

INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Jaboatão dos Guararapes

Análise e Desenvolvimento de Sistemas

WILLIAM ROBERTO ANDRADE DA SILVA

**GUIA INTELIGENTE: UMA FERRAMENTA DE ACESSIBILIDADE PARA
PESSOAS CEGAS OU COM BAIXA VISÃO**

Jaboatão dos Guararapes

2022

INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Jaboatão dos Guararapes

Análise e Desenvolvimento de Sistemas

WILLIAM ROBERTO ANDRADE DA SILVA

**GUIA INTELIGENTE: UMA FERRAMENTA DE ACESSIBILIDADE PARA
PESSOAS CEGAS OU COM BAIXA VISÃO**

Trabalho de conclusão de curso apresentado à
Coordenação de Graduação em Análise e
Desenvolvimento de Sistemas do Instituto
Federal de Educação, Ciência e Tecnologia de
Pernambuco, campus Jaboatão dos
Guararapes, como requisito para obtenção da
graduação em Análise e Desenvolvimento de
Sistemas.

Orientador: Prof. Luciano de Souza Cabral

Jaboatão dos Guararapes

2022

FICHA CATALOGRÁFICA

S586g

Silva, William Roberto Andrade da.

Guia inteligente: uma ferramenta de acessibilidade para pessoas cegas ou com baixa visão/ William Roberto Andrade da Silva; orientação de Luciano de Souza Cabral. Jaboatão dos Guararapes, 2022.

86f.; il.

Trabalho de Conclusão de Curso (Tecnólogo em Análise e Desenvolvimento de Sistemas) – IFPE - Campus Jaboatão dos Guararapes.

Inclui Referências.

1. Tecnologia da Informação e Comunicação 2. Base de dados 3. Tecnologia assistiva 4. Audiodescrição 5. Inteligência artificial I.Cabral, Luciano de Souza. II. IFPE. III. Título

CDD 004.21

Catálogo na fonte elaborada pelo Bibliotecário Thiago Leite Amaro da Silva - CRB4/ 2.021

INSTITUTO FEDERAL DE CIÊNCIA E TECNOLOGIA DE PERNAMBUCO

Campus Jaboatão dos Guararapes

Análise e Desenvolvimento de Sistemas

WILLIAM ROBERTO ANDRADE DA SILVA

**GUIA INTELIGENTE: UMA FERRAMENTA DE ACESSIBILIDADE PARA
PESSOAS CEGAS OU COM BAIXA VISÃO**

Trabalho de conclusão de curso apresentado à
Coordenação de Graduação em Análise e
Desenvolvimento de Sistemas do Instituto
Federal de Educação, Ciência e Tecnologia de
Pernambuco, campus Jaboatão dos
Guararapes, como requisito para obtenção da
graduação em Análise e Desenvolvimento de
Sistemas.

Orientador: Prof. Luciano de Souza Cabral

Trabalho aprovado. Local, data.

Professor Orientador

Convidado 1

Convidado 2

Jaboatão dos Guararapes

2022

Dedicatória

Dedico este trabalho de conclusão de curso à minha mãe,
sem ela eu não chegaria tão longe.

AGRADECIMENTOS

Gostaria de agradecer primeiramente a todos os professores do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco IFPE, campus Jaboatão dos Guararapes. Em especial, gostaria de agradecer ao Professor Luciano de Souza Cabral por todo apoio e orientação que recebi no último ano. Também gostaria de agradecer a todos os meus colegas de turma, todos sem exceção, foram cruciais para que eu chegasse até aqui.

E por fim, mas não menos importante, gostaria de agradecer a Mathias Cordeiro de Oliveira e Thiago José Amaral Oliveira da Silva, alunos de graduação em Análise e desenvolvimento de sistemas, do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, campus jaboatão dos Guararapes, que participaram ativamente no processo de produção da base de dados simulada, por meio da gravação de vídeos, dentro das dependências internas do referido campus, sempre de forma muito participativa e proativa. Muito Obrigado a todos.

RESUMO

Este trabalho de conclusão de curso, Propõe a construção de bases fundamentais para a criação de uma ferramenta de acessibilidade, também conhecida como tecnologia assistiva ou tecnologia adaptativa, para pessoas com deficiência visual, sejam elas cegas ou com baixa visão, proporcionando para estas pessoas um meio facilitador de integração social, principalmente no que diz respeito a sua formação acadêmica, tendo por objetivo auxiliá-las durante os seus deslocamentos, nas dependências internas do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, campus Jaboatão dos Guararapes, sendo que, para cumprir tal objetivo, se faz necessário, a integração de vários conhecimentos, técnicas e métodos, facilitadores da acessibilidade para pessoas com deficiência visual, sendo eles: a Orientação e mobilidade, audiodescrição de cenários, Tecnologias assistivas, funcionando de forma integrada com as mais atuais técnicas de Visão computacional e do Processamento de linguagem natural, sendo estas, subcampos da inteligência artificial, por fim, este trabalho será construído com base na metodologia Design Science Research de pesquisa e desenvolvimento.

Palavras-chave: audiodescrição. base de dados, inteligência artificial.

ABSTRACT

This course conclusion work, proposes the construction of fundamental bases for the creation of an accessibility tool, also known as assistive technology or adaptive technology, for people with visual impairments, whether blind or with low vision, providing for these people a means of facilitating social integration, especially with regard to their academic training, with the objective of helping them during their displacements, in the internal dependencies of the Federal Institute of Education, Science and Technology of Pernambuco, Jaboatão dos Guararapes campus, and, To fulfill this objective, it is necessary to integrate various knowledge, techniques and methods, facilitating accessibility for people with visual impairments, namely: Orientation and mobility, audio description of scenarios, Assistive technologies, working in an integrated way with the most current techniques of Computer Vision and Natural Language Processing, s being these, subfields of artificial intelligence, finally, this work will be built based on the Design Science Research methodology of research and development.

Keywords: audio description. database, artificial intelligence.

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Delimitação do Tema	19
1.2 Motivação e Justificativa	20
1.3 Materiais e Métodos	21
1.4 Objetivos e Contribuições	21
1.4.1 Objetivos gerais	22
1.4.2 Objetivos específicos	22
1.4.3 Contribuições esperadas	23
1.5 Estrutura do Trabalho	23
2. REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS	25
2.1 DEFICIÊNCIA VISUAL	25
2.1.1 Cegueira	26
2.1.2 Baixa Visão	27
2.2 ORIENTAÇÃO E MOBILIDADE (OM)	28
2.2.1 Guia-vidente	30
2.2.2 Bengala Longa	30
2.2.2.1 Técnica de Varredura	32
2.2.3 Autoproteção	32
2.2.3.1 Familiarização	32
2.3 TECNOLOGIAS ASSISTIVAS (TA)	34
2.3.1 Legislação Brasileira	34
2.4 AUDIODESCRIÇÃO (AD)	35
2.4.1 Princípios Gerais	36
2.4.2 Histórico e a Legislação Brasileira	37
2.4.3 Audiodescrição Simultânea (ADS)	37
2.5 PROCESSAMENTO DE LINGUAGEM NATURAL	38
2.5.1 Reconhecimento de Voz	39
2.5.2 Síntese de Voz	39
2.6 VISÃO COMPUTACIONAL	40
2.6.1 OpenCV	41
2.6.1.1 Histogram of oriented gradients (HOG)	42
2.6.1.2 Classificadores Haar Cascades	43
2.6.2 You Only Look Once (YOLO)	44

2.6.3 Legendas em Imagens	45
2.6.3.1 Flickr 8k	46
2.6.4 Reconhecimento óptico de caracteres (OCR)	47
3. METODOLOGIA DE DESENVOLVIMENTO PROPOSTA	49
3.1 Design Science Research (DSR)	54
3.2 Síntese e o Reconhecimento de voz	55
3.3 Detecção e o Rastreamento de Pedestres	56
3.4 Detecção e o Reconhecimento Facial	62
3.5 Estimar Distância e a Antecipação de Obstáculos	62
3.6 Detecção de Linhas e a Detecção de Cantos	64
3.7 Flickr 8k	65
3.8 Tesseract (OCR)	67
4. ANÁLISE DOS RESULTADOS OBTIDOS	69
4.1 Speech_recognition e o Google Text-to-Speech (gTTS)	69
4.2 OpenCV	70
4.2.1 Histogram of Oriented Gradients (HOG)	70
4.2.2 Haar Cascades	71
4.3 You Only Look Once (YOLO)	72
4.4 Flickr 8k	73
4.5 Tesseract (OCR)	74
5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS	76
5.1 Considerações Finais	76
5.2 Limitações	77
5.3 Trabalhos Futuros	77
REFERÊNCIAS	79

Índice de abreviaturas, símbolos e siglas

IFPE	Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco
ABNT	Associação Brasileira de Normas Técnicas
AD	Audiodescrição
OM	Orientação e Mobilidade
PLN	Processamento de Linguagem Natural
CIF	Classificação Internacional de Funcionalidade, Incapacidade e Saúde
OMS	Organização Mundial da Saúde
TA	Tecnologia assistiva
IBGE	Instituto Brasileiro de Geografia e Estatística
CAT	Comitê de ajudas técnicas
NCE	Núcleo de Computação Eletrônica
IA	Inteligencia artificial
OCR	Reconhecimento óptico de caracteres
ICR	Reconhecimento inteligente de caracteres
HOG	Histograma de gradiente orientado
DMRI	Degeneração molecular relacionada à idade
ADA	American with Disabilities Act
SVM	Máquina de vetores de suporte
DSR	Design Science Research
TTS	Text To Speech
UFRJ	Universidade Federal do Rio de Janeiro
UERJ	Universidade do Estado do Rio de Janeiro
ESDI	Escola Superior de Desenho Industrial da UERJ
GTTS	Google Text-To-Speech

Índice de Figuras

Figura 1	– Falhas em decorrência da falta de manutenção adequada dos pisos.....	20
Figura 2	– As principais causas de cegueira no mundo.....	27
Figura 3	– Os dois tipos de pisos podotáteis disponíveis no mercado.....	30
Figura 4	– Representação do Alfabeto Braille.....	34
Figura 5	– Campos de estudo baseados no Processamento de Linguagem.....	38
Figura 6	– Fluxo para o tratamento dos textos dentro do TTS.....	39
Figura 7	– Áreas do conhecimento que se beneficiam da visão computacional.....	41
Figura 8	– Utilização do método thresholding para segmentação de objetos.....	42
Figura 9	– Classificador Haar Cascade para detecção facial.....	44
Figura 10	– Rastreamento de objetos com YOLO.....	45
Figura 11	– Disposição das 5 legendas atribuídas a cada uma das 8 mil imagens..	46
Figura 12	– Exemplos de imagens presentes no conjunto de dados Flickr 8k.....	47
Figura 13	– Representação do funcionamento do (OCR).....	48
Figura 14	– Estrutura do projeto “GuiaInteligente”.....	51
Figura 15	– Série de vídeos gravados no campus Jaboatão dos Guararapes.....	52
Figura 16	– Representação física da ferramenta “GuiaInteligente”.....	53
Figura 17	– Representação lógica da ferramenta “GuiaInteligente”.....	53
Figura 18	– Código que implementa a biblioteca gTTs.....	56
Figura 19	– Código que implementa o método HOG na linguagem de programa...	57
Figura 20	– Múltiplos Bounding box (caixas delimitadoras) geradas nas imagens..	57
Figura 21	– Diferentes tipos de Classificadores Haar Cascades disponíveis.....	58
Figura 22	– Código que implementa os Classificadores Haar Cascades na.....	59
Figura 23	– Bounding box contendo a detecção de dois alunos.....	59
Figura 24	– Imagem carregada na plataforma Google Drive.....	60
Figura 25	– Código que implementa o YOLO na plataforma Google Colab.....	61
Figura 26	– Resultado para a detecção com YOLO na plataforma Google Colab...	61
Figura 27	– Detecção da face de um aluno com a geração de uma caixa.....	62
Figura 28	– Imagem de teste utilizado para se estimar a distância.....	63
Figura 29	– Imagem de teste para antecipação de obstáculos.....	64
Figura 30	– Imagem de teste utilizada para detecção de linhas.....	64
Figura 31	– Aplicação prática do Harris Corner Detector.....	65

