

## 非特定人手语数据生成及其有效性检测\*

倪训博<sup>1+</sup>, 赵德斌<sup>1</sup>, 高文<sup>1,2</sup>, 姜峰<sup>1</sup>, 姚鸿勋<sup>1</sup>

<sup>1</sup>(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

<sup>2</sup>(中国科学院 计算技术研究所, 北京 100190)

### Data Generation and Its Validity Inspection of Signer-Independent Sign Language

NI Xun-Bo<sup>1+</sup>, ZHAO De-Bin<sup>1</sup>, GAO Wen<sup>1,2</sup>, JIANG Feng<sup>1</sup>, YAO Hong-Xun<sup>1</sup>

<sup>1</sup>(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

<sup>2</sup>(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: nixunbo@foxmail.com

**Ni XB, Zhao DB, Gao W, Jiang F, Yao HX. Data generation and its validity inspection of signer-independent sign language. *Journal of Software*, 2010,21(5):1153-1170. <http://www.jos.org.cn/1000-9825/3476.htm>**

**Abstract:** This paper proposes the combination of sign language linguistics with human kinematics to generate and detect the data of SISL (signer independent sign language) according to the characteristics of gesture sign language (GSL). An improved Mean-Shift algorithm is applied to the generation of hand shape data channels without losing the linguistic features of GSL, and then the key hand shape phonetic notation is used to detect the effectiveness of data. In order to enrich the kinematic characteristics of GSL, an improved genetic algorithm is applied to the generation of movement related data channels. Moreover, Labannotation is adopted to inspect the effectiveness of data. Finally, an experimental inspection framework is established based on an original sample to make the proposed detection method adapt to multi-classes data inspection of linguistics. Experimental results show that the proposed method for the generation and detection of SISL data is effective and feasible.

**Key words:** sign language recognition (SLR); phonetic notation; Labannotation; sign language linguistics (SLL); human kinematics; Mean-Shift algorithm; genetic algorithm

**摘要:** 根据手势手语的特点,提出了手语语言学和人体运动学相结合的非特定人手语数据的生成和检测方法。首先,Mean-Shift 算法有控制生成强度的优点,将改进的 Mean-Shift 算法应用于手形数据通道的生成,以保持手势手语的语言学特性,并应用关键手形的音韵标记进行数据有效性的检测;其次,为了丰富手语手势动作的运动特性,将改进的遗传算法应用于与运动相关的数据通道进行数据生成,并应用拉班舞谱对其进行数据有效性检测;最后,提出了基于原始样本的检测实验框架,使得所提出的检测方法适用于语言类的多类别数据检测问题。实验结果表明,所提出的非特定人手语数据的生成和检测方法是有效的。

**关键词:** 手语识别;音韵标记;拉班舞谱;手语语言学;人体运动学;Mean-Shift 算法;遗传算法

中图分类号: TP391 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant Nos.60603023, 60533030 (国家自然科学基金); the Beijing Municipal Natural Science Foundation of China under Grant No.4061001 (北京市自然科学基金)

Received 2008-06-30; Accepted 2008-10-09; Published online 2009-11-24

孤立词手语识别的主要问题是解决手语识别中非特定人、大词汇量的问题.建立、选择合适的模型与准则以及有效利用充分、准确的训练数据集是解决这个问题的关键.在模型的建立和准则选择方面,已经有很多研究者做过大量有价值的工作<sup>[1,2]</sup>.对于识别系统来说,识别的效果不仅依赖于识别对象内在的特性以及分类器的设计,而且受训练集的规模和训练数据的质量的影响.传统统计学所研究的是一种渐进理论,由此提出的各种方法只有当样本数目趋向于无穷大时,其性能才有理论上的保证.在模式识别领域,统计模型需要大量的训练数据才能获得较为令人满意的识别性能,对手语识别的研究能够实现大词汇量的特定人识别,甚至一定程度上的非特定人识别<sup>[3,4]</sup>.然而某些根本性问题并未解决,其中包含非特定人的数据获取问题.传统的采集方式收集训练数据存在以下缺陷:(1) 工作量大,采集难度和开销大;(2) 数据手套数量有限且昂贵;(3) 无法直接判断所采集数据的正确性及有效性.手语识别训练数据的缺乏导致训练复杂模型困难,已使手语识别研究工作,特别是大词汇量的非特定人手语识别工作的开展异常艰难.在模式识别领域,很多实验已经验证大训练集可以提高识别系统的识别效果,由大量数据训练的识别器将更为出色.文献[5]中分析了训练集规模增长并超出通常可获数据的范围对识别效果的影响.在人脸识别领域,Lu<sup>[6]</sup>利用重采样技术来生成原始训练集的若干个子集,基于每个子集构造一个人脸识别器,对这些识别器的识别结果进行不同策略的集成,得到最终的识别结果.Chen<sup>[7]</sup>在一张照片的基础上,利用 Enhanced (PC)<sup>2</sup>A 方法生成新的人脸图像,从而增大训练集,进行人脸识别.Gao 等人<sup>[8]</sup>利用现有的少量照片生成出不同姿态、不同光照下的人脸图像作为训练样本解决样本不足的问题.Chen<sup>[9]</sup>利用基因遗传算法生成新的人脸图像作为训练集,解决人脸检测问题.在手写体识别领域中,Varga 和 Bunke<sup>[10]</sup>用 Perturbation model 生成新的训练样本并应用于手写识别,达到了很好的效果.在文本分类领域,Sarah 提出了利用“background knowledge”方法<sup>[11]</sup>进行文本识别;Martin<sup>[12]</sup>提出了生成样本点被用来提取新的特征并将其利用到原始的训练数据集中,以提高分类器的效率.在静态手势识别领域,Wang 等人<sup>[13]</sup>利用遗传算法生成新的样本作为训练集,取得了较好的效果,但手语数据的动态及多数据流特性使得这一方法难以向手语数据生成扩展.Jiang 等人在文献[14]中提出了一种基于静态手势量化与离散余弦变换(discrete cosine transformation,简称 DCT)相结合的手语生成方法.该方法能够根据现有手语者的数据生成新的手语数据,但只能有条件地将打手语者的特征参数化,且训练过程初始化对生成数据有逆转的影响,因此存在其固有的局限性.姜峰等人<sup>[15]</sup>应用基于 Mean-Shift 方法对手语数据作了全通道的数据生成,但未给出手语生成数据有效性的论证.

在数据检测方面,目前研究热点集中于人脸检测.近几年,许多文献<sup>[16-18]</sup>使用肤色与其他特征相结合来检测人脸区域.张洪明等人<sup>[19]</sup>作了基于肤色模型、神经网络和人脸结构模型的平面旋转的人脸检测.文献[20]提出了利用人脸器官几何特征建立人眼位置的判定准则和阈值递增方法定位复杂背景下灰度图像中的人眼,但该方法受光线改变的影响较大.文献[21]在人眼定位方面采用人脸的几何分布特征和灰度信息特征相结合的方法,该方法限于用人眼的低灰度信息,定位人眼会有相应的误判.Li 等人<sup>[22]</sup>提出的由粗到精、由简单到复杂的金字塔式检测结构在很大程度上改善了检测速度.Viola 和 Jones<sup>[23]</sup>提出了一种通过对简单特征的 Adaptive Boosting 来获得一个级联分类器的有效检测方法,使得人脸检测真正达到了实时的效果.Xiao 等人提出的 Boosting Chain<sup>[24]</sup>训练算法通过将上一层强分类节点作为下一层的一个弱分类器,有效地降低了强分类节点阈值调整时分类精度的损失,并减少了下一层的特征数和训练时间.Wu 等人也提出了 Nested Boosting Cascade<sup>[25]</sup>的类似改进.Li 等人提出了称为 FloatBoost<sup>[26]</sup>的训练算法改进,而 Viola 自己也提出了一种称为 Asymmetric AdaBoost<sup>[27]</sup>的改进训练算法.最新的一些文献<sup>[28-32]</sup>对上述的检测方法作了改进.

