

# 基于虚拟人合成技术的中国手语合成方法\*

王兆其, 高文

(中国科学院 计算技术研究所, 北京 100080)

E-mail: {zqwang, wgao}@ict.ac.cn

http://www.ict.ac.cn

**摘要:** 介绍了一种中国手语合成方法, 实现了文本到中国手语的自动翻译, 并使用虚拟人合成技术, 实现了中国手语的合成与显示, 以此帮助聋人与听力正常人之间实现自然交流. 在该方法中, 首先应用两只数据手套和 3 个 6 自由度位置跟踪器, 基于运动跟踪的原理, 记录真实人体演示每个手语词的运动数据, 建立一个初始的手语词运动数据库. 然后, 应用一种基于控制点的人体运动编辑方法, 对每个手语词的运动数据进行编辑与微调, 最后得到一个高质量的手语词运动数据库. 当给定一个文本句子时, 应用人体运动合成方法, 对每个手语词的手语运动片段进行拼接合成, 最终生成一个完整的手语运动, 并基于 VRML 的人体运动显示方法将合成的运动逼真地显示出来. 基于该方法, 在 PC/Windows/VC6.0 环境下实现了一个中国聋人手语合成系统. 该系统采集了《中国手语》(含续集)中收录的 5 596 个手语词, 可以合成一般生活与教学用语. 经聋校的老师和学生确认, 合成手语准确逼真, 可以广泛应用于教学、电视、Internet 等多种大众媒体, 帮助聋人参与其他听力正常人的活动, 具有广泛的应用前景和重要的社会意义.

**关键词:** 中国手语; 机器翻译; 虚拟人

中图法分类号: TP31 文献标识码: A

手语是一种人体运动语言, 通过手的运动表达意思, 主要用于聋人之间的相互交流. 据统计, 中国目前患听力语言障碍的聋人约有 2 400 万, 这些聋人之间及他们与其他人之间的交流主要通过中国手语进行. 我国政府为关心聋人, 促进他们之间的交流, 早在 20 世纪 50 年代后期, 就开始了中国聋人手语的规范化工作. 到 1994 年, 曾先后出版了《聋人通用手语草图》(4 辑, 2 000 余个手语草图)、《中国手语》(3 330 个词条)、《中国手语》续集(2 266 个词条).

手语合成的目的就是自然语言文本句子自动翻译成聋人手语, 并由计算机生成的虚拟人逼真地显示出来, 是智能人机接口的一个重要研究课题. 合成手语不仅有助于聋人使用各种信息, 而且可以帮助聋人参加各项社会活动(如观看电视), 具有十分重要的应用价值和社会意义.

手语合成的研究始于 1982 年<sup>[3]</sup>, Shantz 和 Poizner 实现了一个合成美国手语的计算机程序. 之后, 中国、美国<sup>[6]</sup>、日本<sup>[7]</sup>、德国<sup>[8]</sup>等许多国家都进行了自己国家的手语合成研究, 并取得了许多重要的研究成果. 本文主要介绍我们对聋人手语进行合成的主要研究方法和研究成果.

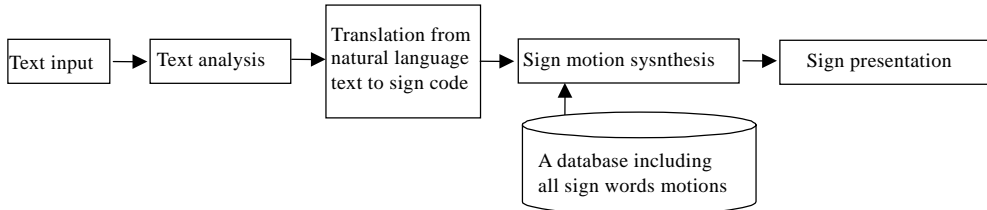
## 1 中国手语及其合成研究

一般来说, 手语合成的系统结构如图 1 所示.

