力比 HMM 强,那么在词集  $C$  上的识别能力一定比 HMM 强.

(3) 如果不附加任何识别手段,如果要想提升 HMM 的识别能力的话,那么易混词的各个子集内部所含的元素应当“化简”为 1 个.

### 4.2 基于语义的易混词集的构建

为了便于说明基于语义的易混词集的构建,对图 14 中“词”的称谓做一下明确的定义.词集  $M$  中识别错误的手语词称为误识词(图中灰度圆点表

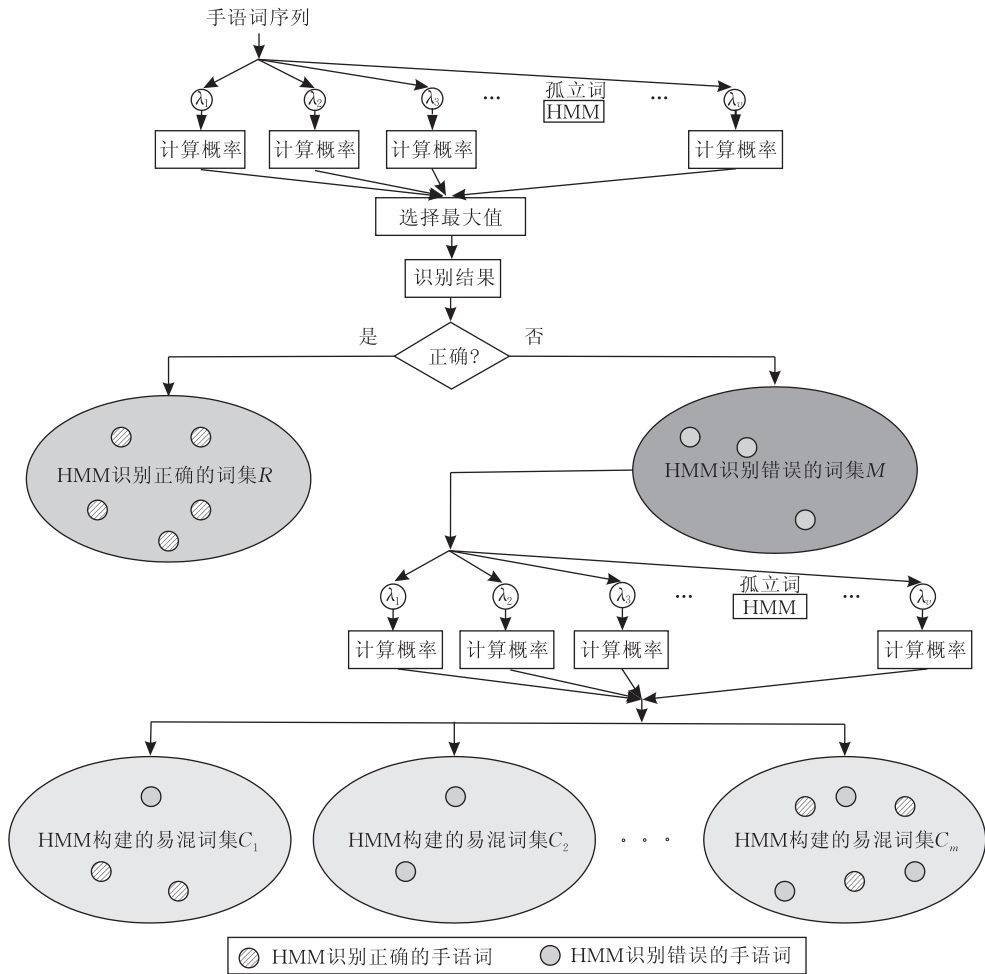


图 14 基于统计识别的易混词集的构建过程

示);识别正确的词(词集  $R$  中的斜纹圆点)导致识别错误的词称为误导词(词集  $C$  中的斜纹圆点). 首先我们对词集  $S$  的所有词做中国手语音韵标记;然后我们依照 4.1 节的结论(3)来构建基于语义的易混词集,使待识别的易混词集具有先验知识;对易混词集的各个子集内部元素做中国手语音韵标记的匹配,“化简”掉易混词集的各个子集中的误导词,使得该易混词子集尽可能的依概率可分;如果对易混词集的各个子集内部的误识词做中国手语音韵标记的匹配,如不一致则标记两个误识词依语义是可分并“化简”掉与误识词相关的误导词,否则不做任何处理. 基于语义的易混词集的构建过程如图 15 所示.

