

# Study of Skeleton Sign Language Based on the Hand Shape

Huaqiang Jiang, Jie Xiang

Institute of Service Engineering, Hangzhou Normal University, Hangzhou  
Email: jhq@hznu.edu.cn

Received: Apr. 12<sup>th</sup>, 2013; revised: May 8<sup>th</sup>, 2013; accepted: May 12<sup>th</sup>, 2013

Copyright © 2013 Huaqiang Jiang, Jie Xiang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract:** Building of sign language thesaurus is the basis of sign language recognition and synthesis. Through analyzing the characteristics of Chinese sign language, this paper proposes a human skeleton model which is based on expansion Kinect, and built skeleton sign language thesaurus with a combination of hand shape. It achieves an editing system of Chinese skeleton sign language by acquisition of the key state in sign language expression, then recording data of hand shapes, positions and orientations, and forming key frame data of sign language at last. It effectively improves synthesis efficiency for sign language. Data generates from which can make dynamic combination through the key frame interpolation algorithm, and lays the foundation for the recognition and synthesis of sign language thesaurus.

**Keywords:** Sign Language Library; Kinect; Editing System of Skeleton Sign Language

## 基于手形的骨骼手语库构建研究

姜华强, 项洁

杭州师范大学国际服务工程学院, 杭州  
Email: jhq@hznu.edu.cn

收稿日期: 2013年4月12日; 修回日期: 2013年5月8日; 录用日期: 2013年5月12日

**摘要:** 手语词库的构建是手语识别与合成的基础。本文通过分析中国手语的特点, 提出了一种基于扩展 Kinect 人体骨骼模型, 以手形组合构建骨骼手语库的方法。通过采集手语表达过程中的关键状态, 记录手形、位置和方向等数据, 形成手语词关键帧数据, 实现了一个中文骨骼手语编辑系统, 有效的提高手语词汇的合成效率。生成的数据通过关键帧插值算法可实现手语词的动态组合, 为基于骨骼手语库的手语识别及合成奠定了基础。

**关键词:** 手语库; Kinect; 骨骼手语编辑系统

### 1. 引言

手语(手势语)作为自然而直观的人际交流方式自然、直观、易于学习, 更是聋人进行信息交流的最自然的方式和工具。近十几年来, 国内外手语合成系统进行了广泛的研究。2003年由高文领导开发的中国手语合成系统, 运用运动捕获技术, 实现了对 5119 个手势词的编码。系统采用数据手套和 3D 跟踪器采集

手形、位置和方向, 最后进行人工编辑调整, 形成手语库<sup>[1]</sup>。这样形成的手语库成本高, 工作量大, 并且大量的数据也对后期手语合成造成了速度方面的影响。

骨架手语库构建研究, 主要是从减少手语库构建成本, 提高手语合成效率角度, 通过研究和分析手模型, 以 Kinect 人体骨骼模型为基础, 扩充手骨骼模型, 然后, 通过分析《中国手语》词典当中的手语图片, 形成骨骼手语库。

\*基金项目: 本课题研究得到了浙江省教育厅科研项目(编号: Y200803197)的资助支持。

## 2. 中国手语词汇中的基本手形特点

手语是聋人的语言，世界各地的聋人手语都有所不同，且存在较大的差异。本研究的是《中国手语》书为代表的大陆地区聋人群体的通用手语。手势就相当于一般语言的词。从语言学上看，词是句子中最小的、独立而有意义的单位<sup>[2]</sup>。在手语中，手语词即手势是最小的、有意义的单位<sup>[3]</sup>。但是，《中国手语》书当中定义的手语词达到 5200 多个，而且还有扩展趋势。这给开发连续手语识别系统和手语合成系统带来了较大的难度。

通过手语语言学的研究可知，手势是由手形、位置、动作和方向共同构成。类似于语言发音一样，舌头通过变换不同的形状和位置来控制气流的进出，从而形成不同的语音。手在手语“发音”中，扮演着“舌形”的角色。

根据手语语言学家对手形研究，各国手语基本只有有限数量的手形存在。比如：美国手语语言学家斯多基最早确定了美国手语有 19 种类型，英国学者借鉴斯多基的理论将英国手语分为 23 个手形。美国的史文汉(Wayne Smith, 1979)借鉴了斯多基的符号系统列举了台湾手语的 51 个手形。国内学者杨军辉曾统计中国手语有 60 多个基本手形。

顾定倩(2005)等人通过对中国手语的动作、打法的分析归纳出中国手语的基本手形，从而发现中国手语在手形构造上的特点和规律<sup>[3]</sup>。手语词包括手指数量、手指组合和手指形态三个结构特征。

