

Leonardo Yuji Fuiguti
Pedro Henrique Fuzário Custódio
Orientador: Maximilian Jaderson de Melo
Orientador: Guilherme Figueiredo Terenciani

Reconhecimento de Textos e sua Sintetização em Linguagem Natural Humana

INÍCIO: 01/04/2018
TÉRMINO: 01/04/2019
IFMS *campus* Naviraí
Rua Hilda 203 Conjunto Habitacional Boa Vista
2018

Leonardo Yuji Fuiguti
Pedro Henrique Fuzário Custódio
Orientador: Maximilian Jaderson de Melo
Orientador: Guilherme Figueiredo Terenciani

Reconhecimento de Textos e sua Sintetização em Linguagem Natural Humana

Relatório do projeto de pesquisa apresentado a 17ª feira de Ciências e engenharia FEBRACE.

Instituto Federal de Mato Grosso do Sul – IFMS *Campus* Naviraí
Educação Profissional de Nível Médio Integrado em Informática para Internet

INÍCIO: 01/04/2018
TÉRMINO: 01/04/2019
IFMS *campus* Naviraí
Rua Hilda 203 Conjunto Habitacional Boa Vista
2018

Lista de ilustrações

Figura 1 – Total de pessoas com deficiência visual por milhão de pessoas	13
Figura 2 – Comportamento da tangente hiperbólica	19
Figura 3 – Imagem original vs ROI detectada e caracteres segmentados.	27
Figura 4 – Segmentação dos caracteres.	28
Figura 5 – Visão geral do <i>software</i>	28

Lista de tabelas

Tabela 1 – Cronograma de atividade de 2018.	24
Tabela 2 – Cronograma de atividade de 2019.	25
Tabela 3 – Materiais necessários para a construção do projeto	25

Lista de abreviaturas e siglas

ROI	Região da Imagem que Apresenta Texto
IBGE	Instituto Brasileiro de Geografia e Estatística
OMS	Organização Mundial de Saúde
CMOS	Complementary Metal Oxide Semiconductor
OCR	Optical Character Recognition
PDI	Processamento Digital de Imagens
KNN	K-Nearest Neighbor
CNN	Convolutional Neural Network
SVM	Support Vector Machine

Sumário

1	INTRODUÇÃO	13
2	OBJETIVOS E RELEVÂNCIA DO TRABALHO	17
2.1	Objetivo geral	17
2.2	Objetivos específicos	17
2.3	Relevância do trabalho	18
2.4	Fundamentação teórica	18
2.4.1	Processamento digital de imagens	18
2.4.2	Rede neural artificial	19
2.4.3	Extração de características	19
2.4.4	Revisão da literatura	20
3	DESENVOLVIMENTO DO PROJETO	23
3.1	Materiais e métodos	23
3.2	Procedimentos	24
3.3	Cronograma	24
3.4	Custos	25
4	RESULTADOS	27
4.1	Resultados experimentais	27
	Considerações finais	29
4.2	Conclusão	29
	REFERÊNCIAS	31

Resumo

Este trabalho de pesquisa propõe o desenvolvimento de uma ferramenta computacional nos moldes de um produto dedicado ao apoio de pessoas com deficiência visual ou pessoas analfabetas. A ferramenta objetiva a identificação de textos e sua sintetização em formato de áudio, facilitando e incentivando assim, o hábito da leitura para esse perfil de pessoas.

Visto a dimensão do problema, nesta primeira parte, é apresentado um reconhecedor de caracteres em imagens utilizando a ferramenta computacional *Tesseract* que usa o OCR para identificação de caracteres em imagens e o sintetizador de voz *E-speak* para gerar o áudio dos caracteres reconhecidos pela primeira ferramenta.

Em continuidade ao trabalho apresentado, será feita a própria ferramenta para obtenção dos caracteres da imagem utilizando Redes Neurais, e será comparada com a *accuracy* da ferramenta *Tesseract*. Visando assim ter um parâmetro de excelência da ferramenta proposta.

No desenvolvimento do software será utilizada a linguagem de programação *Python*, a qual irá integrar diversas ferramentas de visão computacional, inteligência artificial e sintetização de voz. Ao fim do desenvolvimento de ambas as partes, será desenvolvido um protótipo funcional visando um teto de gastos de R\$400,00 reais e disponibilizado para voluntários para testes. Avaliando assim a qualidade do produto proposto.

