

基于关键动作双重转移概率的连续手语语句识别算法

李 晨¹ 黄元元¹ 胡作进²

(南京航空航天大学计算机科学与技术学院 南京 210016)¹

(南京特殊教育师范学院数学与信息科学学院 南京 210038)²

摘 要 目前,连续手语识别的最大难点在于如何对其中包含的词汇进行有效分割。本文将关键动作看作手语的基元,提出了一种基于关键动作双重转移概率的连续手语识别算法。在获得连续手语基元序列的前提下,根据相邻基元的词内及词间转移关系,可以有效地寻找到词汇边界,从而对基元序列做分割,并逐一识别出各基元分组的候选词汇。最后,根据不同基元分组的候选词汇间的转移概率,计算出对应合成句子的概率,并按照最大概率原则输出连续手语的最终识别结果。该算法容易实现,执行效率高,经实验验证其可以面向非特定人群。

关键词 手语语句,关键动作,转移概率

中图分类号 TP391.41 文献标识码 A

Continuous Sign Language Sentence Recognition Based on Double Transfer Probability of Key Actions

LI Chen¹ HUANG Yuan-yuan¹ HU Zuo-jin²

(College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China)¹

(College of Math and Information Science, Nanjing Normal University of Special Education, Nanjing 210038, China)²

Abstract At present, the most difficult problem in continuous sign language recognition is how to split out the words effectively. In this paper, key actions were regarded as the basic units of sign language and an algorithm based on double transfer probability of key actions was proposed. After acquiring the sequence of basic units from continuous sign language, the boundaries of words can be effectively found by judging the intra-word and inter-word transfer relations of all adjacent basic units. Then the sequence of basic units are segmented by these boundaries and the candidate words of each group of basic units can be identified. Finally, according to the transfer probabilities between candidate words of different groups, the probability of corresponding synthetic sentence is calculated and then the final recognition result is output by the principle of maximum probability. The algorithm is easy to implement and has high execution efficiency. It can be applied to non-specific population through experimental verification.

Keywords Sign language sentence, Key actions, Transfer probability

1 引言

在当前众多的人机交互技术中,手势具有生动、形象、直观的特点,且符合用户的操作习惯,成为用户输入信息的重要方式;同时,手语作为聋哑人交流时使用的语言,是他们传达思想、表达情感、获取信息的重要工具。连续手语识别一直是手势研究中较为复杂和关键的部分,所以对它进行研究具有十分重大的应用价值和社会意义。

目前,国内外研究最多的是孤立手语词的识别,对连续手语识别的研究相对较少,原因在于手语单词的边界检测问题一直是个难点问题。在国外,美国佐治亚理工学院的 Starnier 与 Pentland^[1]采用普通摄像机拍摄手语视频,借助于隐马尔可夫模型(Hidden Markov Model, HMM)对由 40 个标明词性的短语组成的简单美国手语语句进行识别,其识别率达到了 91%,开启了连续手语识别的先河。新加坡国立大学的 Kong 等^[2-3]利用局部最小运动速度和局部最大方向角来联合

确定手语边界,实验结果对于特定人的连续手语识别的召回率和精确度分别为 95.7%,96.6%,对于非特定人的连续手语识别的召回率和精确度分别为 86.6%,89.9%。然而,数据的获取需要依赖数据手套和磁传感器,数据采集不够便捷,并且利用局部最小运动速度和局部最大方向角检测到的手语边界数量较多,需要做进一步的处理判断。韩国高丽大学的 Yang 等^[4]首先针对所有的独立手语词汇样本训练了一个条件随机场 CRF(Conditional Random Fields),并进一步建立了一个基于 CRF 的阈值模型,在由 48 个手语词组成的美国手语语句库中达到了 93.5%的识别率;但是由于阈值模型的建立只针对特定的实验参与者,而不是非特定人群,所以不能快速建立适用于非特定人群的通用模型,因而该算法具有一定的局限性。美国南佛罗里达的 Yang 等^[5]首先采用动态时间规整(Dynamic Time Warping, DTW)算法计算候选手语词序列与手语模型之间的距离,接着采用逐层构筑(level-building)算法查找最小句子级别的匹配距离,并通过回溯算法完