在非特定人手语识别中,训练样本的缺乏是一个亟待解决的问题.由于非特定人手语数据采集困难,因而生成非特定人手语数据是有必要的.根据手势手语的特点,本文提出了手语语言学和人体运动学相结合的非特定人手语数据的生成和检测方法.首先,Mean-Shift 算法有控制生成强度的优点,我们将 Mean-Shift 算法应用于手形数据通道的生成.在关键手形的保持段减小生成强度,在关键手形的过渡段增加生成强度,以减少对关键手形无谓的改变来保持手势手语的语言学特性;应用关键手形的音韵标记进行数据有效性的检测,确保生成数据不丢失语义.其次,我们在遗传算法中加入一些相似性动作数据作为变异因子,以丰富手语手势动作的运动,防止手势动作的“近亲繁殖”;应用拉班舞谱对其进行数据有效性的检测,可以防止生成数据的动作过于“夸大”,因而

变异成其他非母体的手语词或非手语词数据.对于手语信号本身特征的认识还处于初级阶段,目前还没有像语音信号那样的较为系统的共性特征,因而数据检测不能以特征为依据,只能以原始样本数据为依据.与人脸检测不同,语言类数据的检测属于多类别数据的检测,因而不能简单地判断生成数据是或不是手语数据.最后,我们提出了基于原始样本的检测实验框架,使得我们所提出的检测方法适用于语言类的多类别数据检测问题.

## 1 结合手语语言学 and 人体运动学的中国手势手语标记

### 1.1 手语语言学的中国手势手语标记































手语中不存在文字,对手语的记录描述十分困难.为此,手语语言学根据各自的研究需要提出了各种手语标记方法,形成了手语语言学的音韵理论.针对中国手语没有自己的符号标记体系的现状,我们在日本手语标记体系<sup>[33]</sup>和汉语拼音的基础上提出了中国手势手语的音韵标记法,以解决中国手语无符号标记的现状.从另一个方面看,目前的音韵标记理论均停留在理论阶段<sup>[33-41]</sup>,并没有一个行之有效的实现.针对这个问题,我们提出滑动窗结合模板匹配来实现对中国手势手语的关键手形进行自动标记.这些工作是为了后面对手形通道的数据有效性的检测做准备.

#### 1.1.1 中国手势手语的关键手形音韵标记

手语语言学中最为主要的标记就是手形的标记,也是表达手语语义信息的主要特征之一.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,并对手势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,对键手形作了如下的整理和归类,见表 1.

Table 1 Key hand shape notation of numbers and pinyin













































表 1 数字与拼音的关键手形标记

Number noting symbol	Pinyin noting symbol							
 4	 a	 b	 c	 d				
 5	 e,3	 f,v,2	 g,i,l	 h				
	 j,9	 k	 l <sub>1</sub>	 m				
	 n	 o,0	 p	 q,7				
	 r,l <sub>2</sub> ,8	 s	 t	 u				
	 w	 x	 y,6	 z				
	 zh	 ch	 sh	 ng				

“组合手形”来源于手势手语关键手形,根据手形与汉语拼音的相似程度进行分类.为了便于说明,我们将此类手形称谓“组合手形”,见表 2.

在手语语言学中,为了表达出手语的完整意思,对方向、位置及运动作了系统的标记,但这些标记是着眼于语言学研究的.对于手语识别系统而言,实现起来比较困难.为此,引入人体运动学的标记来刻画手语运动、方向、位置的信息,便于手语识别系统的实现.手语语言学的方向、位置及运动的标记不是本文的重点,在此不作过多描述,只是说明在手语语言学中,一种完整的手语语言是包括关键手形、方向、位置及运动信息的.

**Table 2** Key hand shape notation of combined hand shapes  
**表 2** 组合手形的关键手形标记

Similar combined hands of a class	 a <sub>1</sub>	 a <sub>2</sub>	 a <sub>3</sub>
Similar combined hands of c class	 c <sub>1</sub>	 c <sub>2</sub>	 c <sub>3</sub>
	 bc	 4c	 4c <sub>1</sub>
Similar combined hands of e class	 e <sub>1</sub>	 e <sub>2</sub>	 e <sub>3</sub>
Similar combined hands of h class	 h <sub>1</sub>	 h <sub>2</sub>	 h <sub>3</sub>
Similar combined hands of j class	 j <sub>1</sub>	 j <sub>2</sub>	 j <sub>3</sub>
	 fj	 gj	 oj
	 rj	 gj	
Similar combined hands of k class	 k <sub>1</sub>	 k <sub>3</sub>	
Similar combined hands of n class	 n <sub>1</sub>		
Similar combined hands of o class	 o <sub>1</sub>	 o <sub>2</sub>	
Similar combined hands of q class	 q <sub>1</sub>	 q <sub>2</sub>	
Similar combined hands of r class	 r <sub>1</sub>	 r <sub>3</sub>	
Similar combined hands of t class	 t <sub>1</sub>		
Similar combined hands of x class	 x <sub>1</sub>	 x <sub>2</sub>	 x <sub>3</sub>
Similar combined hands of z, zh class	 z <sub>1</sub>	 zh <sub>1</sub>	
Similar combined hands of ch <sub>1</sub> , sh class	 ch <sub>1</sub>	 ch <sub>2</sub>	 ch <sub>3</sub>
	 sh <sub>1</sub>		
Similar combined hands of ng <sub>1</sub> class	 ng <sub>1</sub>	 ng <sub>2</sub>	

1.1.2 滑动窗结合关键手形模板对手语语言学的中国手势手语标记的实现

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.滑动窗的工作原理如图 1 所示.其中,滑动窗  $W_s$  是已训练好的手语关键手形的模板(每个模板均对应一个音韵标记).这里,  $t_s$  和  $t_e$  分别表示手手语关键手形数据帧的开始点和结束点.用算法 1 可以对手语数据的手形数据通道进行关键手形的标记.

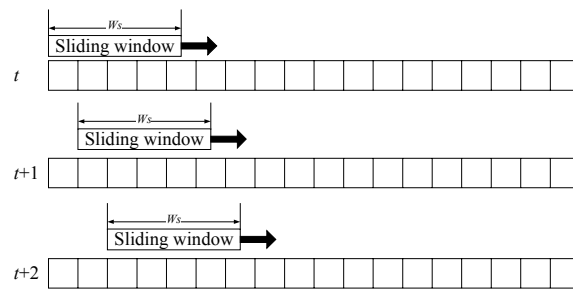


Fig.1 Working principle of sliding window

图 1 滑动窗的工作原理

**算法 1.** 手语语言学的中国手势手语标记算法.

初始化:  $totalDis[t]=0, index[t]=1, 1 \leq t \leq T; T$  为模板的总数,  $O$  为观察序列的总数据帧数,  $1 \leq i \leq O$ .

for  $i=1$  to  $O$

{if  $i$  不为  $W_s$ , then

for  $t=1$  to  $T$

{ $totalDis[t] += \min\{d(i, t_{index[t]}), d(i, t_{index[t]+1})\}; t_k$  表示模板  $t$  中的第  $k$  帧

$index[t] = \operatorname{argmin}\{d(i, t_{index[t]}), d(i, t_{index[t]+1})\};$ }

else { $t_s=i, t_e=W_s; result[t, t] = \operatorname{argmin}\{totalDis[t]\}; t_e=W_s+1$ }

}

## 1.2 人体运动学的中国手势手语标记

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作了概念上的定义,仅用模板匹配的方法对运动做标记是困难的.用 HMM(hide Markov model)实现动作容易失去先验信息,不容易区分动作的细节.为此引入了目前人体运动学上较为流行的拉班舞谱来对运动进行刻画.拉班舞谱<sup>[42]</sup>是记录和分析人体运动的标记体系.拉班舞谱是一种既精密又合理地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.拉班舞谱可以记录所有的人体运动,但不能记录孤立、不连续的舞蹈风格(例如英国古典的芭蕾舞).

### 1.2.1 拉班舞谱对中国手势手语运动标记的实现

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,所能用到的关节仅有腕部(W).根据我们的具体问题,对拉班舞谱作了改造.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.12 个方向标记如图 2 所示.

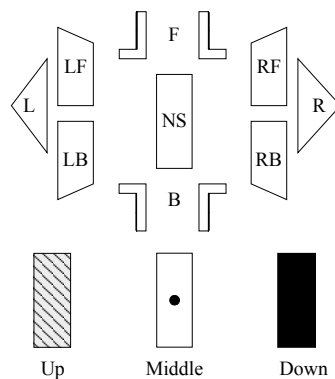


Fig.2 Direction notation of Laban movement

图 2 拉班运动的方向符号

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空间的位置.我们假设前方向定义为正向的  $y$  轴,右方向定义为正向的  $x$  轴.高度定义为  $z$  轴.为了简化计算,对上述的 3 维方向特征向量赋予符号化的度量,见表 3.上半身的关节记号如图 3 所示,其中圆圈所表示的即为我们所用到的左、右手的腕部(W)关节.图 4 所示为一个手势动作的拉班运动记号表示.