\* 收稿日期: 2001-01-17; 修改日期: 2001-04-03

基金项目: 国家自然科学基金资助项目(60103007; 69789301); 国家高技术研究发展计划资助项目(2001AA115131); 中国科学院知识创新工程资助项目

作者简介: 王兆其(1966 - ), 男, 湖南安化人, 博士, 研究员, 博士生导师, 主要研究领域为虚拟现实, 人工智能; 高文(1956 - ), 男, 辽宁大连人, 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.



输入文本, 文本分析, 自然语言到手语码转换, 手语运动合成, 显示, 手语库.

Fig.1 Process of Chinese sign language synthesis

图1 中国手语合成过程

文本输入后,经文本分析,应用自然语言处理的理论与方法将自然语言句子转换成没有歧义的文本句子,再经自然语言到手语码的转换,将文本句子转换成为表示手语的符号序列.序列中的每个单元或者是一个独立的手语词,或者是一个表示手语运动的状态参数,如语气、语调.这样,一个手语符号序列经过手语运动合成,将每个手语符号转换成对应的手语运动(手的关节运动数据),将词与词之间的运动进行平滑的衔接,并根据语气语调等参数对运动节奏进行适当调整,最后生成完整的手语运动,由手语显示模块用三维虚拟人显示出来.本文主要介绍我们建立手语库的方法、基于手语词的手语运动合成方法及三维手语显示方法.

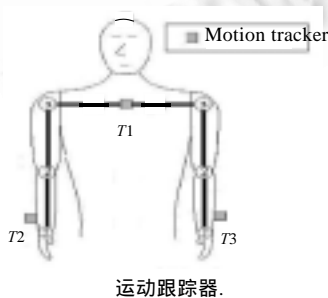
## 2 手语词库的建立

手语词库记录了每个手语词的手语运动信息,是手语合成的基础.建立手语词库不仅工作量大,而且其质量也直接影响到合成到手语的结果.

目前建立手语词库的方法有两种:运动跟踪方法<sup>[10,11]</sup>和手工编辑方法<sup>[6]</sup>.运动跟踪方法主要使用数据手套与传感器记录每个手语词对应的手语运动.这种方法的优点是方便、快捷,可以减少手语库建立的工作量,因此可以建立更加细致的手语库.另外,用这种方法录制的手语包含了丰富的时间信息,使得手的运动变化更加逼真.这种方法的缺点是易受干扰,使得录制的一些手形不准确,而且用常规的方法不易修改<sup>[12]</sup>.手工编辑方法是人为地确定一个手语词的手势个数,并使用软件工具,确定每个手势各关节的角度以及手势之间的变化时间.这种方法的优点是生成的手形准确,但用这种方法建立手语词的工作量太大,而且由人工确定手语词中手势变化的快慢,因此手语运动不够逼真.

结合上述两种方法,我们首先使用运动跟踪的方法,记录每个手语词的真实人体运动数据,建立一个初始的手语词库.然后针对运动跟踪方法记录的运动数据难以修改的问题,提出了一种基于控制点的手语运动编辑方法,对那些不准确的手语运动进行编辑修改.这种方法不仅可以获得一个准确、逼真的手语词库,而且可以减少工作量.

### 2.1 基于数据手套与传感器的手语运动采集



运动跟踪器.

Fig.2 Arms model and sensors placement

图2 虚拟人的上肢模型与传感器的安放方案

我们实现了一种快速的人体上肢运动跟踪方法.这种方法采用了基于关节角度的人体上肢模型,使用了两只CyberGlove数据手套和3个6自由度位置跟踪器(位置跟踪器的安放方法如图2所示),可以实时跟踪记录双手的运动<sup>[12]</sup>.基于此原理,我们记录了《中国手语》中收录的全部5596个手势语和30个手指语,建立了一个初始的手语词运动库.