Figura 32 – Pré-processamento das legendas no Flickr 8k.....	66
Figura 33 – Visualização das legendas atribuídas às imagens de referência.....	67
Figura 34 – Múltiplas caixas delimitadoras geradas na imagem de referência.....	68
Figura 35 – Representação gráfica dos resultados com o método HOG.....	71
Figura 36 – Representação gráfica dos resultados com o Haar Cascades.....	72
Figura 37 – Representação gráfica dos resultados com o YOLO.....	73
Figura 38 – Exemplo de output para a geração de legendas.....	74

Índice de Tabelas

Tabela 1 – As três principais técnicas trabalhadas pela OM.....	29
Tabela 2 – Distinção entre os dois tipos de pisos táteis.....	31
Tabela 3 – Comparativo dos resultados para escolha do método mais eficiente.....	35
Tabela 4 – Legislação Brasileira referente a audiodescrição.....	37
Tabela 5 – Descrição dos diferentes tipos de imagens utilizadas pelo classifica.....	43
Tabela 6 – Sequência das funcionalidades que serão abordadas no Capítulo 03....	50
Tabela 7 – Recursos utilizados ou para futura integração.....	55

1. INTRODUÇÃO

Na história da humanidade é possível encontrar diversos registros sobre a existência de pessoas com deficiência visual, como por exemplo, a imposição de penas, positivadas no código de Hamurabi. Na bíblia, no livro do Êxodo, eram citados homens caminantes, que caminhavam com a ajuda de cajados em suas mãos, entre muitos outros registros históricos. A visão humana compreende um conjunto complexo de processos fisiológicos compreendendo não apenas a captação da luz refletida pelos objetos, por meio do olho humano, como também, um conjunto ordenado de processos cerebrais para identificação e memorização das informações presentes nas imagens. (Instituto dos Cegos, 2023)

Já a deficiência visual, tem como característica principal, a perda total ou parcial, seja ela, congênita ou adquirida, dos sentidos visuais. Havendo na literatura médica dois grandes grupos fundamentais que compartilham semelhanças e ao mesmo tempo diferenças entre si, estas diferenças serão elencadas a seguir.

- A cegueira nada mais é do que a perda total da visão, ou quase nenhuma capacidade de enxergar, entende-se como a ausência total de resíduo visual assim como da percepção de vultos e da luminosidade refletida pelos objetos ou do ambiente.
- Um outro grupo é constituído pelas pessoas com baixa visão ou também denominadas, pessoas com visão subnormal, que tem como característica principal, o comprometimento da visão, de forma persistente, mesmo após tratamento médico especializado, havendo ainda, algum resquício de captação de luz por pelo menos um dos olhos.

Segundo dados da Organização Mundial da Saúde (OMS), cerca de 36 milhões de pessoas no mundo são totalmente cegas e outras 217 milhões de pessoas têm baixa visão. No Brasil, com base nos dados coletados no censo, realizado no ano de 2010, pelo Instituto brasileiro de geografia e estatística (IBGE), um total de 528.634 pessoas são incapazes de enxergar, e outras 6.056.654 pessoas têm baixa visão. (OMS, 2019)

Doenças como Glaucoma, retinopatia diabética, atrofia do nervo óptico e a retinose pigmentar constituem as principais causas da cegueira na população de um modo geral, ocorrendo, na grande maioria dos casos, na vida adulta destas pessoas. Um tema, extremamente relevante no que diz respeito a promoção da acessibilidade para pessoas com deficiência visual, a Orientação e Mobilidade (O&M) representa uma área da educação especial, voltada para reabilitação de pessoas cegas ou com visão subnormal, utilizando para este fim, uma ampla gama de técnicas, métodos e procedimentos, trabalhando em consonância com os sentidos remanescentes destas pessoas.

São desenvolvidos os sentidos remanescentes, como o olfato, tato, audição, percepção vestibular e a visão residual, com o intuito de desenvolver habilidades capazes de proporcionar a estas pessoas, a compreensão de pontos de referência, a incidência de pistas no decorrer do trajeto, a presença de obstáculos, entre muitas outras possibilidades. De forma isolada a orientação é a habilidade de reconhecer o meio ambiente e estabelecer sua posição em relação ao mesmo, significa consciência do espaço, dos objetos e do seu próprio corpo, se desenvolvendo por meio de repetidas experiências **sensorio-motoras** no meio físico/ambiente. A mobilidade de forma isolada é a habilidade física para se mover de forma determinada e segura, pelos espaços, da maneira mais independente e autônoma possível.

A fim de abordar os inúmeros impactos sociais que a deficiência visual, pode representar na vida de uma pessoa, a Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF) ressalta em suas normativas, os aspectos sociais da deficiência visual, e propõe um mecanismo para mensurar o impacto que o ambiente social e físico têm sobre as capacidades funcionais da pessoa com deficiência. Assim, fazendo com que, a deficiência possa deslocar-se da pessoa com deficiência para o ambiente em que vive. (CIF, 2001)

1.1 Tecnologias Assistivas

As tecnologias assistivas ou simplesmente tecnologias adaptativas, são recursos que tem por objetivo principal minimizar os impactos que suas deficiências podem causar em suas vidas.

Em outras palavras, as tecnologias assistivas visam proporcionar condições favoráveis ao desenvolvimento e promoção da acessibilidade em diferentes contextos, com isso, visa aumentar as **capacidades funcionais** da pessoa com deficiência, seja ela qual for. No Brasil, o Comitê de ajudas técnicas (CAT), da secretaria de direitos humanos da presidência da república, descreve as tecnologias assistivas como sendo.

“A tecnologia assistiva é uma área do conhecimento científico, com característica interdisciplinar, que engloba produtos, recursos, metodologias, estratégias, práticas e serviços que objetivam promover a funcionalidade, relacionada à atividade e participação de pessoas com deficiência, incapacidade ou mobilidade reduzida, visando sua autonomia, independência, qualidade de vida e inclusão social” (CAT, 2000).

Sendo assim, as tecnologias assistivas podem referir-se a um equipamento, item ou produto que seja utilizado para aumentar as capacidades funcionais das pessoas com deficiência.

1.2 Audiodescrição

Audiodescrição (AD) representa um recurso de acessibilidade que propicia a ampliação do entendimento das pessoas com deficiência visual, tanto dos diversos elementos quanto dos acontecimentos presentes no ambiente em que se encontram, a audiodescrição de cenários se dá por meio da transformação da linguagem visual em linguagem verbal. Sendo uma atividade de mediação linguística, uma modalidade de **tradução intersemiótica** (conversão de informações visuais em informações textuais ou sonoras, e vice-versa). Pode ser empregada em diferentes contextos, os mais comuns são, nos filmes, mostras culturais, exposições artísticas, eventos esportivos, palestras, congressos e etc. (Audiodescrição, 2023)

A descrição dos cenários e dos acontecimentos nos espaços, como filmes, mostras teatrais, exposições de museus, entre outras, seguem algumas premissas básicas, que serão descritas a seguir:

- 1º Deve-se descrever o que vê, o que ou quem aparece.
- 2º Citar características físicas, roupas, cores e outros elementos visuais.
- 3º Delimitar onde as ações acontecem, assim como o tempo em que acontecem.

- 4º O enquadramento, priorizando elementos de maior importância, de acordo com o contexto.

Nestas etapas, deve-se priorizar a utilização de uma linguagem clara, simples e objetiva, evitando assim interpretações subjetivas que possam prejudicar o entendimento das cenas e de seus acontecimentos.

1.3 Inteligência Artificial

Inteligência artificial (IA) é a capacidade conferida às máquinas de simular a inteligência humana, a fim de solucionar problemas demasiadamente complexos ou repetitivos. O processamento de linguagem natural (PLN), constitui um ramo da inteligência artificial, responsável por unir a aprendizagem de máquina e a linguística, sendo também de sua responsabilidade estudar os diversos problemas referentes à geração e compreensão da linguagem natural humana pelas máquinas. (eveo, 2023)

As tecnologias de fala, referem-se aos processos de produção e a percepção dos sons utilizados na linguagem falada, existindo basicamente duas tecnologias neste sentido, sendo elas ferramentas para o reconhecimento de voz, e as ferramentas para a síntese de voz. Especificamente a síntese de voz, constitui um elemento importante para o desenvolvimento de ferramentas de tecnologias assistivas. O reconhecimento de voz é o processo por meio do qual uma máquina identifica as palavras faladas. (Text-To-Speech, 2008)

A visão computacional é a capacidade conferida às máquinas de simular a visão humana, que decorre por meio de algoritmos específicos e técnicas apropriadas para processamento de imagens. Proporcionando o desenvolvimento de soluções e ferramentas úteis à vida das pessoas de um modo geral, em especial, as pessoas com deficiência visual.

“A visão computacional, representa um corpo de conhecimentos que buscam a modelagem artificial da visão humana, com o intuito de replicar suas funções, por meio do desenvolvimento de software e hardwares avançados”. (Bradski, 2008)

1.1 Delimitação do Tema

Este trabalho está delimitado geograficamente, às dependências internas do Instituto Federal de Educação Ciência e Tecnologia de Pernambuco, especificamente o campus com sede na cidade de Jaboatão dos Guararapes - PE.

Os resultados obtidos através dos dados coletados durante as gravações feitas no referido campus e apresentados neste trabalho de conclusão de curso foram desenvolvidos através de simulações planejadas pelo seu autor.

1.2 Motivação e Justificativa

Este trabalho de conclusão de curso, tem como uma de suas principais justificativas para o seu desenvolvimento, as inúmeras dificuldades que poderão ser enfrentadas por pessoas com deficiência visual, durante o seu deslocamento, nos espaços de livre acesso, nas das dependências internas do já referido campus, sejam elas, pessoas completamente cegas, ou pessoas com visão subnormal.

