

手语识别方法初探

龚晓敏 鞠海军

(清华大学计算机学院 湖南 衡阳 421000)

摘 要 随着人机交互领域的兴起和发展,手语识别也已成为科研领域的热点。手语识别在深度学习、人机交互、模式识别等领域具有重要的意义和研究价值,在未来人类文明的发展中占有不可或缺的地位——消除了听障人士和健听人士之间的交流障碍,使得他们能够生活正常化。本文着重介绍了手语识别技术发展过程中的三个关键阶段,分别为基于数据手套的手语识别、基于视觉的手语识别和基于深度学习的手语识别。然后对手语识别中图像处理的去噪算法进行实验研究,其中基于偏微分方程的图像去噪算法能有效去除噪声,也能够很好地保持细节,是手语识别算法中去噪过程的最佳选择。

关键词 手语识别; 计算机视觉; 图像去噪

中图法分类号 TP39 DOI:10.16707/j.cnki.fjpc.2020.09.003

Preliminary Study of Sign Language Recognition

GONG Xiaomin, JU Haijun

(School of Computer Science, University of South China, Hengyang, China, 421000)

Abstract With the rising and the development of human-computer interaction, sign language recognition has become a hot topic in the scientific research. Sign language recognition is of great significance and research value in the fields of in-depth learning, human-computer interaction, pattern recognition and so on. It plays an indispensable role in the development of human civilization in the future which eliminates the communication barriers between the hearing-impaired and the normal hearing people and enables their lives to be normalized. This paper focuses on three critical stages in the development of sign language recognition technology, including three sign language recognition respectively based on data gloves, vision and in-depth learning. Experimental results show that the image denoising algorithm based on partial differential equation is the best in those three ones, which can effectively remove the noise and keep the details well.

Keywords Sign Language Recognition; Computer Vision; Image Denoising

1 引言

手语就是通过手势表示动作,根据手势的不同位置、手型和其运动轨迹表达不同的含义,是听障人士或无法言语的人士之间的一种沟通方式。中国作为听障人士数量最多的国家,其手语教育却不够全面,导致大多数健听人士甚至一些听障人士都未能接受系统的手语学习,手语的标准化、普遍化程度过低,听障人士和健康人群之间存在着巨大的沟

通障碍。而手语识别是依靠计算机技术将手语转换成大众简单易懂的信息。通过手语识别技术在一定程度上可消除听障人士和健听人群之间的交流障碍。因此,通过计算机来实现手语识别就显得格外重要。

手语识别涉及到人机交互、计算机视觉、图像处理、深度学习、模式识别及自然语言处理等多学科的综合研究领域^[1],是一项极具挑战性的综合性研究。从社会学的角度来看,手语识别通过计算机提供一种符合人类语言习惯的翻译方式,将手语翻

本文得到课题“基于计算机视觉的手语识别系统的设计与开发”(No. X2019037)资助。龚晓敏,男,1998年生,本科,主要研究领域为软件工程。E-mail: 1678691400@qq.com。鞠海军,男,1998年生,本科,主要研究领域为软件工程。E-mail: 1433493077@qq.com。

译成本文或语音,能够清晰地表达出手势的含义。

从上世纪六十年代以来,研究者们就对手语识别技术展开研究。1983年,美国格莱姆斯通过“数字数据手套”利用传感器等硬件设备可以识别72个简单字母^[2];1992年,Starner以摄像头作为输入设备,让使用者佩戴颜色手套作为辅助手段,在此种情况下可达到99.92%的识别率^[3]。在2017年,刘小健和张元提出利用深度学习基于多特征提取来完成对手语识别的检测,达到了最高接近94%的识别率^[4]。

2 手语识别的算法研究

纵观手语识别技术的发展历程,对手语识别的研究主要经历过三个阶段。从输入的获取方式来看,可分为三大类,分别为基于数据手套、基于计算机视觉和基于其他的数据装置。从最初的“数据手套”依赖硬件设备来获取手语信息,到之后的“手势标记”利用高亮的视觉手套来获取手势信息^[4],到现在的利用深度学习的技术来获取图像作为计算机系统的输入,手语识别技术在其发展中既经历过瓶颈也迎来了巨大的机遇。