这样一个初始的手语词运动库含有逼真的手语运动信息,但由于运动跟踪方法容易受金属物及其他电磁场的干扰,因此,有一些录制的手语运动数据不太准确,需要进一步编辑和修改.

## 2.2 基于控制点的手语运动编辑方法

为了编辑手的不准确的手语词运动,本节给出了一个基于控制点的手语运动编辑方法.这种方法充分利用了手语运动连续帧之间的运动相关性,即在手势运动数据中,相邻两个手势之间差别不大,只需手工修改手语运动的某几个关键帧就可以了,并将这些手工修改的帧作为控制点,根据连续帧之间的运动相关性,计算出对其他帧的修改.该方法不仅减少了工作量,而且使得修改后的手语运动仍然保持连续帧之间的运动相关性.

### 2.2.1 手语运动及关节自由度运动曲线

我们在表示手语运动时,每只手用 27 个自由度表示,两只手共有 54 个自由度.因此,一个手势可由一个 54 元向量表示,一个手语运动可以由一个从时间到手势集合的向量函数表示:

$$G(t): R \rightarrow R^{54}. \quad (1)$$

对于单个自由度,我们用  $G_i(t)$  表示,它所对应的函数曲线称为自由度运动曲线.根据手语运动中帧与帧之间的运动相关性可以知道,自由度运动曲线是连续曲线.如图 3 所示是手语词“大家”的手语运动中,右手肩关节  $X, Y, Z$  方向的 3 条自由度运动曲线.

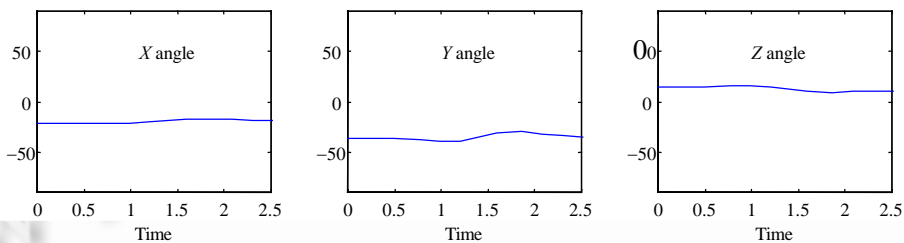


Fig.3 Curves of each DOF's of Chinese sign word "All of you"

图 3 中国手语词“大家”的自由度运动曲线

### 2.2.2 控制点及其对运动曲线的修改

对手语运动的编辑,实际上就是对手语运动的自由度运动曲线进行编辑修改.当曲线上某一点被修改时,曲线附近的区域也会受到影响,只有这样,曲线的连续性(即手语运动中连续帧之间的相关性)才可以得到保证,而且不需要修改曲线上的每一点(帧),从而可以减少编辑手语运动的工作量.

在编辑一个手语词的手语运动时,如果对某一帧进行了修改,则我们称这一帧为一个控制点.每个控制点不仅修改了所在帧的手形,而且这种修改还会影响到与它相邻帧的手的状态,每个控制点的影响区域是它的前一个控制点与它的下一个控制点之间的所有帧.两个控制点之间的中间帧受这两个控制点的共同影响.

假设  $C_1$  和  $C_2$  是两个相邻的控制点,分别是对  $f_1$  和  $f_2$  两帧的修改.对于自由度曲线  $G_i(t)$ ,设  $C_1$  对  $f_1$  帧的修改量为  $\Delta d_1$ ,  $C_2$  对  $f_2$  帧的修改量为  $\Delta d_2$ .那么,对于  $f_1$  与  $f_2$  之间的每一帧  $f'$ ,其自由度  $G_i$  的修改量可以按以下线性插值方法得到,式中  $t_{f'}$ ,  $t_{f_1}$  和  $t_{f_2}$  分别表示  $f'$ ,  $f_1$  和  $f_2$  对应的时刻.