Palavras-chave: Processamento Digital de Imagem; Reconhecedor de textos; Reconhecedor Ótico de Caracteres; Acessibilidade.

1 Introdução

No Brasil existem, segundo dados do IBGE 2010, aproximadamente 6,5 milhões de pessoas com algum grau de deficiência visual sendo 528.624 pessoas que não tem 100% da visão. No gráfico da Figura 1 podemos observar que a cada um milhão de pessoas, 7.283 mil tem algum tipo de deficiência visual. Com ainda mais preocupação, quando olhamos para a Figura 1 em relação as pessoas que tem mais de 60 anos, vemos que a proporção das pessoas com algum tipo de deficiência visual aumenta, chegando a aproximadamente 11,5%.

Em evidência desses fatos, grande parte dessas pessoas tem certa dificuldade em realizar grande parte das atividades cotidianas. Tarefas como: navegar pela internet, interagir com o celular, procurar itens de vendas em supermercados, ler placas de anúncios pela cidade, leitura de livros, entre outras. Atividades que para a maioria das pessoas seriam consideradas simples, para deficientes visuais acabam se tornando uma tarefa complicada.

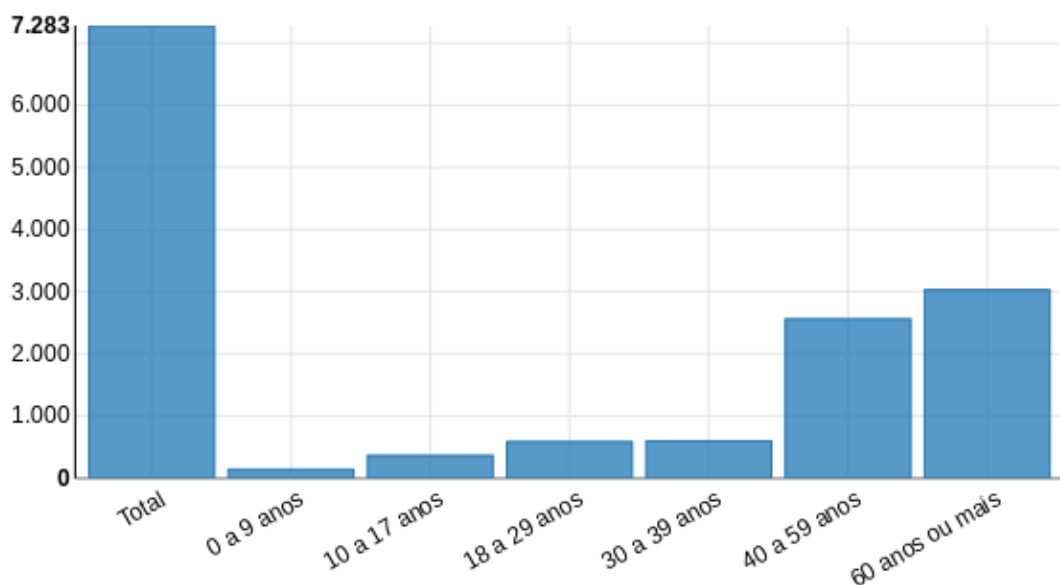


Figura 1 – Total de pessoas com deficiência visual por milhão de pessoas

Fonte: IBGE - Pesquisa Nacional de Saúde - 2013

Deficiência visual é definida com sendo a perda parcial ou total da visão. Doenças como Miopia, Hipermetropia e astigmatismo não pode ser considerado como uma deficiência visual, pois elas podem ser ajustadas com instrumentos, como óculos, e cirurgias. Segundo a OMS existem diferentes graus de deficiência visual que são classificados em:

- Baixa visão: Que pode ser compensada com uso de lentes, lupas ou aparelhos auxiliares como bengala;
- Próximo à cegueira: Quando a visão tem uma alta debilitação, porém ainda existe uma certa percepção de luz, ainda que ínfimo;
- Cegueira: Quando percepção de luz torna-se nula, necessitando assim do uso de aparelhos como bengala, braile ou orientadores.

Ainda vale ressaltar que no Brasil, segundo dados da IBGE do censo de 2016, existem 11,8 milhões de analfabetos. Esse número representa 7,2% da população brasileira. Essas pessoas analfabetas, por sua vez, apresentam uma dificuldade para exercer grande parte das práticas cotidianas, uma vez que muitas atividades tem algum envolvimento com a leitura.

