

基于OCR的视力障碍人群头戴式阅读器的研究

潘惠德, 刘明良, 任爽, 胡校, 韦新宇, 吴梦玲, 梁豫鄂

汉江师范学院数学与计算机科学学院, 湖北 十堰

收稿日期: 2024年2月13日; 录用日期: 2024年3月12日; 发布日期: 2024年3月20日

摘要

本文致力于解决视力障碍人群在阅读文档或书籍时面临的困难。通过整合先进的OCR图像文字识别技术、光学设计中的光源和透镜系统、智能的移动系统以及灵敏的光电传感器, 以此研究并设计出一款独特的视力障碍人群阅读器。该阅读器不仅通过OCR技术将印刷材料转化为可听的语音, 还在光学和移动系统方面做出了创新。设计的过程中不仅仅是技术的整合, 更是一种关注和解决视力障碍人群实际需求的努力。通过提供更便捷的阅读工具, 达到改善视力障碍人群的生活质量的目的, 并在社会中产生积极影响。

关键词

文字识别, OCR, 视力障碍, 阅读器, 语音合成

OCR-Based Head-Mounted Reader for Visually Impaired People

Huide Pan, Mingliang Liu, Shuang Ren, Xiao Hu, Xinyu Wei, Mengling Wu, Yu'e Liang

School of Mathematics and Computer Science, Hanjiang Normal University, Shiyan Hubei

Received: Feb. 13th, 2024; accepted: Mar. 12th, 2024; published: Mar. 20th, 2024

Abstract

This paper is dedicated to solving the difficulties faced by visually impaired people when reading documents or books. By integrating the advanced OCR image and text recognition technology, the light source and lens system in optical design, the intelligent mobile system, and the sensitive photoelectric sensors, in this way, a unique reader for visually impaired people is researched and designed. The reader not only converts printed materials into audible speech through OCR technology, but also makes innovations in the optical and mobility systems. The design process is not just an integration of technology, but an effort to focus on and address the actual needs of the vi-

sually impaired. By providing a more convenient reading tool, it achieves the goal of improving the quality of life of visually impaired people and creating a positive impact in society.

Keywords

Text Recognition, OCR, Visually Impaired, Reader, Speech Synthesis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

视力是人类获取信息和与世界进行互动的重要感知方式之一。然而，对于视力障碍人群来说，阅读成为了一项具有挑战性的任务。据世界卫生组织(World Health Organization)估计，全球约有 2.2 亿人面临不同程度的视力障碍。视力障碍包括失明和低视力，以及其他视觉障碍，如色盲和视野缺损。这些障碍对视力障碍人群的日常生活、学习和就业产生了重大影响。

来自调查分析十堰市城区学龄前儿童弱视流行病学特点。1248 例学龄前儿童中，视力低下率为 13.70%；弱视发病率为 3.21%；而 3 岁组的视力低下率为 30.77%，显著高于其他年龄组[1]，从数据中看出视力障碍的情况仍是不容乐观。可见全球仍有大量的视力障碍人群，且长期处于递增状态，调查研究发现视力障碍也会对生活的多个方面产生重大影响。

图书馆	实体无障碍资源数量(件)
天津图书馆	1029
石家庄市图书馆	1300
秦皇岛市图书馆	5800
沧州市图书馆	100
上海市图书馆	4000
南京图书馆	3000
南通市图书馆	661
济南市图书馆	1000
青岛市图书馆	400
深圳图书馆	3000
广州图书馆	3617
山西省图书馆	300
郑州图书馆	1999
黄石图书馆	300
湖南省图书馆	3000
内蒙古图书馆	264
通辽市图书馆	300
贵州省图书馆	878
陕西省图书馆	1500
辽宁省图书馆	4000
哈尔滨市图书馆	114
齐齐哈尔市图书馆	580

Figure 1. Statistics on the number of physical accessibility resources

图 1. 实体无障碍资源数量统计

我国公共图书馆无障碍资源数量的统计数据来源于调研中获取到的 22 个图书馆的实体无障碍资源数量, 主要包括盲文、大字版图书以及有声读物, 不包括各类网络资源。首先, 馆际间的实体无障碍资源数量差距大, 从几百到几千不等; 其次, 无障碍资源馆藏量在 2000 件/册以下的占一半以上, 而在 4000 件/册以上的寥寥无几。无障碍资源是图书馆视障者阅读服务的基础, 资源的匮乏成为制约图书馆视障者阅读服务的关键问题[2]。如图 1 所示。