Deve-se destacar que, as instalações do referido campus, apresentam-se em relativa conformidade com critérios mínimos para a promoção da acessibilidade em Instituições Federais de ensino. Porém é possível de se perceber, que mesmo após dois anos de sua inauguração, ainda persistem em determinados pontos da instituição, problemas que podem prejudicar potenciais alunos com necessidades especiais, de frequentarem a instituição de forma regular, prejudicando sua autonomia e segurança, um destes problemas será exemplificado na **Figura 01** a seguir.

Figura 01 – Falhas em decorrência da falta de manutenção adequada dos pisos táteis.



Fonte: Elaboração própria.

Descrição: A **Figura 01**, demonstra as falhas existentes nos pisos táteis, instalados no campus Jaboatão dos Guararapes, como por exemplo a falta do referido piso, em certos pontos onde foram instalados.

Em decorrência disto, vale a pena ressaltar, que segundo dados do Instituto Brasileiro de Geografia e Estatística IBGE, no estado de Pernambuco, 100 mil pessoas são consideradas cegas, sendo que apenas 5% destas pessoas frequentam o ensino público de forma regular. Este percentual é ainda menor, quando se leva em consideração a taxa de frequência de alunos com deficiência visual em instituições de ensino técnico e de nível superior.

1.3 Materiais e Métodos

Para este trabalho de conclusão de curso, foi utilizada uma câmera de celular LG k22 <<https://www.tudocelular.com/LG/fichas-tecnicas/n6462/LG-K22.html>> para a captura das imagens por meio da gravação de vídeos curtos. No que diz respeito a ferramentas de software, foram utilizados a biblioteca de visão computacional de código aberto OpenCV, Yolo e a base de dados Flickr 8k, assim como a ferramenta tesseract para o reconhecimento óptico de caracteres e a biblioteca de processamento de linguagem natural Speech_recognition e Google Text-to-Speech.

Para Execução dos Códigos, que implementam as bibliotecas acima citadas, foi utilizado um Notebook Samsung Essentials e30, com processador i3 7th Geração, 8GB RAM, de poder de processamento. alguns códigos foram implantados e adaptados a partir da plataforma Google Colab.

1.4 Objetivos e Contribuições

Devido a grande relevância da acessibilidade, em especial para as pessoas com deficiência visual, almeja-se criar as bases fundamentais de estudo, para criação de uma tecnologia assistiva voltada para este público. Mais adiante, serão apresentados os objetivos gerais e em seguida, serão apresentados os objetivos específicos que nortearam o desenvolvimento deste documento, tanto a nível de fundamentação teórica, quanto a aplicação prática com base nos testes realizados com a base de dados simulada, produzida na instituição, assim como os seus resultados, tendo como base, a metodologia de pesquisa e desenvolvimento proposta. Além disso, serão discutidos os resultados com base em algoritmos avançados de visão computacional e processamento de linguagem natural, estando estes algoritmos, disponíveis através de links, anexados neste documento.

1.4.1 Objetivos gerais

O principal objetivo deste trabalho de conclusão de curso é propor a análise e integração de diversos conhecimentos, capazes de promover a acessibilidade para pessoas com deficiência visual, sendo eles, a Orientação e Mobilidade (OM), as Tecnologias assistivas (TA), e a audiodescrição (AD), trabalhando de forma integrada com ferramentas de processamento de linguagem natural e ferramentas de visão computacional.

1.4.2 Objetivos específicos

Buscando alcançar o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

- Testar o funcionamento da biblioteca OpenCV, e seus métodos HOG (Histogramas de Gradientes Orientados), assim como o funcionamento dos classificadores Haar Cascades para detecção e rastreamento de pedestres entre outras detecções em imagens e fluxos de vídeo, pré-gravados.
- Mensurar as capacidades de detecção e rastreamento da biblioteca de visão computacional Yolo.
- Estudar e avaliar os resultados dos testes com a base de dados de imagens estáticas, utilizando o conjunto de dados Flickr 8k para criação de legendas em imagens estáticas.
- Avaliar o funcionamento da biblioteca Tesseract para o reconhecimento óptico de caracteres.

1.4.3 Contribuições esperadas

Entre as contribuições esperadas em razão do desenvolvimento deste trabalho de conclusão de curso está a possibilidade de que pessoas com deficiência visual possam se locomover com maior segurança e autonomia nas dependências internas do Instituto Federal, da forma mais eficiente possível, por meio da detecção e rastreamento de pedestres, assim como a identificação de obstáculos, podendo evitar possíveis colisões com os mesmos.

1.5 Estrutura do Trabalho

No **capítulo 02**, está representando o referencial teórico assim como os trabalhos relacionados, pertinentes aos temas, da pesquisa realizada, sendo estes, essenciais para a sua boa compreensão, a ordem em que, estes temas estão estruturados, da seguinte forma a seguir.

- Deficiência visual;
- Orientação e Mobilidade (OM);
- Tecnologias Assistiva (TA);
- Audiodescrição (AD);
- Processamento de Linguagem Natural (NLP);
- Visão Computacional;

Mais adiante, no **capítulo 03**, será trabalhada, a metodologia de desenvolvimento proposta para este trabalho de conclusão de curso, que tem por objetivo, buscar a solução de um ou mais problemas, a partir do entendimento destes problemas, construindo e avaliando artefatos, permitindo transformar situações, a fim de encontrar caminhos capazes de modificá-las, para condições mais satisfatórios.

O **capítulo 04**, apresentará a análise dos resultados, com base nos testes feitos, utilizando a base de dados simulada, produzida no Instituto Federal, em consonância com as tecnologias descritas no **capítulo 03** Por fim, o **capítulo 05** apresentará as considerações finais sobre o desenvolvimento deste trabalho de conclusão de curso.

2. REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

Neste capítulo, serão abordados os principais conceitos e termos, assim como a fundamentação teórica referente aos temas propostos neste trabalho de conclusão de curso. Em um primeiro momento, serão descritos os principais aspectos envolvendo a visão humana e os inúmeros problemas físicos e sociais ocasionados pela deficiência visual.

Logo em seguida, será abordada as diferentes formas de emprego, de técnicas e métodos de orientação e mobilidade (OM) para pessoas com deficiência visual. Em seguida, será explanado, conceitos referentes às principais tecnologias assistivas disponíveis no mercado, que tenham por objetivo principal, promover a acessibilidade em locais públicos, para pessoas com cegas ou com baixa visão. Mais adiante, serão descritos, os principais pontos referentes a audiodescrição de cenários, a fim de facilitar o entendimento das pessoas com deficiência visual, sobre o ambiente a sua volta.

Na sequência, serão abordados os principais conceitos referentes ao processamento de linguagem natural, assim como seus diferentes casos de uso nos dias atuais. Por fim, serão descritos de forma teórica, o funcionamento, das tecnologias de visão computacional trabalhadas neste documento, sendo eles, descritas na ordem em que serão apresentadas.

- Biblioteca OpenCV para detecção e rastreamento de pedestres.
- Biblioteca YOLO para o rastreamento de pedestres.
- Base de dados, Flickr 8k, para a criação de legendas em imagens.
- Biblioteca Tesseract para o reconhecimento óptico de caracteres.

2.1 Trabalhos Relacionados

O projeto intitulado “UTILIZAÇÃO DE VISÃO COMPUTACIONAL E DETECÇÃO DE CARACTERÍSTICAS PARA AUXILIAR NA NAVEGAÇÃO DE PESSOAS COM DEFICIÊNCIA VISUAL EM AMBIENTES INTERNOS” (UFFS, 2016), apresentado na VI Jornada de iniciação científica e tecnológica que decorreu de 17 a 18 de Outubro de 2016 na Universidade Federal da Fronteira do Sul - UFFS - Campus Chapecó, propôs a utilização da tecnologia da informação para auxiliar a

navegação de pessoas com deficiência visual, propondo a construção de uma sistema de baixo custo para o reconhecimento de objetos.

Neste sentido, seria desenvolvido um aplicativo que seria instalado no smartphone do usuário, onde o usuário faria uma gravação com a câmara do dispositivo, do ambiente em que esta, e por meio de sinais sonoros, receberia o resultado das detecções dos objetos encontrados, como portas, escadas, elevadores, entre outros, estas detecções seria possível através de técnicas de visão computacional previamente definidos.

Tendo isto em mente, optou-se por utilizar a biblioteca de visão computacional OpenCV, onde foram avaliados três métodos, considerados pertinentes aos objetivos pretendidos pelo projeto, sendo eles: Features Detection, Template Matching e Histogram Comparison, devido a baixo grau de assertividade do método de Template Matching optou-se em descartá-lo antes da execução dos testes. Os resultados dos testes estão representados na tabela a seguir.

Tabela 03 – Comparativo dos resultados para escolha do método mais eficiente.

<i>Método</i>	<i>Resultado</i>	Objetos (72)	Percentual
Features Detection	Verdadeiro Positivo	36	50
	Verdadeiro Negativo	33	45,8
	Falso Positivo	3	4,2
	Falso Negativo	0	0
Histogram Comparison	Verdadeiro Positivo	28	38,9
	Verdadeiro Negativo	21	29,2
	Falso Positivo	23	31,9
	Falso Negativo	0	0

Fonte: UFFS (Adaptado).

Descrição: A **Tabela 03**, contém os resultados dos testes realizados com os métodos Features Detection e Histogram Comparison.

2.2 DEFICIÊNCIA VISUAL

Os seres humanos são seres visuais em suas ações e comportamentos, em razão disto, uma grande parte das informações, presentes no ambiente, são captadas pelos olhos. Enxergar é um processo fisiológico, de alta complexidade, que envolve inúmeras estruturas do organismo humano. A luz em um primeiro momento, passa pelos olhos, em seguida é captada pela retina e enviada para o cérebro, por

meio do nervo óptico, por fim, os neurônios são responsáveis pelo processamento destas informações. (Flavio, 2019)

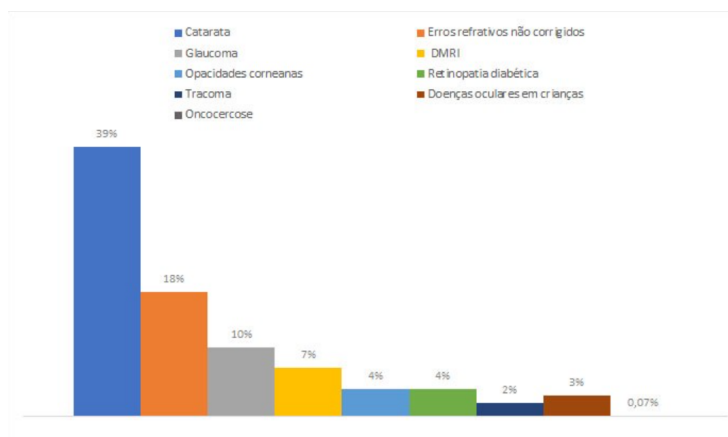
Segundo dados do World Report on Disability (Relatório Mundial sobre as Deficiências, sendo o primeiro documento a abordar de forma ampla, as necessidades e barreiras que estas pessoas enfrentam em seu dia-a-dia), divulgados no ano de 2010, a cada 5 segundos, 1 pessoa se torna cega no mundo, sendo que, do número total de novos casos de cegueira, próximo a 90% destes casos ocorrem em países subdesenvolvidos ou emergentes. (WHO, 2011)

A deficiência visual caracteriza-se como sendo, a perda total ou parcial, da visão, seja por motivo de doença congênita ou doença adquirida. Os acidentes oculares representam, um número expressivo de fatores, que podem ocasionar novos casos de perda total ou parcial, da visão, resultado em decorrência de lesões por meio do contato direto, com objetos ou, através do contato com substâncias químicas, como produtos domésticos de limpeza, ou até mesmo, do contato direto com produtos químicos inflamáveis como combustíveis e produtos químicos corrosivos como ácidos entre outras possibilidades.