例如,中间位置向前的方向可以被映射成向量  $\left[ \cos \frac{\pi}{2}, \sin \frac{\pi}{2}, 0 \right]$ ,中间位置向右前方的方向可以被映射成向量  $\left[ \cos \frac{\pi}{4}, \sin \frac{\pi}{4}, 0 \right]$ .两个手势之间的差异可以由两个拉班节拍  $B_i$  和  $B_j$  来计算,如下所示:

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

$v_{i,k}$  和  $v_{j,k}$  是三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. $w_k$  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.

**Table 3** Laban direction notation and its metric  
**表 3** 拉班方向标记及符号化度量

Direction	Right	Left	Middle	Up	Down	Forward	Backward	Left-Forward	Left-Backward	Right-Forward	Right-Backward
Laban-Notation	R	L	M	U	D	F	B	LF	LB	RF	RB
Notation metric	$\sin 0$	$\sin \pi$	0	1	-1	$\cos \frac{\pi}{2}$	$\cos \frac{3\pi}{2}$	$\cos \frac{3\pi}{4}, \sin \frac{3\pi}{4}$	$\cos \frac{5\pi}{4}, \sin \frac{5\pi}{4}$	$\cos \frac{\pi}{4}, \sin \frac{\pi}{4}$	$\cos \frac{7\pi}{4}, \sin \frac{7\pi}{4}$

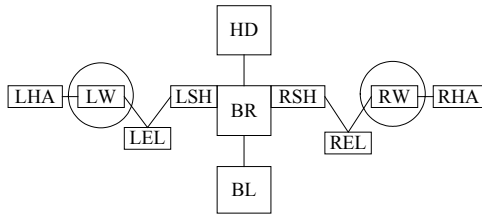


Fig.3 Joint notation of upper body of Labanotation  
图 3 拉班舞谱上半身的关节记号

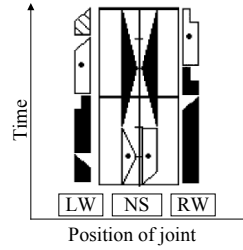


Fig.4 Labanotation of a sign language movement  
图 4 一个手势动作的拉班舞谱表示

手语手势运动的拉班运动标记方法描述如下:对拉班舞谱规定的 12 个方向(包括中立位置)设定一个检测阈值  $\theta_r$ ,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 个方向对应的拉班运动标记为: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  $O = \{o_1, o_2, \dots, o_T\}$ ,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.

1.2.2 拉班运动序列相似性的判别

假设有两个拉班运动序列  $Q$  和  $T$ ,从两个拉班运动子序列中分别提取相似性测度.两个子序列被表示为  $Q_{ab}=[q_a, q_b]$ 和  $T_{cd}=[t_c, t_d]$ ,这里的  $q_a$  和  $q_b$  是第  $a$  个和第  $b$  个序列  $Q$  的节拍, $t_c$  和  $t_d$  是第  $c$  个和第  $d$  个序列  $T$  的节拍.这两个拉班运动序列  $T$  和  $Q$  的 DTW 距离需满足下面的公式,才能说明两个拉班运动序列  $T$  和  $Q$  是相似的.

$$LB(Q_{ab}, T_{cd}) = \sum_{i=0}^{b-a} \min_{-r \leq k \leq r} (d(i+a, j+c+k)) \leq \sum_{k=1}^L w_k \tag{2}$$

如图 5 所示,一个弯曲的权值曲线  $W$  连接  $[q_a, q_b]$ 和  $[t_c, t_d]$ .该曲线由单元格构成.假设权值  $W$  的第  $k$  个点是  $w_k=(q, t)_k$ ,则权值曲线可以被定义为  $W=w_1, \dots, w_k, \dots, w_L$ ,这里的  $L$  是曲线的长.图 5 中的带条由两条直线构成.两条直线以 1 为斜率,两条直线的距离为  $2r$ .

人体运动学的标记可以细致地刻画手势的运动,但对于手势手语而言则缺少相应的语义信息.关键手形是手势手语中表达语义的重要组成部分.基于拉班舞谱的拉班运动符号序列的 DTW 匹配算法如算法 2 所示.拉班舞谱结合关键手形才能完整地表达出手势手语的含义.关键手形也是手语动作区别于其他手势运动的标志.

**算法 2.** 拉班运动符号序列的 DTW 匹配算法.

输入:拉班运动符号序列  $[q_a, q_b]$ 和  $[t_c, t_d]$ ,阈值  $\theta, LB(Q_{ab}, T_{cd})=0$ .

For  $i=a$  to  $b$  {

If  $\sum_{i=0}^{b-a} \min_{-r \leq k \leq r} (d(i+a, j+c+k)) > \sum_{k=1}^L w_k$  return false; Else  $LB(Q_{ab}, T_{cd}) = \sum_{i=0}^{b-a} \min_{-r \leq k \leq r} (d(i+a, j+c+k))$

```

If  $LB(Q_{ab}, T_{cd}) / (b - a + c - d + 2) > \theta$  return false
} return true

```

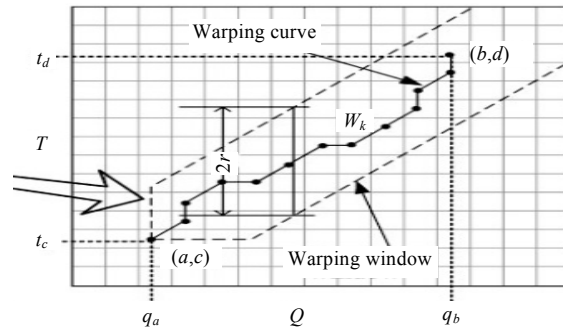


Fig.5 DTW match of Laban movement sequences  $T$  and  $Q$

图5 拉班运动序列  $T$  和  $Q$  的 DTW 匹配

## 2 非特定人手势手语数据的生成

手语是具有肢体动作的视觉语言.相对于人脸图像等单类别的数据,手语数据是多类别的数据,因而手语的数据生成和检测相对于人脸图像来说更为困难.手语的一个单词若被视为人脸图像,其他手语词就应当被视为其他类别的图像(诸如其他动物的面部图像或是风景图像),所以说手势手语的数据生成是一个多类别的数据生成问题.

### 2.1 基于手语语言学的手语数据生成

在手语中,手形包含最主要的手语语言学信息,是手语手势动作区别于一般的手势动作的重要标志,因此手语数据的生成过程中应当尽可能地避免手语语义信息的丢失,即尽可能地减小手形通道上生成数据与原始数据的差异.从数据生成算法的角度上来看,遗传算法使得数据交叉剧烈,不可避免地会造成手形通道的语义信息的丢失,因而应当选择使语义信息丢失小的数据生成算法.从数据有效性检测的角度上看,手形通道的关键手形数据不像手语手势动作的数据那样变化频繁,即关键手形数据的变化是有停顿的,而不是连续变化的,因而无法再应用拉班舞谱<sup>[42]</sup>对手形通道上的生成数据进行有效的检测.

#### 2.1.1 Mean-Shift 算法简介

Mean-Shift 算法的基本计算模型是一个经典的模式识别过程,算法中所需要的唯一参数就是处理的分辨率.Comaniciu 等人<sup>[43]</sup>不但证明了 Mean-Shift 迭代的收敛性,而且发现了迭代轨迹光滑的特性.由于它是一种没有嵌入假设的非参数方法,因此能够适用于任意结构的特征空间的分析.Mean-Shift 算法的非参性,很恰当地满足了手语数据生成过程的要求,而使用参数化的方法解决更为复杂的数据生成问题是行不通的.Mean-Shift 过程是一种核密度估计方法,它基于模式识别中估计概率密度函数的 Parzen 窗方法.仿照 Parzen 核密度估计的思想方法,我们可以根据样本估计概率密度的梯度.其基本思路是,先以无参估计获取样本空间概率密度分布,然后对其进行微分从而得到梯度信息.

基于 Mean-Shift 方法的样本生成可以从类内生成和类间生成两种方式去考虑.类内生成是利用类内已有的样本生成新的样本.类内生成又包括内向和外向两部分:内向部分使生成的样本朝向类内样本分布密度增大的方向;外向部分朝向类内样本分布密度稀疏的方向.类间生成是利用类内样本与其易混类的样本生成新的样本.类内生成的样本具有较高的有效性,但广泛的代表性较弱;类间生成的数据具有广泛的代表性,但有效性难以得到保证.类间生成的情况比类内生成要复杂,一方面要使生成的样本涵盖非特定人的数据,另一方面还要规定扩张的边缘,即界,并保证扩张在界内进行.关于 Mean-Shift 方法的详细介绍,可参见姜峰等人<sup>[15]</sup>的研究工作.

