

北京航空航天大学“冯如杯 创意大赛参赛论文

动作语言转换器

2012 年 3 月

摘要

在这样一个和谐的时代，很明显，依然存在很多不和谐，不公平的事情发生。在社会上有这样一个群体，他们因为语言听力能力的障碍，被一些人瞧不起，在工作方面也有很多不公平。我预想设计的这一款“动作语言转换器”，可以很好的缓解，甚至解决这一问题。它主要通过图像识别，语音识别及传感器的应用，将动作分解成图像信号，将他们所熟悉的手语与我们的语言之间进行对应的翻译。可以说是一个随身携带的翻译，从而实现我们之间的无障碍交流。从而解决一些问题，同时可以使那部分人，在其他方面的才能得以发掘。该作品的体积小是他的一大优点，同时需要的技术也已经成熟，所以生产成本应该会较低，同时这是一款人性化的产品，创新性强。所以设计，生产这个产品应该会有不错的效果。

关键字： 图像识别，语音识别，传感器，动作语言转换。

Abstract

In such a harmonious time, obviously, there is still a lot of discord, unfair things happen. There is such a group in the community, because the barriers to language listening ability, some people look down at them, there are many unfair things at work. I expected to design a product that named "action language conversion machine" it can be a good remission, or even solve this problem. It is mainly through image recognition, voice recognition and sensor applications, broken down the action into the image signal, to translate the sign language they are familiar with to our language translation. It can be a portable translator, in order to achieve a barrier-free exchange between us. And then, it can solve some problems, while allowing their abilities in other areas can be excavated. The small size of the work is a big advantage, at the same time, the need of the technology is mature, so production costs should be low, and then, this is a user-friendly product, innovation. So, it will be great to design and make it.

Keywords: Image recognition, voice recognition, sensors, movement, language Conversion.

目录

摘要	i
Abstract	i
一：创意说明	1
1.1 创意来源	1
1.2 解决的问题	1
二：当代这一方面取得的进步与缺陷	1
2.1 通过投影仪，手语翻译进行翻译	1
2.2 缺陷	2
三：动作语言转换器主体	2
3.1 主要结构及功能	2
3.1.1 动作捕捉与产生装置	2
3.1.2 语音捕捉与产生装置	2
3.2 主要利用的原理	2
3.2.1 图像识别	2
3.2.2 语音识别	3
3.3.3 传感器的应用	4
3.3 工作过程及原理	4
四：产品特点	5
4.1 作品优点	5
五：细节的改进及其他方面的应用	5
5.1 细节改进	5
5.2 其他方面应用	5
结束语	6
参考文献：	7

一：创意说明

1.1 创意来源

在当代社会，也许你会遇到这样的情况，某个时候，一个语言或是听力能力有所障碍的人跑焦急过来，告诉你似乎很重要的问题，而你看不懂他的“语言”，或者是你问的问题他却听不懂；同时你也会发现在你的身边几乎没什么在语言或是听力方面有障碍的人作为同事，因为他们在某方面的能力缺陷，工作也会受到不公平的待遇。同时对于这样的群体而言，他们也许在某一发面有所缺陷，但是对于他们而言在其他方面也许会有过人之处。于是，首先是为了解决这样的一个不公平，不和谐的问题，再者为了让这一特别群体也能有所表现机会，发挥自己的才能，我想到设计的这样一个动作语言转换器。

1.2 解决的问题

动作语言转换器，顾名思义，就是将你所看到的动作转换为语言，将你所听到的语言转换为你所熟悉的动作。那么可以想象，我的产品就是将一群体熟悉的手语与我们所熟悉的语言之间进行的对应的翻译，不管你的语言障碍是什么，听不见抑或是不能说，只要你能运用手语，能看懂手语，那么该产品就可以将你们的“语言”与我们的语言之间进行对应的翻译。那么这样交流就不会有什么问题了。那么上面所说的有很着急的事却不能听懂，有很重要的问题却不能问，有了工作的机会却因为语言问题而受挫折，得不到应有的机会。或是有什么才能却因为语言上的不自信，而不敢“外露”。抑或是到银行，医院，邮局不能进行正常的交流。这些问题都能因为这一产品得到一一解决。