现有的视障人士阅读辅助工具虽然有益, 但在几个关键领域存在不足, 包括光学字符识别(OCR)的准确性有限、对不同文本格式和语言的适应性不足, 以及人体工程学和可用性方面的挑战。研究旨在通过开发一种头戴式阅读设备来弥合这些差距, 该设备利用先进的 OCR 技术, 集成了创新的光学设计和增强移动性的功能。通过关注实时文本到语音转换的准确性、对多种语言和复杂布局的支持以及以用户为中心的设计, 本研究不仅旨在增强视障用户的阅读体验, 而且为辅助技术领域做出重大贡献。通过技术创新和对用户需求的深刻理解, 致力于显著改善视障人士的生活质量, 为他们提供更大的独立性和获取印刷信息的机会。

2. 系统设计

2.1. 总览系统设计

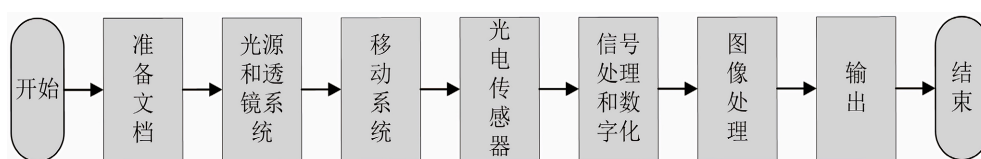


Figure 2. Flowchart showing the overview of the reader system for the visually impaired population
图 2. 视力障碍人群阅读器系统总览流程图

视力障碍人群阅读器系统逻辑流程图如图 2 所示。开始: 启动阅读器系统; 准备文档: 用户放置待扫描的文档或纸质书籍; 光源和透镜系统: 阅读器打开光源, 透过透明玻璃或传感器照亮文档表面。提高文本与环境的差异提高文字识别的准确性; 移动系统: 阅读器的移动系统移动摄像头, 以便涵盖整个文档表面。因为考虑到使用者是视力障碍人群, 该系统设计可以语音提醒使用者移动文档或系统移动摄像头以达到对文档尽可能的识别完整; 光电传感器: 光电传感器检测文档上反射回来的光线; 信号处理和数字化: 使用 OCR 文字识别系统中的信号处理单元将光电传感器输出的模拟信号转换为数字信号; 图像处理: 数字信号经过图像处理单元, 进行校正、去噪和其他图像处理步骤; 输出: 处理后的数字图像通过接口(USB、Wi-Fi、蓝牙等)传输到计算机或其他设备; 结束: 扫描完成, 系统处于就绪状态。

阅读器系统可以搭载移动设备上, 这样可以使阅读器有便携性、低成本等特点。移动设备的特点之一是便携性。用户可以携带自己的阅读器, 使其能在不同场合使用, 可以在不同环境中阅读文档。移动设备通常具有较强的算力资源, 阅读器可以仅搭载摄像头模块, 然后通过对移动设备有线或无线设备连接, 将对算力要求较大的部分交由移动设备完成, 以节省阅读器额外的算力成本。移动设备还有较高的连结性, 可以搭载云服务, 可以使得用户在多个设备之间同步和分享阅读器系统的数据和设置。除去云服务, 移动设备可以连接多个设备如语音播放设备——将文字内容以语音方式输出, 屏幕——将文本文档等较小的字体以更大的形式展现。