2.1 基于数据手套的手语识别

在手语识别技术发展初期,主要利用硬件设备通过传感器技术来获取手势信息作为系统的输入,计算机系统通过数据手套提供的手型和其弯曲程度等数据辨识出手势含义,呈现给用户。从二十世纪七十年代起,就有多家研究机构专注于数据手套的研究,如Zimmerman在1978年研制出的Sayre Glove和Kramer在1989年研制的Cyber Glove等^[5]。2018年尹舒冉在数据手套的基础上^[6],利用模式匹配算法使得手势识别率达到了93.3%,在识别准确率上达到了一定的精度。但是,由于这种方式过于依赖硬件设备,数据手套容易受温度和光照等环境因素影响,获取的手势信息存在较大误差,且可识别的手势种类数量较少,识别率达不到理想的状态。同时,对于一些手势变化不明显的微手势,其识别方法和结果都达不到预期的效果。因此,由于数据手套对于手语的识别存在一定的局限性,未能得到社会各界的广泛应用。

2.2 基于视觉的手语识别

随着计算机视觉的发展,研究人员发现利用视觉分析来对手势信息进行捕捉,能够迅速精确地对

手势完成检测识别。2013年,姚远等人设计了一个基于视觉的开发框架^[7],采用RGB-D信息作为此框架的信息输入,大大提高了计算机视觉对于复杂手势的识别率。2017年,党佩光和李军恒通过高清摄像头对手势动作进行动态捕捉,然后进行帧分离,形成不同的手势图像^[8]。将获得的图像与计算机中的识别算法进行比对处理,将处理所得的信息返回给用户,既提高了手势识别的精度,也在识别速度上有了了一定的提升。诚然,基于视觉的手语识别解脱了对于硬件设备的束缚,让人的注意力集中在人机交互中。但是,基于视觉算法的识别过程十分复杂,而且对于图像的清晰度要求极高,易受光照等环境影响,并且需要提前在计算机中输入大量的手势样本作为手势检测的数据集。因此,基于视觉的手语识别需要对图像进行较好的处理之后才能够得到一个满意的结果。

2.3 基于深度学习的手语识别

由于手势含义丰富且又具有随意性,而且容易受温度、手势背景、运动速度等影响,基于机器学习模式的手语识别的精确度和鲁棒性已经达到了一个顶峰。人们便转而运用深度学习的方式来进行手语识别。2020年,张淑军等人提出手语识别可分为传统的识别方法和基于深度学习的方法,利用循环神经网络(RNN,善于长时间捕获上下文语义)将手语视频中的特征手势进行建模,在孤立词手语集上具有较高的识别率^[9]。基于深度学习的手语识别在精确度和鲁棒性方面都表现出良好的性能。但是对于残障人士来说,运用这一技术还存在一定的困难。

3 手语识别中去噪算法的研究

基于视觉的手语识别,通常需要经过三个过程。首先,通过图像采集设备获取图像;然后,对获取到的图像进行去噪,再进行一系列的图像处理,然后将手势分离;最后,对图像进行建模,根据模型与语义的匹配结果,以语音或文字的形式输出。其识别流程如图1所示。

在图像的采集过程中,所得到的图像有时候并不是理想的状态,会因为环境等因素的影响导致图片中存在一些噪声,而噪声会对后期图像的处理产生很大的影响。因此在基于视觉的手语识别中,通过消除图像的噪声从而提高图像的质量,是很有必要的。本文选择基于偏微分方程的图像去噪算法为

图像处理过程中的去噪算法^[10]，算法可表示为：

$$\begin{cases} \frac{\partial u}{\partial t} = \text{div}[g(|\nabla u|)\nabla u] \\ u(x,y,0) = u(x,y); t \in (0,T) \end{cases}$$