2.2.1 Cegueira

Sendo a incapacidade total de enxergar, a cegueira pode ser de três tipos: congênita (têm início dos 0 ao 1 ano de idade); precoce (se surge entre 1º e o 3º anos de idade); ou adquirida (tem início após os 3 anos de idade). Em razão disso, uma pessoa cega congênita, devido a ausência ou redução de um referencial visual(imagem mental), possui uma representação intelectualizada do mundo (ambientes, volumes, perspectivas, cores e etc). (Bernardes, 2016)

Figura 02 – As principais causas de cegueira no mundo.



Fonte: (Saúde Plena, 2019) (adaptado) .

Descrição: A **Figura 02**, apresenta um gráfico de colunas, com os valores em porcentagem, indicando as principais causas de cegueira no mundo, segundo dados coletados pelo Conselho Brasileiro de Oftalmologia produzidos no ano de 2010.

Outro ponto importante que vale a pena ressaltar, diz respeito aos casos de cegueira parcial ou total, que surgem em decorrência de acidentes com forte impacto na região da cabeça. Mas de todo modo, fatores determinantes, que podem acarretar em cegueira ou visão subnormal na população adulta, advém de doenças como, glaucoma, retinopatia diabética, atrofia do nervo óptico, retinose pigmentar e a degeneração molecular relacionada à idade (DMRI). Entre as crianças, as principais causas advém de doenças como o glaucoma congênito, retinopatia da prematuridade, catarata congênita, e toxoplasmose ocular congênita.

2.2.2 Baixa Visão

A pessoa com baixa visão ou visão subnormal apresenta uma redução na sua capacidade de enxergar, mesmo após a correção de erros de refração comuns, sendo os tipos mais comuns a miopia, hipermetropia, presbiopia e astigmatismo, o que pode acabar por, limitar as suas capacidades funcionais. Sendo os principais problemas relacionados a baixa visão: baixa acuidade visual, dificuldades para enxergar de perto e/ou de longe, campo de visão reduzido, assim como problemas na visão de contraste, entre outros.

O uso de qualquer resíduo visual, poderá ajudar uma pessoa com baixa visão a executar inúmeras tarefas, tais como:

- Na mobilidade e locomoção.
- Na coordenação dos movimentos.
- Na aprendizagem por meio da leitura e escrita.
- No contato e relação com o ambiente.
- Na Constituição e organização do espaço.

2.3 ORIENTAÇÃO E MOBILIDADE (OM)

Promover a acessibilidade em locais públicos para todos os indivíduos que possuam limitações físicas para o seu deslocamento de forma segura e eficiente, é de responsabilidade de toda a administração pública, seja no âmbito municipal, estadual ou federal. Assim como, também é de responsabilidade das entidades privadas, implementar em seus espaços, todas as normas descritas na **Lei Nº 10.098, de 19 de dezembro de 2000**, que estabelece regras gerais e critérios básicos para a promoção da acessibilidade das pessoas portadores de deficiência ou com mobilidade reduzida. (Felippe, 2001)

“Orientação e Mobilidade nada mais é do que uma disciplina que tem a finalidade de auxiliar as pessoas visualmente deficientes a desenvolver ou estabelecer a capacidade de movimentação independente, eficiente e segura pelos espaços, para satisfazer suas próprias necessidades”. (Felippe, 2001)

Em razão disto, existem diversas técnicas que são trabalhadas e ensinadas por profissionais especializados no estudo e ensino da orientação e mobilidade, sendo as 3 principais: A técnica de OM que ocorre, com o apoio de um guia-vidente (pessoas com visão, podendo ser um parente, um amigo ou até mesmo um desconhecido), uma outra seria, por meio da utilização de uma bengala longa, sendo, esta, aquela que representa o maior fator de autonomia e integração social, para as pessoas com deficiência visual, e por fim mas não menos importante a técnica que utiliza um cão-guia, sendo um animal, devidamente treinado para executar esta tarefa.

Tabela 01 – As três principais técnicas trabalhadas pela OM.

Técnica	Descrição
Guia-vidente	Qualquer pessoa que, <u>após aceitação prévia do deficiente</u> , a auxiliará durante o seu deslocamento em locais desconhecidos. Caracteriza-se fundamentalmente como uma forma dependente de locomoção.
Bengala longa	Capaz de oferecer ao deficiente visual, informações cruciais sobre o ambiente, por meio de técnicas como: Varredura, detecção de objectos, passagem por portas, rastreamento com técnica diagonal, subir e descer escadas, técnicas de toque, técnica de toque e deslize, técnica de deslize, rastreamento com técnica de toque, rastreamento em três pontos e etc.
cão-guia	Animal treinado, com a finalidade de oferecer assistência e direcionamento durante os deslocamentos, antecipando obstáculos, e consequentemente, minimizando potenciais acidentes.

Fonte: Elaboração própria.

Descrição: A **Tabela 01**, contém as descrições em suas linhas, das três principais técnicas de orientação e mobilidade, amplamente trabalhadas nos dias atuais.

O profissional especializado no ensino da orientação e mobilidade tem como uma de suas principais responsabilidades elaborar um plano de trabalho de acordo com as capacidades funcionais de cada aluno. O treinamento divide-se em várias etapas e no decorrer destas etapas serão, abordados alguns aspectos fundamentais para o ensino da O&M, estes aspectos serão listados a seguir.

- Aspectos cognitivos, que dizem respeito a segmentação de conceitos, com a natureza, as funções dos objetos, abstração e solução de problemas diversos.
- Aspectos psicomotores, que proporcionam aos alunos de O&M, experiências no desenvolvimento de suas capacidades perceptivas e de movimentos básicos.
- Aspectos emocionais, que visam ajudar o aluno, a adquirir maior autoconfiança, incremento na auto-estima e motivação.

Outro ponto relevante para o ensino de O&M, diz respeito à utilização dos sentidos remanescentes como um meio facilitador da mobilidade. como por exemplo, o reconhecimento de sinais sonoros comuns a determinada região, ou até

mesmo, a familiarização com determinados odores característicos de uma local, constituindo um importante elemento no processo de ensino-aprendizagem da OM.

2.3.1 Guia-vidente

Observações Importantes: A técnica de O&M que acontece com o auxílio de um Guia-vidente, só poderá acontecer com o consentimento prévio do deficiente visual. Já a técnica de O&M com o auxílio de um cão-guia, é de extrema importância que em momento algum, se distraia o animal com situações que prejudicam a execução de suas atividades.

2.3.2 Bengala Longa

Instrumento facilitador da acessibilidade para deficientes visuais, a bengala longa, funciona como uma espécie de, “extensão do próprio corpo”, tendo por objetivo principal, antecipar o contato direto com possíveis obstáculos, desvios no trajeto, também sendo capaz, de proporcionar a quem a utiliza, a percepção de forma empírica de uma direção a ser seguida, ou seja, um caminho livre de obstáculos a sua plena locomoção.

Através da utilização da bengala longa é possível beneficiar de um outro recurso de acessibilidade, de extrema importância para o ensino da OM, este recurso, encontra-se instalado em locais de acesso público como de acesso privado, são os chamados **pisos táteis**. Presentes em estações de metrô, terminais de ônibus, ruas, praças e diversos outros locais, assim como, em estabelecimentos comerciais privados, lojas, shopping, entre outros. Esta tecnologia se faz presente nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes. (Bittencourt, 2006)

São também conhecidos como pisos podotáteis, existindo dois tipos fundamentais, que são fisicamente semelhantes entre si, mas com propósitos distintos que se complementam, devendo sempre, serem utilizados em conjunto. A **tabela 02** a seguir apresenta a diferença fundamental entre os dois tipos de pisos táteis.

Tabela 02 – Distinção entre os dois tipos de pisos táteis.

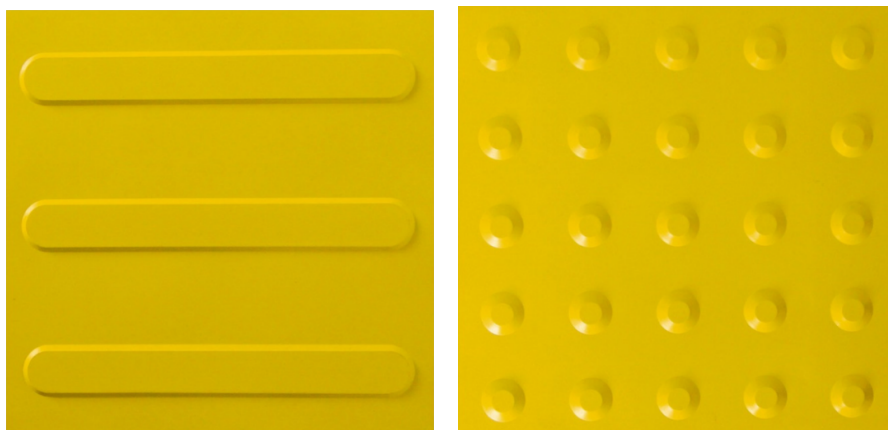
Tipos	Descrição
Piso tátil direcional	Tem a função de indicar uma direção ou caminho, livre de obstáculos a ser seguido, é formado exclusivamente por linhas em alto relevo sobre sua superfície.
Piso tátil de alerta	Tem por objetivo, Indicar obstáculos à frente, como escadas, entrada de elevador, pilastras e etc.

Fonte: Elaboração Própria.

Descrição: A **Tabela 02**, descreve basicamente, o propósito para qual, cada tipo de piso tátil foi criado.

O conjunto de técnicas, ferramentas e procedimentos, que atuam como facilitadores da acessibilidade para pessoas com deficiência visual, acabam por resultar em um outro conceito, de extrema importância para o estudo e ensino da OM, diz respeito a existência de uma “Linha-guia” imaginária, referindo-se a percepção de indicação de uma direção a ser seguida durante o deslocamento.

O campus Jaboatão dos Guararapes, dispõe de forma ampla, deste importante recurso de acessibilidade, A forma física dos pisos táteis, baseia-se em pontos de alto relevo, fazendo com que, desta maneira, seja possível sentir a presença deste piso especial, através do tato, com o auxílio de uma bengala longa ou com os próprios pés, os dois tipos de pisos podotáteis serão representados na **Figura 03** a seguir.

Figura 03 – Os dois tipos de pisos podotáteis disponíveis no mercado.

Fonte: (safe park sinalização, 2023) (adaptado).