我们在易混词集内各个子集中,对误识词和误导词进行中国手语音韵标记的匹配(包括手形和运动的匹配),其目的之一是在易混词集内各个子集中词汇量相对于  $S$  较少,减少了计算量;目的之二是尽可能从易混词集中去掉误导词,以减小误导词对统计模型的干扰.

基于语义的易混词集生成算法.

1. 输入基于统计识别的易混词表:  $Table(r, i), r, i = 1, 2, \dots, M$ .
2. if  $r=1$  then {if(误识词语  $O_1$  与误导词  $O_i$  做关键手形匹配,若不一致) then 从  $Table(r, j)$  删除  $O_i$ ; else {if(误识词语  $O_1$  与误导词  $O_i$  用 3.4 节的算法做拉班运动的相似性判别,若不一致) then 从  $Table(r, j)$  删除  $O_i$ ; }},
3. if  $r \geq 2$  then {if(误识词语  $O_1$  与误识词  $O_r$  做关键手形匹配,若不一致) then 标记误识词语  $O_1$  与误识词  $O_r$  为语义可分; else {if(误识词语  $O_1$  与误识词  $O_r$  做拉班运动的相似性判别,若不一致) then 标记误识词语  $O_1$  与误识词  $O_r$  为语义可分; }},

已标记语义可分的误识词与其相关的误导词,按照步 2 的方式处理.