Estes são alguns dos problemas que vem trazendo um grande prejuízo dentro de instituições acadêmicas, mercado de trabalho, no âmbito doméstico, entre outros. Durante a realização de uma prova, por exemplo, uma pessoa com deficiência visual precisaria de um auxiliar para que ele possa responder as questões de sua prova. Além disso não é grande o número de livros disponíveis em Braille (escritas com pontos em relevos).

Em todos os dois problemas citados acima, a falta da capacidade de leitura, tanto pelo fato de ter uma deficiência visual, quanto pelo fato do analfabetismo, tem levado as pessoas a serem excluídas da sociedade. Deixando assim de garantir a plena execução da constituição federal no seu Art. 3º ¹ onde devemos diminuir a desigualdade social dessas pessoas marginalizadas pela falta da capacidade de leitura.

Atualmente, existem projetos que auxiliam pessoas com deficiências, como: o relógio inteligente em Braille² que utiliza ímãs e um conjunto de pinos para criar caracteres em Braille, a bengala sensorial de (RIGOLON; ALMEIDA, 2016) acessório que detecta o ambiente por meio de sensores que emitem sinais transformados em vibração, e o *bookreader*, dispositivo que identifica e converte o texto para a pronúncia, tem a mesma estética de uma impressora, possuindo objetivos semelhantes ao deste projeto.

O projeto propõe apresentar o conteúdo de um objeto que contenha texto, por exemplo um livro, através de um código de reconhecimento e identificação de texto e a sintetização do áudio deste. Este projeto poderá auxiliar pessoas com variados graus de deficiência na visão, principalmente as com deficiência total. O dispositivo pode até substituir os óculos quando tratamos de questão de leitura de texto. Este projeto pode também auxiliar pessoas, que não tem deficiência visual, que não tem aptidão à leitura. O projeto também promove a inserção dessas minorias representativas na sociedade, por

¹ https://www.senado.leg.br/atividade/const/con1988/con1988_a_tual/art3.asp

² <https://olhardigital.com.br/noticia/veja-como-funciona-um-relogio-inteligente-em-braile/67013>

conta das possibilidades entregues pelo projeto, possibilitando-os realizar a leitura de algum texto, na qual representa uma das atividades mais importantes no dia a dia.

Esse projeto tem como objetivo final desenvolver um dispositivo computacional que irá identificar e interpretar um texto captado e pronunciar este texto selecionado para o usuário. Para isso será utilizado uma câmera, um fone de ouvido e um código para reconhecimento de imagem. Para isso será necessário a utilização do OpenCV (biblioteca) e aprendizagem de processamento de imagens e inteligência artificial. Para o desenvolvimento do projeto será utilizado a linguagem de programação Python.

2 Objetivos e relevância do trabalho

2.1 Objetivo geral

O objetivo deste projeto é o desenvolvimento de um produto em formato de armação de óculos voltada a acessibilidade e a melhoria da qualidade de vida das pessoas as quais apresentam algum tipo de deficiência visual e/ou algum grau de dificuldade de leitura de textos.

2.2 Objetivos específicos

Compreender até que ponto uma pessoa com deficiência visual e/ou algum grau de analfabetismo precisa de auxílio na leitura de textos.

Obter imagem a partir do sensor CMOS, câmera.

Pré-processar imagem, de modo a filtra-la, a fim de remover toda informação desnecessária, como os ruídos. Em seguida converter a binarizar a imagem para uma escala cinza.

Efetuar o processamento na imagem, de modo a isolar todas as regiões de interesse(ROI's). A detecção da ROI demanda de operações de processamento digital de imagens diversas, como por exemplo, *Sobel* e *Canny*.

Armazenar o resultado das ROI's detectadas, em um arquivo para processamento posterior. Detectar as palavras presentes na ROI ,anteriormente armazenada, através da ferramenta especializada (implementa o reconhecimento OCR).

Sintetizar e reproduzir as palavras reconhecidas, em formato de áudio.

Estudar e desenvolver uma aplicação especializada no reconhecimento de caracteres através de características diversas da mesma, como classificadores usando extratores de características, baseados em formas, como o momentos de HU e de (SOUZA; PISTORI, 2004).