无论何种手形，都是由手指构成的，但手形在表达时总是有选择地展现各个手指，有的手形只使用一个手指，有的使用两个、三个、四个或者五个全部使用。使用手指数量不同，手形就不同。为此，可将手形归纳为一指手形、二指手形、三指手形、四指手形和五指手形等类型。在手指数量相同的情况下，手指组合成为区分手形的主要特征。手指形态包括手指间关系和手指弯曲度两方面的结构特征。

为了方便阐述，将一手五指分别用数字 1~5 来表示：1 代表拇指，2 代表食指，3 代表中指，4 代表无名指，5 代表小指。使用五个手指的手形可用“五”表示。利用这些数字可以对各手形所使用的手指或手指组合进行标记。如：字母 I 的手指可记为 2；字母 R 的手指可记为 1 + 2；字母 Q 的手指可记为 1 + 2 + 3；字母 B 的手指可记为 2 + 3 + 4 + 5；字母 U 的手指可标记为五。

确定《中国手语》书中手语有 61 个基本手形，如表 1 所示。

Table 1. Base gesture of Chinese sign language  
表 1. 中国手语基本手形

序号	手形	序号	手形	序号	手形	序号	手形
1	1-伸	17	1+5-捏	33	2+3+4-伸	49	五-S
2	1-弯	18	1+5-弯	34	2+3+5-伸	50	五-捏
3	2-伸	19	2+3-伸	35	3+4+5-伸	51	五-O
4	2-弯	20	2+3-并	36	3+4+5-弯	52	五-D
5	3-伸	21	2+3-弯	37	1+2+3+4-M	53	五-CH
6	5-伸	22	2+3-交	38	2+3+4+5-伸	54	五-OK
7	5-弯	23	2+5-伸	39	2+3+4+5-并	55	五-P
8	1+2-伸	24	1+2+3-伸	40	2+3+4+5-弯	56	五-WC
9	1+2-并	25	1+2+3-捏	41	五-伸	57	五-兰花指
10	1+2-弯	26	1+2+3-弯	42	五-并	58	五-床
11	1+2-平	27	1+2+3-警察	43	五-聚	59	五-姜
12	1+2-环	28	1+2+3-N	44	五-开	60	五-仿 d
13	1+2-捏	29	1+2+3-K	45	五-侧开	61	五-毛笔
14	1+2-十字	30	1+2+3-SH	46	五-弯		
15	1+2-半	31	1+2+3-除号	47	五-平		
16	1+5-伸	32	1+2+5-伸	48	五-C		

### 3. 扩展 Kinect 人体骨骼模型

Kinect 体感器是微软为 XBOX 系统开发的一款外部设备，自 2010 年 11 月 4 日发售以来，在短短四个月的时间积累了一千万的销量，被吉尼斯认证成为历史上销售速度最快的消费电子产品。Kinect 不但价格低廉，而且重量轻巧，有两个前置摄像头，分别是 3D 深度摄像头和 RGB 摄像头。除此之外还内置一个马达用来调整俯仰角度。Kinect 体感器将捕捉到的影像与本身内部存有的人体模型相对照，建立起具有二十个骨骼点的人体骨骼模型<sup>[4]</sup>，如图 1 所示。

这一模型简化了人体运动的细节模型，对于识别效率提高具有显著的作用。但是，对于手语识别的应用，缺少了技术支持。为此，本研究针对手语识别的特点，扩展了 Kinect 人体骨骼模型。

从人体解剖学知识可知，人手由 27 块骨骼组成，分为腕骨，掌骨和指骨，其中 19 块组成了手掌以及手指<sup>[5,6]</sup>。我们可以把手看成一个多肢节的系统，由 4 个相邻的手指，一个大拇指以及手掌组成。定义手部骨骼模型的根节点是 Kinect 人体骨骼模型当中的 Hand\_Right 和 Hand\_Left 节点(即腕关节)，如图 2 所示。

扩展 Kinect 人体骨骼模型，以原有 20 个骨骼节点为基础，采用基于关节角度的手部骨骼模型来记录手形，共 30 个手部节点(左手 15 个，右手 15 个)。每