### 2.1.2 手形通道应用 Mean-Shift 算法的数据生成过程

由于类内外向数据生成对驱动效果的影响与类内内向数据生成相比更加敏感,  $\eta$  作为一个动量算子控制生成强度. 当生成的手形通道数据处于平稳段(一个关键手形)时, 减小  $\eta$  并使用类内内向数据生成; 当生成的手形通道数据处于过渡段(一个关键手形向另一个关键手形转换)时, 增强  $\eta$  并使用类内外向数据生成.

设  $I = \{O_1, O_2, \dots, O_N\}$  为一个由  $N$  段手形通道数据组成的集合, 则样本的衍生过程生成一个由  $N$  段生成手形通道数据组成的新集合  $\Gamma = \{\hat{O}_1, \hat{O}_2, \dots, \hat{O}_N\}$ , 其中,  $\hat{O}_i$  与  $O_i$  具有相同的帧数. 手形通道数据样本生成算法如算法 3 所示.

**算法 3.** 手形通道数据样本生成算法.

初始化阶段:

对相邻手形通道数据段  $O_i, O_{i+1}$  使用 DTW 计算规整路径并计算最优数据帧的匹配, 建立匹配间连接.

生成阶段:

For 每个观测值  $O_i$ , 计算生成样本  $\hat{O}_i$

For  $O_i$  中的帧  $O_{ij}$  (循环步长:  $j=1:\text{legtm}(O_i)$ )

计算以  $O_{ij}$  为根节点, 建立一个树状结构  $\Gamma$ , 对于  $\Gamma$  中每个节点赋予权值, 其中, 根节点  $O_{ij}$  权值  $C_{ij}=1$ , 其他节点权值初始为 0;

按从根节点到叶子节点的顺序传播权值, 对于任意节点  $O_{mn} \in \Gamma$ , 如果其权值为 0 且其父节点权值  $C \neq 0$ , 则权值更新为  $C/(1+Y)$ ,  $Y$  为  $O_{mn}$  在树结构中兄弟的个数;

For  $k=1:N$

计算第  $k$  个规整手语段的第  $j$  帧, 规整之后第  $k$  个样本  $I_k$  第  $j$  帧为

$$I_{kj} = \frac{\sum_{n=1}^{\text{length}(O_k)} O_{kn} C_{kn}}{\sum_{n=1}^{\text{length}(O_k)} C_{kn}} \quad (3)$$

End

End

结合  $g(x) = -k'(x)$  计算生成的  $\hat{O}_i$

If 手形通道数据处于平稳段 then 减小  $\eta$

内向:

$$\hat{O}_i = \frac{\sum_{k=1}^N I_k \times g \left\| \frac{O_i - I_k}{h} \right\|^2}{\sum_{k=1}^N g \left\| \frac{O_i - I_k}{h} \right\|^2} \quad (4)$$

If 生成的手形通道数据处于过渡段 then 增强  $\eta$

外向:

$$\hat{O}_i = (1 + \eta) O_i - \eta \times \frac{\sum_{k=1}^N I_k \times g \left\| \frac{O_i - I_k}{h} \right\|^2}{\sum_{k=1}^N g \left\| \frac{O_i - I_k}{h} \right\|^2} \quad (5)$$

End

### 2.1.3 基于手语语言学的手语生成数据有效性检测

手形通道的关键手形数据不像手语手势动作的数据那样变化频繁, 即关键手形数据的变化是有停顿的, 而不是连续变化的, 因而无法再应用拉班舞谱对手形通道上的生成数据进行有效的检测. 这里, 我们应用前面叙述的“中国手势手语的关键手形音韵标记”对手形通道的生成数据进行检测, 即将关键手形音韵标记做成数据模



板,并对手形通道的生成数据应用关键手形音韵标记的数据模板做模板匹配,如果对手形通道的生成数据检测出来的关键手形音韵标记与原始数据出现顺序一致,则认为手形通道的生成数据是符合手语语言特征的.在应用这种方法之前必须作如下假设:忽略两个关键手形间的过渡动作.作出上述假设的主要原因是动作数据不适合用数据模板做模板匹配,虽然使用拉班舞谱可以检测关键手形间的过渡动作,但必须对手指关节做出拉班运动标记.由于手上的关节较多,这样势必导致数据维数的膨胀.从手语语言学的角度看,检测两个关键手形间的过渡动作对于表征手势手语的语义的意义并不是很大,但由此导致数据维数增加的代价却很大.应用上述算法对原始数据和生成数据进行标记,然后对原始数据和生成数据的标记进行符号上的匹配.如果两者的标记符号在顺序上一一对应,则说明生成的手语数据未丢失相应的语义信息,生成的手语数据是有效的(如图 6 所示).

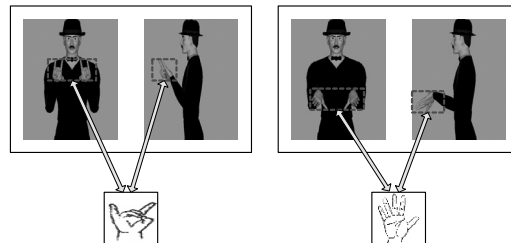


Fig.6 Validity inspection of generation data for hand shape channels

图 6 手形通道的生成数据检测

## 2.2 基于人体运动学的手语数据生成

所谓的保持语义信息不变是指我们对手势手语的手形通道数据不作任何变动,保持手形通道数据是原始采集的;仅对那些相似的人体手势动作进行数据生成处理,这些数据是除了手形通道数据以外的数据,并且包含了大量的人体手势动作信息.在对相似的人体手势动作的选择上,我们是参考后文的表 4 作为选择依据的.

### 2.2.1 基本思想

遗传算法仿真自然界的现象.同一个手势的两个原始样本可被看作是父样本,它们可通过交叉生成子样本.为了获得更多的训练数据,可采用遗传算法中的交叉思想,通过现有的样本生成新的样本作为训练数据.我们可以将手势分为若干个通道,每个通道可被看作是手势的一条基因.为了使生成的手语数据不丧失语义信息,我们对左、右手形通道的数据不作任何处理.受遗传算法的启发,我们利用交叉思想生成新的样本数据.从直觉上考虑,可以采用如下的分割策略:

$S=\{Position,Orientation\}$ ,通道数  $Channel\ Number=2$ ,保持 Left Hand Shape, Right Hand Shape 的原始性.

由于我们为使生成的手语数据不丧失语义信息而对手形通道的数据不作任何处理,我们的检测方法主要是应用拉班舞谱进行手势动作的运动检测.通过比较生成的人体手势动作数据与目标词汇的人体手势动作数据是否相似来判断生成的手语数据的有效性.图 7 和图 8 分别显示了手势动作的不同和相似的拉班舞谱.

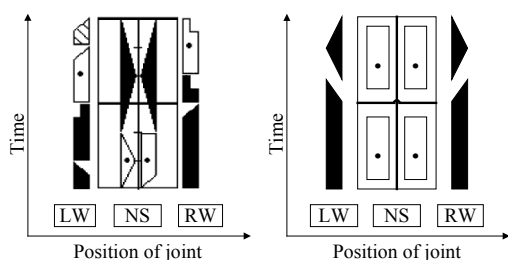


Fig.7 Labanotation of totally different gesture movements

图 7 手势动作完全不同的拉班舞谱

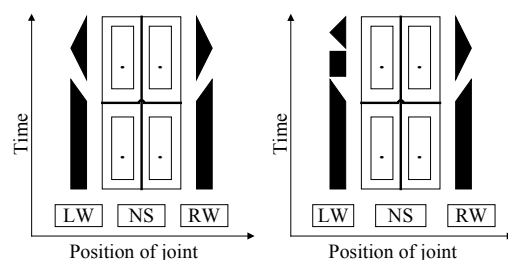


Fig.8 Labanotation of similar gesture movements

图 8 手势动作相似的拉班舞谱

打手势是一个时间相关的过程,同一个手势词的不同样本可能会有不同的时间长度,并且速度也会不同.可以采用动态规划的方法来解决时序对齐的问题,如语音识别领域中的动态时间规整算法(dynamic time warping,简称 DTW).DTW 算法对于手语识别领域的具体应用和描述,可参见吴江琴<sup>[44]</sup>的工作.

### 2.2.2 中国手语基本词(基本动作)类型

手语语言学研究者对中国手语字典手语词的动作做了细致的归类<sup>[45]</sup>.通过对 2003 年 4 月《中国手语》中所有手势动作的分析,中国手语基本词的类型可以划分为两类,共 15 种,并对中国手语各类各种基本词数量占总数的比例做了比较详细的统计,统计分析的结果见表 4.