Desenvolver um protótipo com a morfologia de um óculos convencional contendo sensores de captura de imagem, por exemplo CMOS(FOSSUM, 1997), que efetue comunicação *Wireless* com algum dispositivo de processamento, o qual deve executar a parte de software. Esse protótipo deve visar a usabilidade e um design discreto, que é um critério fundamental, visto a dependência do senso estético do usuário.

Refinar o protótipo mirando a facilidade de uso, de modo que qualquer pessoa

com o mínimo de conhecimento acerca de tecnologia consiga interagir com o produto sem grandes dificuldades. O custo de produção deve ser levado em consideração, de modo a torna-lo acessível à população em geral.

2.3 Relevância do trabalho

A proposta do trabalho se foca na melhoria na vida de pessoas com deficiência visual e/ou pessoas com certo grau de dificuldade na leitura, de modo a eliminar um dos principais problemas que permeiam em seu cotidiano: a leitura de textos diversos. A proposta do trabalho constitui um produto tanto com viés científico quanto com viés comercial, podendo culminar, inclusive, em patente de produto.

2.4 Fundamentação teórica

2.4.1 Processamento digital de imagens

O interesse nos métodos de processamento digital de imagens provém de duas áreas principais de aplicação: melhora das informações visuais para a interpretação humana e processamento de dados de imagens para armazenamento, transmissão e representação, considerando a percepção automática por máquinas. Este capítulo tem vários objetivos: (1) definir o escopo da área que chamamos de processamento de imagens; (2) apresentar uma perspectiva histórica das origens dessa área; (3) dar uma ideia do que há de mais avançado em processamento de imagens analisando algumas das principais áreas nas quais ele é aplicado analisar brevemente as principais abordagens utilizadas no processamento digital de imagens; (5) mostrar uma visão geral dos componentes contidos em um sistema de processamento de imagens típico; e (6) oferecer um referencial bibliográfico e outras fontes nas quais o trabalho relativo ao processamento de imagens normalmente é relatado. (GONZALEZ; WOODS, 2000).

Segundo o livro o processamento digital de imagens refere-se ao processamento de imagens digitais em uma máquina digital. Uma imagem digital é formada por conjunto de *pixels*, na qual este é composto por coordenada espacial (x, y), valor específico de intensidade das matizes rgb, entre outros.

O livro (GONZALEZ; WOODS, 2000) foi utilizado no projeto (dublagem de texto) com a intenção de aprendizagem sobre como remover ruídos que podem se tornar obstáculo quando se trata de reconhecimento de caracteres.

Ele se torna a referência principal sobre toda operação inerente ao PDI que será empregado no projeto.

2.4.2 Rede neural artificial

Inicialmente, existe a entrada de um sinal, que contém um módulo de peso, no “neurônio”. Na estrutura do neurônio existe uma função agregadora, que realizará a soma dos produtos dos sinais. Os sinais podem passar, ser interrompido ou até mesmo alterar a saída conforme a entrada do neurônio. Além disso existe o limiar de ativação, constante decisória na passagem do sinal. Quando o sinal passa, significa dizer que ele tem o potencial de ativação. Esse por sua vez, atravessa a função g , também conhecida como função de ativação, que permite ou não a passagem do sinal, além de poder alterar o módulo que foi dada inicialmente.

A função de ativação pode ser modificada conforme o contexto. Sendo assim existem variados tipos de função de ativação. Uma das funções mais recorrentes é a função Degrau. Nesta terá, inicialmente, um valor de entrada “ u ”, e esse u terá como saída o valor de “ g ”. Para o u maior ou igual a um, terá um valor de saída igual a um, que poderá passar, enquanto que para o valor de u menor que zero terá um valor de saída igual a zero, que não poderá passar.

Outra função é a função degrau bipolar. Se o valor de u for maior que zero, g recebe um. Se o valor de u for igual a zero, g recebe zero. Por fim, caso o u for menor que zero, g recebe menos um.

As funções que serão usadas no projeto, são provavelmente, a função tangente hiperbólica e a gaussiana. A função hiperbólica, quando colocado sobre o gráfico transcreve uma morfologia similar a de um “s”, conforme mostra a Figura 2. Ela varia de um a menos um.