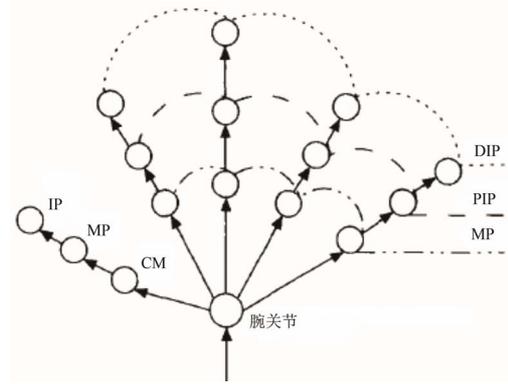


Figure 2. Hand skeleton model  
图 2. 手部骨骼模型

一个手指(2~5)均具有四个自由度。其中手指的基部(MP)有两个自由度，弯曲和旋转。手指的中间关节处(PIP)和末端关节处(DIP)分别各有一个自由度，主要是弯曲运动。大拇指除了与其他四个手指一样具有四个自由度外，还有一个外展运动。所以大拇指具有五个自由度<sup>[5]</sup>。外加手掌的前后左右运动二个自由度，所以手运动总具有 23 个自由度。要正确计算手指节点的三维坐标，需要使用变换矩阵模型，设一个节点的当前位置  $p$ ，通过与变换矩阵相乘后得到目标位置  $p1$ ，由于变换矩阵既包括平移变换，也包括旋转变换信息，则节点  $p$  的坐标计算公式为：

$$p1 = R \times T \times p$$

其中： $R$  为旋转变换矩阵， $T$  为平移变换矩阵。

由于扩展 Kinect 人体骨骼模型可由树结构来表示，因而在三维场景中所有非根节点的位置都可以通过计算该节点相对于父节点的旋转量及偏移量，再乘以父节点当前的模型矩阵获得，而其父节点的位置则是通过计算相对于父节点的父节点的旋转量及偏移量，再乘以父节点的父节点当前的模型矩阵获得，如此向上回溯计算，直至到根节点<sup>[7]</sup>。

### 4. 骨骼手语库结构

根据对中国手语词汇当中的手形分析，建立手语词的骨骼手语库。采用关键帧法记录最能体现手语动作本质特征的状态。每个关键帧保存的是每个手语词所对应的手形、位置、动作和方向等信息<sup>[7]</sup>，同时还记录了关键帧变化的时间间隔。骨骼手语库中的每个词条由手语词编号、手语词内容、关键帧总数、关键帧时间间隔及所指示的扩展 Kinect 人体骨骼模型 50

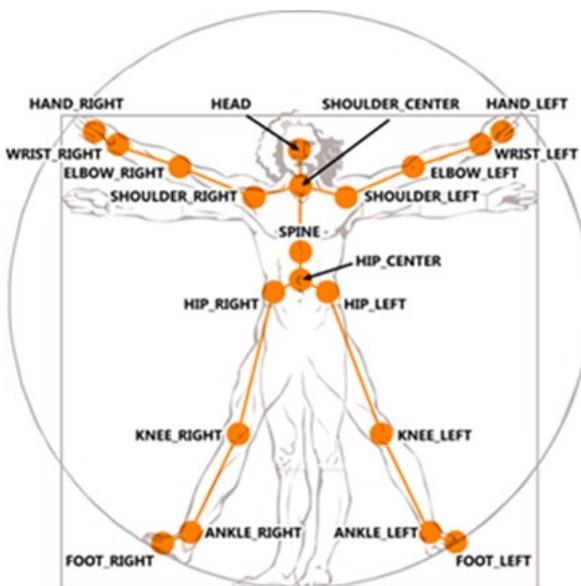


Figure 1. Kinect skeleton model  
图 1. Kinect 人体骨骼模型

个节点的关键帧数据组成,表结构关系如图3所示<sup>[8]</sup>。其中每个关键帧数据都包括了所有节点的自由度<sup>[9]</sup>。

表 Humanoid 定义了人体骨骼数据的根节点信息,表 Joint 存储每个骨骼节点的坐标数据等内容,表 Animation 定义了主要骨骼动作的信息,表 Animation\_Starts\_With\_Joint 和 Animation\_Ends\_With\_Joint 定义了骨骼动作的起始和结束节点参数。

### 5. 系统设计与实现

通过对中国手语的表达特征分析,基于手形的手语词汇生成系统构建了一系列的手语词生成模块,包括手语词关键帧数据生成模块,手语词关键帧基础数据快速生成模块,手语词关键帧数据调整模块。

#### 1) 手语词关键帧数据生成模块

手语词关键帧数据生产模块主要通过选择手形,手形位置和方向等基本信息,产生手语词关键帧数据,如图4所示。通过调整产生的关键帧数据上下关系和插入间隔时间片段,手语词的所有关键帧数据。