**Table 4** Proportion statistics of basic words (basic movements) in Chinese sign language

**表 4** 中国手语各类基本词(基本动作)比例统计

Category	Type	Number	Ration in total (%)
Basic vocabulary of hand shape (basic motions)	Hand shape of single alphabet	29	1.0
	Hand gesture of finger writing alphabet	15	0.5
	Hand shape of alphabetical variant	88	3.1
	Hand shape of alphabetical syllable	19	0.7
	Alphabet+hand gesture of finger writing	3	0.1
	Hand shape Imitation Chinese Character	31	1.1
	Hand gesture of finger writing Chinese character	16	0.5
	Chinese character+hand gesture of finger writing	21	0.7
	Hand shape of Chinese Character variant	12	0.4
	Hand gesture of finger writing symbol	16	0.5
	Hand shape of single action	749	26.1
	Hand shape of complex actions	1 685	60.7
	Hand shape of continuous actions	30	1.1
	Out coming hand shape	80	2.9
	Local hand shape	17	0.6
	Total	2 801	100.0

上述动作分类是基于中国手语语言学的层面而提出的.在具体实现上,我们应用拉班舞谱对上述的手语词做运动标记,然后根据拉班运动的相似性测度进行聚类.将表中相似的人体手势动作作为“变异因子”加入到基于遗传算法的手语数据的生成过程.

### 2.2.3 基于人体运动学的手语数据生成及检测算法

我们将每一帧数据分成 4 个部分:位置、方向、左手手形、右手手形.为了不改变每个样本主要的语义信息.我们保持父样本与子样本的左、右手手形数据的一致性,即父样本与子样本的左、右手手形数据不参与交叉,仅对父样本的位置和方向数据进行交叉.为了避免遗传算法在手语数据生成的过程中出现“近亲繁殖”的现象,将一些动作相近的手语词样本的位置和方向数据作为“变异因子”引入到上述交叉的过程中,以增强鲁棒性,防止后代生成数据的退化.交叉、变异的过程如图 9 所示.

我们根据上述“拉班符号序列的 DTW 匹配算法”将两个原始手语数据的手势运动做拉班符号序列的 DTW 匹配,并将  $LB(Q_{ab}, T_{cd})$  保存下来,作为两个手语手势运动相似程度的检测阈值.获得两个手语手势运动相似程度的检测阈值  $\bar{\theta}$ .  $\bar{\theta}$  的存在是由于非特定人手语数据差异性矛盾所决定的.我们对原始的手语数据样本成对地应用上述算法,这样我们可以获得一系列的给予原始手语数据的手语手势运动相似程度的检测阈值  $\bar{\theta}$ .将这些阈值根据手语手势运动的差别大小分别得到  $\bar{\theta}_{\max}$  和  $\bar{\theta}_{\min}$ .如果手语生成数据与原始数据的拉班运动相似性测度落在  $(\bar{\theta}_{\max}, \bar{\theta}_{\min})$  的开集之内,则认为生成的手语数据在运动检测的层面上是有效的.手语生成数据的拉班运动有效性检测算法如算法 4 所示.该算法的原理如图 10 所示.

**算法 4.** 手语生成数据的拉班运动有效性检测算法.

输入:生成的手语数据的手势运动的拉班运动符号序列  $[q_a, q_b]$  和原始手语数据的手势运动的拉班运动符号序列  $[t_c, t_d]$ .

输出:手语数据在运动检测的层面上是否有效.

$LB(Q_{ab}, T_{cd})=0$ ;

For  $i=a$  to  $b$  {

If  $\sum_{i=0}^{b-a} \min_{-r \leq k \leq r} (d(i+a, j+c+k)) > \sum_{k=1}^l w_k$  return false; Else  $LB(Q_{ab}, T_{cd}) = \sum_{i=0}^{b-a} \min_{-r \leq k \leq r} (d(i+a, j+c+k))$   
 If  $LB(Q_{ab}, T_{cd}) / (b-a+c-d+2) > \bar{\theta}_{max}$  OR  $LB(Q_{ab}, T_{cd}) / (b-a+c-d+2) < \bar{\theta}_{min}$  return false}  
 return true

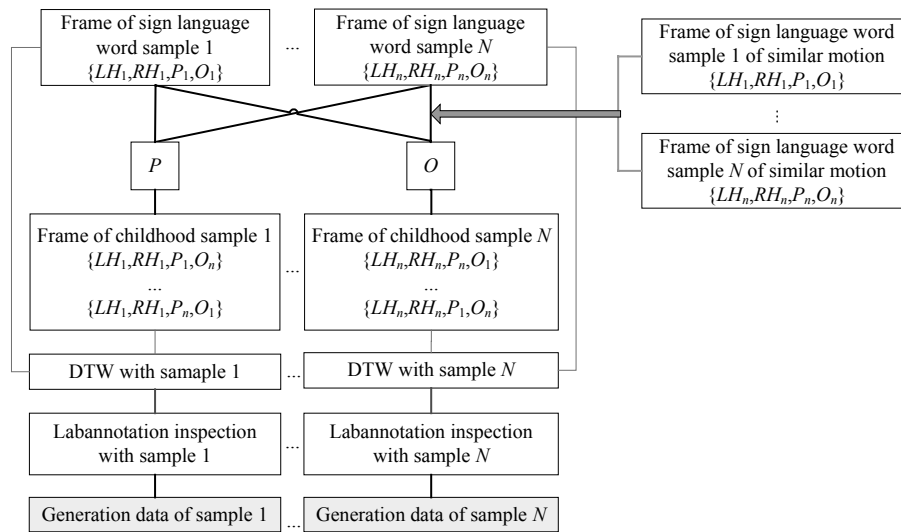


Fig.9 Sign language data generation and detection algorithm based on human kinematics

图 9 基于人体运动学的手语数据生成及检测算法

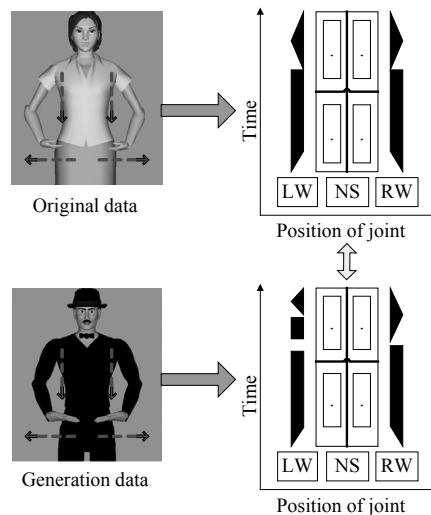


Fig.10 Principle demonstration for the validity inspection of Laban movement of sign language generation data

图 10 手语生成数据的拉班运动有效性检测的原理图

### 2.3 基于手语语言学与人体运动学相结合的手语数据生成

基于手语语言学与人体运动学相结合的原则仍是以保持语义信息不变为主,其次是保持运动信息不变.也就是说,在确保手形数据正确的前提下,再对与动作相关的数据进行遗传算法的衍生.生成算法和检测算法如前所述.基于手语语言学与人体运动学相结合的手语数据的生成和检测过程如图 11 所示.

