

# 基于 Kinect 的动态手势识别研究

邵天培<sup>1,2</sup>, 蒋刚<sup>1,3</sup>, 留沧海<sup>1,2</sup>

- (1. 西南科技大学 制造科学与工程学院, 四川 绵阳 621010;
2. 制造过程测试技术省部共建教育部重点实验室, 四川 绵阳 621010;
3. 成都理工大学 核技术与自动化工程学院, 成都 610059)

**摘要:** 针对人机交互领域中基于视觉的传统动态手势识别方法准确率不高、易受不同强度光照影响等问题, 对动态手势识别方法进行了研究; 首先利用 Kinect 传感器采集的深度图像对手势进行分割, 并基于矩和链码进行手势质心与手势轨迹特征的计算, 再利用动态时间规整算法进行手势轨迹特征识别, 最后将识别结果传输给六足机器人进行人机交互实验, 实现了动态手势对六足机器人的控制; 实验结果证明: 该方法识别准确率最高可达 97%, 且不易受光照影响, 具有较强的鲁棒性, 同时也满足了人机交互需求。

**关键词:** 动态手势; 人机交互; 动态时间规整; 六足机器人

## Research on Dynamic Gesture Recognition Based on Kinect

Shao Tianpei<sup>1,2</sup>, Jiang Gang<sup>1,3</sup>, Liu Canghai<sup>1,2</sup>

- (1. College of Manufacturing Science and Engineering, Southwest University of Science and Technology, Mianyang 621010, China;
2. Ministerial Key Laboratory of Testing Technology for Manufacturing Process, Mianyang 621010, China;
3. College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu 610059, China)

**Abstract:** The traditional dynamic gesture recognition method based on vision in the field of human-computer interaction is not accurate enough and is easily affected by different intensity of light. To solve this problem firstly, the depth image collected by Kinect sensor is used to segment the gesture, and the gesture centroid and gesture trajectory feature are calculated based on moment and chain code, and the gesture trajectory feature is recognized by dynamic time warping algorithm. Finally, the recognition results are transmitted to hexapod robot for human-computer interaction experiment, and the control of hexapod robot by dynamic gesture is realized. Experimental results show that the recognition accuracy of this method is up to 97% and is not affected by the light, and it has strong robustness and meets the requirements of human-computer interaction.

**Keywords:** dynamic hand gestures; human-computer interaction; dynamic time warping; hexapod robot

## 0 引言

随着人类社会和机器人领域的不断发展, 人类与机器人在各个方面也在不断地融合, 机器人逐渐融入到了人类的生产制造过程和生活中, 人类与机器人的关系日益密切, 这就迫切地需要人类与机器人进行交互, 从而能够比较顺利地实现人机合作与人机共融<sup>[1]</sup>。在人机交互领域中基于视觉的手势识别因为具有简单、直观、表达能力突出且自然的优势逐渐成为研究热点<sup>[2-5]</sup>。Bertsch 等人<sup>[6]</sup>在仿人机器人平台上构建了基于视觉的人机交互系统, 该研究使用隐马尔可夫模型识别动态手势, 但是这种方法准确率不高且易受光照影响。Think 等人<sup>[7]</sup>在机器人上安装手机, 利用手势通过手机上的摄像头与机器人进行交互。随着 Kinect 的问世, 进一步弥补了人机交互方式上硬件的不足<sup>[8]</sup>。王兵

等人<sup>[9]</sup>基于 Kinect 采用指尖检测算法并针对运动轨迹中运动方向的变化, 提取了动态手势的特征并识别各个手势。Y. Gu 等人<sup>[10]</sup>利用 Kinect 并基于骨骼特征, 使用隐马尔可夫模型训练进行测试, 但该方法平均识别率不高且数据训练复杂。本文针对动态手势识别准确率不高且易受光照影响等问题对动态手势识别进行了研究。

## 1 动态手势识别

动态手势识别通过关注分析视频序列中的连续手势轨迹信息来判断手势语义, 而不关注手型<sup>[11]</sup>。本文将基于 Kinect 传感器进行动态手势识别, Kinect 会首先采集视频序列中的用户手势, 然后依次进行手势分割、手势质心计算、手势轨迹特征计算, 最后进行手势轨迹特征识别。