本文受江苏省“双创”项目资助。

李 晨(1995-),女,硕士生,主要研究方向为模式识别、图像处理;黄元元(1975-),女,博士,副教授,主要研究方向为多媒体技术、图像处理、模式识别;胡作进(1965-),男,博士,教授,主要研究方向为数据处理、机器学习, E-mail:805861040@qq.com。

成连续句子的划分。该系统在150个简单美国手语语句上取得了83%的识别率,但是该算法的时间复杂度很高,难以实现手语的实时识别。亚琛工业大学的Koller等^[6]将卷积神经网络CNN(Convolutional Neural Network)嵌入到隐马尔可夫模型中,先基于CNN求解出每帧图像属于各隐状态的概率,再选用连续手语对应的词汇序列的后验概率作为目标函数,结合 n 元语法模型(N-Gram)、一阶隐马尔可夫过程及维特比算法,识别出连续手语最优的词汇序列。在PHOENIX 2012,PHOENIX 2014,SIGNUM这3个手语库中,该算法的单词错误率分别为32%,32.5%,7.4%,分类效果相较于对比算法取得了超过15%的改善。但是,该算法的识别原理复杂,还需要对3个超参数进行网格搜索来确定其取值,在一定程度上增加了时间损耗。

在国内,哈尔滨工业大学的方高林等^[7-8]把精简循环网(Simple Recurrent Network,SRN)作为连续手语的段边界检测器,在由208个词组成的100个中国手语语句上实现了91.9%的识别率;但是在SRN的训练过程中没有考虑到运动连接词的存在,训练数据输入SRN的输出只有左边界、右边界和段内部3个结果,对于运动连接词的归属问题定义比较模糊。中国科学技术大学的杨文文等^[9]在逐层构筑算法的基础上结合HMM,同时辅以手语词帧长约束和二元语法模型(bigram),最终在由66个手语词组成的44个中国手语句子库上获得了90.2%的识别率。但是,该算法运行时的时间复杂度仍然较高,且对大规模数据的处理比较缓慢。南京航空航天大学徐鑫鑫等^[10]利用加重的关键动作实现对连续手语语句的识别,算法执行效率较高,但是如果高权值的关键动作出现错误或漏检,将无法给出正确的识别结果。中国科学技术大学的黄杰等^[11]提出一种新的连续手语识别框架LS-HAN,它由3部分构成:用于生成视频特征表示的双流卷积神经网络,用于缩小手语视频和词汇序列语义差距的潜在空间(Latent Space,LS),以及基于识别的分层注意力网络(Hierarchical Attention Network,HAN)。该算法没有对连续手语做时间上的分割,可以避免时间上的错误分割给识别带来的影响,但是它对训练样本库中不包含的句子的识别效果较差。

由此可以看出,现有的连续手语识别算法在寻找手语边界的问题上,仍然有很大的提升空间。目前已有的边界检测算法大多设计比较复杂,并且对于非特定人群没有较好的鲁棒性。设计复杂的原因之一是现有的模型大部分是针对所有的手语样本进行处理,样本的数据量比较大,而不同人的手语动作的幅度和习惯不一样也加大了识别的难度。因此,本文提出了一种基于关键动作双重转移概率的连续手语语句识别算法。经实验验证,本文提出的算法简单易行,且准确度较高,针对非特定人群也可以有较好的识别效果。

2 算法设计

关键动作可以看作是手语动作的基元^[12]。之前,我们已经做过大量的研究和实验,证明了利用关键动作描述手语词汇以及连续语句的可行性与有效性^[10,13-15]。利用基元及基元的组合描述连续手语,不仅可以极大地减少数据量,而且由于基元的相对稳定性,可以有效地去除那些手语者自身以及手语者之间存在的稳定性与差异性给连续手语数据带来的波