$$\Delta d_{f'} = \Delta d_1 + \frac{t_{f'} - t_{f_1}}{t_{f_2} - t_{f_1}} (\Delta d_2 - \Delta d_1). \quad (2)$$

图 4 显示了手语词“客观”的手语运动修改前后左手肘关节的  $X, Y$  方向自由度运动曲线.不带点的曲线是修改前的运动曲线,带点的曲线为修改后的运动曲线.对这样一个运动只需手工修改其中的 3 帧,其他各帧都将自动完成修改,修改后的运动曲线仍然是连续的运动曲线.图 5 列出了手语词“客观”修改前后的手语运动.其中上图表示修改前的手语运动“客观”,下图表示修改后的手语运动“客观”.

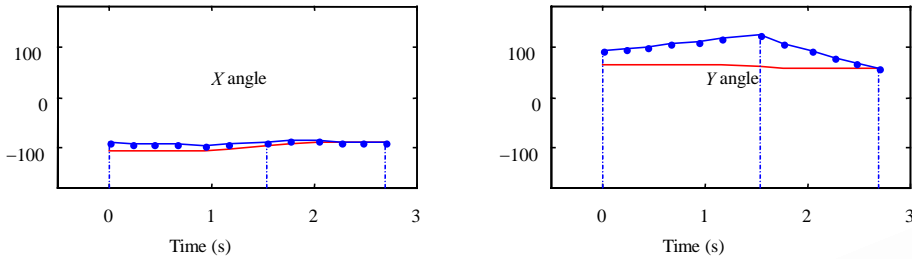


Fig.4 Modifies to sign word “fact” using control points

图 4 控制点对手语词“客观”的手语运动的修改



Fig.5 Modifies to motion of sign words “fact” using control points

图 5 基于控制点方法对手语运动“客观”的修改

### 3 手语运动合成方法

在经自然语言处理和手语自动翻译后的手语符号序列中,只包含两种单元:手语词和手语运动参数.手语运动合成的目的是将这样的符号序列转换成手语运动.

手语符号序列中的每个手语词的手语运动数据,可以直接从手语词运动数据库中查询得到,因为它们是用运动跟踪方法采集得到,并基于控制点的方法编辑而成的,因此,其内部具有很好的运动逼真性和连续帧之间的运动相关性.如果每个关节的旋转变用 Euler 角表示,每个手语词内部帧之间直接对各个 Euler 角进行插值,即可以很好地满足各种不同播放速度和语气语调变化的要求.

但手语词之间的平滑过渡则要复杂些,因为两个连续的手语词之间,前一个手语词的最后一个手势(最后一帧)与下一个手语词的第 1 个手势(第 1 帧)之间一般不存在运动相关性.它们之间的差别可以是任意的.为了实现这种不连续帧之间的平滑过渡,我们对一些比较复杂的关节(如肩关节)采用基于四元组的运动插值方法<sup>[13]</sup>.图6 列出了合成手语“大家好”从手语词“大家”的最后一帧到手语词“好”的第 1 帧之间的平滑过渡.



(a) The first frame of the CSL word “all of you”  
(a) “大家”的最后一帧

(b) All generated frames in between  
(b) 平滑过渡

(c) The first frame of the CSL word “are great”  
(c) “好”的第 1 帧

Fig.6 Motion concatenation between two sign words “all of you” and “are great”

图 6 从“大家”到“好”的平滑过渡

当然,更有效的平滑过渡方法是基于人体运动学与动力学的仿真方法,并实现碰撞检测与基于物理特性的碰撞响应<sup>[14]</sup>.但目前这些方法难以满足合成速度的要求.

#### 4 基于 VRML 的虚拟人手语显示

手语是一种可视语言,合成的手语需要用图形或图像的方法显示出来,观察者才能“读”取手语的信息与意义.为了方便手语信息的传递,我们采用了一种基于 VRML(virtual reality modeling language)的手语显示方法,这种方法采用了 VRML 的人体表示模型.因此,适合于在 Internet 上表示的手语信息.