### 1.1 手势分割

手势分割是指将人手准确地从简单或复杂背景环境中

收稿日期: 2020-06-19; 修回日期: 2020-07-10。

基金项目: 四川省重大科技专项(2020ZDZX0019); 四川省科技厅重点研发计划(19ZDYF1083)。

作者简介: 邵天培(1994-), 男, 四川德阳人, 硕士研究生, 主要从事机电一体化、图像处理、人工智能方向的研究。

蒋刚(1978-), 男, 四川广安人, 博士生导师, 教授, 主要从事机器人技术、机电一体化方向的研究。

留沧海(1966-), 男, 福建泉州人, 博士, 教授, 主要从事机电一体化、机器人技术方向的研究。

引用格式: 邵天培, 蒋刚, 留沧海. 基于 Kinect 的动态手势识别研究[J]. 计算机测量与控制, 2021, 29(2): 161-165.

分割出来<sup>[12,13]</sup>，这是极其重要的一步，因为手势分割的好坏会影响后续环节。手势分割部分的难点在于目标手势易受光照和环境的影响。常用的方法，比如单纯利用肤色进行手势分割<sup>[14]</sup>，则会把类肤色物体包括进来，在室内不同光照下利用肤色进行分割，效果如图 1 所示：从图 1 (a) 可知，肤色分割虽然完整分割出了手势，但也将人的面部和类肤色物体一起包括。从图 1 (b) 可知，在室内较弱光照条件下，分割出的手势甚至不完整，所以光照条件的改变会令手势分割变得异常困难。



(a) 较强光照下分割结果



(b) 较弱光照下分割结果

图 1 不同光照下基于肤色的手势分割结果

而利用 Kinect 采集的深度图像信息则会在进行手势分割时有效减少光照条件带来的影响。这是因为 Kinect 采集的深度图像存储在一个二维数组里<sup>[15]</sup>，数组中的每一个像素点表示该像素点到传感器的距离，由于做一组动态手势时人手通常离传感器最近，所以可从绘制出的深度图像直方图中找到较小的灰度级，设定不同的阈值就可以准确地从背景中分割出感兴趣区域，再将感兴趣区域转换成二值图像即可。经过上述操作，就可以完整地分割出目标手势。如图 2 (a) 右侧所示为较强光照条件下的深度图像，图 2 (b) 左侧为深度图像直方图，右侧为手势分割结果，其中，深度直方图中圆圈所在横坐标为选定阈值的大概范围，该范围根据不同距离选取不同的值。本文中阈值范围为 35~44 左右，下同。