2.2. 系统设计

2.2.1. OCR 技术的应用

OCR 技术是一种将图像或文档中的印刷或手写字符转换为可编辑文本的技术, 主要原理设计图像获

取、预处理、文本区域检测、文本分割、特征提取、字符识别、后处理最终输出文本。传统的 OCR 识别流程如下图 3 所示。首先获取包含文本的图像，这部分交由阅读器外置部分获取，由于获取的图像在一定程度上可能包含噪音、失真或其他干扰，所以需要获取的图像进行预处理，采用灰度转换、二值化、去噪、图像增强等[3]，以提高后续字符识别的准确性。对处理后的图像需要版面分析确定图像中文本的区域，这一步可以采用计算机视觉技术边缘检测、轮廓检测等。在文本区域检测之后，需要将文本分割成单个字符。这一步骤是为了确保每个字符都能独立识别，而不受相邻字符的影响。在特征提取后，使用机器学习算法或模式匹配技术进行字符识别。这可能包括神经网络、支持向量机(SVM)等方法，用于训练模型以识别不同的字符。之后将识别出的字符按识别前的版面恢复。一些 OCR 系统可能会进行后处理步骤，如纠正错误或补充缺失的字符。这有助于提高整体识别的准确性[4]。

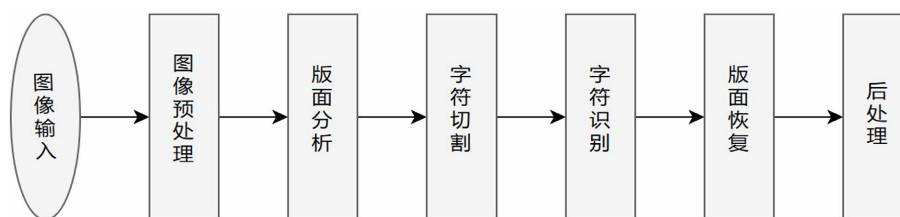


Figure 3. Traditional OCR Recognition Process

图 3. 传统 OCR 识别流程

2.2.2. 光源和透镜系统的设计

光源选择和布局，本文考虑到阅读器的多环境使用情况，光源需要均匀性、可调性、稳定并具有较低的反射性。所以光源类型选择冷阴极荧光灯，它具有体积小、亮度高、寿命长等特点，但它工作前需要预热。第二类是发光二极管，它的优点在于功耗小、噪音低、发热量小且无需预热，但亮度低，亮度均匀度较差[5]。第三类是卤素灯光源，他的亮度很高，且预热时间短更换容易，但发热量较高。从亮度、寿命、预热时间等方面，本设计采用冷阴极荧光灯，在后续的使用过程中，我们将会根据用户体验反馈调整光源。

透镜系统设计是为了实现特定的光学效果包括聚焦、成像、色彩校正等，透镜系统主要分为两部分，一部分是添加在光源前，将光源调整为更均匀、更稳定的光源。一部分是添加在摄像头前，将光线汇聚或分散来形成图像。对于凸透镜，它可以将平行光线汇聚到焦点形成实像；对于凹透镜，光线分散后看似汇聚在一个虚拟焦点形成虚像。透镜系统调整倍率，可以将或远或近的文档放大与缩小，保证文档文字的完整性。光圈控制透镜的入射光量，影响曝光和景深。光圈大小由透镜的开口控制。透镜系统的景深是指图像中各点的清晰范围。通过调整光圈大小和焦距可以改变景深。综上，本设计采用多透镜，相互协同工作以实现更复杂的光学效果。

2.2.3. 移动系统的工作原理

移动系统负责移动摄像头或传感器，其文档文本能够居于识别正中间，确保文档的完整性并附带稳定系统，保证拍摄图片的清晰，从而完成对文本的扫描。对于移动系统，因为本设计的便携性，所以移动系统不能太大，所以不能考虑相机的三轴框架，因为它会占用较大的体积，故本设计采用类似与市面上口袋云台设计成像模块分离布局，即将 CCD 底板和镜头安装在云台上，让其信号通过排线与相机主板连接，以实现云台体积的最小化，而且镜头规格基本使用较小的板机或者手机这种尺寸的镜头，通过 MIPI 通信输出。