动。因此,在对手语语句提取了关键动作,并对关键动作进行了有效识别与分类后,本文设计了一种基于关键动作双重转移概率的识别算法,可以对连续语句进行有效分割,找到其中包含的词汇边界,进而对语句做出识别。在关键动作出现漏检或者错误的情况下,本文的算法依然有可能给出正确的识别结果。

2.1 数据模型设计

若当前有 M 个手语单词 $W_i(1 \leq i \leq M)$,这些词汇一共包含 N 个关键动作,即基元 $k_j(1 \leq j \leq N)$,那么理论上由这些词汇组成的连续语句均可以做出识别。首先,需要设定4种数据模型,这些数据模型均在离线情况下创建。

(1) 词汇模板

每个词汇均可以看作是基元的组合,这些基元可以通过分类器进行识别^[14-15]。若将每个基元以其所属类别的标签表示,那么每个词汇 $W_i(1 \leq i \leq M)$ 即可看作一个字符串,字符串中的每个字符对应基元的类别标签。

(2) 单词之间的转移概率矩阵

构造一个 $M \times M$ 的方阵 B ,其中的元素 b_{ij} 表示在实际的连续语句中单词 W_i 出现在单词 W_j 之前的概率,称为 W_i 到 W_j 的词间转移概率,即:

$$b_{ij} = p(W_i, W_j) \quad (1)$$

这个概率可以通过对大量的机读样本进行学习 and 训练,并以统计的方式得到,即:

$$p(W_i, W_j) = \frac{\text{sum}(W_i, W_j)}{\sum_{j=1}^M \text{sum}(W_i, W_j)} \quad (2)$$

其中, $\text{sum}(W_i, W_j)$ 表示在所有学习样本中,单词 W_i 出现在单词 W_j 之前的次数。

(3) 基元的词内转移概率矩阵

构造一个 $N \times N$ 的方阵 Q ,称其为关键动作的词内转移概率矩阵。其中的元素 q_{ij} 取决于类别标签为 i 的基元 k_i 与类别标签为 j 的基元 k_j 是否存在同一个单词内部存在先后的时序关系,即:

$$q_{ij} = \begin{cases} 1, & \text{词汇模板中存在“ij”这个子串} \\ 0, & \text{词汇模板中不存在“ij”这个子串} \end{cases} \quad (3)$$

有些词汇包含多个基元,也有少量的词汇仅包含一个基元。因此,当 $i=j$ 即对角线元素的数值被设定为:

$$q_{ii} = \begin{cases} 1, & \text{存在仅包含类别标签“i”的词汇} \\ 0, & \text{不存在仅包含类别标签“i”的词汇} \end{cases} \quad (4)$$

(4) 基元的词间转移概率矩阵

构造一个 $N \times N$ 的方阵 H ,称其为关键动作的词间转移概率矩阵。其元素 h_{ij} 表示类别标签为 i 的基元 k_i 与类别标签为 j 的基元 k_j 是否存在不同单词之间的转移概率,即:

$$h_{ij} = \begin{cases} b_{xy}, & \text{存在以 } k_i \text{ 为尾基元的词汇 } W_x, \\ & \text{且存在以 } k_j \text{ 为首基元的词汇 } W_y \\ 0, & \text{其它} \end{cases} \quad (5)$$

其中, b_{xy} 为矩阵 B 中 W_x 到 W_y 的词间转移概率。若 W_x 或者 W_y 是词汇集合,那么取 $h_{ij} = \max\{b_{xy} | x \in \omega_1, y \in \omega_2\}$ 。

在方阵 H 中,设定一个阈值 ρ ,若 h_{ij} 大于阈值 ρ ,表示 k_i 与 k_j 之间存在强词间转移,否则是弱词间转移。对于 ρ 的取值,经过大量实验发现,对所有非零 h_{ij} 求其大津(Ostu)阈值,得到的实验效果最好。