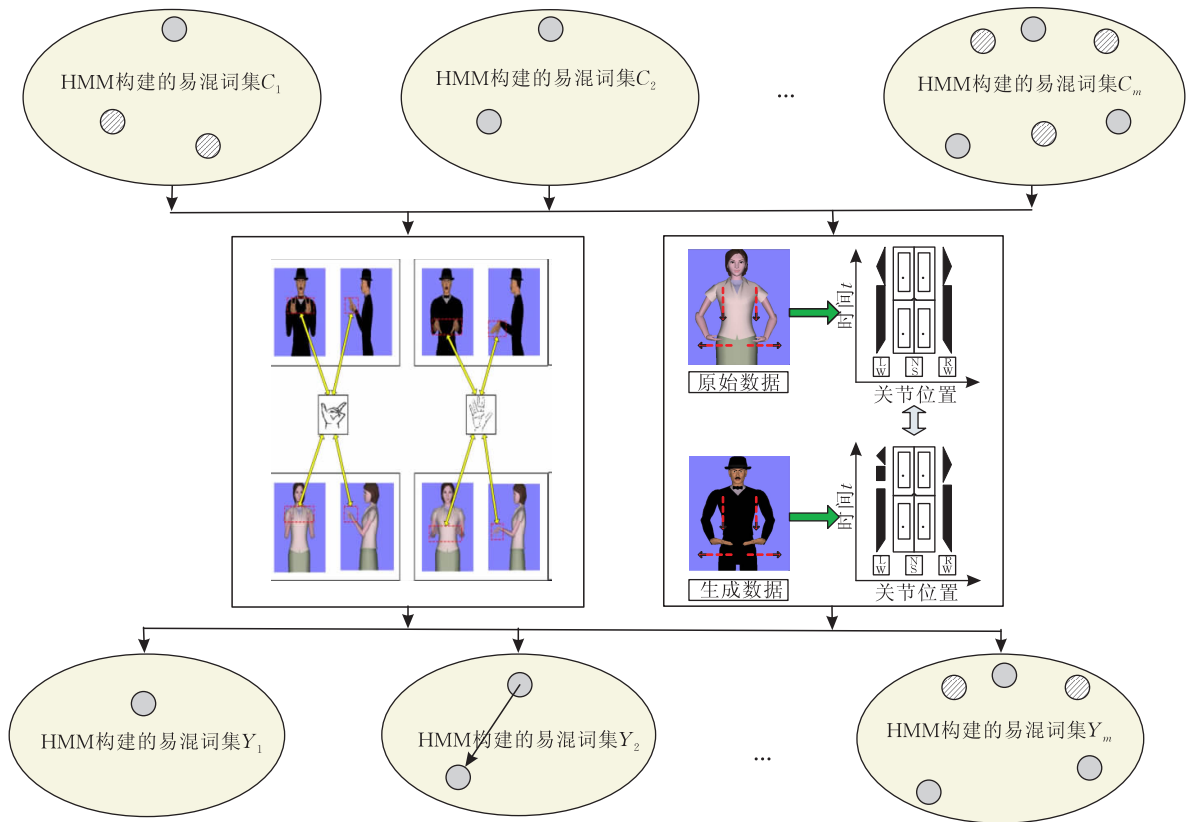


图 15 基于语义信息的易混词集的构建过程

4. 对所有的  $Table(r, i), r, i=1, 2, \dots, M$  均做步 2 和 3 的处理。

$Y \leftarrow Table(1, j) \cup \dots \cup Table(r, j), r, j=1, 2, \dots, M.$

5. 返回易混词集  $Y$ 。

综上所述,在基于语义的易混词集  $Y$  过程中,说明了中国手语音韵标记及其匹配算法是具备语义识别能力的.正是依靠这种语义识别能力消除易混词集  $C_i (i=1, 2, \dots, M)$  中的一些误导词,使一些原本依概率不可分的误识词变得依概率可分,以此来辅助统计识别模型或方法提高识别率。

### 4.3 在上述两种易混词集上进行识别实验的结果与分析

实验是在大词汇集上进行的.实验数据库是由 61356 个样本组成,分别由 6 位不同的手语老师 (Signer) 将 5113 个手语词分别采集两遍.6 个人中的一组作为已注册测试集,剩下的 10 组数据作为 HMM 的训练集.使用交叉证实的方式,5 个人的数据 (共 10 组) 进行训练,剩下 1 人的数据作为未注册测试集 (UnReg).

#### 4.3.1 在易混词集 $C$ 上的实验结果

这 6 个人的 5113 个手语词均用 {若干关键手形:拉班手势运动} 的形式,应用前文所叙述的音韵标记算法进行标记,在测试的过程中应用 4.1 节的

算法,构造出若干个基于统计识别的易混词集  $C_i, i=1, 2, \dots, M. C_i$  合并后的结果就是易混词集  $C$ .表 7 是不同的统计模型在易混词集  $C$  上进行大词汇量非特定人手语识别,而得到的识别错误的手语词的个数,即  $M$  词集内部元素的个数。

表 7 不同识别方法在易混词集  $C$  上识别后  $M$  词集内部元素个数变化情况

Signer	HMM	SOFM/HMM	Effort
A	1060	912	796
B	1252	1104	988
C	1099	951	835
D	1124	976	860
E	984	836	720
Average	1104	956	840

从表 7 的实验结果可以看出,在未引入语义信息的易混词集  $C$  上进行识别时,识别方法呈现出  $HMM < SOFM/HMM < Effort$  的强弱关系。

#### 4.3.2 在易混词集 $Y$ 上的实验结果

我们在易混词集  $C$  的基础之上,应用 4.2 节所叙述的算法,构建了基于语义的易混词集  $Y$ ;在构建基于语义的易混词集  $Y$  的过程中,我们发现每个易混词集  $C_i (i=1, 2, \dots, M)$  中手形相似但手势运动不同的手语词居多.在易混词集  $Y$  上的识别过程如图 16 所示。