二：当代这一方面取得的进步与缺陷

2.1 通过投影仪，手语翻译进行翻译

在某些大型会议，有聋哑人参加的会议，会通过一个大的投影仪，将会议的内容通过手语翻译，将翻译成的手语投影到屏幕上。聋人发言则通过速录人员，将手语所表达的的记录，再通过手语翻译，翻译成对应的文字，打在大屏幕上，让会议人员第一时间了解他们所说的话，这样就能较为正常通常的进行有聋人参与的会议。同时手机短信的发明，电影字幕的设计也让聋哑人的生活变得鲜艳，交流的得到了一定的缓解。

2.2 缺陷

虽然这样的设备有解决一些问题，但是这样庞大的阵容，豪华的设施，总不能带回家，带到日常生活中吧。这样的话，每一个要说话的要听明白的聋哑人都需要一个速录人员，都需要一个手语翻译，想想这样的条件根本就是不可能的，这样的他们只是一时的会“说话”，没能真正解决问题。再说手机短信的交流，虽然某方面来说，还不错，但是发短信的信息量小，交流缓慢，还容易受到信号的干扰，例如，在火车穿过山洞的时候就没法交流。所以这样的方法还不够解决问题。

三：动作语言转换器主体

3.1 主要结构及功能

该设备外观上主要有两个重要的装置，一个是负责动作方面的装置，外形与眼镜是一样的，内部主要结构是手语的存储器；另一个是负责语言接收产生的，外形就是一个麦的形状，内不是语言存储器。可以说是外形小巧，携带方便。

3.1.1 动作捕捉与产生装置

首先介绍一下负责动作的装置，主要用它进行动作图像的捕捉，对手语进行捕捉，信号传送至手语存储器，同时将由语言存储器所传送过来的语言信号所对应的手语动作表现出来，如同通过该眼镜你能看到前面有一个手语翻译正在一字一句的翻译给你看而不是将手语聚焦在眼睛上。既然是捕捉器，那么上面肯定有一个类似微型摄像头的东西，他可以将动作非常准确的记录下来，如同将语言一个字一个字的记录下来一样，将动作类似于分解成一个一个图像，然后记录下来。

3.1.2 语音捕捉与产生装置

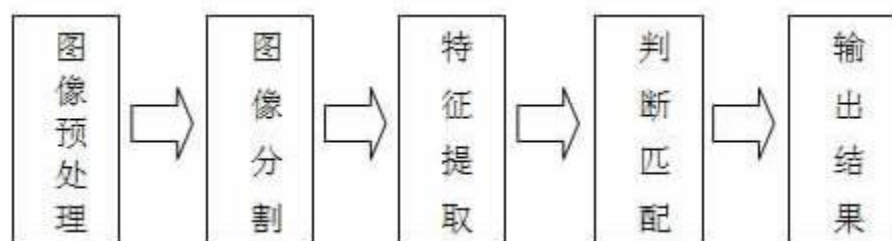
再说一下该装置的另外一个装置，那就是负责语言的装置。他和动作装置的作用结构基本相同，内部有一个语言存储器，储存了很大一部分我们所了解的，使用率较高的字词，语言等。语言装置通过麦上的语言接收器将语言接收并转换为信号传送，同时也能进其逆过程，将已翻译的动作所对的语言发出声来。

3.2 主要利用的原理

3.2.1 图像识别

图像识别是当代图像处理研究的热点课题之一，是众多研究人员的研究重点，图像识别经历了三个阶段：文字识别，数字图像处理，物体识别。在我所设想的作品中会用到数字图像处理，与物体识别。因为数字图像识别比与模拟图像相别

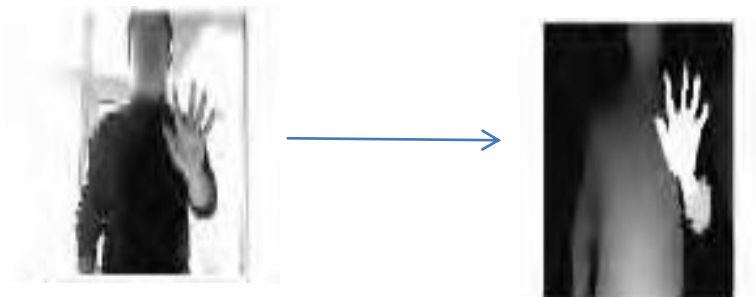
具有储存传输方便，可压缩，传输过程不易失真，处理方便的优势。物体识别是对三维世界的认识与识别。图像识别关键步骤主要有下面几步：