(1)

式（1）中， ∇u 中是梯度算子， $g(|\nabla u|)$ 表示扩散系数函数； div 是散度算子； $u(x,y,0)$ 是随时间变化的图像；其中 $g(|\nabla u|)$ 具有如下性质：

$$g(|\nabla u|) = \frac{1}{1+(|\nabla u|/k)^2}$$

(2)

根据式（1）和式（2）易知，边缘保持是因为梯度在边缘处扩散的系数增大。而基于偏微分方程的去噪算法通过控制扩散系数，可以很好地保持边缘。

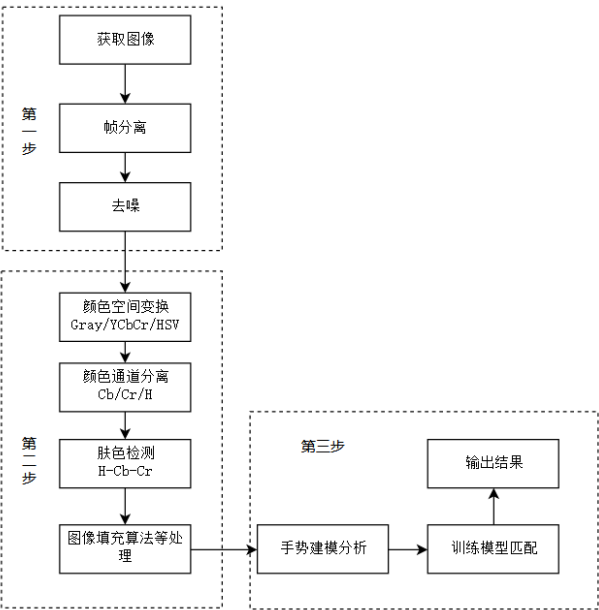


图 1 手语识别流程

为了进一步确定基于偏微分方程的图像去噪算法是否能够满足手语识别对图像去噪的要求，进行了如下实验。

3.1 实验环境

实验平台为 NS3 仿真工具。为确保实验可行性，本文通过对 5 幅大小为 512*512 的图像叠加高斯白噪声，并以峰值信噪比（以下简称 PSNR）及结构相似性（以下简称 SSIM）来研究算法去噪效果^[11-12]。

为了验证不同去噪算法去噪效果的有效性，将实验图像设置标准方差为 40 的高斯白噪声，研究 4 种算法对带有噪声的图像去噪效果，另外三种方

法分别为基于小波阈值的去噪算法^[13]、基于小波变换的图像去噪算法^[14]、非采样下轮廓波变换快速算法^[15]。实验结果如表 1 所示。

为了验证基于偏微分方程的去噪算法在边缘保持上的性能，对上述提到的五幅图片分别加以大小为 20dB、30dB、40dB 的噪声强度，并分析其去噪性能。将其与软阈值算法^[16]、硬阈值算法^[17]进行比较，去噪图片效果对比如图 2 所示，去噪效果如表 2 所示。

3.2 实验结果

表 1 4 种方法对高斯白噪声的 PSNR 及 SSIM 对比

测试图像	基于偏微分方程的图像去噪算法		基于小波阈值的去噪算法	
	PSNR		SSIM	
	PSNR	SSIM	PSNR	SSIM
1	26.205	0.204	22.807	0.227
2	23.608	0.415	21.636	0.356
3	24.66	0.32	22.058	0.264
4	24.025	0.387	21.769	0.235
5	23.802	0.349	21.588	0.296

测试图像	基于小波变换的图像去噪算法		非采样下轮廓波变换快速算法	
	PSNR		SSIM	
	PSNR	SSIM	PSNR	SSIM
1	21.044	0.157	19.129	0.146
2	20.087	0.269	18.73	0.255
3	20.773	0.191	19.007	0.182
4	20.226	0.224	18.873	0.237
5	20.246	0.213	18.828	0.214