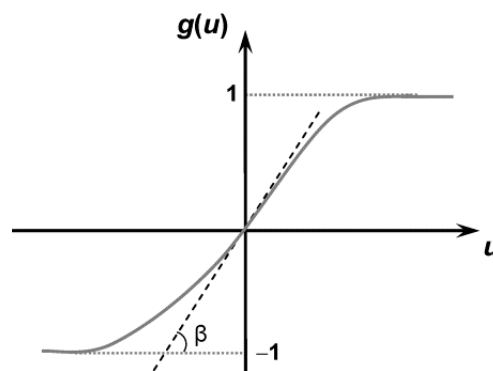


Figura 2 – Comportamento da tangente hiperbólica

Fonte: [Silva, Spatti e Flauzino \(2010\)](#)

2.4.3 Extração de características

Quando se fala em reconhecimento de padrões é fundamental desenvolver um método de extração de característica eficiente. Quando se obtém um extrator de

características eficientes, temos a premissa de que iremos conseguir um classificador eficiente. Uma das principais características de uma imagem é a sua textura.

A característica da textura de uma imagem é observada na distribuição de *pixels* de determinada região com seu brilho cor e tamanho. Estes são alguns fatores que designam certo objeto contido em uma imagem. Quando se encontra algum grau de padrão entre duas ou mais texturas pode-se classificar como sendo um objeto de uma mesma classe.

Uma textura, quando analisado visualmente, pode-se qualificar em fina, grossa, granulada, lisa, entre outras. Essas são uma das características que tornam possível a padronização do objeto, lembrando que quanto mais característica forem extraídas mais preciso será o resultado.

2.4.4 Revisão da literatura

Com o avanço da computação, muito tem se falado de desenvolver técnicas capazes de facilitar a vida de pessoas por meio da tecnologia. Existem muitas pesquisas no ramo da computação envolvendo trabalhos que possam melhorar a qualidade de vida da população (BRUCKSCHEN; RIGO; FAGUNDES, 2007; OSSADA; RODRIGUES, 2015; RIGOLON; ALMEIDA, 2016). Visando tal avanço e buscando maneiras de melhorar a qualidade de vida de pessoas com deficiência visual.

Resolver os problemas de detecção de caracteres em textos é uma tarefa com alta complexidade. A tarefa de detectar caracteres em textos digitalizados ou textos contidos em fotografias é uma tarefa árdua. Muito se tem estudado sobre formas de conseguir identificar de forma correta caracteres sobre texto. Na literatura (AGARWAL; KAUSHIK, 2015; NEUMANN; MATAS, 2012; NEUMANN; MATAS, 2010; COATES et al., 2011; CHEN; ODOBEZ; BOURLARD, 2004), utilizam técnicas de inteligência artificial, mais especificamente, na área de visão computacional para resolver tal problema. Agarwal e Kaushik (2015), em seu trabalho, utilizam redes neurais artificiais (Artificial Neural Network) e algoritmos genéticos (Genetic Algorithm) para extrair de forma eficientes caracteres e textos contidos em arquivos digitalizados de hospitais. No trabalho, os resultados quanto à extração de textos em imagens têm se mostrado promissores.

Em (ANTONELLO, 2017), os autores realizam a automatização de uma cancela, regulando a entrada e saída de veículos em uma instituição. No trabalho, os autores utilizaram de amplo arsenal de visão computacional e inteligência artificial para o cumprimento dos objetivos propostos, dentre os quais destacam-se o reconhecimento e segmentação da placa e reconhecimento dos caracteres pertencentes à placa. Para tanto, ferramentas para o auxílio de OCR (Reconhecimento óptico de caracteres) como a *tesseract* foram empregadas, garantindo uma taxa de acerto de 93,54% nos testes com um conjunto de testes contendo 10 placas cadastradas.

Em (HAZRA; SINGH; DAGA, 2017; CHANG; GUPTA; ZHANG, 2016; KANG et al., 2014), os autores utilizaram investigaram diversas abordagens para realizar OCR com as mais diversas finalidades, como por exemplo reconhecimento de *handwritten text* (texto escrito à mão) e equações. Foram usadas estratégias que envolveram o uso de redes neurais artificiais diversas, como KNN (K-Nearest Neighbor) e CNN (Convolutional Neural Network), além de SVM (Support Vector Machine), com taxas de acerto superiores a 90%.