Descrição: A **Figura 03**, demonstra os dois tipos de pisos táteis disponíveis no mercado, sendo o primeiro, o piso tátil direcional, com linhas sobressalentes, que podem ser facilmente sentidas por qualquer pessoa que estiver com os pés sobre ele. Já o segundo, diz respeito ao piso tátil de alerta ou o popularmente conhecido como “piso de bolinhas”, que como o próprio nome sugere, tem a função de alertar o deficiente visual sobre a presença de obstáculos à sua frente ou mesmo mudanças de direção.

2.3.2.1 Técnica de Varredura

Esta técnica, proporciona a pessoas com deficiência visual uma exploração, quase que imediata e abrangente do terreno na área próxima ao seu corpo.

2.3.3 Autoproteção

O termo autoproteção compreende uma gama de procedimentos que, se aplicados de forma correta, possibilitam à pessoa com deficiência visual, uma maior proteção durante o seu deslocamento. Estes procedimentos são realizados com o auxílio do próprio corpo, utilizando para os segmentos corporais (membros superiores e inferiores, tronco, cabeça), estabelecendo relações posicionais e relacionais, sendo a autoproteção comum às outras técnicas de OM citadas neste documento, como o guia-vidente, a bengala longa e o cão-guia.

2.3.3.1 Familiarização

Proporciona a pessoa com deficiência visual, um importante, meio de estudo, sobre o ambiente em que esta pessoa se encontra, ocorrendo de forma segura e eficiente em ambientes diversos e desconhecidos.

2.4 TECNOLOGIAS ASSISTIVAS (TA)

De modo geral, as tecnologias assistivas representam um conjunto de itens, peças ou equipamentos, cuja sua utilização tem o intuito de aumentar as

capacidades funcionais das pessoas com deficiência. As tecnologias adaptativas, referem-se a itens que foram desenvolvidos especificamente para serem utilizados por pessoas com deficiência. (Assistiva Tecnologia e Educação, 2023)

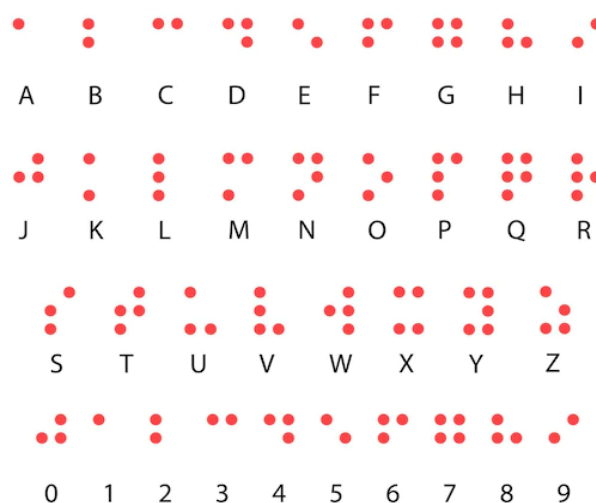
As tecnologias assistivas tem como característica a transdisciplinares, o que significa que englobam profissionais de diversas áreas do conhecimento, vale a pena citar algumas destas áreas, a saber:

- Educação.
- Psicologia.
- Engenharia.
- Design.
- Fisioterapia.
- Técnicos e profissionais de outras áreas.

Com relação as normas existentes que tratam deste tema, a American with Disabilities Act (ADA), que é uma lei criada em 1990 nos estados unidos da america, versa sobre os direitos civis das pessoas com deficiência, vedando qualquer forma de discriminação, considerando o rol de tecnologias assistivas especializadas no auxílio de pessoas com deficiência visual, sejam elas cegas ou com baixa visão. (ADA, 2023)

Diante disso, desde sua criação há mais de um século, a linguagem Braille é um sistema tátil de leitura e escrita, que até os dias de hoje representa um grande marco na busca de integração social, por parte das pessoas com deficiência e de todos aqueles que trabalham para este fim, principalmente no que diz respeito a inserção destas pessoas na educação básica até a educação de nível superior, assim como, em sua inserção no mercado de trabalho. O sistema Braille é constituído, basicamente, por um conjunto de pontos em alto relevo, estruturados em unidades chamadas células Braille, que podem ser facilmente sentidas através do contato com os dedos das mãos, geralmente a leitura é feita com o dedo indicador. O alfabeto Braille será representado na **Figura 04** a seguir. (Colégio gtm, 2023)

Figura 04 – Representação do Alfabeto Braille.



Fonte: (freepik, 2023) (adaptado).

Descrição: A **Figura 04**, contém todos os símbolos que compõem o alfabeto Braille, cada célula tem uma letra correspondente, onde estão dispostos de A a Z, e com as numerações dispostas 0 a 9 nesta ordem.

2.4.1 Legislação Brasileira

Em 16 de novembro de 2006, sob a competência da Secretaria Especial dos Direitos Humanos da Presidência da República - SEDH/PR, por meio da portaria nº 142, instituiu o Comitê de Ajudas Técnicas - CAT, a fim de reunir especialistas e representantes de órgãos governamentais em volta de uma agenda de trabalho comum.

“Para as pessoas sem deficiência a tecnologia torna as coisas mais fáceis
Para as pessoas com deficiência, a tecnologia torna as coisas possíveis”.
(RADABAUGH, 1993)

2.5 AUDIODESCRIÇÃO (AD)

Sendo um importante recurso de acessibilidade, a audiodescrição de cenários amplia o entendimento das pessoas com deficiência visual, sobre o ambiente a sua volta e seus acontecimentos, seja em: eventos culturais, peças de teatro, filmes e exposições artísticas, dentre muitos outros exemplos possíveis, estas informações

chegam até o usuário deste recurso por meio sinais sonoras, geralmente falados por um indivíduo, representando uma faixa narrativa adicional, que pode ser reproduzida de forma pré gravada ou gerada em tempo real em quando os fatos nos exemplos citados acima, se desenvolvem. (Motta, 2010)

2.5.1 Princípios Gerais

A audiodescrição consiste basicamente, na clara e objetiva descrição de todos os elementos visualmente presentes no ambiente, permitindo o recebimento das informações contidas em imagens, estáticas, e dinâmicas (vídeos). Devendo o profissional audiodescritor deve descrever:

- Informações sobre o ambiente.
- Expressões faciais e corporais.
- Mudanças de tempo e espaço.

De um modo geral, a audiodescrição deve ser clara e objetiva, não devendo conter inferências e interpretações a fim de prejudicar o entendimento do que se é descrito.

2.5.2 Histórico e a Legislação Brasileira

O símbolo internacionalmente reconhecido, para a audiodescrição, é constituído pelas letras A e D, sendo que, à direita da letra D, existem três sinais de parênteses, fazendo alusão a ondas sonoras se propagando. Nos Estados Unidos, a audiodescrição surgiu de forma gradativa a partir do ano de 1975, no Brasil, a audiodescrição teve início no ano de 2003.

Tabela 04 – Legislação Brasileira referente a audiodescrição.

Legislação aplicada à Audiodescrição	
Lei nº 10.098, de 19 de dezembro de 2000	Estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências.
Decreto Federal nº 5.296, de 2 de dezembro de 2004	Regulamenta as Leis nos 10.048, de 8 de novembro de 2000, que dá prioridade de atendimento às pessoas que especifica, e 10.098, de 19 de dezembro de 2000, que estabelece normas gerais e critérios básicos para a promoção da acessibilidade das pessoas portadoras de deficiência ou com mobilidade reduzida, e dá outras providências.
Portaria nº 310, de 27 de junho de 2006, do Ministério das Comunicações	Aprova a Norma Complementar nº 01/2006 - Recursos de acessibilidade, para pessoas com deficiência, na programação veiculada nos serviços de radiodifusão de sons e imagens e de retransmissão de televisão.
Decreto nº 6.949, de 25 de agosto de 2009	Promulga a Convenção Internacional sobre os Direitos das Pessoas com Deficiência e seu Protocolo Facultativo, assinados em Nova York, em 30 de março de 2007.
Portaria nº 188, de 24 de março de 2010	Altera o subitem 3.3 e o item 7 da Norma Complementar nº 01/2006 – Recursos de acessibilidade, para pessoas com deficiência, na programação veiculada nos serviços de radiodifusão de sons e imagens e de retransmissão de televisão, aprovada pela Portaria nº 310, de 27 de junho de 2006.
Portaria nº 312, de 26 de junho de 2012 (DOU de 29/06/12, página 63)	Altera texto do item 7.1 da Norma Complementar nº 1/2006, estabelecendo valor mínimo de horas para veiculação obrigatória do recurso de legenda oculta para emissoras do serviço de sons e imagens e de retransmissão de televisão.
Instrução Normativa Ancine nº 116, de 18 de dezembro de 2014	Dispõe sobre as normas gerais e critérios básicos de acessibilidade a serem observados por projetos audiovisuais financiados com recursos públicos federais geridos pela Ancine ; altera as Instruções Normativas nº 22/03, 44/05, 61/07 e 80/08, e dá outras providências.
Lei nº 13.146, de 6 de julho de 2015	Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência).

Fonte: (Enap, 2023) (Adaptado).

Descrição: A **Tabela 04**, apresenta uma tabela com os principais instrumentos legais, que versam sobre a audiodescrição no Brasil.

2.5.3 Audiodescrição Simultânea (ADS)

Na audiodescrição simultânea todas as fases da audiodescrição acontecem ao mesmo tempo em que os acontecimentos decorrem. Geralmente acontece quando o audiodescritor não tem conhecimento prévio dos acontecimentos. Recomenda-se usar verbos específicos que indiquem a maneira de realização das ações ex: correr, pular, agachar, entre outros.

Durante a audiodescrição de atributos físicos de um personagem é recomendável a seguinte sequência: gênero, faixa etária, etnia, cor da pele, estatura, compleição física, olhos, cabelos e demais características relevantes.

2.6 PROCESSAMENTO DE LINGUAGEM NATURAL

O Processamento de Linguagem Natural, tem como principal intuito, unir o aprendizado de máquina, com os elementos da linguagem humana (linguística). Em outras palavras, tem o propósito de entender a língua humana e habilitar as máquinas a compreender e responder de forma inteligente e compreensível, dentro desta mesma linguagem. De modo geral, as fases do processamento de linguagem natural, segmentam a linguagem humana em partes menores e essenciais, possibilitando o entendimento e relacionamento entre estas partes menores, a fim de criar um significado maior para estas partes. (Jéssica Rodrigues, 2017)

Figura 05 – Campos de estudo baseados no Processamento de Linguagem Natural.



Fonte: (Jones Granatyr, 2020) (adaptado).

Descrição: A **Figura 05**, contém um conjunto de 7 imagens dos principais campos de estudo do processamento de linguagem natural, sendo elas, na ordem em que foram dispostas na **Figura 05**: a transcrição de fala, a tradução automática neural, chatbots (chat inteligente), Q&A, resumo de textos, legendas em imagens e por fim, as legendas em vídeos.