#### 2) 手语词关键帧数据调整模块

由手语词关键帧数据生成模块产生的关键帧数据,由于手形、位置和方向的固定性,导致数据较生硬。因此,系统提供了手语词关键帧数据调整模块,

提供了从手臂开始的主要节点的自由度参数调整控件,通过改变控件的值来驱动关节运动,进而实现各种手语数据的修改,图5给出了数据调整模块的主界面。

#### 3) 手语词关键帧基础数据快速生成模块

通过手语词关键帧数据生成模块和调整模块就可以完全实现生成各种手势的功能,但是,由于手势运动的复杂性,手工生成手语词数据的工作就显得非常繁杂,为此,系统提供了手语词关键帧基础数据快速生成模块。模块通过导入《中国手语》上的手语词图片,采用骨骼提取算法,自动匹配选择61种手形,分析其位置和方向,形成手语词关键帧基础数据。然后,由人工采用手语词关键帧数据调整模块进行检查和修改,形成可用数据。

通过本系统我们收集了50个中文手语词汇及32个字母的手势运动数据,这为后续手语合成与识别奠定了基础。

### 6. 结束语

本文从中国手语语言学出发,采用中国手语词汇当中的基本手形,以扩展Kinect人体骨骼模型为基础,提出了一种构建了一个骨骼手语库的方法。根据这一方法,编制了用于生成各种手势的手语词汇生成系