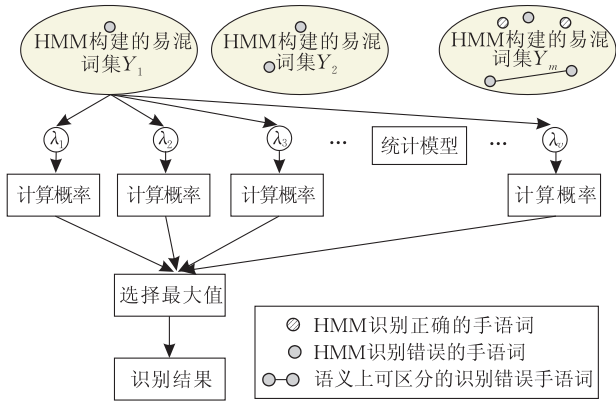
图 16 易混词集  $Y$  上的识别框架图

表 8 是不同的统计模型在易混词集  $Y$  上进行大词汇量非特定人手语识别的实验结果,而得到的识别错误的手语词的个数,即  $M$  词集内部元素的个数。

表 8 不同识别方法在基于语义的易混词集  $Y$  上识别后  $M$  词集内部元素个数变化情况

Signer	HMM	SOFM/HMM	Effort
A	827	711	772
B	1019	903	964
C	866	750	811
D	891	775	836
E	751	635	696
Average	871	755	816

从表 8 的实验结果可以看出,在引入语义信息的易混词集  $C$  上进行识别时,识别方法呈现出  $HMM < Effort < SOFM/HMM$  的强弱关系。

#### 4.3.3 对上述实验结果的分析

通过图 15 和表 7 与表 8 的实验结果进行分析,可以得出如下结论。

(1) 由于有了如同易混词集  $Y_1$  这样一些易混词集的出现,使得原本在  $C_1$  上依概率不可分的识别错误,应用 4.2 节所叙述的算法步 2 消除掉  $C_1$  内的误导词(图 15 中的斜纹圆点),使得易混词集  $Y_1$  的误识词(灰度圆点表示)依概率可分. 在不对 HMM 做任何改动的情况下,减少了 HMM 的识别错误,识别错误平均减少了 233 个. 在这里需要说明的是,由于传统的 DTW 和 HMM 对手语词的识别是在全通道进行计算的,手形特征维在手语数据通道中所占比例较大,运动特征维在手语数据通道中所占比例较小,于是在识别计算过程中“大”特征淹没了“小”特征,导致了识别错误的发生. 由于每个易混词集  $C_i (i=1, 2, \dots, M)$  中手形相似但手势运动不太相似的手语词居多,我们在易混词集  $C_i$  内部做标记匹配时,对手形特征维的标记匹配是粗略的,防止了在识

别计算过程中“大”特征淹没了“小”特征,但也带来了不良的影响,这种影响我们将在下面进行说明;我们对运动特征的标记匹配是精细的,由于拉班舞谱将运动特征维手语数据进行了量化(见表 7)并映射成了 3 维的方向特征向量,从计算的角度来讲,相当于对运动特征维上增加了计算维数,计算量相对于传统的 DTW 和 HMM 增大了,但在运动特征维上的计算准确性比传统的 DTW 和 HMM 要强。

(2) SOFM/HMM 模型将数据转换成一个紧凑、重要的低维表示形式,并使用隐含的方法进行特征提取,这种做法有些机械化,但从某种程度上防止了在识别计算过程中“大”特征淹没了“小”特征. 从表 7 与表 8 的对比中,可以看出 SOFM/HMM 在易混词集  $Y$  上的识别错误数比在易混词集  $C$  上的识别错误数要少,与前述①中的原因是一致的. SOFM/HMM 相对于 HMM 识别错误数的减少,其原因在于 SOFM/HMM 相对于 HMM,可以区分图 15 中诸如易混词集  $Y_2$  中语义上可区分的误识词(带连线的两个灰度点),这种误识词在 4.2 节所叙述的算法步 3 做了相应的标记处理;SOFM/HMM 相对于 HMM 区分能力要好,本质原因在于 SOFM/HMM 数据转换成一个紧凑、重要的低维表示形式,从某种程度上防止了在识别计算过程中“大”特征淹没了“小”特征. 在这里需要指出的是,我们不将中国手语音韵标记的匹配方法应用于识别过程,其目的是为了减少计算量;中国手语音韵标记的匹配方法应当说是引入手语语言学与人体运动学信息的 DTW,在大词汇量信号类的语言识别领域, HMM 识别效率比 DTW 好已经成为一种共识. 因而,我们仅把中国手语音韵标记的匹配方法应用于基于语义的易混词集  $Y_i$  的构建上,原因在于  $C_i$  内的词汇量小。