2.2 基于关键动作双重转移概率的连续手语识别

连续手语语句经过关键动作的提取与识别后,可被看作

是一个关键动作序列^[10]在该序列中,只要确定词汇边界,即可完成识别。假设从一个连续手语语句中提取到 L 个关键动作,即 L 个基元 $\{k_1, k_2, \dots, k_L\}$,这 L 个基元对应的类别标签是 $\{c_1, c_2, \dots, c_L \mid 1 \leq c_i \leq N, 1 \leq i \leq L\}$ 。首先,可以创建一个长度为 L 的标记序列 $S = \{s_1, s_2, \dots, s_L\}$,初始情况下标记序列中所有元素值设定为 0。每个标记 s_i 对应着基元 k_i 与 k_{i+1} 之间的关系,最后根据 s_i 的取值实现对基元序列的分割,从而完成对话语句语义的识别。确定 s_i 的过程为:

(1)对于 k_i 与 k_{i+1} ,首先判断若词内转移概率 $q_{c_i c_{i+1}} = 1$,则 s_i 仍取值为 0,表明 k_i 与 k_{i+1} 属于同一个词汇。

(2)若 k_i 与 k_{i+1} 不存在词内转移概率,但存在词间转移概率,且是强词间转移,即 $q_{c_i c_{i+1}} = 0$,且 $h_{c_i c_{i+1}} \geq \rho$,则判定 $s_i = 2$,表明在 k_i 与 k_{i+1} 之间存在一个词汇边界。

(3)若 k_i 与 k_{i+1} 不存在词内转移概率,但存在弱词间转移,即 $q_{c_i c_{i+1}} = 0$,且 $h_{c_i c_{i+1}} < \rho$,此时分几种情况来考虑。

1)由于关键动作的提取与分类有可能存在漏检或者错误的情况,因此首先在基元的词内转移概率矩阵 Q 中观察是否存在某个 $x (1 \leq x \leq N)$,使得 $q_{c_i x} = 1$,且 $q_{x c_{i+1}} = 1$,同时在词汇模板中也确实存在由 k_i, k_x 以及 k_{i+1} 这 3 个基元的类别标签组成的子串。如果存在,说明在基元 k_i 与 k_{i+1} 之间漏掉了一个基元 k_x ,补充上 k_x 后, k_i 与 k_{i+1} 依然属于同一个词汇内部,此时 s_i 取值为 1。

2)若不存在这样的 x ,那么观察是否存在某个 $y (1 \leq y \leq N)$,使得 $q_{c_i y} = 1$,且 $h_{y c_{i+1}} \geq \rho$ 。若存在,则需要在 k_i 与 k_{i+1} 之间补充 k_y , k_y 与 k_{i+1} 之间存在词汇边界, s_i 取值为 3。

3)若找不到这样的 y ,那么继续观察是否存在 $z (1 \leq z \leq N)$,使得 $h_{c_i z} \geq \rho$,且 $q_{z c_{i+1}} = 1$ 。若存在,则需要在 k_i 与 k_{i+1} 之间补充 k_z , k_i 与 k_z 之间存在词汇边界, s_i 取值为 4。

4)若上述 3 种情况均不满足,则判定 $s_i = 2$,表明在 k_i 与 k_{i+1} 之间存在一个词汇边界,但该边界是弱边界。

(4)若 k_i 与 k_{i+1} 不存在词内转移概率,也不存在词间转移概率,即 $q_{c_i c_{i+1}} = 0$,且 $h_{c_i c_{i+1}} = 0$,则首先按照上述的 1), 2), 3)种情况来判定是否存在漏掉的某个基元,若均不满足,则认为 k_{i+1} 有可能是一个错误的基元,将其从原始序列中去掉,后序的基元序号与类别标签序号向前依次递减,继续判断 k_i 与新的 k_{i+1} 之间的关系。