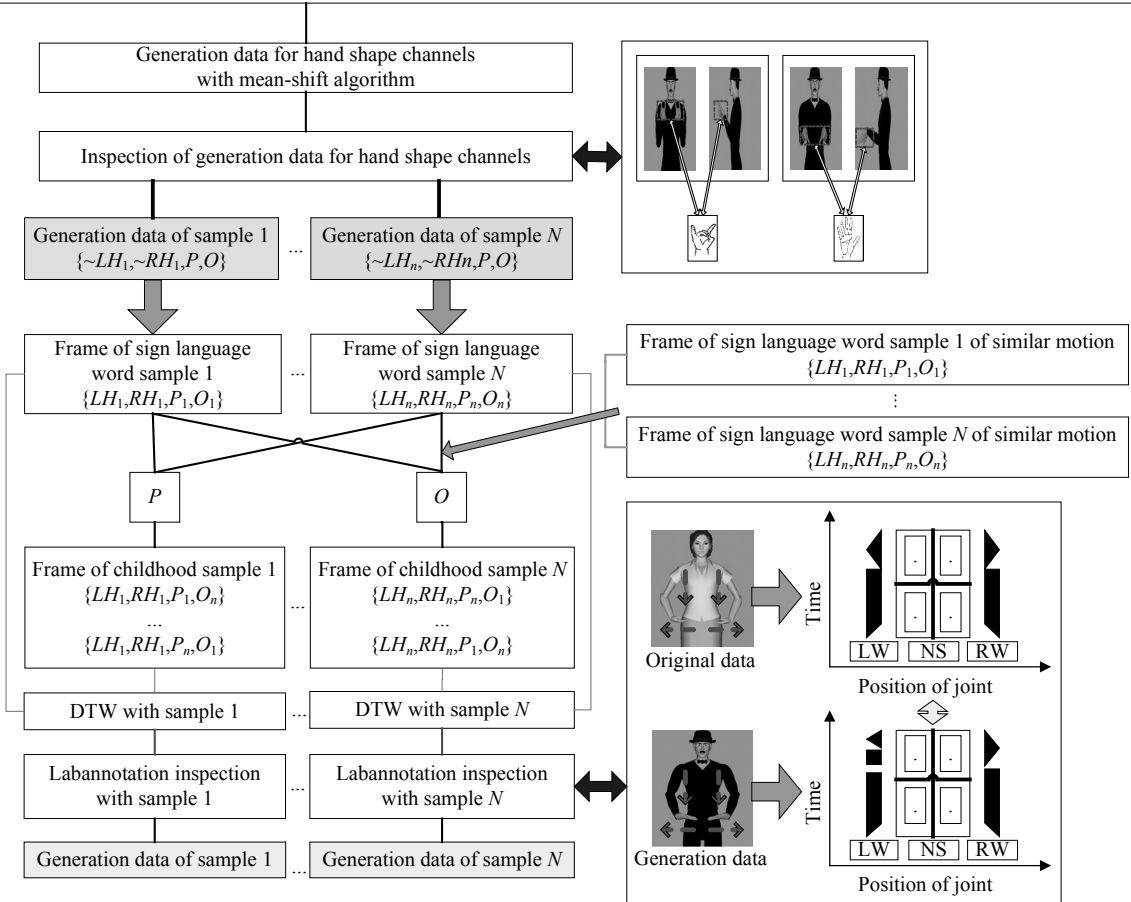
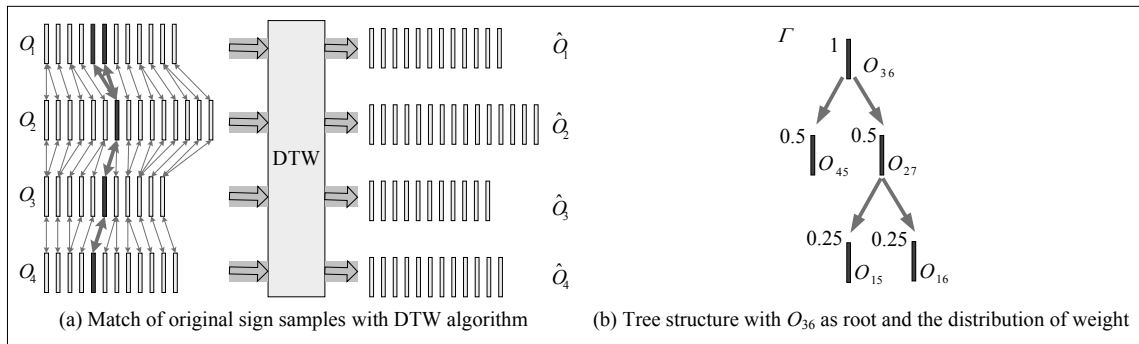


Fig.11 Principle demonstration for the integration of sign language linguistics and human kinematics

图 11 基于手语语言学与人体运动学相结合的原理图

### 3 实验

为了验证所提出方法的泛化能力,我们设计并进行了一些实验.我们依据表 4,从中选取 2 435 个手势手语词作为原始词汇集.这些手语词的数据分别来自于 6 位手语老师,每个手势手语词分别采集了两遍.

#### 3.1 非特定人的手势手语的有效性检测的实验

我们根据上述手势手语的生成方法,在这里也分别采用了 3 种策略.我们对 3 种策略作如下的定义:

1. S1 策略采用基于人体运动学的手语数据生成方法,使用拉班舞谱对手语生成数据的拉班运动有效性检测.
2. S2 策略采用基于手语语言学的手语数据生成方法,使用左、右手关键手形模板对手语生成数据进行标记和有效性检测.
3. S3 策略采用保持语义信息不变与保持运动信息不变相结合的手语数据生成方法,使用左、右手关键手形模板与拉班舞谱相结合的方法对手语生成数据进行有效性检测.

由于手势手语数据检测没有已有的研究成果作为借鉴,因而在实验架构的设计上,我们借鉴人脸检测的实验架构设计.由于人脸检测是单类别的检测问题,而手势手语是多类别的检测问题,因而我们根据手势手语数据的特点提出了下述基于手势手语数据检测的统计实验架构.我们首先作如下的定义:

- (1) 一次检测率:以原始手势手语数据作为基准进行检测,是正确检测到手语生成数据样本数所占所有手语生成数据样本数的百分比.
- (2) 二次检测率:一次误检率所构成的手语生成数据在选取的原始词汇集范围内进行识别,将识别结果的原始手势手语数据作为基准进行检测,是正确检测到手语生成数据样本数所占一次误检率所构成的手语生成数据样本总数的百分比.
- (3) 误检报警数:检测正确而识别率低的样本个数.

如图 12 所示,数字“1”代表原始数据样本,用不同的形状来表示不同的手语词汇.数字“2”代表由原始数据样本“1”(用三角形表示)所生成的各个样本,数字“2”表示正确生成的样本,即由一次检测获得的生成数据样本;圆圈数字“2”表示由于变异等因素的存在,使得生成的数据样本扩张到同一个原始抽样词汇集中的其他手语词,即由二次检测获得的生成数据样本;方块数字“2”表示由于变异等因素的存在,使得生成的数据样本扩张到同一个原始抽样词汇集以外的其他手语词,或者可能已经不是手语词了.

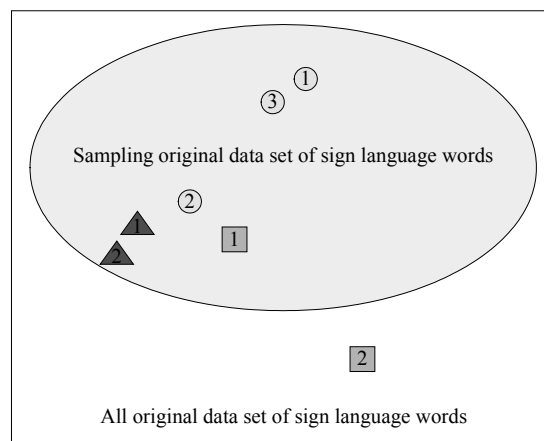


Fig.12 Original sample based inspection principle

图 12 基于原始样本的检测原理图

由表 5 的实验结果可以看出,手形通道随着生成样本的规模增大,其误检报警数也随之增多.之所以存在二次检测,是由于在数据样本的生成过程中存在“变异”或“生成强度”等因素的影响,产生了数据样本的“基因突变”或“关键手形之间的类别跨越”.由于这种非常规的变化发生在易混词之间,因而它们所占的比率并不是很大.存在二次检测的另一个原因是,我们的检测依据是原始数据的一个手语词,一次检测仅能说明生成的数据是否为该手语词,并不能判断生成的数据是否“变异”到原始抽样词集的其他手语词,因而只能先借助非特定人识别确定是原始抽样词集中的哪个词(无法从手语数据中直接判断),再利用所确定词的原始数据样本在应用上述的检测方法重新检测.

**Table 5** Statistics for the inspection experimental results of some generation strategies**表 5** 几种生成策略的检测实验结果的统计

Generation statistics	Size of generation samples	One time detection rate (%)	Two time detection rate (%)
S1	613 621 (average 69 human/word)	86.30 (1 error-detected number)	2.02 (3 error-detected number)
S2	525 960 (average 7 human/word)	74.48 (20 error-detected number)	0.49 (33 error-detected number)
S3	8 006 279 (average 137 human/word)	79.91 (31 error-detected number)	3.96 (50 error-detected number)

为了说明检测方法的有效性,我们制定了如下两种用识别系统作为检测方法的生成策略:

1. S4 策略为 6 个相互独立的通道建模.将这 6 个通道作为基因片段,应用遗传算法进行手势手语的数据生成.这 6 个通道是:左手位置、左手方向、左手手形、右手位置、右手方向、右手手形.
2. S5 策略不对原始手语数据进行通道建模,直接应用 Mean-Shift 算法进行手势手语的数据生成.

表 6 中的非特定人识别系统得到的检测率即为生成数据的识别率.对上述以识别系统作为检测方法的生成策略所生成的手语数据经识别后认为正确的样本,我们应用左、右手关键手形模板与拉班舞谱相结合的方法再进行检测,经检测发现,错误的样本与所有生成样本的比率在表 6 中称为检测错误率.S4 实验平台与文献[14]一致,而 S5 实验平台与文献[15]一致.