其中最重要的是图像分割，图像分割结合了一些特定的理论，方法，道具。实现模拟空间到类别空间的映射。方法也有很多，如边缘检测方法，区域提取方法等。

当然一般的图像识别是不能满足识别手语的要求的，毕竟手语不是图像而是动作，再者这样的误差也会很高。

近些年，一些学者在基于视觉的手势上进行了一定的研究，那么我们也可以利用他们的研究结果，对我们所需要的图像识别系统进行改进，毕竟手语也是一种手势。我们可以利用吧B. Kaufmann等提出的智能进化算法，C. Weng等采用的多特征融合的方法，杨波提出的针对复杂背景下采用空间分布特征的方法等，提高在复杂情况下，动作图像捕捉器的精确度，从而提高动作转换为语言的精确度。同时可以利用李瑞峰等所研究的深度图像成像原理，利用微软开发的Xbox的Kinect，体感器之类的系统，进行图像采集。该系统采集 640×480 的深度图像速度可达30帧每秒，深度分析率为5mm左右。该系统通过深度摄像头采集光编码技术，利用人眼看不见的红外线，通过镜头前的扩散片将红外线均匀的分布投射在测量空间，



在利用红外图像传感器读取空间的编码点信息，然后进行解码，生成深度信息。这样就可以

进行下面一系列的翻译工作了。

实际上，要做出这样的作品最为重要的是实现图像识别，实际上说的更为具体应该叫做动作识别。预想，将手语这一系列的动作通过一个拍照速度超级快的，类似于 x-box360 的体感器一样的装置，将动作进行分解成小细节的图像，在进行图像的分割，从而传送到图像存储器进行匹配，翻译。

3.2.2 语音识别

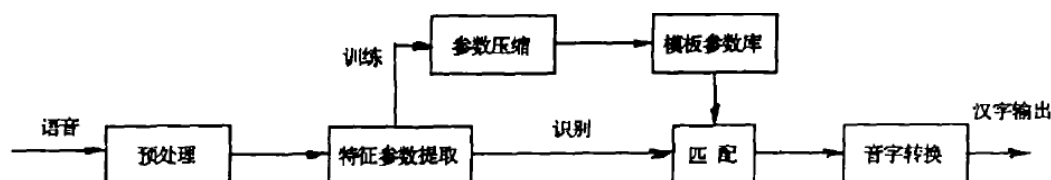
和图像识别一样，语音识别也是该作品的核心工作原理。可以说语音识别与图像识别各占技术性的一半。

首先，什么事语音识别，简单地说就是让机器代替人的耳朵，代替人的听觉，实现听的部分。

国内外的语音识别技术经过了几十年的光阴，现在已经取得了一定的成就。就国内而言，中科院声学所，清华大学，西安电子科技大学，中国科技大学，哈工大，国防科大与总参 51 所联合，四达公司，星河公司都相继设计出汉语全音

节实时语音识别系统，单音节识别率达到 90% 左右。1992 年清华大学在深圳计算机展览会上展出了 208 个军事用语的实时非特定人的语音识别系统，识别率高达 100%。北京星河智能计算机研究所研发的 220 旅游用语（2-8 个字）语音识别，识别率也达到了 100%。在大词汇特别人识别方面也从几千词到了几万词，识别率也高达 90%。

识别过程同图像识别基本类似：



语音识别重要的是预处理，因为由于外界环境的影响，一般汉语的起音部分的辅音能量较弱，有处于高频，在环境和设备噪音较大时，这一部分很难分辨，因此去除环境和设备引起的噪音对整个语音识别是一个不可忽视的问题，目前大多数的识别系统采用高质全的采集系统和安静环境。。然而不管怎么说语音识别却比图像识别更加成熟，现在很多智能手机都用了语音识别的系统，从而体现人机结合的概念，所以讲语音识别的技术应用于我所设想的产品应该不会很难。

3.3.3 传感器的应用

相比与图像识别与语音识别，传感器的应用就会显得很简单了，传感器在该产品中主要用于信号数据的传输。将非电物理量，通过敏感元件，再通过转换器件，转换电路，最后变成电学量。在该产品中，传感器相当于内部的翻译机，将图像信息，转变为电信号，进行后续处理。同时也能将语音信息转变为电信号进行处理。可以说，传感器将两大技术紧密的结合，才能够成为这样的一个产品。