3 Desenvolvimento do projeto

3.1 Materiais e métodos

Para a execução do projeto, será utilizada a linguagem de programação *Python*. Embora a linguagem ainda se mostrar menos eficiente em tempo de execução em comparação com outras linguagens, como a linguagem C, *Python* conquistou um grande espaço no que se diz respeito à inteligência artificial e PDI, visto o amplo ferramental nativo e diversas outras bibliotecas abertas, dedicadas à área. O amplo suporte, com uma sintaxe clara e simples, além de bibliotecas de classificação de imagens amplamente usadas na literatura, foram os pontos mais promissores e que mais contribuíram na escolha da linguagem.

Além da escolha da linguagem Python, este trabalho também utilizará a biblioteca OpenCV4 (*Open Source Computer Vision Library*). A biblioteca *OpenCV* traz inúmeras ferramentas de visão computacional, mais especificamente na área de PDI (Processamento Digital de Imagens), importantes para a extração de características de imagens e será de suma importância para o desenvolvimento deste trabalho. Com mais de 47 mil usuários espalhados pelo mundo a biblioteca ainda conta com otimizações. A biblioteca permite ainda a captura e processamento a partir de fluxo de vídeo, obtido por qualquer tipo de sensor de captura de imagens. Tal capacidade torna a biblioteca extremamente adequada ao projeto.

A ferramenta *Tesseract* realizará detecção de caracteres via OCR, a partir de imagens capturadas pelo fluxo de vídeo de uma *webcam*, acoplada a um computador pessoal. A biblioteca, que é amplamente empregada na literatura, será empregada para o desenvolvimento de um software comparativo da taxa de acertos. Esse comparativo terá a finalidade de validação da eficiência do projeto.

Também serão estudadas ferramentas para a classificação de textos em imagens. Após a extração dos textos em imagens, serão utilizadas ferramentas para transformar as sub imagens de textos extraídos em textos reais. Para transformar os textos extraídos, serão utilizadas as bibliotecas do scikit-learn (PEDREGOSA et al., 2011). E também serão estudadas redes neurais como KNN (HAZRA; SINGH; DAGA, 2017) e CNN (CHANG; GUPTA; ZHANG, 2016; KANG et al., 2014), além da biblioteca TensorFlow¹ e levar esse projeto para dispositivos móveis.

A construção do protótipo da armação está atualmente em aberto. Estuda-se a aplicação da prototipação com ferramentas de modelagem em 3D, como *Blender*² e

¹ <https://www.tensorflow.org/?hl=pt-br>

² <https://www.blender.org/>

posterior impressão com uma impressora 3D. Para a captura de imagens pretende-se usar um sensor CMOS e para comunicação um circuito de comunicação sem fio, como por exemplo o ESP8266-1 ³.

3.2 Procedimentos

Pesquisa bibliográfica, fase inicial de todo trabalho científico com o propósito de procurar dados ou informações em livros, artigos, etc que servirão de base com intuito de solucionar o problema proposto por tal projeto. Os conteúdos procurados foram: "RECO-NHECIMENTO DE GÊNEROS MUSICAIS UTILIZANDO ESPECTROGRAMAS COM COMBINAÇÃO DE CLASSIFICADORES"([COSTA et al., 2013](#)) e "PROCESSAMENTO DIGITAL DE IMAGENS 3º edição"([GONZALEZ; WOODS, 2000](#)).

3.3 Cronograma

O Cronograma de desenvolvimento deste projeto pode ser visualizado nos Quadros [1](#), .




Tabela 1 – Cronograma de atividade de 2018.

Atividade	A	M	J	J	A	S	O	N	D
Levantamento bibliográfico	X	X	X	X	X	X			
Instalação de ferramentas e configuração do ambiente de desenvolvimento	X								
Aprendizado sobre PDI e OpenCV	X	X							
Implementação de filtragem de imagens em <i>Python</i>			X	X					
Implementação da detecção de ROI em <i>Python</i>				X	X				
Segmentação dos caracteres da ROI				X	X	X			
Utilização do reconhecedor de textos <i>Tesseract</i>					X	X			
Utilização de sintetizador de áudio <i>eSpeak</i>						X			
Desenvolvimento de um software com os recursos desenvolvidos						X	X		
Escrita de artigos					X	X	X	X	X

Tabela 2 – Cronograma de atividade de 2019.