**Table 6** Results of inspection experimental for generation data S4 and S5 after recognizing the system of identification sign language recognition**表 6** S4 和 S5 的生成数据经非特定人识别系统识别后的检测实验结果

Generation statistics	Size of generation samples	Detection rate after recognizing the system of identification sign language recognition (%)	Inspection-Error rate (%)
S4	12 739 922 (average 218 human/word)	84.78	34.40
S5	10 460 758 (average 179 human/word)	80.30	19.55

由表 6 的实验结果可以看出,生成数据经识别系统识别后存在着较大的检测错误率,这说明识别率较高的生成数据未必是手语数据,即一些带有“手形的手势动作”也会“累积”一个较高的识别概率.对于一个鲁棒性强的识别系统,检查其识别对象也是必须的.对手语的生成数据来讲,只有生成非特定人的手语数据才有意义,而目前非特定人的手语识别系统的识别效果并不是很理想.

### 3.2 非特定人手势手语的有效生成数据对识别系统的影响

生成非特定人的手语数据的最终目的是提高非特定人的手语识别系统的识别率.传统的隐马尔可夫模型(HMM)的训练方法基于统计概率的最大似然准则(maximum likelihood estimator,简称 MLE),在训练样本数目足够大的情况下,这种方法在理论上可以得到最优的结果.因此,我们把 S1,S2,S3 策略经一次检测所生成的有效数据样本加入到非特定人的手语识别系统的训练集中,并将非特定人的原始数据作为测试集,同时采用交叉验证的方式.5 个人的数据加入训练集,剩下 1 人的数据作为未注册测试集(UnReg).本实验与方高林<sup>[1]</sup>所做实验的不同点是测试数据不参与训练.

由表 7 可以看出,S1 策略有效生成数据加入到训练集,在非注册集上平均识别率提高了 0.17%.S2 策略有效生成数据加入到训练集,在非注册集上平均识别率提高了 0.02%.S3 策略有效生成数据加入到训练集,在非注册集上平均识别率提高了 0.39%.由此可见,我们的生成策略和检测方法是有效的,但由于原始数据样本容量的限制,生成的有效数据的规模不可能达到足够大,因而识别率提高的程度很小.另一个方面也说明非特定人手语数据所存在的差异性当中,动作的影响比手形所造成的影响要大得多.

由于传统的基于 MLE 准则的 HMM 识别系统需要在训练样本数目足够大的情况下才能取得理想的识别效果,由表 7 我们得出了生成样本经识别系统检验是有效的这一结论.然而,生成有效样本的规模并未达到足够大.为了提高识别率,我们采用可区分性训练(discriminative training)结合 HMM 的方法(DT/HMM)<sup>[46]</sup>来解决由于训练样本的不足造成识别率低的问题.具体的实验结果见表 8.

**Table 7** Recognition results for the unregistered sets of MLE identification system

表 7 MLE 识别系统的非注册集识别结果

Signer	MLE/HMM (%)	Recognition rate after S1 effective generation data was added to training set (%)	Recognition rate after S2 effective generation data was added to training set (%)	Recognition rate after S3 effective generation data was added to training set (%)
A	78.30	78.47	78.33	78.65
B	74.23	74.43	74.25	74.59
C	75.26	75.48	75.28	75.61
D	76.19	76.41	76.21	76.62
E	79.57	79.59	79.60	80.02
Average	76.71	76.88	76.73	77.10

**Table 8** Recognition results for the unregistered sets of DT/HMM identification system

表 8 DT/HMM 识别系统的非注册集识别结果

Signer	DT/HMM (%)	Recognition rate after S1 effective generation data was added to training set (%)	Recognition rate after S2 effective generation data was added to training set (%)	Recognition rate after S3 effective generation data was added to training set (%)
A	86.32	91.10	86.50	91.63
B	82.26	84.65	82.45	87.51
C	83.29	85.69	83.47	88.57
D	84.23	86.73	86.92	89.74
E	87.61	90.01	87.79	93.02
Average	84.74	87.64	85.43	90.09

对比表 7 和表 8 可以看出,有效的生成样本对于识别率的提高是不依赖于识别系统的,但不同的识别系统由于采用不同的参数训练方法,对于训练样本的依赖程度也有所不同,因而导致识别率提高的幅度不一致.在原始数据集上,DT/HMM 比传统 HMM 识别率要高 8.03%.在基于 MLE 准则的传统 HMM 识别系统中,将 3 种策略有效生成数据加入到训练集中,非注册集的识别率提高幅度分别为 0.17%,0.02%,0.39%.在 DT/HMM 识别系统中,由于  $h$  参数对于参数模型的收敛具有促进作用,非注册集的识别率提高幅度分别为 2.90%,0.69%,5.35%.

#### 4 结束语

训练样本缺乏及数据的有效性给识别系统性能带来了挑战.针对手语识别研究中训练样本缺乏及如何检测数据的有效性问题,根据手势手语的特点,本文提出了基于手语语言学和人体运动学相结合的非特定人手语数据的生成和检测方法.Mean-Shift 算法具有便于控制生成强度的优点,我们将改进的 Mean-Shift 算法应用于手形数据通道的生成.在关键手形的保持段减小生成强度,在关键手形的过渡段增加生成强度,以减少对关键手形无谓的改变来保持手势手语的语言学特性;应用关键手形的音韵标记进行数据有效性的检测,确保生成数据不丢失语义.其次,我们在遗传算法中加入一些相似性动作数据作为变异因子,以丰富手语手势动作的运动,防止手势动作的“近亲繁殖”;应用拉班舞谱对其进行数据有效性的检测,可以防止因生成数据的动作过于“夸大”而变异成其他非母体的手语词或非手语词数据.根据语言类数据样本的检测属于多类别数据样本的检测这一特点,我们提出了基于原始样本的检测实验框架.从该框架下得到的实验结果表明,我们的手势手语的生成策略和检测方法是有效的.但由于原始数据样本容量的限制,有效的生成样本的容量并不能达到“足够大”;应用传统的以 MLE 为准则的 HMM 识别系统进行识别,识别效果并不理想;在 DT/HMM 识别系统中,由于可区分性训练对于参数模型的收敛具有促进作用,因而非注册集的识别率提高幅度较为理想.

#### References:

- [1] Fang GL. Research on statistical models for sign language recognition [Ph.D. Thesis]. Harbin: Harbin Institute of Technology, 2004 (in Chinese with English abstract).
- [2] Christian PV. American sign language recognition reducing the complexity of the task with phoneme based modeling and parallel hidden Markov models [Ph.D. Thesis]. Pennsylvania: University of Pennsylvania, 2003.