3.3 工作过程及原理

在之前的讲述过程中，我基本上都多多少少有涉及到该作品的基本工作的过程及原理。实际上，该创意的工作过程很简单，只要图像识别与语音识别已经做得够好的话。现在如果有一个不会说话也不能听见说话的A人，欲与一个正常人B交流：A先“说话”，通过眼镜上的动作捕捉器，接收动作，再将动作快速分解成微笑的细节图像，进行预处理，然后进行图像分割，通过传感器将图像信号转变为电信号传输至图像存储器（存储了基本所有手语的动作图像），进行匹配，提取其特征性，利用相似性进行匹配，再将对应的图像信号变为电信号传至语音存储器进行匹配，然后在重新组合所对应的翻译的语言通过简单地原理“说”出来，那么B就能很轻松地并且及时的听到A想要表达的意思。同样的，当B要说话给A“听”时，先将语言接收（非特定人，大词汇，连续语音），通过语音识别的过程将语音信号转换为电信号再转换为对应的图像信号，从而进行组合合成动作，在眼镜的“前方”类似形成一个手语翻译在进行手语发翻译的动作。工作原理就是基于上述所说的一系列技术，通过传感器转化，处理信号实现动作语言的转换。

四：产品特点

4.1 作品优点

该产品不得不说的优点，必须是体积小，便于携带，不必再把投影仪，手语翻译带在身边，也能进行正常的交流，他就是一个小型的手语翻译，可以随身携带的手语翻译。同时该产品的外观非常正常，必要时还可以当作装饰品，没有一点什么不同，或是奇怪的感觉。同时该产品集合了投影仪，手语翻译于一身，可以说是性价很高，并且是属于人性化的设计，并且是基于人的需求而考虑设计的，所以进行研究，投入生产，肯定会受到欢迎，解决很多的因为语言障碍所产生的很多问题。

五：细节的改进及其他方面的应用

5.1 细节改进

首先我们会很容易想到外界的干扰，不管是对于动过的接收还是语音的接收，外界的干扰都将成为最大的影响准确度的因素。那么怎么最大的解决这一问题呢？首先对于图像的接收，可以采用当前很多相机采用的自动聚焦的效果，将图形动作捕捉器聚焦在你所面对的人的手上或是聚焦在你自己的手语动作上，那么就可以尽量会减少，由于周围的东西而引起很大的误差。而对于语音接收器，通过怎样的调节才能让其尽量只接收你所要听的那些声音呢？我预想在语音接收装置中在安装一个语调记忆器，对话之前，将你所交流的人的语调进行记录，然后通过很多如同保密工作一样的进行筛选，将不通的语调淘汰，从而只翻译先前纪录语调的人的声音，交流完后，将记录的语调删除以便下一次的交流。当然要进行多人的对话，那最好你们就一个人一个人的说，别集体说话了。

5.2 其他方面应用

如果能实现这样的装置，即手语与正常的语言之间的互相转换。那么，我认为经过一系列的改进或改装，可以利用类似的原理，实现唇语与语言的交流，将表情转换为语言，如此这样，就能够实现秘密的对话，保障了我们的隐私，同时对于刑事案件的侦破也会有一定的效果。同时在某些关键的时刻，不能说话，或是已经没力气说出话的时候，那么如果他的表情或是唇语被能转换为语言，那岂不是会解决很多问题。相信在当代社会科技的发展速度之下，加之当今已经有一部分技术相当成熟，相信加以改进，继续创新，实现这样的设想绝不是空想。

结束语

现在的市场上还没有这样的类似的产品出现，所以它的市场还处于未开发的状态，但是他肯定会是一个潜力股，因为他真的是以为人们解决问题而设想出来的，真的希望能够在某一天解决那些棘手的问题，解决那些干扰的因素，真正的改进，更加完善后得到设计，真的希望他能够帮助那些真的需要帮助的人，实现他们“说话”，“听见”人家说话的愿望，毕竟每个人都差不多，在痛苦快乐面前我们都是平起平坐的，还有那么多人在寂寞，希望这个愿望终有一天能实现。

参考文献:

- 1 富志伟. 我看“聋人信息无障碍”. [M]. 中国残疾人 2007.10
- 2 张家怡. 图像识别的技术现状与发展趋势. [M]. 电脑知识与技术 2010.7
- 3 李瑞峰, 曹维清, 王丽. 基于深度图像和表观特征的手势识别. [M]. 华中科技大学学报 2011.4
- 4 王成友, 梁句农, 孔铁生, 陈辉注, 唐朝京. 自动语言识别技术评述. [M]. 声学与电子工程 1996 第三期
- 5 梁振福, 李平, 胡正元. 环境噪声对语言识别率的影响. [M]. 声学技术 1998.12