移动系统中电机也是较为重要的部分，使用不同与航模上的不同绕线方式，采用星形，并选用内阻

较大的漆包线(> 10 ohm)获取更平顺转动效果。商业化的云台稳定器电机闭环(FOC)是十分必要的,最常用的电机闭环的位置检测一般有电位器、非接触式磁编码编码(AMS 系列)、光电编码器、霍尔。为了更小小体积口袋云台使用了2个布局夹角为120°的SOT-23封装的霍尔元件,直接输出2个相位差90°的正弦波电压值。霍尔元件焊接在电机底部电路板上,通过HRS或者Panasonic这些品牌的板对板端子连接,当然霍尔是直接测量转子的电角度,当电角度 > 360°(一对极代表一个电角度)时,电机在每次开机时都需要自检机械限位。而磁编码通常采用的是外部磁铁,一圈360°机械角度同时也对应1个电角度。所以移动系统在开机的时候无需自检。

在OCR(光学字符识别)图像扫描过程中,文档图像总是在一定程度上倾斜放置,当倾斜度足够大时,会影响文档分析的效果并降低识别精度,因为布局分析和字符识别算法对页面倾斜非常敏感[6]。使用本系统足够解决这样的问题。

2.2.4. 光电传感器

光电传感器是利用光的各种性质,检测物体的有无和表面状态的变化的传感器,光电传感器主要由发光的投光部和接受光线的受光部构成[7]。如果投射的光线因检测物体不同而被遮掩或反射,到达受光部的量将会发生变化。受光部将检测出这种变化,并转换为电气信号,进行输出。大多使用可视光(主要为红色,也用绿色、蓝色来判断颜色)和红外光。通过光电传感器可以实现本设计在不同距离调整获取图像的能力[8]。

本设计的光电传感器采用红外光电传感器,通过红外发射器发射红外光束,然后用红外接收器接受被发射的红外光,并根据接收到的光的特性进行响应,最后红外接收器检测到光的变化,并将其转换为电信号。变化的模式可以用来检测物体的存在、距离、运动等。通过把文档文本的存在和距离传输给移动系统,使其调整好摄像部分,获取更清晰的图像,保证较准确的图像识别。

2.2.5. 信号处理和数字化与图像处理

信号处理与图像处理对阅读器的算力较高,为了节约成本并获取更加强大的算力资源,本设计将这部分交由移动设备完成,阅读器的外置部分任务主要是将图像拍摄的尽可能的清晰,并通过有线或无线(WiFi、蓝牙等)实现图像传输。

通过光电传感器获取的模拟信号是连续的波形,信号处理从这一阶段开始,需要将模拟信号转换为数字信号,并交给移动系统完整对焦等操作。到图像处理时,从图像获取、图像校正、去噪、增强、分割、OCR处理等一系列手段保证识别的准确性,阅读器获取图像,图像应采取适当的采样率和分辨率对于保留图像细节和确保准确的数字表示,在图像获取后,可能需要进行校正,以纠正扭曲、畸变或其他可能影响文字识别的问题。图像可能受到噪声的影响,去噪处理有助于提高文字识别的准确性。这包括去除图像中的杂色、斑点或其他干扰。图像增强技术有助于提高图像的对比度和清晰度,使文字更容易被OCR系统识别。文档中的文字可能存在于不同的区域,分割技术有助于将图像分割成单个字符或单词,以便进一步处理。最终,通过OCR引擎对处理后的图像进行处理,将图像中的文字转换为计算机可读的文本[9]。

2.2.6. 输出

将用户的实体文本文档识别为计算机可读的文本时,就需要输出为用户可以接受的文本,本设计将着重以两个形式输出,一是考虑到一部分视力障碍人群并不是完全丧失对光源的获取,只是因为现存的书籍文字过小而导致无法阅读,本设计将以更大的文字输出,二是完全丧失或难以获取文字信息的用户,本设计才用语音输出的方式。因为现今移动设备的普及与拓展性,这两种输出方式也可以交由移动设备

连接的设备完成，例如拓展屏幕、蓝牙音响、手机听筒等。

3. 系统界面设计及设计部分代码

用户界面设计

本研究设计的主要对象倾向于视力障碍人群，所以设计上注重简洁易用，尽可能让用户避免对图形界面的强依赖。通过语音提示向用户提供指导，或使用声控功能，集成语音控制，使用户能够通过声音指令执行一些操作，如开始阅读、暂停等。或采用手势控制。在整个设计过程中，理解用户需求，特别是视力障碍人群的特殊需求，并采用无障碍设计原则，以确保阅读器在实际使用中是易于操作、友好的。定期进行用户测试和反馈收集，以便不断优化和改进系统的设计。如下图 4 所示为本设计使用的用户界面，在进入使用之前，点击画面上任意一部分即可开启语音控制，并同过语音使用户知晓并同意。