Atividade	J	F	M	A	M	J	J	A	S	O	N	D
Levantamento bibliográfico		X	X	X	X	X						
Implementar um classificador para reconhecimento de textos		X	X									
Comparar os classificadores desenvolvido e <i>Tesseract</i>			X	X								
Desenvolvimento da modelagem do protótipo em formato de óculos				X	X	X						
Confecção do circuito de comunicação entre câmera e <i>software</i>			X	X	X	X						
Montagem do protótipo					X	X	X	X	X			
Testes com voluntários com deficiência visual							X	X	X			
Escrita de artigos					X	X	X	X	X	X	X	X

Tabela 3 – Materiais necessários para a construção do projeto

Material	Descrição	Preço
	Módulo WiFi ESP8266	R\$26,90
	Raspberry Pi Noir, Infrared Camera	R\$269,00
	Filamento ABS PREMIUM	R\$85,00

3.4 Custos

Os itens citados na Tabela 3 podem ser encontradas nos sites: <<https://3dlab.com.br/produto/filamento-abs-premium/>>, <<https://3dlab.com.br/produto/filamento-abs-premium/>> e <<https://www.soldafria.com.br/wifi-modulo-esp8266-1-para-arduino-p-8055.html>>. O custo para a criação do protótipo de óculos para o projeto será aproximadamente R\$390,00 reais. Sendo que, além do gasto com os materias citados na tabela, também será

³ <https://en.wikipedia.org/wiki/ESP8266>

necessário a utilização de uma impressora 3D para construção da armação do protótipo.

4 Resultados

4.1 Resultados experimentais

Até o presente momento, foi elaborado somente o *software* e as etapas concluídas foram:

- Reconhecimento do retângulo(folha): encontrar a ROI (região da imagem que apresenta texto) e isolá-la do restante, graficamente representado por um retângulo. Algumas vezes ocorre erro em reconhecer a região, devido a angulação, escala, incidência variada de luz e sombra, etc.
- reconhecimento dos caracteres: fase na qual ocorre mais erros, por conta do *Tesseract* apresentar alguns falsos positivos.
- sintetização do texto: o texto reconhecido é reproduzido pelo *eSpeak*¹, um *software* de sintetização de áudios, .

Estas etapas são representadas pelas Figuras.

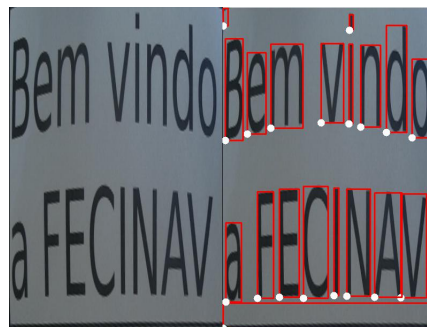


Figura 3 – Imagem original vs ROI detectada e caracteres segmentados.

¹ <http://espeak.sourceforge.net/>

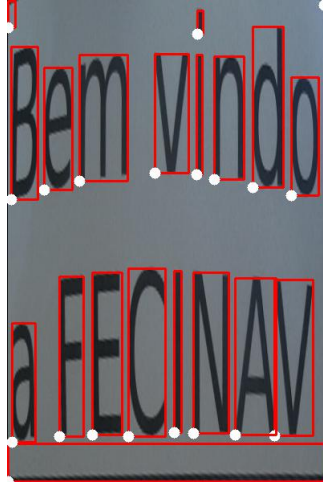


Figura 4 – Segmentação dos caracteres.

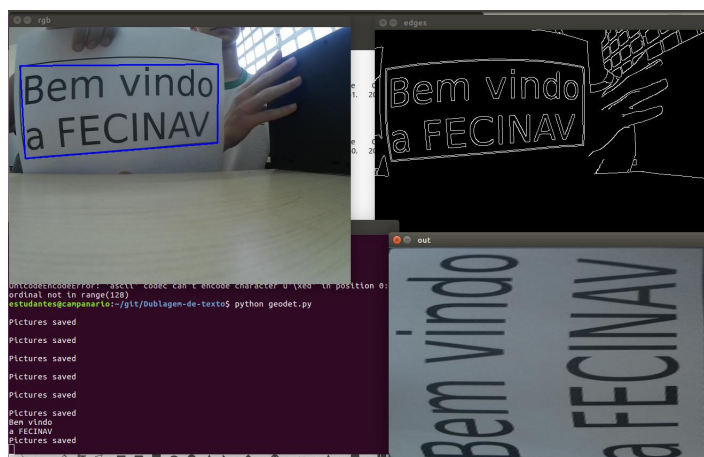


Figura 5 – Visão geral do *software*.