(5)若两个连续的标记 s_{i-1} 和 s_i 均小于 2,说明 k_{i-1}, k_i 以及 k_{i+1} 均属于同一个词汇,此时回溯一下,判断 k_{i-1} 到 k_{i+1} 的类别标签组合是否是词汇模板中的子串。如果是,则继续向下进行,如果不是,则说明 k_i 与 k_{i+1} 之间应该存在边界,但是存在漏掉的基元,应该按照上述 2), 3), 4)种情况来重新进行判定,若均不满足,则说明 k_{i+1} 是一个错误基元,可以从序列中去掉。

(6)利用上述步骤,得到标记序列 $S = \{s_1, s_2, \dots, s_{L'}\}$,其中 L' 表示去除所有错误的基元后基元序列的长度。显然,在 $s_i \geq 2$ 的地方均存在词汇边界。

(7)根据词汇边界,先将基元序列分割成多个组,再根据标记序列补充各分组缺失的基元。此时,每个分组内基元的类别标签组成一个字符串,根据我们先前创建的词汇模板,找到所有包含该字符串的词汇,将其作为该分组的候选词汇。对所有的基元分组均做此处理。最后根据矩阵 B ,将不同基元分组的候选词汇间的转移概率累乘,计算出对应合成句子的概率,并按照最大概率原则输出手语句子的最终识别结果。

在实际处理过程中,为了确保能输出一个有意义的语句,上述过程先从 k_1 开始向后处理一遍,然后再从 k_L 开始向前处理一遍,基于两个标记序列,按照最大概率准则输出最终结果。

3 实验结果与分析

Kinect 为我们录制手语视频的工具。硬件环境为:CPU: Inter(R)Core(TM)i7-4790 CPU@3.60GHz 3.60GHz; 安装内存(RAM):8.00GB。操作系统为 Windows 10 旗舰版 64 位,在 Microsoft Visual Studio 2010 中选用 C# 作为开发语言进行实验。

我们邀请了 3 位手语者,录制了 47 个手语单词作为样本,那么理论上由这 47 个单词所构成的连续语句均可以做出识别。针对这 47 个词汇,一共包含 65 个不同的关键动作,创建数据模型——词汇模板、 47×47 的词间转移概率矩阵 B 、 65×65 的基元词内转移矩阵概率 Q ,以及 65×65 的基元词间概率转移矩阵 H 。通过对样本的学习和训练,可以实现对关键动作即手语基元的提取与分类识别。

假定现在以手语者做出的语句“我们在礼堂见面”为例,对其提取基元并识别后^[13-14],得到一个包含 8 个基元的序列 $\{k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8\}$,这 8 个基元对应的类别标签序列是 $\{c_1 = 59, c_2 = 40, c_3 = 39, c_4 = 51, c_5 = 30, c_6 = 21, c_7 = 22, c_8 = 23\}$ 。但实际上,该句子包含 4 个词汇,这 4 个词汇包含的基元应该是如表 1 所列的情况,也就是 k_3 是一个错误的基元,其类别标签应该是 $c_3 = 65$,但实际的识别结果是 $c_3 = 39$;然后在 k_5 和 k_6 之间漏检了一个基元,也就是类别标签为 31 的基元。由此可见,实际的基元提取与识别存在一定的误差,但是本文设计的算法依然可以做出识别。

表 1 词汇与词汇基元

词汇	包含的关键动作的类别标签
我们	["59" "40" "65"]
在	["51"]
礼堂	["30" "31" "21"]
见面	["22" "23"]

首先,创建标记序列 $\{s_1, s_2, \dots, s_8\}$,所有的 s_i 初始值均为 0,获取标记序列值的过程为:

(1)从 k_1 开始, k_1 与 k_2 的 $q_{59,40} = 1$,因此 $s_1 = 0$ 。

(2)判断 k_2 与 k_3 的关系,由于 $q_{40,39} = 0$,且 $h_{40,39} = 0$,因此首先按照 2.2 节中的步骤(4)判断是否存在漏检的情况,由于找不到任何满足条件、可以补充的 k_x, k_y 或者 k_z ,因此判定 k_3 是一个错误识别的基元,将其从原始序列中去掉。当前的基元序列变为 $\{k_1, k_2, k_3, k_4, k_5, k_6, k_7\}$,类别标签序列变为 $\{c_1 = 59, c_2 = 40, c_3 = 51, c_4 = 30, c_5 = 21, c_6 = 22, c_7 = 23\}$,标记序列变为 $\{s_1, s_2, \dots, s_7\}$ 。