Sistemas baseados em Processamento de Linguagem Natural, nos dias atuais, são desenvolvidos com base em premissas que levam em consideração não somente a clara necessidade de entender o significado literal das palavras em um texto ou fala, mas também, buscam apreender como interpretar os textos de uma forma satisfatória, sendo capaz de analisar as diversas intenções dos autores, para como o leitor, como também, ser capaz de detectar possíveis sentimentos expressos nestes textos de forma oculta ou de difícil compreensão. (Leandro Kovacs, 2021)

2.6.1 Reconhecimento de Voz

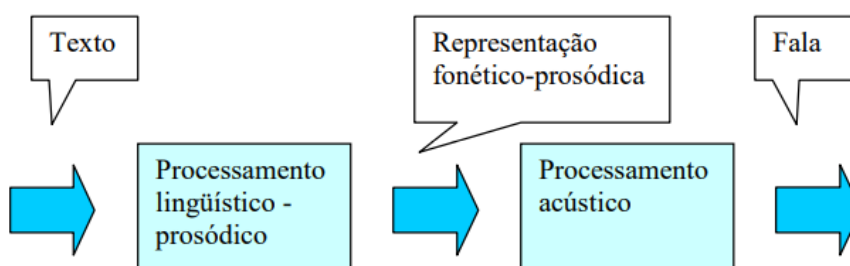
Os sistemas de reconhecimento de voz (ASR) podem ser separados em diferentes tipos, que determinam a maneira como o reconhecimento das falas serão realizadas, podendo ser por meio de (Calebe et al, 2008):

- Palavras isoladas;
- Palavras conectadas;
- Discurso contínuo;
- Fala espontânea;
- Verificação / identificação de voz.

2.6.2 Síntese de Voz

Síntese de voz compreende a criação artificial da voz humana. Diz respeito a um sistema que utiliza para este fim, um sintetizador, e pode ser implementado em software e em equipamentos de hardware, podendo ser criadas pela fusão de trechos de discursos pré gravados, armazenados em um banco de dados, específico para este fim, ou por meio da simulação da voz humana, de forma sintética.

Figura 06 – Fluxo para o tratamento dos textos dentro do TTS.



Fonte: (Calebe et al, 2008) (Adaptado).

Descrição: A **Figura 06**, representa o fluxo de tratamento de uma texto para convertê-lo em fala por meio da síntese de voz.

2.7 VISÃO COMPUTACIONAL

A visão computacional e a capacidade conferida às máquinas, de processamento das imagens captadas do mundo real, fazendo com que, estas máquinas, interpretem as informações contidas nestas imagens de uma forma lógica e coesa. Mesmo em um estágio em que, a visão computacional ainda possa ser considerada incipiente, ela vem possibilitando o aparecimento cada vez mais frequente de aplicações desenvolvidas com base nestes conhecimentos.

Aplicações desenvolvidas para aplicações militares, comerciais, de segurança pública, empresarial ou desenvolvimento de processos produtivos mais ágeis e eficientes. Estes avanços, ocorreram em maior monta, principalmente nos últimos anos, valendo a pena ser destacado, as duas principais empresas, líderes neste segmento, sendo elas a Tesla, e a Waymo do Google, ambas focadas no desenvolvimento e aprimoramento de sistemas embarcados presentes nos carros autônomos. (Bradski, 2008)

Entre os diversos tipos de ferramentas e/ou funcionalidades que foram desenvolvidas nos últimos anos, com base em visão computacional, destacam-se alguns delas:

- Detecção de tumores cancerígenos com base na análise de tomografias computadorizadas.
- Análises qualitativas de produtos nos processos produtivos de grandes indústrias.
- Análise de raio-x pulmonar com base em caso de covid-19.
- Detecção e reconhecimento facial.
- Rastreamento de pedestres e de objetos.
- Automação veicular.

Figura 07 – Áreas do conhecimento que se beneficiam da visão computacional.



Fonte: (PrograMaria, 2020) (adaptado) .

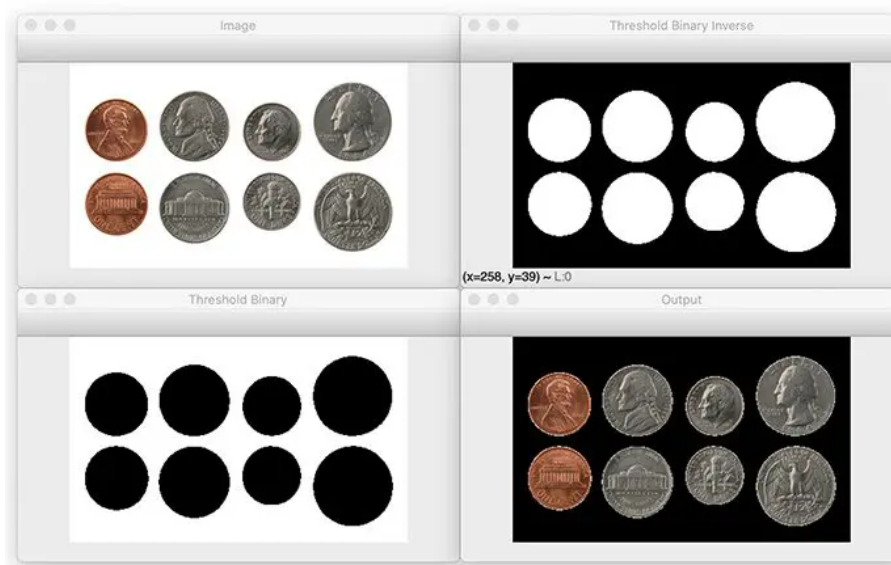
Descrição: A **Figura 07**, apresenta um círculo maior, centralizado que representa o campo da visão computacional, fazendo intersecção com este círculo maior, estão outros campos do conhecimento, como a matemática, a física, a inteligência artificial, dentre outros.

2.7.1 OpenCV

OpenCV é uma biblioteca de código aberto, para o desenvolvimento de aplicações com base na visão computacional. Tendo seu suporte disponível em linguagens de programação como Python, C ++ e Java. Atualmente OpenCV possui em seu acervo mais de 500 funções, de fácil implementação, sendo uma das bibliotecas mais utilizadas para o desenvolvimento de soluções neste segmento. É muito utilizada para o tratamento de imagens e no pré-processamento de imagens, dentre muitos outros casos de uso. (OpenCV, 2023)

Entre estes mais de 500 métodos diferentes, o método de thresholding tem a função de segmentar uma ou mais características presentes em uma imagem, separando o objeto de interesse que encontra-se, geralmente no primeiro plano, do plano de fundo desta mesma imagem.

Figura 08 – Utilização do método thresholding para segmentação de objetos.



Fonte: (Adrian Rosebrock, 2021) (adaptado) .

Descrição: A **Figura 08**, apresenta os diferentes tipos de resultados obtidos através da utilização do método thresholding em uma imagem, na imagem em questão, existem moedas com tamanhos e cores diferentes, segmentando e separando estas moedas, com relação ao plano de fundo desta mesma imagem.

2.7.1.1 Histogram of oriented gradients (HOG)

Antes de tratar dos principais conceitos referentes ao método HOG se faz necessária uma breve contextualização do que são os descritores visuais para a visão computacional. Os descritores visuais, tem o objetivo de descrever os recursos visuais presentes em um imagem, descrevendo suas características elementares, sendo eles por exemplo: a distribuição de cores, a presença de formas geométricas determinadas, a existência de diferentes tipos de texturas, dentre muitas outras características. (SILVA, 2018)

Diante disso, torna-se possível a extração de características específicas de uma ou mais pessoas em uma imagem, tais como: cabeça, braços, pernas, mãos, cor da pele, cor dos olhos, dentre outras. Estas características podem ser utilizadas para a construção e treinamento de um novo modelo de aprendizagem de máquina, a fim de obter resultados mais aprimorados para as detecções, aprimorando a obtenção de resultados a cada novo treinamento executado.

2.7.1.2 Classificadores Haar Cascades

Os classificadores Haar Cascades representam uma abordagem baseada no aprendizado de máquina, onde as imagens positivas e negativas são usadas para o treinamento do classificador. As diferenças fundamentais entre, imagens positivas e negativas serão descritas na **Tabela 05** a seguir. (Arnaldo Gualberto, 2018)

Tabela 05 – Descrição dos diferentes tipos de imagens utilizadas pelo classificador Haar Cascades.

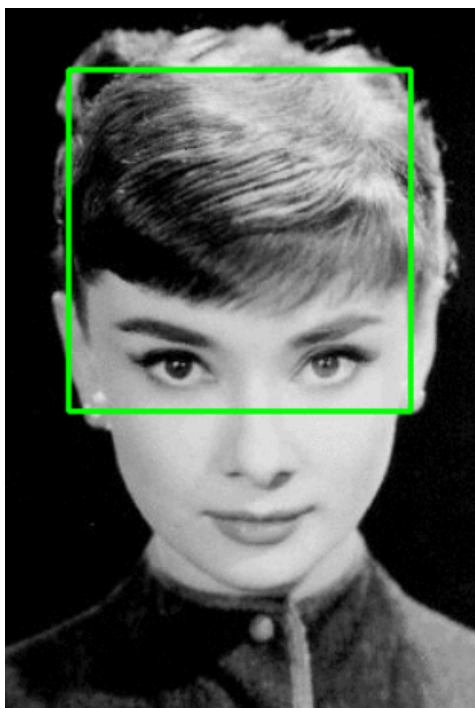
Imagens Positivas	Imagens Negativas
Contém as imagens que se deseja que o classificador identifique.	Representam todo o resto, ou seja, imagens que não contém o objeto/item de interesse.

Fonte: Elaboração própria.

Descrição: A **Tabela 05**, contém uma descrição das diferenças entre as imagens positivas e negativas no treinamento do classificador Haar Cascades.

Em seu artigo, intitulado, “Rapid Object Detection using a Boosted Cascade of Simple Features”, (Paul Viola e Michael Jones, 2001), propuseram a criação de um algoritmo capaz de detectar os objetos presentes em imagens, independentemente de sua localização ou escala, ainda sendo capaz de detectar os objetos em fluxo de vídeo pré-gravados ou em tempo real. Na **Figura 09** a seguir, está exemplificado o funcionamento deste algoritmo em uma imagem. (Silva, 2018)

Figura 09 – Classificador Haar Cascade para detecção facial.



Fonte: (Adrian Rosebrock, 2021) (adaptado).

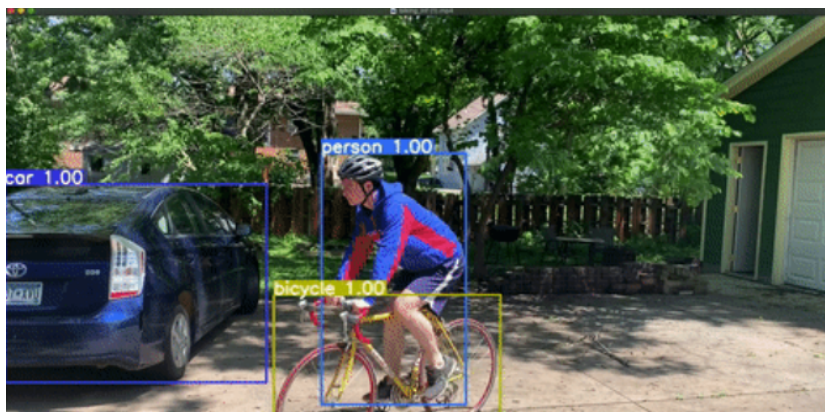
Descrição: A **Figura 09**, demonstra como o funcionamento do algoritmo por trás do classificador Haar Cascades percorre a imagem, a fim de detectar o objeto/item de interesse.