Figure 4. User interface

图 4. 用户使用界面



Figure 5. Device connection interface

图 5. 设备连接界面

如上图 5 所示，将阅读作为类似外置摄像头或是音响等设备连接使用。

在 OCR 视力障碍阅读器中，用到了比较重要的算法和技术，阅读器对于实现有效的文本识别和阅读

辅助功能至关重要。以下是项目中被用到的关键算法：

光学字符识别(OCR)算法：

OCR 算法用于从图像中提取文本信息，将图像中的字符转换为可识别的文本。OCR 是整个阅读器的核心部分，决定了系统对图像中文本的准确性和效率。其部分代码如下：

```
from PIL import Image
import pytesseract
def ocr(image_path):
    """
    使用 Tesseract OCR 进行光学字符识别。
    Parameters:
    - image_path (str): 图像文件路径。
    Returns:
    - str: 识别到的文本。
    """
    img = Image.open(image_path)
    text = pytesseract.image_to_string(img)
    return text
```

图像预处理算法：

在进行 OCR 之前，对输入图像进行预处理，可能包括图像增强、降噪、二值化等操作。良好的图像预处理能够提高 OCR 算法的稳定性和准确性，尤其对于原始图像质量较差的情况。其部分代码如下：

```
import cv2
def image_preprocessing(image):
    # 图像增强、降噪、二值化等预处理步骤
    # ...
    return preprocessed_image
```

自然语言处理(NLP)算法：

NLP 算法用于处理 OCR 输出的文本，可能涉及分词、词性标注、语法分析等任务。通过 NLP 算法，系统可以更好地理解文本内容，提高对用户的友好程度，例如更自然的语音输出。其部分代码如下：

```
import nltk
def nlp_processing(text):
    """
    使用 NLTK 进行自然语言处理，包括分词、词性标注等。
    Parameters:
    - text (str): 输入文本。
    Returns:
    - processed_text: 处理后的文本。
    """
    tokens = nltk.word_tokenize(text)
    pos_tags = nltk.pos_tag(tokens)
    # 其他 NLP 处理步骤
    # ...
    return processed_text
```

移动设备上的实时图像处理算法：

适应移动设备的实时图像处理算法，确保在移动设备上实时且有效地处理摄像头捕获的图像。对于

头戴式阅读器等移动设备，实时性是关键，算法需要在有限的计算资源下高效运行。其部分代码如下：

```
import cv2
def real_time_image_processing(frame):
    """
    对实时图像帧进行处理。
    Parameters:
    - frame: 输入图像帧。
    Returns:
    - processed_frame: 处理后的图像帧。
    """
    # 实时处理图像帧
    # ...
    return processed_frame
```

语音合成算法：

语音合成算法用于将文本转换为自然语音输出，使得用户能够听到文本内容。语音合成提供了对视力障碍用户友好的交互方式，是提升阅读器易用性的关键因素。其部分代码如下：

```
from gtts import gTTS
import os
def text_to_speech(text, language='en'):
    """
    使用 gTTS 进行文本到语音合成。
    Parameters:
    - text (str): 输入文本。
    - language (str): 语音合成的语言，默认为英语。
    Returns:
    - None
    """
    tts = gTTS(text, lang=language)
    tts.save("output.mp3")
    os.system("start output.mp3")
```

4. 关于创新性的讨论

本设计与当前视力障碍人群中使用的产品相比更具有创新性。集成有先进的 OCR 技术：提高 OCR 的准确性和速度，将印刷文本转换为语音，专为视障人士量身定制。以用户为中心的基础上设计：在用户界面与交互方面上进一步创新，为满足视障用户的需求，使技术更易于访问和用户友好。移动性和无障碍解决方案：开发可在各种环境中使用的便携式和易于使用的设备，提高视障人士的移动能力。光学和信号处理改进：在光源、镜头系统设计和信号处理方面进行创新，以提高不同照明条件下的读数精度。