较弱光照条件下，利用深度信息对目标手势进行分割的结果如图 3 所示。

在不同环境条件下，利用深度信息对目标手势进行分割的结果如图 4 所示。

从上述实验结果可知，在不同环境和不同光照条件下，本文所采用的方法依然可以准确地分割出目标手势，证明了本文方法的可靠性和优势。

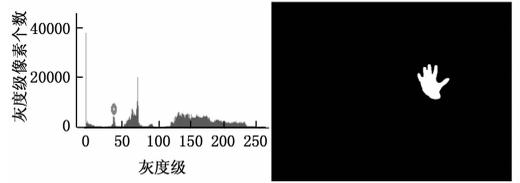
### 1.2 手势质心计算

矩代表一幅图像或一个轮廓的全局特征，是描述图像特征的算子<sup>[16]</sup>。得到手势分割图像后，可计算出图像和轮廓的矩。

$m_{00}$  表示图像零阶矩， $v(i, j)$  表示图像在点的灰度，



(a) 原始图像与深度图像

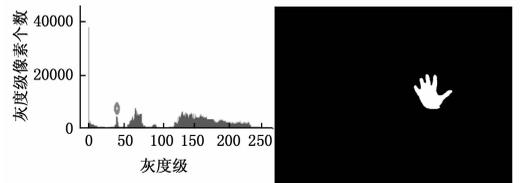


(b) 较强光照下深度直方图与分割结果

图 2 较强光照下基于深度信息的手势分割结果



(a) 原始图像与深度图像

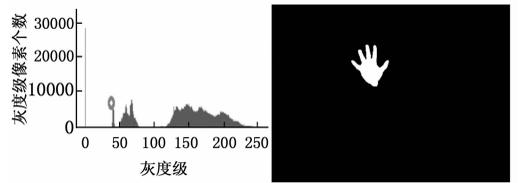


(b) 较弱光照下深度直方图与分割结果

图 3 较弱光照下基于深度信息的手势分割结果



(a) 原始图像与深度图像



(b) 不同环境下深度直方图与分割结果

图 4 不同环境下基于深度信息的手势分割结果

$(i, j)$  当图像为二值图像时， $m_{00}$  代表图像中白色区域的总和，因此可以用来求二值图像和轮廓的面积。如式 (1) 所示：

$$m_{00} = \sum_i \sum_j V(i, j) \tag{1}$$

$m_{10}$  和  $m_{01}$  表示图像的一阶矩，当图像为二值图像时，

$m10$  表示图像所有白色区域坐标的累加,  $m01$  表示图像所有白色区域坐标的累加。如式 (2)、(3) 所示:

$$m10 = \sum_i \sum_j i \cdot V(i, j) \quad (2)$$

$$m01 = \sum_i \sum_j j \cdot V(i, j) \quad (3)$$

再根据式 (5), 式 (6):

$$x = \frac{m10}{m00} \quad (4)$$

$$y = \frac{m01}{m00} \quad (5)$$

就可得到目标区域即手势的质心点坐标  $(x, y)$ 。最后绘制出轮廓的矩形边界。

利用图 2 (b) 中的手势分割结果得到的手势轮廓矩形边界、零阶矩和一阶矩的值以及手势图像质心坐标如图 5 所示。



图 5 矩形轮廓与手势质心坐标

当每帧手势图像传入时, 都按照上述方法操作, 得到每帧图像中的手势质心, 这样便得到了一连串的手势质心点坐标。综上所述, 完成手势质心计算。

### 1.3 手势轨迹特征计算

前文提到, 动态手势识别关注的是手势轨迹信息, 而不关注手形。本文要做的就是从连续的视频帧里得到动态手势的轨迹信息。当一个动态手势完成后, 假设相邻两帧图像手势质心点坐标分别为  $(x_1, y_1)$  和  $(x_2, y_2)$ , 则每帧图像中的手势质心坐标结合在一起可以表示成一个手势质心轨迹  $p$ :

$$p = [(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)] \quad (6)$$

同时, 每相邻两帧手势质心点连成的直线与水平方向都会形成一个角度, 设该角度为  $\theta_i$ :

$$\theta_i = \arctan \frac{\Delta y_i}{\Delta x_i}$$

$$(\Delta x_i, \Delta y_i) = (x_{i+1} - x_i, y_{i+1} - y_i) \quad (7)$$

则多个不同的  $\theta_i$  可以组成个一维向量  $T$ :

$$T = [\theta_i, \theta_{i+1}, \dots, \theta_{i+n}] \quad (8)$$

将一维向量  $T$  中的每个角度都利用链码进行编号, 其

中,  $\theta_i$  的值决定了  $\theta_i$  所在链码编号, 依次类推。如此便可得到一个由链码编号组成的一维向量  $X$ , 而  $X$  正是动态手势的轨迹特征。

如图 6 所示, 链码<sup>[17]</sup>把一个平面平均分成了 12 个部分, 并对这 12 个部分进行编号, 编号分别为 1~12。

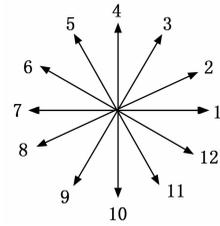


图 6 链码示意图

设  $0 \sim 180^\circ$  范围内, 相邻编号所属直线之间的角度为  $30^\circ$ 。在  $0 \sim -180^\circ$  范围内, 相邻编号所属直线之间的角度为  $-30^\circ$ 。

当平面被平分的部分越多, 通过角度  $\theta_i$  的值量化得到的编号越准确, 则动态手势的轨迹特征越精确, 所以这里将平面划分成 12 个部分。

例如, 利用第一帧手势图像质心坐标点与第二帧手势图像质心坐标点计算出两点形成的直线与水平方向所成角度为, 再比较是否在  $0 \sim 30^\circ$  范围之内, 若为  $0^\circ$  则添加编号 1 到一维向量  $X$  中, 若在  $0 \sim 30^\circ$  之间, 则添加编号 2 到一维向量  $X$  中。由此可以得到动态手势轨迹特征  $X$ 。