表 2 不同噪声强度下的去噪效果

噪声强度 /dB	基于偏微分方程的图像去噪算法	软阈值算法	硬阈值算法
	PSNR	SSIM	PSNR
20	28.28	27.04	22.53
30	31.52	29.52	24.22
40	34.13	33.05	25.33

3.3 实验结果分析

由表 1 可知，基于偏微分方程的图像去噪算法在实验中对 5 幅标准图像中的 PSNR 和 SSIM 均高于其他算法。由此可见，使用基于偏微分方程的图像去噪算法能够有效地去除图像中的噪声。

由图 2 可知，硬阈值算法去噪能力较差；软阈值算法在边缘保持上效果较差；而基于偏微分方程

的去噪算法在这两方面则都有较为优秀的性能。从表 2 实验数据对比可知, 基于偏微分方程的去噪算法性能都优于硬阈值算法和软阈值算法。

综上所述, 基于偏微分方程的图像去噪算法能有效去除噪声, 也能够很好地保持细节, 因此该算法是在手语识别中去噪过程的最佳选择。



图 2 去噪图片效果对比图

4 总结与展望

本文明确了手语识别的流程。通过对当前主流算法的对比分析发现: 基于偏微分的图像去噪算法在去噪性能上具有优异表现, 选择基于偏微分方程的图像去噪算法作为手语识别过程中的去噪算法。但在研究过程中也发现, 对图像进行去噪后的下一个环节图像填充在目前研究的算法中暂未发现时间性能较好的算法。因此接下来的研究重点将会放在如何选择和优化图像填充算法, 提高算法整体流程的运行效率, 从而达到实时识别的效果。此外, 也将通过对手部模型的进一步优化来提高手语识别的准确率。

参 考 文 献

- [1] 王民,郝静,要趁红,史其琦.基于优化全卷积神经网络的手语语义识别.激光与光电子学进展,2018,55(11):214-220
- [2] 张良国,陈熙霖.手语识别研究综述.信息技术快报,2009,7(3):1-12
- [3] 江勇军.基于Kinect的孤立词手语识别系统研究[硕士学位论文].中国科学技术大学,合肥,2015
- [4] 刘小建,张元.基于多特征提取和SVM分类的手势识别.计算机工程与设计,2017,38(04):953-958
- [5] 代雨锟.基于可穿戴数据手套的手势识别方法研究[硕士学位论文].华南理工大学,广州,2019
- [6] 尹舒冉.基于石墨烯柔性传感器的手势识别手套系统[硕士学位论文].重庆邮电大学,重庆,2018
- [7] 姚远,张林剑,乔文豹.RGB-D图像中手部样本标记与手势识别.计算机辅助设计与图形学学报,2013,25(12):1810-1817

- [8] 党佩光,李军恒.基于计算机视觉的手势检测识别技术.电脑迷,2017(11):39
- [9] 张淑军,张群,李辉.基于深度学习的手语识别综述.电子与信息学报,2020,42(04):1021-1032
- [10] 王洋.基于多尺度分析和偏微分方程的图像处理算法研究[博士学位论文].吉林大学,长春,2012
- [11] Welstead, Stephen T. Fractal and wavelet image compression techniques. Washington DC: SPIE Optical Engineering Press, 1999: 155-56
- [12] Wang, Zhou, Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13 (4): 600-612
- [13] 池明旻.小波软阈值算法在SAR图像去噪处理中的应用研究[硕士学位论文].厦门大学,厦门,2002
- [14] 关雪梅.一种基于中值滤波和小波变换的图像去噪处理算法研究.中州大学学报,2020,37(01):121-124
- [15] 赵春晖,郭蕴霖.一种快速的基于稀疏表示和非下采样轮廓波变换的图像融合算法.电子与信息学报,2016,38(07):1773-1780
- [16] 丁西明,段汉根.基于软阈值的图像增强.河北北方学院学报(自然科学版),2011,27(04):28-31
- [17] 张守成,张玉洁,刘海生.一种改进的EMD硬阈值去噪算法.计算机测量与控制,2014,22(11):3659-3661