本研究的核心创新之处在于开发了一种专为视障人士设计的阅读辅助技术，该技术在多个方面显著优于市面上现有的文档扫描设备，如科密(Comet) GP-1000 高拍仪。科密 GP-1000 提供了 1400 万像素的分辨率和基本的 OCR 文字识别功能，主要面向一般办公应用。然而，本研究通过采用更高精度的图像捕获技术和先进的 OCR 算法，大幅提升了文本识别的准确性和速度，尤其是在处理多样化文本和复杂布局时的表现，更加符合视障人士的阅读需求。

进一步地，与 GP-1000 的标准用户界面相比，本研究的设计重点优化了用户交互体验，引入了语音反馈和触觉指引功能，使得视障用户能够更直观、更方便地操作设备。此外，考虑到视障人士的特殊需求，设计了更为轻便和便携的形态，确保用户能够在不同环境下，包括光线不足的情况下也能高效使用。

本设计的技术还突破了兼容性限制，不仅支持国产系统，还能与多种操作系统和辅助技术设备无缝集成，进一步拓宽了产品的应用范围。这种高度的适配性和兼容性，特别是与屏幕阅读器等辅助设备的集成，为视障人士提供了前所未有的阅读体验。

综上所述，通过在图像处理精度、用户体验设计、便携性、以及系统兼容性等方面的创新，本研究所开发的阅读辅助技术显著优于科密 GP-1000 高拍仪，特别是在满足视障人士特定需求方面展现了巨大的潜力和价值。

5. 结论

在基于 OCR 的视力障碍人群阅读器的设计与实现中，通过采用 OCR 技术，成功地实现了对文档和书籍等文字场景的快速识别，为视力障碍人群提供了更便捷的阅读体验。研究结果表明，该阅读器在提高文本识别准确性和实时性方面取得了显著的进展。

然而，研究也有一些局限性。首先，系统在处理手写文本和特定字体时可能表现不如预期。此外，阅读器对于复杂的文档格式和布局的适应性仍有改进的空间。

未来的研究方向可以包括进一步优化 OCR 算法，改进系统的用户界面和交互体验，以及扩展阅读器的适用范围，例如应用于教育领域。

综上所述，研究为解决视力障碍人群面临的阅读困难问题提供了一个切实可行的解决方案，同时也为相关领域的进一步研究提供了有益的启示。

基金项目

汉江师范学院 2023 大学生创新创业省级项目：基于 OCR 的视力障碍人群头戴式阅读器的研究与设计，编号：S202310518009。

参考文献

- [1] 邵杰, 彭承斌, 梁永霞, 等. 十堰城区学龄前儿童弱视状况调查[J]. 湖北医药学院学报, 2015, 34(6): 556-558.
- [2] 张曼曼, 范并思. 我国公共图书馆视障者阅读服务研究[J]. 图书情报工作, 2017, 61(13): 31-36.
- [3] 王佳. 基于图像处理的无人机姿态测量技术研究[D]: [硕士学位论文]. 南京: 南京航空航天大学, 2019.
- [4] 刘树春, 贺盼, 马建奇, 王佳军, 等. 深度实践 OCR 基于深度学习的文字识别[M]. 北京: 机械工业出版社, 2020: 5.
- [5] 曹菁, 陈康, 齐宁, 等. 基于 OCR 和图像检测的盖章文书图像自动审核方法[J]. 应用科学学报, 2023, 41(6): 1058-1067.
- [6] 双精准示范专业建设教材编写组. 工业机器人基础操作与编程[M]. 北京: 化学工业出版社, 2021: 4.
- [7] 周奇才. 中国战略性新兴产业研究与发展·智能化立体车库[M]. 北京: 机械工业出版社, 2021.
- [8] 邹建生. 基于机器视觉系统对磁瓦进行在线检测和分级的研究[D]: [硕士学位论文]. 江门: 五邑大学, 2010.
- [9] 沈旭, 曾竟明, 王新政. 基于 OCR 文档扫描与语音阅读技术的记事本应用研究与实现[J]. 软件, 2021, 42(6): 126-129+132.