Considerações finais

4.2 Conclusão

Os resultados obtidos até o momento são satisfatórios e promissores. Como trabalhos futuros destacam-se estudos e aplicação de técnicas baseadas em inteligência artificial para a detecção dos caracteres contidos nas imagens, o desenvolvimento do *hardware* (prototipação de uma armação em forma de óculos, o circuito de comunicação com um dispositivo externo, que executa o nosso *software*), portar o *software* para uma versão *mobile*, para integrá-lo ao celular da pessoa assistida.

Referências

- AGARWAL, M.; KAUSHIK, B. Text recognition from image using artificial neural network and genetic algorithm. In: IEEE. *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*. [S.l.], 2015. p. 1610–1617. Citado na página 20.
- ANTONELLO, L. L. R. Identificação automática de placa de veículos através de processamento de imagem e visão computacional. 2017. Citado na página 20.
- BRUCKSCHEN, M.; RIGO, S. J.; FAGUNDES, É. Desenvolvimento de software educacional livre e inclusão de alunos com deficiência visual. *RENOTE*, v. 5, n. 2, 2007. Citado na página 20.
- CHANG, J.; GUPTA, S.; ZHANG, A. Painfree latex with optical character recognition and machine learning. 2016. Citado 2 vezes nas páginas 21 e 23.
- CHEN, D.; ODOBEZ, J.-M.; BOURLARD, H. Text detection and recognition in images and video frames. *Pattern recognition*, Elsevier, v. 37, n. 3, p. 595–608, 2004. Citado na página 20.
- COATES, A. et al. Text detection and character recognition in scene images with unsupervised feature learning. In: IEEE. *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. [S.l.], 2011. p. 440–445. Citado na página 20.
- COSTA, Y. M. et al. Reconhecimento de gêneros musicais utilizando espectrogramas com combinação de classificadores. 2013. Citado na página 24.
- FOSSUM, E. R. Cmos image sensors: Electronic camera-on-a-chip. *IEEE transactions on electron devices*, IEEE, v. 44, n. 10, p. 1689–1698, 1997. Citado na página 17.
- GONZALEZ, R. C.; WOODS, R. E. *Processamento de imagens digitais*. [S.l.]: Edgard Blucher, 2000. Citado 2 vezes nas páginas 18 e 24.
- HAZRA, T. K.; SINGH, D. P.; DAGA, N. Optical character recognition using knn on custom image dataset. In: IEEE. *Industrial Automation and Electromechanical Engineering Conference (IEMECON), 2017 8th Annual*. [S.l.], 2017. p. 110–114. Citado 2 vezes nas páginas 21 e 23.
- KANG, L. et al. A deep learning approach to document image quality assessment. In: IEEE. *Image Processing (ICIP), 2014 IEEE International Conference on*. [S.l.], 2014. p. 2570–2574. Citado 2 vezes nas páginas 21 e 23.
- NEUMANN, L.; MATAS, J. A method for text localization and recognition in real-world images. In: SPRINGER. *Asian Conference on Computer Vision*. [S.l.], 2010. p. 770–783. Citado na página 20.
- NEUMANN, L.; MATAS, J. Real-time scene text localization and recognition. In: IEEE. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. [S.l.], 2012. p. 3538–3545. Citado na página 20.

OSSADA, S. A. R.; RODRIGUES, S. C. M. Uma análise de softwares para inclusão de deficientes auditivos na educação à distância. *Reverte-Revista de Estudos e Reflexões Tecnológicas da Faculdade de Indaiatuba*, n. 13, 2015. Citado na página 20.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011. Citado na página 23.

RIGOLON, D.; ALMEIDA, H. V. S. d. *Projeto de equipamento sensorial para orientação e mobilidade de deficientes visuais*. Dissertação (B.S. thesis) — Universidade Tecnológica Federal do Paraná, 2016. Citado 2 vezes nas páginas 14 e 20.

SILVA, I. d.; SPATTI, D. H.; FLAUZINO, R. A. Redes neurais artificiais para engenharia e ciências aplicadas. *São Paulo: Artliber*, v. 23, n. 5, p. 33–111, 2010. Citado na página 19.

SOUZA, K. P.; PISTORI, H. Implementação de um extrator de características baseado em momentos da imagem. 2004. Citado na página 17.