O Haar Cascade pode ser utilizado para detectar diversos tipos de objetos, contudo, se faz necessário a existência de um arquivo com a extensão “.xml”, devidamente treinado com objeto que se deseja detectar.

2.7.2 You Only Look Once (YOLO)

Desenvolvido para detecção e rastreamento de objetos com base em redes neurais convolucionais de disparo único, o YOLO, que é uma abreviação de “you only look once”, que traduzindo para o português, significa “você só olha uma vez”. Foi construído e é mantido em uma arquitetura conhecida como Darknet. O YOLOv5 (versão número cinco, que é a mais atual) representa o primeiro dos modelos YOLO, desenvolvido na estrutura PyTorch <<https://pytorch.org/docs/stable/index.html>>, sendo consideravelmente mais leve e de fácil utilização. (Joseph, 2015)

Figura 10 – Rastreamento de objetos com YOLO.



Fonte: (Jobu, 2023) (adaptado).

Descrição: A **Figura 10**, exemplifica como ocorre o rastreamento de diferentes tipos de objetos, em uma imagem, utilizando o YOLO. A imagem em questão está no formato "gif", contendo um homem andando em uma bicicleta, com um carro ao fundo, estes três objetos distintos são detectados em simultâneo.

O YOLOv5 funciona com uma velocidade significativamente maior, quando comparado a outros algoritmos contemporâneos para detecção e rastreamento de objetos, funcionando a 45 FPS (quarenta e cinco quadros por segundo, representando a quantidade de imagens exibidas na tela em apenas um segundo).

2.7.3 Legendas em Imagens

Através da utilização do conjunto de dados, é possível atribuir a uma imagem uma legenda apropriada, contendo as características principais daquela imagem em questão. Uma ferramenta desenvolvida dentro deste conceito é o Flickr, que está disponível de forma gratuita na plataforma Kaggle (Plataforma para competições de Datascience), sendo capaz de funcionar de maneira satisfatória, em computadores com poder de processamento reduzido. (HSANKESARA, 2018)

2.7.3.1 Flickr 8k

O Flickr 8k é uma base de dados, capaz de atribuir legendas apropriadas em imagens, contém em sua base de dados 8000 imagens, divididos em 3 conjuntos. Sendo o primeiro, o conjunto principal de treinamento contendo 6000 imagens de

referência. Um outro conjunto específico para desenvolvimento contendo um total de 1000 imagens, e por fim, um conjunto de teste, composto por outras 1000 imagens. Cada uma destas imagens, tem atribuídas a elas, 5 legendas distintas que as descrevem. (HSANKESARA, 2018)

Um ponto importante, que vale a pena ser observado com relação a criação desta base de dados, diz respeito a sua forma de seleção, tendo em vista que as imagens tendem a não conter pessoas ou locais conhecidos do público em geral, é estas imagens foram selecionadas manualmente a fim de retratar uma variedade maior de cenários e situações.

Especificações com relação aos números do Flickr 8k, destaca-se, que deste de sua última grande atualização, que ocorreu no ano de 2020, foram contabilizados 104 contribuições diretas feitas a base de dados. O somatório de transferências que indicam a quantidade de downloads feitos desde sua criação, ultrapassa o número de 20 mil.

Ainda com relação ao conjunto de dados e sua formação, as duas Figuras que eram apresentadas a seguir neste capítulo, sendo elas a, **Figura 11** e a **Figura 12**, demonstraram respectivamente a maneira como foram distribuídas as 5 legendas a cada imagem de referência, em seguida, os diferentes cenários e situações representados nas imagens presentes na base de dados.

Figura 11 – Disposição das 5 legendas atribuídas a cada uma das 8 mil imagens.

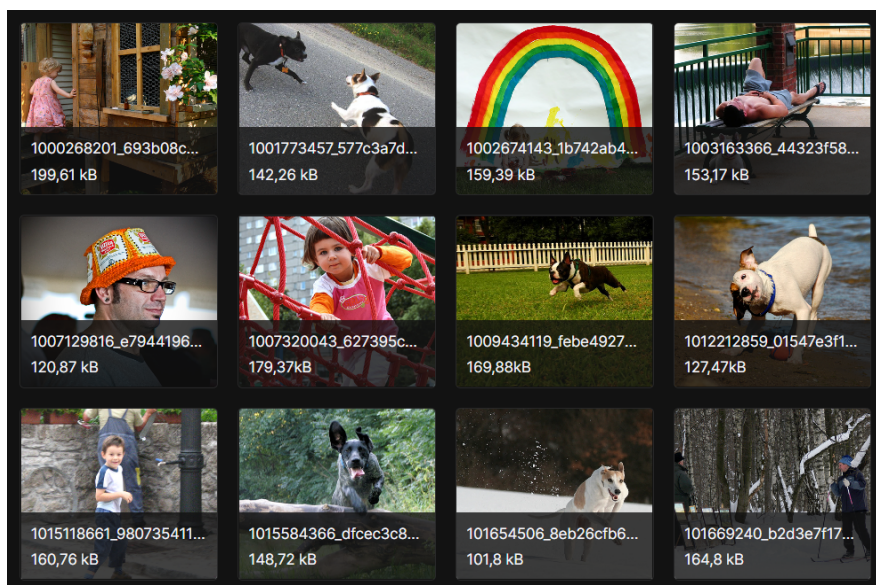
```
image,caption
1000268201_693b08cb0e.jpg,A child in a pink dress is climbing up a set of stairs in an entry way .
1000268201_693b08cb0e.jpg,A girl going into a wooden building .
1000268201_693b08cb0e.jpg,A little girl climbing into a wooden playhouse .
1000268201_693b08cb0e.jpg,A little girl climbing the stairs to her playhouse .
1000268201_693b08cb0e.jpg,A little girl in a pink dress going into a wooden cabin .
1001773457_577c3a7d70.jpg,A black dog and a spotted dog are fighting
1001773457_577c3a7d70.jpg,A black dog and a tri-colored dog playing with each other on the road .
1001773457_577c3a7d70.jpg,A black dog and a white dog with brown spots are staring at each other in the
1001773457_577c3a7d70.jpg,Two dogs of different breeds looking at each other on the road .
1001773457_577c3a7d70.jpg,Two dogs on pavement moving toward each other .
1002674143_1b742ab4b8.jpg,A little girl covered in paint sits in front of a painted rainbow with her hair
1002674143_1b742ab4b8.jpg,A little girl is sitting in front of a large painted rainbow .
1002674143_1b742ab4b8.jpg,A small girl in the grass plays with fingerpaints in front of a white canvas w
1002674143_1b742ab4b8.jpg,There is a girl with pigtails sitting in front of a rainbow painting .
1002674143_1b742ab4b8.jpg,Young girl with pigtails painting outside in the grass .
1003163366_44323f5815.jpg,A man lays on a bench while his dog sits by him .
1003163366_44323f5815.jpg,a man sleeping on a bench outside with a white and black dog sitting next to h
1003163366_44323f5815.jpg,A shirtless man lies on a park bench with his dog .
```

Fonte: (HSANKESARA, 2018) (adaptado).

Descrição: A **Figura 11**, diz respeito a um print screen, contendo a disposição das 5 legendas atribuídas a cada uma das 8 mil imagens de referência, presentes no conjunto de dados, esta imagem em questão, foi extraída da página

oficial do conjunto de dados Flickr 8k, na plataforma, kaggle para competição em ciência de dados.

Figura 12 – Exemplos de imagens presentes no conjunto de dados Flickr 8k.



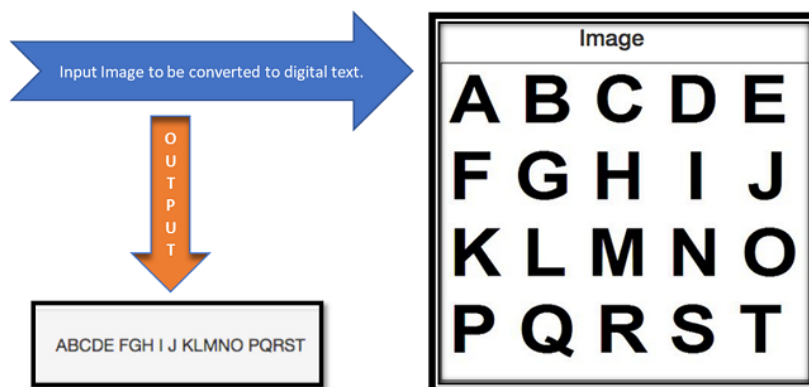
Fonte: (HSANKESARA, 2018) (adaptado).

Descrição: A **Figura 12**, diz respeito a um print screen, contendo algumas imagens de referência, que tem por objetivo representar diferentes tipos de cenários reais, contento tanto pessoas de nacionalidades diferentes quanto paisagens em países diferentes, esta imagem em questão, foi extraída da página oficial do conjunto de dados Flickr 8k, na plataforma, kaggle para competição em ciência de dados.

2.7.4 Reconhecimento óptico de caracteres (OCR)

O surgimento da visão computacional, teve como grande fator de motivação as necessidades de criação de novas e eficazes ferramentas que possibilitam às pessoas com deficiência visual adentrarem na vida educacional e no mercado de trabalho, de forma inclusiva, o que por consequência impulsionou pesquisas referentes a este tema, mais especificamente no desenvolvimento de ferramentas especializadas no reconhecimento óptico de caracteres (OCR). Direcionadas para aumentar as capacidades funcionais das pessoas com deficiência visual. (Lucas et al, 2021)

Figura 13 – Representação do funcionamento do (OCR).



Fonte: (edureka, 2021) (adaptado).

Descrição: A **Figura 13**, contém uma representação do funcionamento do OCR, ilustrando uma imagem de entrada contendo diversas letras do alfabeto, em seguida esta imagem é convertida digitalmente, é sua saída, apresentando todas as letras contidas na imagem de forma digitalizada.

3. METODOLOGIA DE DESENVOLVIMENTO PROPOSTA

Tem crescido cada vez mais o desenvolvimento de ferramentas com base na visão computacional e no processamento de linguagem natural, isto acontece, graças a tecnologias construídas em linguagens como, C++ e Java, em especial, aplicações construídas em Python, sendo responsável por abstrair grande parte dos complexos cálculos matemáticos necessários para o funcionamento destes sistemas. (mordorintelligence, 2023)

Neste capítulo, será apresentado de forma prática, o funcionamento de um sintetizador de voz, e uma biblioteca para o reconhecimento de voz, em seguida, será abordado o funcionamento das bibliotecas de visão computacional, OpenCV e YOLO para Detecção e Rastreamento de Pedestres, Detecção e Reconhecimento Facial, Estimar Distância e a Antecipação de Obstáculos, Detecção de Linhas e a Detecção de Cantos, também, será demonstrado, algoritmos que implementam a base de dados Flickr 8k para geração de legendas em imagens. E por fim, o reconhecimento óptico de caracteres por meio da ferramenta Tesseract.