(3) 应用 Effort 的非特定人手语识别,先对手语数据进行全通道的 Effort 分析,然后对测试集和训练集的手语数据加以处理,在一定程度上消除了非特定人手语数据的差异性;处理后的手语数据含有 Effort 信息,详细的叙述可参考姜峰等<sup>[36]</sup>的文献. 我们所构建的基于语义的易混词集,不对原始的手语数据(包括测试集和训练集)做任何的变动,仅通过符号化的标记方式将手语语言学与人体运动学的先验信息加入到了原始的手语数据当中. 这是基于语义的易混词集与 Effort 处理后的测试集的不同之处. 由于 Effort 分析一定程度上消除了手语数据的差异性,由表 7 的实验结果可以看出, Effort 的

识别能力要高于 HMM 和 SOFM/HMM; 由于应用 Effort 方法对易混词集  $C$  和  $Y$  进行识别时, 易混词集  $C$  和  $Y$  中的数据均做了 Effort 分析处理, 所以平均的识别错误数相差仅为 24 个. 在这里需要说明的是 Effort 分析处理手语数据是在全同上进行的, 因而也存在“大”特征淹没了“小”特征的问题. 而中国手语音韵标记的匹配方法的缺点在于, “大”特征即手形特征计算粗略, 无法对关键手形之间的动作进行细致的计算. 所幸依靠关键手形动作来区分的手语词所占比率很小, 手语词的动作分类可参考顾定倩等<sup>[37]</sup>的研究工作, 这也是我们不对手形通道的数据建立拉班舞谱的原因之一; 而另外的一个原因是引起手形特征维的维数膨胀, 计算量大幅度的增加.

#### 4.3.4 实验结论

综合上述实验结果与分析我们得出如下结论:

##### (1) 优点

首先, 应用中国手语音韵标记的标记体系建立了基于语义的易混词集, 以加入先验知识的形式, 在不对统计模型做任何改动的情况下, 辅助统计模型减少识别错误. 其次, 在不对原始手语数据做任何变动的情况下, 使得基于语义的易混词集  $Y$  上的识别能力呈现出 SOFM/HMM 略强于 Effort, HMM 则接近于 Effort. 最后, 在基于语义的易混词集  $Y$  的构建过程中, 中国手语音韵标记的匹配算法在每个易混词集  $C_i (i=1, 2, \dots, M)$  中具有较强的语义识别能力, 说明了中国手语音韵标记及其匹配算法是具备语义识别能力的.

##### (2) 不足

应用中国手语音韵标记的标记体系建立了基于语义的易混词集, 在一定程度上辅助了统计识别模型或方法提高了识别率, 但是识别错误仍然存在很多. 通过上述实验分析, 可知是由于统计识别模型或方法本身所具有的局限性所造成的. 要解决这个问题就需要通过以下途径进行解决: ① 对诸如 HMM 或 SOFM/HMM 这样的统计模型或方法加以改进; ② 用更多的样本训练统计模型参数, 来改善统计模型或方法的识别性能, 这也是我们今后研究工作的重点.

## 5 结束语

基于统计模型的识别系统缺乏先验知识, 一定程度上影响了系统的识别性能. 本文提出了中国手语音韵标记的标记体系, 并将上述符号标记在计算

机上进行了实现. 为了说明中国手语音韵标记的标记体系的有效性, 我们应用上述符号标记建立了基于语义的易混词集. 从而将带有语义信息的先验知识加入到统计模型的识别系统, 提高了识别系统的性能. 从中国手语音韵标记及其匹配算法有效性的说明与分析当中以及相应的实验结果均表明, 中国手语音韵标记方法是具备语义上识别性能的. 因而, 在基于语义的易混词集上应用传统 HMM 还是 SOFM/HMM 进行识别, 在不对统计模型做任何改动的情况下, 都有效地较少了识别错误. 在不对原始数据做任何改动的情况下, 在基于语义的易混词集上进行识别时, HMM 的识别错误接近 Effort, SOFM/HMM 的识别错误略少于 Effort. 本文是手语语言学、人体运动学以及计算机科学相结合的一次大胆的尝试, 将手语语言学的音韵标记在计算机上进行有效地实现, 对于手语语言学以及在计算机科学的识别领域都有着一定的借鉴意义.