VRML 是一种虚拟现实建模语言,主要用于在 Internet 上表示并传递三维虚拟场景.VRML 有一部分是专门用于描述三维人体模型的,这个标准又称为 H-Anim.按照该标准表示的人体模型可以在国际互联网范围内交互共享.我们实现了一个 H-Anim 虚拟人体模型阅读器,可以将符合该标准的三维人体模型读入,并使用此模型显示手语运动.图7 是使用 3 个不同的虚拟人体模型显示的手语“好”.

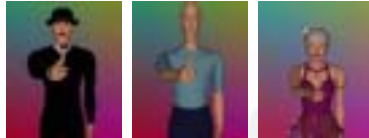


Fig.7 Present the same gesture “are great” by different models  
图 7 用不同虚拟人显示的手语“好”

#### 5 总 结

我们基于运动跟踪的方法,在 PC/Windows/VC6.0 环境下开发了一个人体上肢运动采集程序.该程序以每秒 15 帧的速度实时地采集人体运动.我们用该系统采集了《中国手语》(含续集)中收录的 5 500 个手语词,然后在一个支持基于控制点编辑方法的手语编辑系统中对这些手语词进行了编辑.经聋校的老师和学生确认,绝大部分手语词准确逼真.

在 5 500 个手语词的基础上,适当使用同义词替换,并结合使用手指语,我们开发了一个手语合成系统,可以合成一般的教学与日常生活用语.图8 显示了“大家好”对应的合成手语运动.



Fig.8 Generated Chinese sign language of sentence “all of you are great”  
图 8 “大家好”的合成手语运动

手语是一种人体运动语言,手语合成的研究是智能人机接口的研究内容,可以促进聋人之间的交流,并且可以广泛应用于教学、电视、Internet 等多种大众媒体,帮助聋人参与各种正常人活动,具有广泛的应用前景和重要的社会意义.本文介绍了我们对中国聋人手语合成的研究成果,其中的主要理论方法与技术不仅适用于中国聋人手语的合成,同样适合其他语种的手语合成,具有很好的通用性.在今后的工作中,我们将进一步研究手语合成中的唇动与表情的合成、语气语调的合成,使合成手语真正实用化.

#### References:

- [1] Chinese Deaf Associate. Chinese Sign Language. Beijing: Hua Xia Press, 1988 (in Chinese).
- [2] Chinese Deaf Associate. Chinese Sign Language. 2nd ed., Beijing: Hua Xia Press, 1988 (in Chinese).
- [3] Shantz, M., Poizner, H. A computer program to synthesize American sign language. Behavior Research Methods and Instrumentation, 1982,14(5):467~474.
- [4] Gao, Wen. The framework of multimodal human-computer interface. In: Wu, Quan-yuan, Gao, Wen, eds. Proceedings of the 2nd Conference of Computer Intelligent Interface and Applications of China. Beijing: Tsinghua University Press, 1995. 17~20 (in Chinese).
- [5] Shu, Hong-xian, Wang, Shen-kang, Zhang, Gao-qing. Study and implementation of hypermedia computer-aided educational system. In: Wu, Quan-yuan, Gao, Wen, eds. Proceedings of the 2nd Conference of Computer Intelligent Interface and Applications of China. Beijing: Tsinghua University Press, 1995. 407~412 (in Chinese).
- [6] Noriko, T., et al. An alternative method for building a database for American sign language. In: Proceedings of the Technology for Persons with Disabilities Conference 2000. Los Angeles, CA, 2000. <http://www.csun.edu/cod/conf/2000/proceedings/0071Tomuro.htm>.