(3)继续判断当前 k_2 与 k_3 的关系,由于 $q_{40,51} = 0$,且 $h_{40,51} = 0$,经过判定,可以找到一个 $k_y (y = 65)$,使得 $q_{40,65} = 1$,且 $h_{65,51} \geq \rho$ 。因此,在 k_2 与 k_3 之间需补充 $k_y, s_2 = 3$ 。

(4)对于 k_3 与 $k_4, q_{51,30} = 0$,且 $h_{51,30} \geq \rho$,所以 k_3 与 k_4 之间存在边界, $s_3 = 2$ 。

(5)对于 k_4 与 $k_5, q_{30,21} = 0$,且 $h_{30,21} = 0$,可以检测到存在 $k_x (x = 31)$,使得 $q_{30,31} = 1$,且 $q_{31,21} = 1$ 。因此,在 k_4 与 k_5 之间须补充 k_x, k_4 与 k_5 属于同一个词汇内部,因此 $s_4 = 1$ 。

(6)按照上述步骤,依次可以得到 $s_5 = 2, s_6 = 0$ 。由于 k_7 已经是最后一个基元,因此直接设定最后一个标记 $s_7 = 0$ 。最终得到标记序列 $\{0, 3, 2, 1, 2, 0, 0\}$,由此可以看出基元序列中包含 3 个边界。

(7)根据词汇边界得到基元分组 $G_1 = \{“59”“40”\}, G_2 = \{“51”\}, G_3 = \{“30”“21”\}, G_4 = \{“22”“23”\}$ 。再根据标记序列补充各分组缺失的基元,得到最终的基元分组 $G_1 = \{“59”“40”“65”\}, G_2 = \{“51”\}, G_3 = \{“30”“31”“21”\}, G_4 = \{“22”“23”\}$ 。然后逐一确定各基元分组的候选词汇,并将不同分组的候选词汇间的转移概率累乘,得出对应合成句子的概率,并按照最大概率原则得出识别结果就是“我们在礼堂见面”。对原始基元序列做反向处理,得到相同的输出。

由此可见,基于关键动作的双重转移概率算法在关键动作出现漏检或误识的情况下,依然能够找到正确的手语边界,并且补充相应组内缺失的关键动作,确定边界之间可能存在的手语单词,进而得到连续手语正确的识别结果。此外,本文的算法并不针对特定人群,只要能够检测和识别出关键动作,即可采用本文的算法识别连续手语语句。因此,该算法可以面向非特定人群。

为了验证本文算法的有效性,我们与文献[10]以及文献[6]的算法做了比较。对用这 47 个手语词汇组成的连续语句进行识别,实验结果如表 2 所列。手语者 1 是熟练的手语者,手语者 2 是经过训练、比较熟练的手语者,手语者 3 是经过现场培训、非熟练手语者。其中,文献[10]也是基于关键动作的,它先根据关键动作对手语词语义的贡献大小赋予权值,然后基于加权的键动作识别连续手语。由实验结果可以看出,该算法执行效率较高,但是它比较依赖权值大的关键动作,虽然权值小的关键动作漏检或者错误对识别结果几乎没有影响,但是一旦有大权值的关键动作发生漏检或错误的现象,该算法就无法给出正确的识别结果。因此,它的性能并不稳定,尤其对于非熟练的手语者,可能由于动作不够规范,导致大权值关键动作的错误概率增高,从而影响语句的识别效果。文献[6]提出了一种混合 CNN-HMM 架构,通过最大化连续手语对应的词汇序列的后验概率,并结合 N-Gram 模型、一阶隐马尔可夫过程及维特比算法,得出连续手语的最优词汇序列作为识别结果。但由于该算法识别原理复杂,还需要通过网格搜索来确定 3 个超参数,虽然引入了剪枝操作,但运行时间还是较长,难以满足实际应用的需要。此外,CNN 分类时仅输入右手图像,在一定程度上影响了识别准确率。对比之下,本文算法具有较优的识别率以及更高的容错性,能够实现连续手语的实时识别。