## 参 考 文 献

- [1] Stokoe W, Casterline D, Croneberg C. A dictionary of American Sign Language. Washington, D C: Gallaudet College, 1965
- [2] Liddell S K, Johnson R E. American Sign Language: The phonological base. Sign Language Studies, 1989, 64: 195-277
- [3] Sandler W. Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language. Dordrecht: Foris, 1989
- [4] Simakosi T S. Construction model of continuous sign language: The new trends of sign language phonology theory// Proceedings of Japanese Sign Language Academic Research Association. Japan, 1989, (10): 2-14(in Japanese)  
(島越隆士. 手話言語における継起的構造モデル: 手話音韻論の最近の動向// 日本手話学術研究会論文集. 日本, 1989, (10): 2-14)
- [5] Honnna S K, Kanda T Y, Oda K H, Katou S H K. Notations of sign language// Proceedings of Japanese Sign Language Academic Research Association. Japan, 1985, (7): 2-11(in Japanese)  
(本名信行, 神田和幸, 小田候朗, 加藤三保子. 手話の表記法について 日本手話学術研究会論文集. 日本, 1985, (7): 2-11)
- [6] Kanda T Y. Sign language abbreviation notations of Japanese sign language and handshapes system using phonology notations. Summaries of East Sea Industry Academy, 1986, (7): 55-70(in Japanese)  
(神田和幸. 日本手話の字型の簡略表記システム—手話の音韻的表記法の研究. 東海産業短期大学紀要, 1986, (7): 55-70)
- [7] Honnna S K, Kanda T Y, Oda K H, Katou S H K. Proposals of sign language notation. Documents of Speech Research of Japanese Acoustic Association, 1984, (2): 83-88(in Japanese)

