



Pontifícia Universidade Católica de Minas Gerais
Campus Poços de Caldas
Curso de Ciência da Computação

Proposta de Projeto

Reconhecimento Óptico de Caracteres (OCR) em Imagem para Conversão em Texto com Python

Visão Computacional e Realidade Misturada

Alunos: Leonardo Vilas Boas Mendes, Vinicius Machioni

Orientador: Will Machado

02/2024

1 Resumo do Projeto

No cotidiano, a necessidade de converter documentos impressos, anotações manuscritas ou textos em imagens para formatos digitais é uma tarefa comum, mas que demanda tempo e esforço. Este projeto propõe o desenvolvimento de uma aplicação de Reconhecimento Óptico de Caracteres (OCR) utilizando a linguagem de programação Python para automatizar esse processo. A aplicação visa facilitar a digitalização de textos, reduzindo o tempo para transcrever documentos manualmente e minimizando erros, utilizando técnicas de visão computacional. A solução irá proporcionar uma maior produtividade para usuários em ambientes de trabalho, acadêmico ou pessoal, onde a conversão rápida e precisa de texto é essencial.

2 Objetivo

Este projeto tem como objetivo desenvolver uma aplicação de Reconhecimento Óptico de Caracteres (OCR) utilizando a linguagem de programação Python para a conversão de texto presente em imagens em um formato digital editável. Visando oferecer uma solução eficiente para a digitalização de textos utilizando técnicas de visão computacional e aprendizado de máquina para lidar com alguns desafios, como baixa qualidade de imagem, ruídos e variações de iluminação.

3 Detalhamento das Atividades

3.1 Revisão Bibliográfica

- **Objetivos:** O objetivo dessa etapa é realizar uma revisão da literatura existente para entender as pesquisas anteriores relevantes e as abordagens existentes para problemas semelhantes.

3.2 Metodologia

3.2.1 Aquisição da Imagem

- **Objetivos:** O objetivo dessa atividade é realizar a aquisição da imagem.
- **Descrição do trabalho:** Aquisição da imagem por meio de *upload* na aplicação.
- **Ferramentas/Recursos a serem utilizados:** Imagem em formato PNG (*Portable Network Graphic*) e JPEG (*Joint Photographic Experts Group*).

3.2.2 Pré-processamento da Imagem

- **Objetivos:** O objetivo dessa atividade é realizar o pré-processamento da imagem coletada para transcrever os caracteres encontrados.
- **Descrição do trabalho:** Melhorar qualidade da imagem, conversão para escala de cinza, remoção de ruído e ajuste de contraste.
- **Ferramentas/Recursos a serem utilizados:** Linguagem Python, OpenCV.

3.2.3 Desenvolvimento da Aplicação OCR

- **Objetivos:** O objetivo dessa atividade é implementar uma aplicação de reconhecimento de caracteres.
- **Descrição do trabalho:** Implementar e integrar o sistema de OCR com o módulo de pré-processamento de imagem.
- **Ferramentas/Recursos a serem utilizados:** Biblioteca Tesseract.

3.2.4 Testes e Validação

- **Objetivos:** Realizar testes de funcionalidades para garantir que o OCR reconheça corretamente textos em diferentes condições, utilizando imagens de teste.
- **Descrição do Trabalho:** Avaliar precisão e eficiência da aplicação, e se existir, corrigir bugs.

3.2.5 Geração de Arquivo de Texto

- **Objetivos:** O objetivo dessa atividade é a partir do reconhecimento dos caracteres, gerar para o usuário um arquivo de texto editável, contendo as palavras que foram transcrevidas. O arquivo gerado será um arquivo de formato .txt.

4 Implementação Original Tesseract

4.1 Visão Geral da Biblioteca

Este projeto fará uso da biblioteca Tesseract com implementação referência. O Tesseract é uma biblioteca de OCR de código aberto da Google e é amplamente utilizada. É eficaz para reconhecer textos em uma grande variedade de idiomas e scripts.

4.2 Principais Recursos

- **Suporte Multilíngue e Personalização:** A biblioteca é capaz de reconhecer textos em mais de 100 idiomas, com a opção de treinar novos modelos para fontes e idiomas específicos.
- **API Simples e Flexível:** Oferece uma API compatível com várias linguagens de programação, por exemplo, com Python via *pytesseract*. O que acaba facilitando a integração em diversas aplicações.
- **Deteção de Layout e Direção de Texto:** Realiza a identificação automaticamente do layout de páginas complexas, como documentos com múltiplas colunas e imagens embutidas, além de reconhecer textos em diferentes orientações (vertical, horizontal, etc.).

5 Adaptação da Implementação Referência

O projeto utilizará como base o Tesseract OCR para o reconhecimento de texto, adaptando novas funcionalidades para aumentar a flexibilidade e acessibilidade da aplicação. As adaptações visam proporcionar ao usuário uma experiência mais rica e funcional.

- **Escolha do Tipo de Arquivo de Saída:** O usuário poderá selecionar o formato de saída para o texto transcrito, como PDF, TXT, DOCX ou HTML.
- **Conversão de Texto para Fala:** Será incluída uma opção de conversão de texto reconhecido em áudio, utilizando uma voz sintetizada (*Text-to-Speech*). Isso permitirá ao usuário ouvir o conteúdo transcrito, aumentando a acessibilidade para pessoas com deficiência visual ou com dificuldades de leitura.
- **Deteção de Idioma:** Ao ser detectado o idioma do texto, o usuário contará com uma opção de traduzir o texto transcrito para o idioma desejado.
- **Interface de Usuário Amigável e Interativa:** Uma interface gráfica intuitiva (GUI) para facilitar o carregamento de imagens, configurações de saída do texto, e interação com o sistema, tornando a aplicação mais acessível para usuários não técnicos.

6 Bibliografia

Gonzalez and Woods (1992) GONZALES, R. C. WOODS, R. E. Digital Image Processing. University of Tennessee Perceptics Corporation, 1992.

Castleman (1996) Castleman, Kenneth R. Digital Image Processing. Upper Saddle River: Prentice Hall, Inc. 1996.

MoisesTedeschi. "Python/Transcrição_de_imagem_em_texto_com_Python_e_Tesseract_OCR.i

pyb at Master · MoisesTedeschi/Python.” *GitHub*, 2016,
github.com/MoisesTedeschi/Python/blob/master/Transcri%C3%A7%C3%A3o_de_imagem_e
m_texto_com_Python_e_Tesseract_OCR.ipynb. Accessed 11 Sept. 2024.