表 2 连续手语识别的准确率和平均运行时间

算法	手语者 1 识别 准确率/%	手语者 2 识别 准确率/%	手语者 3 识别 准确率/%	平均运行 时间/s
本文算法	98.85	92.23	90.95	2.5620
文献[10]算法	97.45	88.38	68.78	2.0375
文献[6]算法	92.75	88.12	82.85	13.5822

结束语 针对当前连续手语识别算法中存在的问题,本文提出了一种基于关键动作双重转移概率的连续手语语句识别算法。利用该算法可以较为快速、准确地检测出连续手语关键动作序列中的单词边界,将对连续手语的识别转化为对若干手语词汇的识别,简化了识别过程的同时提高了识别的速度。实验证明了本文算法具有更好的稳定性以及容错性。

参考文献

- [1] STARNER T, PENTLAND A. Visual Recognition of American Sign Language Using Hidden Markov Models[C]//International Workshop on Automatic Face & Gesture Recognition. 1995: 189-194.
- [2] KONG W W, RANGANATH S. Automatic Hand Trajectory Segmentation and Phoneme Transcription for Sign Language [C]//IEEE International Conference on Automatic Face & Gesture Recognition. IEEE, 2008: 1-6.
- [3] KONG W W, RANGANATH S. Towards Subject Independent Continuous Sign Language Recognition [J]. Pattern Recognition, 2014, 47(3): 1294-1308.
- [4] YANG H D, SCLAROFF S, LEE S W. Sign Language Spotting with a Threshold Model Based on Conditional Random Fields [M]. IEEE Computer Society, 2009.
- [5] YANG R, SARKAR S, LOEDING B. Handling Movement E-pentthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(3): 462-477.
- [6] KOLLER O, ZARGARAN S, NEY H, et al. Deep Sign: Enabling Robust Statistical Continuous Sign Language Recognition via Hybrid CNN-HMMs [J]. International Journal of Computer Vision, 2018, 126(12): 1311-1325.
- [7] 方高林, 高文, 陈熙霖, 等. 基于 SRN/HMM 的非特定人连续手语识别系统 [J]. 软件学报, 2002, 13(11): 2169-2175.
- [8] FANG G, GAO W, ZHAO D. Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models [J]. IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans, 2007, 37(1): 1-9.
- [9] YANG W, TAO J, YE Z. Continuous sign language recognition using level building based on fast hidden Markov model [J]. Pattern Recognition Letters, 2016, 78(C): 28-35.
- [10] XU X X, HUANG Y Y, HU Z J. Research on Continuous Sign Language Sentence Recognition Algorithm Based on Weighted Key-Frame [J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2018, 22(4): 483-490.
- [11] HUANG J, ZHOU W, ZHANG Q, et al. Video-based Sign Language Recognition without Temporal Segmentation [C]//32nd AAAI Conference on Artificial Intelligence. 2018: 2257-2264
- [12] 中国残疾人联合会教育就业部, 中国聋人协会. 中国手语 [M]. 北京: 华夏出版社, 2003.
- [13] LI S R, HUANG Y Y, HU Z J, et al. Key Frame Detection Algorithm based on Dynamic Sign Language Video for the Non Specific Population [J]. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2015, 8(12): 135-148.
- [14] SHI M M, HUANG Y Y, HU Z J. Dynamic Sign Language Recognition Algorithm Using Weighted Gesture Units [J]. Journal of Information and Computational Science, 2015, 12(15): 5611-5621.
- [15] LIANG W L, HUANG Y Y, HU Z J. Real-Time Dynamic Sign Language Recognition Based on Hierarchical Matching Strategy [J]. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2017, 10(7): 21-34.