- (本名信行, 神田和幸, 小田候朗, 加藤三保子. 手話の表記法に関する提案. 日本音響学会音声研究会資料(資料番号: 83-88(1984-2)), 1984)
- [8] Kanda T Y. Phonology notations of Japanese sign language. *Sign Language Research Japanese Sign Language Association*, 1991, (12): 84-91(in Japanese)  
(神田和幸. 日本手話の音韻表記法. 手話学研究, 日本手話学会, 1991, (12): 84-91)
- [9] Kanda K Y, Itikawa A R, Nagaashima Y J, Terauchi M N, Kato T C. Sentence notations of sign language//*Proceeding of the 99's Human Interaction System*. Japan, 1999, 2: 745-748(in Japanese)  
(神田和幸, 市川熹, 長嶋祐二, 寺内美奈, 加藤雄士. 手話の文表記法—sIGNDEX V.2. ヒューマンインタフェースシンポジウム'99 論文集. 日本, 1999, 2: 745-748)
- [10] Kanda T Y. Research of cognitive sign language. *The Institute of Electronics, Information and Communication Engineers*, 2007, J90-D(3): 609-616(in Japanese)  
(神田和幸. 手話の認知学的研究. 電子情報通信学会論文誌, 2007, J90-D(3): 609-616)
- [11] Marble C C A. An image processing system for interpreting motion in American sign language. *Journal of Biomedical Engineering*, 1992, 14(15): 419-425
- [12] Fels S S, Hinton G E. Glove-talk: A neural network interface between a data-glove and a speech synthesizer. *IEEE Transactions on Neural Networks*, 1993, 4(1): 2-8
- [13] Kadous M W. Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language//*Proceedings of the Workshop on the Integration of Gesture in Language and Speech*. Wilmington, DE, 1996: 165-174
- [14] Grobel K, Assan M. Isolated sign language recognition using Hidden Markov Models//*Proceedings of the International Conference of System, Man and Cybernetics*. Orlando, USA, 1997: 162-167
- [15] Sagawa H, Takeuchi M, Ohki M. Description and recognition methods for sign language based on gesture components//*Proceedings of the International Conference on Intelligent User Interfaces Florida, USA*, 1997: 97-104
- [16] Kim J S, Jang W, Bien Z. Dynamic gesture recognition system for the Korean sign language (KSL). *IEEE Transactions on Systems, Man and Cybernetics*, 1996, 26(2): 354-359
- [17] Deng J W, Tsui H T. A two-step approach based on PaHMM for the recognition of ASL//*Proceedings of the 5th Asian Conference on Computer Vision*. Melbourne, Australia, 2002: 126-131
- [18] Shamaie A, Sutherland A. Accurate recognition of large number of hand gestures//*Proceedings of the 2nd Iranian Conference on Machine Vision and Image Processing*. School of Computer Applications, Dublin City University, 2003: 308-317
- [19] Chen F S, Fu C M, Huang C L. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing*, 2003, 21(8): 74-78
- [20] Huang C L, Jeng S H. A model-based hand gesture recognition system. *Machine Vision and Application*, 2001, 12(5): 243-258
- [21] Kobayashi T, Haruyama S. Partly-Hidden Markov model and its application to gesture recognition//Taylor F J ed. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. New York; Academic Press, 1997: 3081-3084
- [22] Kadous M W. Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language//Messing L ed. *Proceedings of the Workshop Integration of Gestures in Language and Speech*. Delaware: IEEE Computer Society Press, 1996: 165-174
- [23] Assan M, Grobel K. Video-based sign language recognition using hidden Markov models//*Proceedings of the Gesture-Workshop*. Bielefeld; Springer-Verlag, 1997: 97-109
- [24] Vamplew P, Adams A. Recognition of sign language gestures using neural networks. *Australian Journal of Intelligent Information Processing Systems*, 1998, 5(2): 94-102
- [25] Handouyahia M, Ziou D, Wang S. Sign language recognition using moment-based size functions//*Proceedings of the International Conference on Vision Interface*. Kerkyra; CRC Press, 1999: 210-216
- [26] Su M C. A fuzzy rule-based approach to spatio-temporal hand gesture recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2000, 30(2): 276-281
- [27] Kong W W, Ranganath S. 3-D hand trajectory recognition for signing exact english//*Proceedings of the International Conference on Automatic Face and Gesture Recognition*. Seoul; ACM Press, 2004: 535-540
- [28] Gao Wen, Fang Gao-Lin, Zhao De-Bin, Chen Yi-Qiang. HandTalker II: A Chinese sign language recognition system based on SOFM/SRN/HMM. *Pattern Recognition*, 2004, 37(12): 2389-2402
- [29] Ong S C, Ranganath S W. Deciphering gestures with layered meanings and signer adaptation//*Proceedings of the International Conference Automatic Face and Gesture Recognition*. 2004: 559-564
- [30] Ni Xun-Bo et al. Generation and inspection of signer independent sign language based on the combination of sign language linguistics and human kinematics. *Journal of Software*, to appear(in Chinese)  
(倪训博等. 手语语言学与人体运动学相结合的非特定人手语数据的生成及其有效性检测. 软件学报, 待发表)
- [31] Williams A, Hut C. *Labanotion or Kinetography Laban: The System of Analyzing and Recording Movement*. New-York; A Theaten Arts Book, 1977
- [32] Bahl L R, Padmanabhan M, Nahamoo D, Gopalakrishnan P S. An N-best candidates-based discriminative training for speech recognition applications. *IEEE Transactions on Speech and Audio Processing*, 1994, 2(1): 206-216
- [33] Chow Y. Maximum mutual information estimation of HMM parameters for continuous speech recognition using the N-best algorithm//*Proceedings of the ICASSP'90*. Albuquerque, 1990: 701-704
- [34] Jiang Feng, Ye Jian-Jun, Yao Hong-Xun. Multilayer architecture based on SVM and HMM in Chinese sign language recognition. *Journal of Computer Research and Development*, 2005, 42(Supplement B): 212-217(in Chinese)

(姜峰, 叶剑君, 姚鸿勋. 基于 HMM 和 SVM 二级分类器的大词汇量中国手语识别系统. 计算机研究与发展, 2005, 42 (增刊 B): 212-217)

- [35] Ni Xun-Bo et al. Discriminative training methods of HMM for sign language recognition. CAAI Transactions on Intelligent Systems, 2007, 2(1): 80-84(in Chinese)

(倪训博等. 手语识别中基于 HMM 的区分性训练方法. 智能系统学报, 2007, 2(1): 80-84)

- [36] Jiang Feng, Gao Wen, Yao Hong-Xun, Chen Xi-Lin. LMA approach in person-independent sign language recognition.