本文选择其中稳定的 6 个手势质心点进行计算, 则  $X$  的长度为 5, 即每个完整的手势轨迹特征由 5 个编号组成。

### 1.4 手势轨迹特征识别

在完成手势分割、手势质心与手势轨迹特征计算的基础上就可以进行手势轨迹特征识别。由于不同的人在做同一手势动作时速度不一样, 所需的时间也不一样, 则会造成每个人的手势起终点不确定。若要进行手势起终点判定则会加大计算量和复杂程度, 影响效率。

而动态时间规整算法 (DTW, dynamic time warping) 则通过把手势轨迹特征进行延伸或缩短, 来计算比较两个手势轨迹特征之间的相似性, 所以采用 DTW 算法可以省去手势起终点的判定, 并减少不同的人做相同手势时, 手势轨迹特征不完全相同带来的影响。

假设给定两个序列:

$$X = [x_1, x_2, \dots, x_i], Y = [y_1, y_2, \dots, y_j] \quad (9)$$

其中:  $i=j$ 。在本文中,  $X$  实际为得到的手势轨迹特征,  $Y$  为事先设定好的多个手势轨迹特征模板且每一个模板  $Y$  都拥有唯一序号。当  $X$  与  $Y$  之间的累积距离越小, 则代表这两个序列越相似, 即表示  $X$  与  $Y$  越吻合。

DTW 算法的步骤如下:

- 1) 构造一个  $i * j$  的距离矩阵  $M$ ,  $M$  矩阵中元素  $(i, j)$  表示  $x_i$  和  $y_j$  两个点的距离  $d(i, j)$ ;
- 2) 根据距离矩阵  $M$  生成累积距离矩阵  $N$ , 生成  $N$  的

方法为：矩阵  $N$  中第一行第一列的元素为  $M$  矩阵中第一行第一列的元素。矩阵  $N$  中其他位置的元素通过：

$$N(i, j) = \min(N(i - 1, j - 1), N(i - 1, j), N(i, j - 1) + M(i, j)) \quad (10)$$

计算得出；

3) 矩阵  $N$  中最右上角的元素为累积距离  $Q$ 。

举例说明：假设手势轨迹特征  $X = [3, 5, 6, 7]$ ，模板  $Y = [4, 3, 2, 6]$  现在要比较  $X$  和  $Y$  的相似程度，DTW 会根据  $X$  与  $Y$  之间点到点的距离生成一个距离矩阵  $M$ ，结果如下：

$$M = \begin{pmatrix} 3 & 1 & 0 & 1 \\ 1 & 3 & 4 & 5 \\ 0 & 2 & 3 & 4 \\ 1 & 1 & 2 & 3 \end{pmatrix} \quad (11)$$

然后根据距离矩阵  $M$  生成累积距离矩阵  $N$ ，结果如下：

$$N = \begin{pmatrix} 5 & 3 & 3 & 4 \\ 1 & 4 & 7 & 12 \\ 1 & 3 & 4 & 7 \\ 1 & 1 & 3 & 6 \end{pmatrix} \quad (12)$$

由累积距离矩阵  $N$  中最右上角的元素可知，累积距离  $Q=4$ 。

综上，当完成手势分割，手势质心和手势轨迹特征计算后，会得到一个一维的手势轨迹特征向量  $X$ ，将  $X$  作为输入喂入 DTW 算法模型并依次与模型中的多个手势轨迹特征模板  $Y$  进行模板匹配，得到多个累积距离  $Q$ ，将  $Q$  中最小值所属模板  $Y$  的序号作为输出结果，从而完成手势轨迹特征识别。

DTW 算法<sup>[18-20]</sup>数据训练简单，结构简单明了易于实现，同时也省略了手势起终点的判定且识别准确率高。缺点是模板  $Y$  需要事先定义，且计算量随着模板的增多而加大。

## 2 实验验证与分析

### 2.1 实验硬件与软件

本文实验的硬件主要包括：显卡为 GTX1050，内存为 4 GB 和处理器为 i7 的笔记本电脑、基于 Kinect 开发环境的传感器、一台六足机器人。

软件部分主要包括：PyCharm 编译环境、开源计算机视觉库 (OpenCV, open source computer vision library)、开放式的自然交互框架 (OpenNI, open natural interaction) 以及采用 Python 编写的手势识别算法。