- [7] Lee, J. Using the 3D hand model to recognize stationary hand signs. *International Journal of Modeling and Simulation*, 1999,19(1): 24~32.
- [8] Frank, G., Thomas, S. Modeling and generating sign language as animated line drawings. In: Crawford, D., ed. *Proceedings of the International ACM Conference on Assistive Technologies*. New York, NY: ACM Press, 1998. 78~84.
- [9] Xu, Lin. Study on synthesis of Chinese sign language [Ph. D. Thesis]. Harbin: Harbin Institute of Technology, 2000 (in Chinese).
- [10] Erenshsteyn, R., Laskov, P., Foulds, R., *et al.* Recognition approach to gesture language understanding. In: Kavanagh, M.E., Werner, B., eds. *Proceedings of the 13th International Conference on Pattern Recognition*. Los Alamitos: IEEE Computer Society, 1996. 431~435.
- [11] Fels, S., Hinton, G. Glove-TalkII: a neural network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions on Neural Networks*, 1998,9(1):205~212.
- [12] Wang, Zhao-Qi, Gao, Wen, Xu, Yan. A sensor-based method to capture motion of two arms. *Chinese Journal of Computers*, 2001, 24(6):616~619 (in Chinese).
- [13] Ken, S. Animating rotation with quaternion curves. *ACM Computer Graphics*, 1985,19(3):245~254.
- [14] Zordan, V.B., Hodgins, J.K. tracking and modifying upper-body human motion data with dynamic simulation. In: Thalmann, N.M., Thalmann, D., eds. *Eurographics Workshop on Animation and Simulation'99*. Wien: Springer-Verlag, 1999. 13~22

#### 附中文参考文献:

- [1] 中国聋人协会. 中国手语. 北京: 华夏出版社, 1988.
- [2] 中国聋人协会. 中国手语续集. 北京: 华夏出版社, 1988.
- [4] 高文. 多功能感知机的框架结构. 见: 吴泉源, 编. 第 2 届中国计算机智能接口与智能应用学术会议论文集. 北京: 清华大学出版社, 1995. 17~20.
- [5] 舒宏先, 王申康, 章高清. 超媒体手语计算机辅助教学系统的研究与实现. 见: 吴泉源, 编. 第 2 届中国计算机智能接口与智能应用学术会议论文集. 北京: 清华大学出版社, 1995. 407~412.
- [12] 王兆其, 高文, 徐燕. 一种基于传感器的人体上肢运动实时跟踪方法. *计算机学报*, 2001, 24(6): 616~619.

## A Method to Synthesize Chinese Sign Language Based on Virtual Human Technologies\*

WANG Zhao-qi, GAO Wen

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

E-mail: {zqwang;wgao}@ict.ac.cn

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

**Abstract:** A method to translate Chinese text into Chinese sign language is introduced, and how to present gesture by a 3D virtual human, to help the deaf communicate with the hearing. In this method, an initiate database including all sign words motions is firstly created based on a motion capture method, where two data gloves and three 6-DOF position tractors are used to record the motion of a real person when he or she demonstrates each sign word. Then all motion data can be tuned accurately and easily by modifying the gestures just on some control points. After that, a database that contains motions of all sign words has been built up. Given a sentence of text, the motion of each word can be found in the database, concatenated one by one, and presented smoothly by a virtual presenter who is represented by a VRML virtual human model, then the sentence is translated into Chinese sign language. To help the deaf study, watch TV, and browse Internet, a system that can be used to translate Chinese text into Chinese sign language is developed in the PC/Windows/VC environment, and all most of generating signs by this system are very realistic and smooth as confirmed by a lot of teachers in deaf schools. The database of this system contains motions of all 5 596 Chinese sign words collected by a handbook of Chinese sign language.

**Key words:** Chinese sign language; machine translation; virtual human

\* Received January 17, 2001; accepted April 3, 2001

Supported by the National Natural Science Foundation of China under Grant Nos.60103007, 69789301; the National High-Tech Research and Development Plan of China under Grant No.2001AA115131; the Knowledge Innovation Program of the Chinese Academy of Science