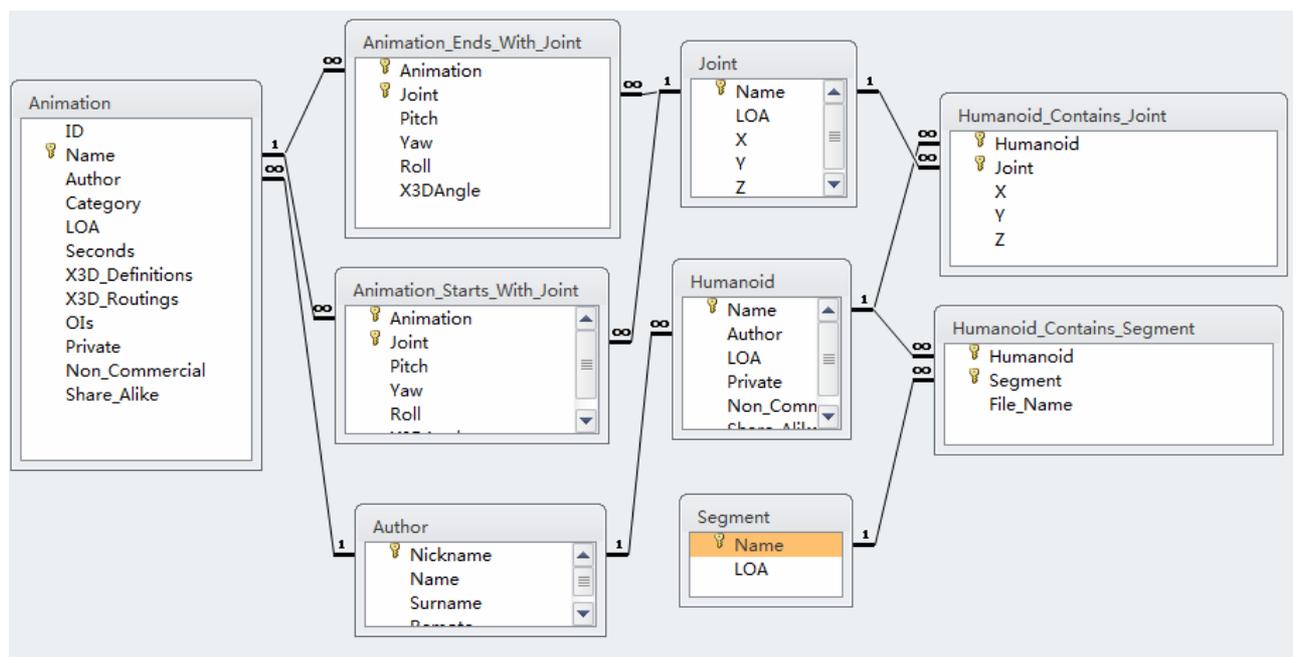


Figure 3. The tables relation of skeleton database  
图3. 骨骼手语库数据表关系

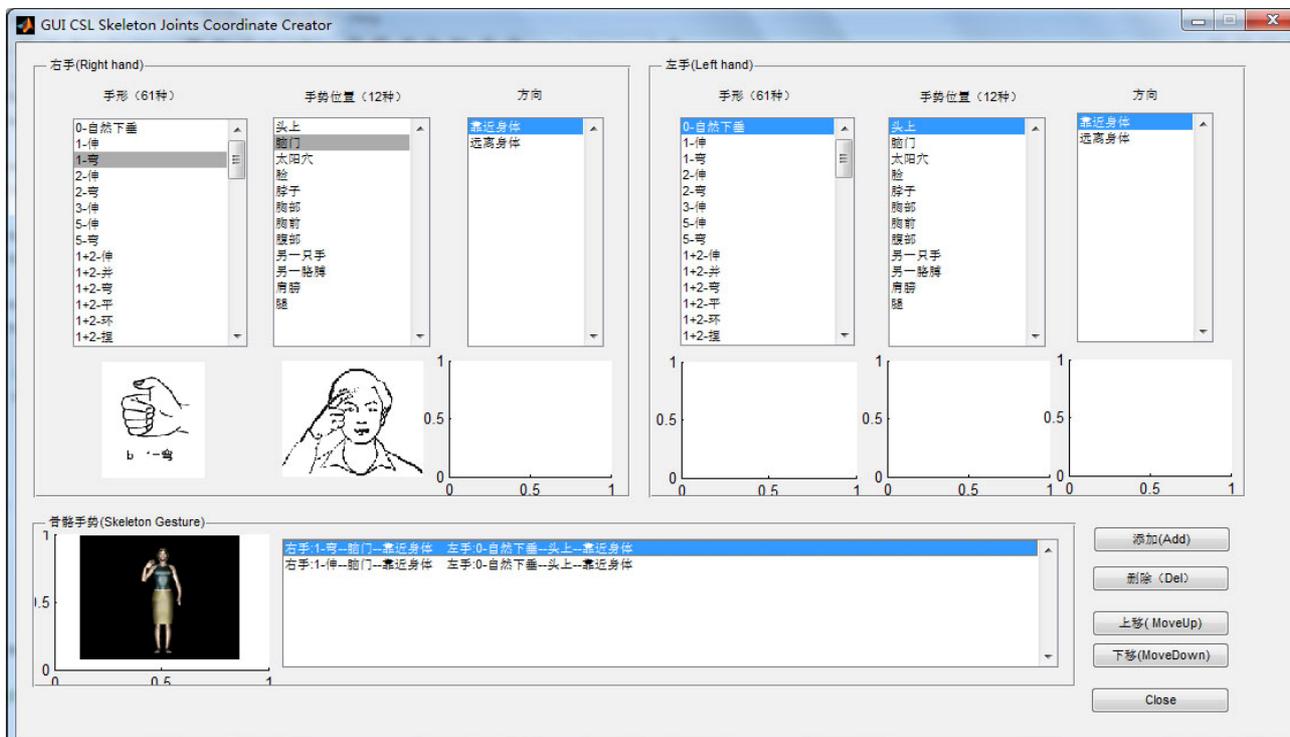


Figure 4. 3D skeleton sign language database create interface  
图 4. 三维骨骼手语库实现界面



Figure 5. Sign language keyframe adjust interface  
图 5. 手语词关键帧数据调整主界面

统，里面包含了基本模块以及快速生成手语词关键帧基础数据的辅助模块等功能，有效地提高了手语词的生成质量和效率。在保存手语词汇信息的时候，以抽取各个手势的关键状态点，记录手形、位置和方向，大大节约了手语词汇信息的储存空间，为手语的检索

带来了极大的方便。

### 参考文献 (References)

- [1] 冯玉花, 阿里甫·库尔班, 陈景超. 基于 VRML 虚拟人的维文手语库的构建[J]. 计算机应用与软件, 2012, (2): 37-39,43.
- [2] 王春立等. 基于词根的中国手语识别方法[J]. 计算机研究与发展, 2003, (2): 150-156.
- [3] 骆维维, 《中国手语》手形研究[M]. 北京: 北京师范大学, 2008.
- [4] 微软公司. 人机界面指南[M]. 2011.
- [5] 曾芬芳等. 手模型分析及手势识别[J]. 华东船舶工业学院学报, 1998, (5): 32-36.
- [6] 吕治国, 李焱, 贺汉根. 基于 Poser 模型的三维人体建模方法[J]. 计算机工程, 2008, (13): 256-258,261.
- [7] 李健, 孔令寅. 基于 H-Anim 的三维手部动画的研究与实现[J]. 计算机工程与设计, 2010, (22): 4866-4869.
- [8] F. Buttussi, L. Chittaro and D. Nadalutti. H-Animator: A visual tool for modeling, reuse and sharing of X3D humanoid animations. New York: ACM Press, 2006: 109-117.
- [9] 王兆其, 高文. 基于虚拟人合成技术的中国手语合成方法[J]. 软件学报, 2002, (10): 2051-2056.