### 2.2 实验流程

首先传感器采集用户动态手势视频序列，将视频序列传给上位机进行处理得到与该用户手势轨迹特征匹配的模板的序号，将该模板序号传给 STM32，驱动板根据 STM32 计算得到的舵机数据进行六足机器人舵机控制，同时机器人自身传感器采集六足机器人状态，将状态信息发送至 STM32，形成闭环控制，完成手势对六足机器人的控制。

### 2.3 实验

首先设定人机交互手势模板，如表 1 所示。该模板设计了 4 种动态手势指令分别控制机器人 4 种动作。该模板存储在 DTW 模型中，其中模板基于链码规则设定。

表 1 自定义人机交互手势模板

模板 Y 序号	机器人动作	手势指令	模板 Y
0	机器人前进	手从左向右水平移动	1 1 1 1 1
1	机器人右转	手从上向下垂直移动	10 10 10 10 10
2	机器人左转	手从下向上垂直移动	4 4 4 4 4
3	机器人后退	手从右向左水平移动	7 7 7 7 7

现在要使机器人前进，则用户将手在六足机器人搭载的传感器前由左向右水平移动。

经过手势分割的移动过程如图 7 所示。



图 7 动态手势移动分割图

经过上位机处理得到如下 6 个手势质心坐标和手势轨迹特征  $X = [1, 2, 1, 2, 1]$ ，如图 8 所示。

```
[[247, 354], [247, 354], [267, 366], [267, 366], [272, 369], [272, 369]]
#####
[1, 2, 1, 2, 1]
```

图 8 质心坐标与手势轨迹特征

随后  $X$  将被送入 DTW 模型中与已经建立好的多个模板  $Y$  进行匹配，匹配结果如图 9 所示。

```
模板Y序号:0
累积距离Q: 2.0
模板Y序号:1
累积距离Q: 43.0
模板Y序号:2
累积距离Q: 13.0
模板Y序号:3
累积距离Q: 28.0
匹配模板序号: 0
```

图 9 匹配结果

由图 9 可知， $X$  与序号为 0 的模板累积距离为 2， $X$  与序号为 1 的模板累积距离为 43， $X$  与序号为 2 的模板累积距离为 13， $X$  与序号为 3 的模板累积距离为 28。

经过筛选最终得到了与  $X$  累积距离最小的模板：序号为 0，累积距离  $Q=2$ 。即手势轨迹特征  $X = [1, 2, 1, 2, 1]$  与模板  $Y = [1, 1, 1, 1, 1]$  最相似吻合。

将该模板序号数字 0 通过无线传输设备传给 STM32 分析处理语义信息，再由 STM32 向六足机器人传输指令控制六足机器人的行动轨迹。以图 10 中黑色叉为标记，采用前进手势进行交互时，六足机器人完成前进动作。

本文在不同环境和不同光照条件下进行了多次动态手

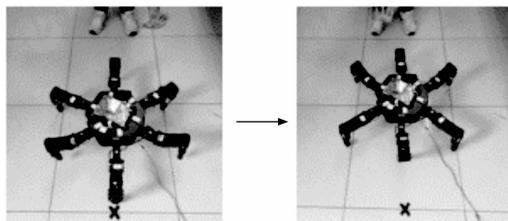


图 10 人机交互结果

势识别实验, 得出结果如表 2~4 所示。

表 2 室外光照识别结果

模板序号	实验次数	正确次数	识别准确率/%
0	300	271	90.3
1	300	264	88
2	300	259	86.3
3	300	270	90

表 3 室内较强光照识别结果

模板序号	实验次数	正确次数	识别准确率/%
0	300	285	95.6
1	300	278	92.6
2	300	283	94.3
3	300	292	97.3

表 4 室内较弱光照识别结果

模板序号	实验次数	正确次数	识别准确率/%
0	300	281	93.6
1	300	274	91.3
2	300	277	92.3
3	300	280	93.3

实验结果表明, 在室内较强光照条件下, 手势识别效果比较弱光照条件下好, 最好的识别准确率能达到 97.3%。在室内较弱光照条件下, 准确率仍然能达到 91% 及以上。在室外的手势识别效果不如室内, 但准确率依然能达到 90% 左右。由此可知在外部环境变换的情况下, 该方法依然能保持一个较好的识别准确率, 也证明了整个动态手势识别方法的鲁棒性较好。