Chinese Journal of Computers, 2007, 30(5): 851-860 (in Chinese)

(姜峰, 高文, 姚鸿勋, 陈熙霖. 手势手语力效分析. 计算机学报, 2007, 30(5): 851-860)

- [37] Gu Ding-Qian et al. The analysis of Chinese sign language's basic words (basic movements). Chinese Journal of Special Education, 2005, 2(56): 56-72(in Chinese)

(顾定倩等. 中国手语基本词(基本动作)类型分析. 中国特殊教育, 2005, 2(56): 56-72)



**NI Xun-Bo**, born in 1979, Ph. D. candidate. His main research interests include pattern recognition and neural networks.

**ZHAO De-Bin**, born in 1963, Ph. D., professor, Ph. D. supervisor. His main research interests include multimedia

## Background

Phonetic analysis of sign language as a kind of vision-sign language, which is completely different with sound language style, should be like that of sound language? The first trial of structural description to basic vocabulary unit of sign language was done by Stokoe. Before Stokoe, people believed that analysis to hand gestures was impossible which, as a whole, had only iconicity without internal structure. Stokoe firstly used the linguistics in sign language research to describe natural linguistic characteristics of sign language. Stokoe called the research on sign language structure the chorology. He thought the sign language consisted of place of articulation, hand shape and movement totally three factors. The structural analysis to sign language made by him laid a foundation for sign language research later on. The sign language linguistics afterwards made "grammar"-like summarization to sign language and hence the theory of phonological representation of sign language was formed. However, these theories of phonological representation of sign language are not realized actually in computer, which are mostly just theoretical description or noted by hand (in the form of video). Besides, these theories of phonological representation of sign language are not able to describe details of sign language movements.

Basing on the situation that there is no notation system for Chinese sign language symbols, firstly, the paper proposes a phonological representation system according to the characteristic of Chinese sign language and language habit; secondly, a method which combined data template of hand shape and Laban spectrum is proposed according to the characteristic of recognition system of Chinese sign language which is based on data glove, and the recognition system above is realized by the computer by means of automatic label; finally, recognition system of Chinese sign language of data

data compression, image processing, computer vision, multimodal interface, and artificial intelligence.

**JIANG Feng**, born in 1978, lecturer. His main research interests include pattern recognition, digital image processing and neural networks.

**CHENG Dan-Song**, born in 1972, Ph. D. candidate, lecturer. His research interests include pattern recognition, digital image processing.

glove is lack of transcendental information, thus a confused word assemble with transcendental semantic information is established through the system of phonological representation, which is recognized with HMM, SOFM/HMM and Effort in the confused word assemble construed by the authors.

Experiment results show that the recognition performance of the confused word assemble basing on statistic recognition comparing with that of the confused word assemble basing on transcendental semantic information, without changing statistic model, the method proposed by the authors helps HMM, SOFM/HMM and Effort to improve recognition rate to different extent. A confused word assemble with transcendental semantic information is established through the system of Chinese sign language phonological representation helps the statistic model or method to a certain extent to improve recognition rate, however, there are still a lot of recognition failures. The ways below should be used to solve the problem: ① make improvements to statistic models or methods such as HMM or SOFM/HMM; ② train model factors with more samples to improve recognition performance of statistic models or methods.

The research work in the paper has been supported by Chinese National Natural Science Foundation Key Project (60533030) and Chinese National Natural Science Foundation Project (60603023). The paper is a bold attempt in the combination of sign language linguistics, kinesiology and computer science, which realizes effectively the phonological representation of sign language linguistics in computer and is meaningful to a certain degree as a reference for sign language linguistics and the recognition field of computer science. Meanwhile, the paper is theoretical basis and practical assurance for effectiveness verification to signer-independent sign language data.