- [3] Fang GL, Gao W, Ma JY. Signer-Independent sign language recognition based on SOFM/HMM. In: Proc. of the IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems. 2001. 90–95.
- [4] Jiang F, Ye JJ, Yao HX. Chinese sign language recognition system with large words based on HMM and SVM two level classifier. Journal of Computer Research and Development, 2005,42(Suppl.):212–217 (in Chinese with English abstract).
- [5] Baird H. State of the art of document image degradation modeling. In: Proc. of the 4th IAPR Workshop on Document Analysis Systems. 2000. 1–16.
- [6] Lu X, Jain AK. Resampling for face recognition. In: Proc. of the Int'l Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA 2003). 2003. 869–877.
- [7] Chen SC, Zhang DQ, Zhou ZH. Enhanced (PC)<sup>2</sup>A for face recognition with one training image per person. Pattern Recognition Letters, 2004,25(10):1173–1181.
- [8] Gao W, Shan SG, Chai XJ, Fu XW. Virtual face image generation for illumination and pose insensitive face recognition. In: Proc. of the ICASSP2003, Vol.IV. 2003. 776–779.
- [9] Chen J, Chen XL, Gao W. Resampling for face detection by self-adaptive genetic algorithm. In: Proc. of the 17th Int'l Conf. on Pattern Recognition (ICPR 2004). Cambridge, 2004. 822–825.
- [10] Varga T, Bunke H. Effects of training set expansion in handwriting recognition using synthetic data. In: Proc. of the 11th Conf. of the Int'l Graphonomics Society. Los Alamitos: IEEE Computer Society Press, 2003. 200–203.
- [11] Sarah Z. Using background knowledge to improve text classification [Ph.D. Thesis]. Rutgers: Rutgers University, 2002.
- [12] Martin S, Tommi J. Kernel expansions with unlabeled data. Advances in Neural Information Processing Systems, 2001,13: 626–632.
- [13] Wang CL, Gao W. Re-Sampling for Chinese sign language recognition by genetic algorithm. In: Proc. of the 6th Int'l Workshop on Gesture in Human-Computer Interaction and Simulation. Berlin: Springer-Verlag, 2005. 200–207.
- [14] Zhang CX, Jiang F, Yao HX. Static gesture quantization and DCT based sign language generation. In: Proc. of the 1st Int'l Conf. on Affective Computing and Intelligent Interaction. Berlin: Springer-Verlag, 2005. 160–167.
- [15] Jiang F, Gao W, Yao HX, Zhao DB, Chen XL. Synthesized data driving: An approach toward singer-independent sign language recognition. Journal of Computer Research and Development, 2007,44(5):873–881 (in Chinese with English abstract).
- [16] Srisuk S, Kurutach W. A new robust face detection in color images. In: Proc. of the 5th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Washington: IEEE Computer Society, 2002. 306–311.
- [17] Albiol A, Torres L, Delp EJ. An unsupervised color image segmentation algorithm for face detection application. In: Proc. of the Int'l Conf. on Image Processing. 2001. 681–684.
- [18] Lewis TW, Powers DM. Lip feature extraction using red exclusion. In: Proc. of the Selected Papers from Pan-Sydney Area Workshop on Visual Information Processing (VIP2000). 2000. 61–67.
- [19] Zhang HM, Zhao DB, Gao W. Face detection under rotation in image plane using skin color model, neural network and feature-based face model. Chinese Journal of Computers, 2002,25(11):1250–1256 (in Chinese with English abstract).
- [20] Ma XY, Yang GS. Locating human eyes algorithm based on the improved progressive increase of the threshold. Journal of He'nan University, 2006,36(4):83–86.
- [21] Li XS, Li GH. A rapid approach for face normalization based on human eyes location. Computer Engineering & Science, 2006, 28(12):63–65 (in Chinese with English abstract).
- [22] Li SZ, Zhu L, Zhang ZQ, Blake A, Zhang HJ, Shum H. Statistical learning of multi-viewface detection. In: Proc. of the 7th Europe Conf. on Computer Vision. LNCS 2353, London: Springer-Verlag, 2002. 67–81.
- [23] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol.1. 2001. 511–518.
- [24] Xiao R, Zhu L, Zhang HJ. Boosting chain learning for object detection. In: Proc. of the 9th IEEE Int'l Conf. on Computer Vision (ICCV 2003), Vol.1. 2003. 709–715.
- [25] Wu B, Ai HZ, Huang C, Lao SH. Fast rotation invariant multi-view face detection based on real Adaboost. In: Proc. of the 6th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. 2004. 79–84.



- [26] Li SZ, Zhang Z. Float boost learning and statistical face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2004, 26(9):1112–1123.
- [27] Viola P, Jones M. Fast and robust classification using asymmetric AdaBoost and a detector cascade. In: *Advances in Neural Information Processing System 14*. 2001. 1311–1318.
- [28] Hao LB, Nlu HJ, Lu CM. Face detection algorithm fusion of the skin-color model and wavelet transform. *Journal of Xidian University (Natural Science Edition)*, 2007,34(6):864–868 (in Chinese with English abstract).
- [29] Fang YY, Huang XY, Shen ZX. New face detection method based on skin color and template matching. *Journal of Chongqing Institute of Technology (Natural Science Edition)*, 2007,21(12):102–106 (in Chinese with English abstract).
- [30] Zeng Y, Chen G, Xu SF. Face detection based on minimum feature number expectation. *Computer Simulation*, 2007,24(12):328–331 (in Chinese with English abstract).
- [31] Chen HY, Huang DR, Lin JB. Face location based on color and geometric invariance. *Computer Measurement & Control*, 2007, 15(12):1688–1690 (in Chinese with English abstract).
- [32] Sun JQ, Wang RG, Hu Q, Li SY. Fast face detection combine both feature-based and image-based approach. *Application Research of Computers*, 2008,25(1):294–296 (in Chinese with English abstract).
- [33] Kanda TY. Phonology notations of Japanese sign language. *Sign Language Research*, 1991,12:84–91.
- [34] Liddell SK, Johnson RE. American sign language: The phonological base. *Sign Language Studies*, 1989,(64):195–277.
- [35] Sandler W. *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*. Berlin: Walter de Gruyter, Inc., 1989.
- [36] Simakosi. Construction model of continuous sign language: The new trends of sign language phonology theory. *Proc. of Japanese Sign Language Academic Research Association*, 1989,(10):2–14.
- [37] Honnna SK, Kanda TY, Oda KH, Katou SHK. Notations of sign language. *Proc. of Japanese Sign Language Academic Research Association*, 1985,(7):2–11.
- [38] Kanda TY. Sign language abbreviation notations of Japanese sign language and handshapes system using phonology notations. *Summaries of East Sea Industry Academy*, 1986,(7):55–70.
- [39] Honnna SK, Kanda TY, Oda KH, Katou SHK. Proposals of sign language notation. *Documents of Speech Research of Japanese Acoustic Association*, 1984,(2):83–88.
- [40] Stokoe W, Casterline D, Croneberg C. *A Dictionary of American Sign Language on Linguistic Principles*. Washington: Gallaudet College, 1965.
- [41] Kanda KY, Itikawa AR, Nagaashima YJ, Terauchi MN, Kato TC. Sentence notations of sign language. *Proc. of Human Interaction System'99*, 1999(2):745–748.
- [42] Williams A, Aun HC. *Labanotation or Kinetography Laban: The System of Analyzing and Recording Movement*. New York: A Theaten Arts Book, 1977.
- [43] Comaniciu D, Ramesh V, Meer P. Kernel-Based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, 25(5):564–575.
- [44] Wu JQ. *Research and implementation on Chinese sign language recognition algorithm [Ph.D. Thesis]*. Harbin: Harbin Institute of Technology, 2000 (in Chinese with English abstract).
- [45] Gu DQ, Song XH, Yu YY. The analysis of Chinese sign language's basic words (basic movements). *Chinese Journal of Special Education*, 2005,2(56):56–72 (in Chinese with English abstract).
- [46] Wang YX, Ni XB, Jiang F. Discriminative training methods of HMM for sing language recognition. *CAAI Trans. on Intelligent Systems*, 2007,2(1):80–84 (in Chinese with English abstract).

#### 附中文参考文献:

- [1] 方高林. 手语识别统计模型的研究[博士学位论文]. 哈尔滨: 哈尔滨工业大学, 2004.
- [4] 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): 212–217.
- [15] 姜峰, 高文, 姚红勋, 赵德斌, 陈熙霖. 非特定人手语识别问题中的合成数据驱动方法. *计算机研究与发展*, 2007, 44(5): 873–881.
- [19] 张洪明, 赵德斌, 高文. 基于肤色模型、神经网络和人脸结构模型的平面旋转的人脸检测. *计算机学报*, 2002, 25(11): 1250–1256.

- [28] 郝琳波,牛海军,卢春梅.融合肤色模型与小波变换的人脸检测算法.西安电子科技大学学报(自然科学版),2007,34(6):864-868.
- [29] 叶艳芳,黄席樾,沈志熙.一种基于肤色和模板匹配的人脸检测方法.重庆工学院学报(自然科学),2007,21(12):102-106.
- [30] 曾义,陈刚,许世峰.基于特征数期望最小化的人脸检测分类器的构建.计算机仿真,2007,24(12):328-331.
- [31] 陈虹宇,黄大荣,林全表.基于颜色信息和几何不变性的人脸检测定位.计算机测量与控制,2007,15(12):1688-1690.
- [32] 孙见青,汪荣贵,胡琼,李守毅.基于特征和图像相结合的快速人脸检测.计算机应用研究,2008,25(1):294-296.
- [44] 吴江琴.中国手语识别算法的研究与实践[博士学位论文].哈尔滨:哈尔滨工业大学,2000.
- [45] 顾定倩,宋晓华,于缘缘.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.
- [46] 王宇轩,倪训博,姜峰.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.



倪训博(1979-),男,黑龙江哈尔滨人,博士生,主要研究领域为模式识别,机器学习,信号处理,手语运动数据检测,手语语言学,运动数据的流形学习.



姜峰(1978-),男,博士,讲师,CCF 会员,主要研究领域为模式识别,机器学习,图像处理.



赵德斌(1963-),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.



姚鸿勋(1965-),女,博士,教授,博士生导师,主要研究领域为视觉语言,多媒体技术,数字水印.



高文(1956-),男,博士,教授,博士生导师,CCF 高级会员,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.