### 3 结束语

综上所述, 本文基于 Kinect 传感器对动态手势识别进行了研究。利用深度信息分割出目标手势, 此方法不易受环境和光照的干扰。基于矩和链码计算出手势质心和手势轨迹特征。在手势轨迹特征识别部分, 传统方法大多数采用隐马尔可夫模型, 但该方法训练非常困难和复杂, 同时要进手势起始点判定。所以采用训练简单、结构易懂、识别准确率高的 DTW 算法并且避免了手势起始点判定, 该算法缺点是不能识别模板中没有的手势。最后将手势识别结果与六足机器人进行交互, 实现动态手势对六足机器人的控制。

### 参考文献:

- [1] 齐静, 徐坤, 丁希仑. 机器人视觉手势交互技术研究进展 [J]. 机器人, 2017, 39 (4): 565-584.
- [2] 李为斌, 刘佳. 基于视觉的动态手势识别概述 [J]. 计算机应用与软件, 2020, 37 (3): 190-197, 208.
- [3] 张莹莹, 郭星. 基于 Kinect 动态手势识别算法的研究与实现 [J]. 计算机技术与发展, 2017, 27 (12): 11-15.
- [4] 姚圆圆. 动态手势识别的关键技术研究及实现 [D]. 重庆: 重庆邮电大学, 2016.
- [5] 易靖国, 程江华, 库锡树. 视觉手势识别综述 [J]. 计算机科学, 2016, 43 (s1): 103-108.
- [6] Bertsch F A, Hafner V V. Real-time dynamic visual gesture-recognition in human-robot interaction [A]. 9th IEEE-RAS International Conference on Humanoid Robots [C]. Piscataway, USA: IEEE, 2009: 447-453.
- [7] Thinh N T, Tuyen N T V, Son D T. Gait of quadruped robot and interaction based on gesture recognition [J]. Journal of Automation and Control Engineering, 2016, 4 (1): 53-58.
- [8] 王健. 基于 Kinect 的手势识别及人机互动 [D]. 南京: 南京邮电大学, 2018.
- [9] 王兵, 董洪伟, 张明敏, 等. 基于 Kinect 的动态手势识别 [J]. 传感器与微系统, 2018, 37 (2): 143-146.
- [10] Gu Y, Do H, Ou Y, et al. Human gesture recognition through a Kinect sensor [A]. 2012 4th International Conference on Intelligent Human-Machine System and Cybernetics (IHMSC) [C]. IEEE, 2012, 1: 274-279.
- [11] 孙靖. 基于视觉的实时手势识别及其应用 [D]. 南宁: 广西大学, 2013.
- [12] 李逸琳, 陶静, 霍艺文, 等. 手势特征提取与图像分割的优化研究 [J]. 计算机应用与软件, 2020, 37 (2): 161-165, 206.
- [13] 黄文静, 马力. 基于 Kinect 手势识别的研究与应用 [J]. 电子设计工程, 2017, 25 (24): 166-169.
- [14] 王俊南, 赖惠成, 杨东旭. 基于肤色分割的改进 Camshift 人脸跟踪算法 [J]. 现代电子技术, 2019, 42 (10): 166-170.
- [15] 孙农亮, 王伟志. 基于 Kinect 深度信息的虚拟手驱动算法研究 [J]. 山东科技大学学报 (自然科学版), 2019, 38 (4): 92-99.
- [16] 韩锐. 基于不变矩特征模型耦合相似度量规则的图像匹配算法 [J]. 包装工程, 2018, 39 (9): 204-211.
- [17] 朱思聪, 周德龙. 角点检测技术综述 [J]. 计算机系统应用, 2020, 29 (1): 22-28.
- [18] 吴彩芳, 谢钧, 周开店. 基于手势识别的人机交互技术研究 [J]. 计算机时代, 2016 (2): 29-32.
- [19] Pouw, Dixon. Gesture networks: introducing dynamic time warping and network analysis for the Kinematic study of gesture ensembles [J]. Discourse Processes, 2020, 57 (4): 301-319.
- [20] 张露. 基于 DTW 的单个手语识别算法 [J]. 现代计算机 (专业版), 2016 (8): 77-80.