Todos os códigos apresentados neste capítulo, estão disponíveis no repositório “GuiaInteligente”, em anexo a este documento. Vale ressaltar que a base de dados de imagens e vídeos, utilizados neste trabalho de conclusão de curso, contidos no repositório “GuiaInteligente”, foram produzidos em caráter de simulação, exclusivamente nas dependências internas do Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco IFPE, campus Jaboatão dos Guararapes.

As gravações foram feitas entre os dias 04/07/2022 a 16/11/2022 em 4 momentos diferentes, sendo todos no período da manhã. Deve-se destacar que todas as pessoas envolvidas na elaboração desta base de dados, são pessoas com visão e não possuem dificuldades significativas para enxergar.

Com o intuito de facilitar o entendimento da metodologia de desenvolvimento proposta neste trabalho, foi feita uma divisão com relação a ordem de apresentação das ferramentas abordadas neste documento. Sendo esta divisão, disposta da seguinte forma, como está descrito na **Tabela 06** a seguir.

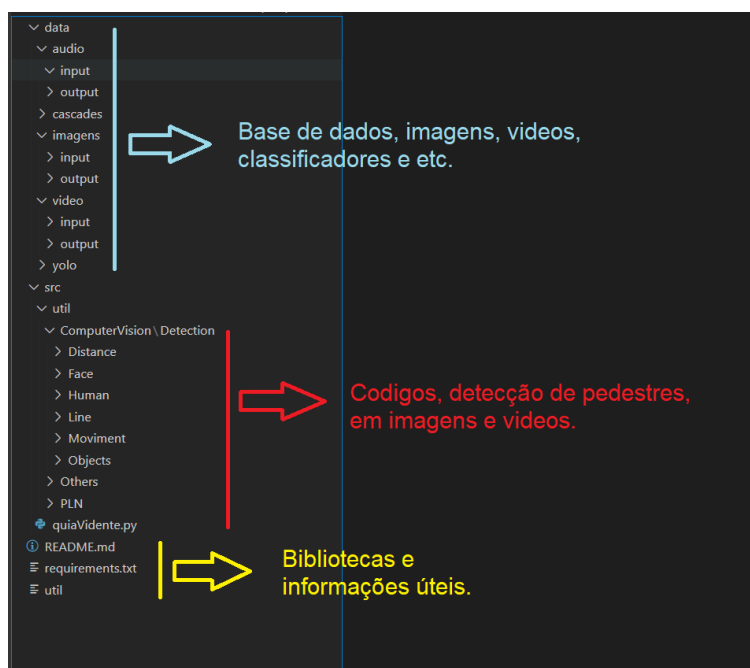
Tabela 06 – Sequência das funcionalidades que serão abordadas no **Capítulo 03**.

OpenCV	You Only Look Once	Speech Recognition	Flickr 8k	Google Text to speech (gTTs)	Tesseract (OCR)
					

Fonte: Elaboração própria

Descrição: A **Tabela 06**, apresenta uma tabela, a fim de simplificar a ordem cronológica dos conteúdos que serão abordados da metodologia proposta neste capítulo.

Para um bom entendimento do projeto, se faz necessário organizar os diversos componentes referente às tecnologias atualizadas, neste sentido, o projeto “GuiaInteligente”, foi organizado sob a seguinte estrutura de diretórios e arquivos. A estrutura em questão será exemplificada na **Figura 14** a seguir.

Figura 14 – Estrutura do projeto “GuiaInteligente”.

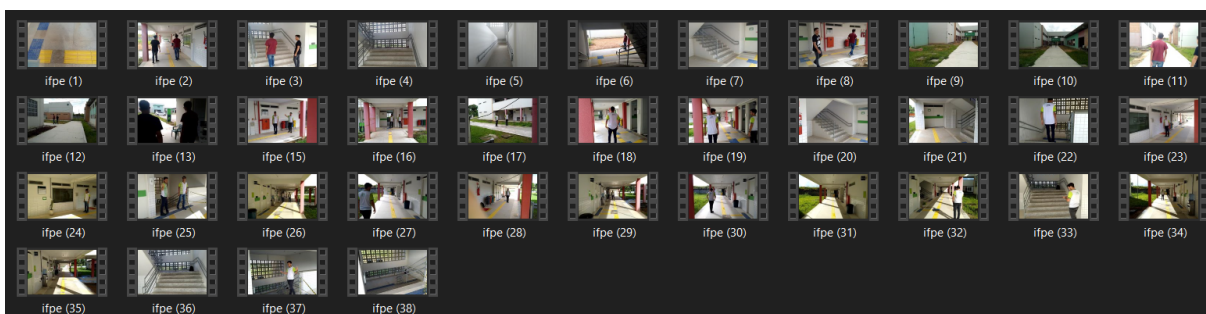
Fonte: Elaboração própria.

Descrição: A **Figura 14**, contém um print screen do projeto “GuiaInteligente”, aberto na IDE visual studio code, contendo a estrutura de pastas e subpastas, assim como os arquivos utilizados no projeto.

A estrutura foi montada com o objetivo de sintetizar os três elementos básicos para a execução dos códigos e a obtenção de seus respectivos resultados com a utilização da base de dados simulada. Na parte superior da imagem, destacada em azul, estão as imagens, os vídeos e os classificadores Haar Cascades. No centro da imagem, está destacado em vermelho, os códigos executáveis “.py” (arquivo no formato python), e por último, na parte inferior da imagem, destacado em amarelo estão as informações de configuração do projeto e a forma de instalação das bibliotecas necessárias.

Foram gravados uma série de vídeos, e a partir destes vídeos, geradas imagens, em locais e dias diferentes, locais que geralmente são acessados com relativa frequência pelos alunos do campus Jaboatão dos Guararapes, sendo estes locais, os corredores, as escadarias de acesso ao primeiro andar dos blocos C e D.

Figura 15 – Série de vídeos gravados no campus Jaboatão dos Guararapes.



Fonte: Elaboração própria.

Descrição: A **Figura 15**, apresenta um print screen de um repositório do projeto “GuiaInteligente”, destinada a armazenar a base de dados simulada utilizada nos testes.

No que diz respeito, a apresentação prática das ferramentas citadas no **Tabela 06**, esta disposição será apresentada a seguir.

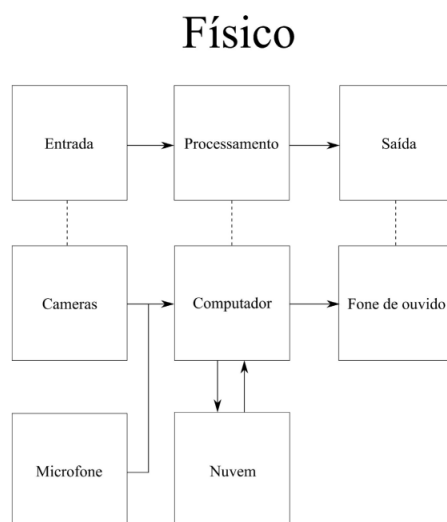
- Detecção e Rastreamento de Pedestres.
- Detecção e Reconhecimento Facial.
- Estimar a Distância e a Antecipação de Obstáculos.
- Detecção de Linhas e a Detecção de Cantos

- Flickr 8k.
- Tesseract (OCR).

Será proposta a integração destas ferramentas e conhecimentos a fim de gerar uma ferramenta capaz de oferecer maior autonomia e segurança para o deslocamento das pessoas com deficiência visual, nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes, este sistema está fundamentado, nos conhecimentos desenvolvidos no **Capítulo 02**, sendo eles, as tecnologias assistivas, a audiodescrição e a orientação e mobilidade, a visão computacional e o processamento de linguagem natural, trabalhando de forma integrada, em favor da acessibilidade. A **Figura 16**, a seguir, apresenta a representação Física por meio de um fluxograma.

Esta representação física da ferramenta divide-se em dispositivos de entrada, processamento e saída de dados. O sistema de entrada e saída de dados são os responsáveis pela comunicação com o usuário e o ambiente, o que se torna possível através de um dispositivo de processamento capaz de relacionar as interações por meio de um circuito eletrônico.

Figura 16 – Representação física da ferramenta “Guiainteligente”.

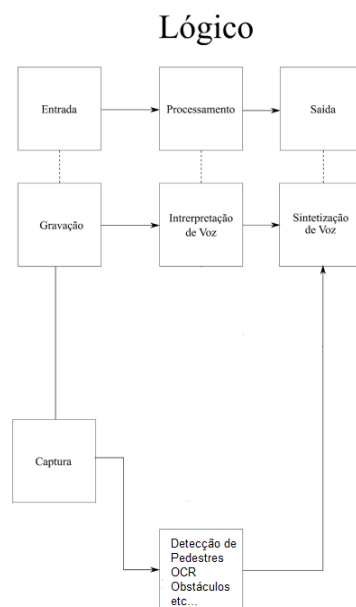


Fonte: (Scientia Prima, 2020) (Adaptado).

A representação lógica da ferramenta, descrita na **Figura 16**, se encarrega de interpretar os sinais visuais/gravações vindos do ambiente a serem reconhecidos, realizando a conversão destes dados em sinais de áudio, por meio de uma sintetizador de voz, responsável pela interação com o usuário. É ele quem interpreta

as informações sonoras obtidas e tomará a melhor descrição de locomoção com base nestas informações, observando de forma primordial, a segurança e integridade física do usuário.

Figura 17 – Representação lógica da ferramenta “GuiaInteligente”.



Fonte: (Scientia Prima, 2020) (Adaptado).

Descrição: A **Figura 17**, contém a representação em uma fluxograma, com a estrutura lógica do projeto “GuiaInteligente”.

3.1 Design Science Research (DSR)

Design Science é o paradigma para método de pesquisa, que dá suporte ao desenvolvimento de artefatos inovadores para resolução de problemas demasiadamente complexos na área de sistemas da informação (S.I). Sendo a maioria dos trabalhos teóricos com base em DSR para a pesquisa científica, contidos no campo de sistemas de informação.

“Área do conhecimento que trata do planejamento, da programação e do projeto dos objetos com os quais o homem lida em seu cotidiano, assim como dos ambientes em que mantém seu espaço de vida”.(Escola Superior de Design Industrial da UERJ - ESDI)

Segundo Alan Hevner et al, 2004, a pesquisa DSR se baseia em 7 diretrizes, sendo fundamentais, sendo elas:

- Diretriz 1: Design como um artefato.
- Diretriz 2: Relevância do problema.
- Diretriz 3: Avaliação em Design Science Research.
- Diretriz 4: Contribuição da pesquisa.
- Diretriz 5: Rigor em Design Science Research.
- Diretriz 6: Design como um processo de busca.
- Diretriz 7: Comunicação dos resultados.

Seguindo as diretrizes, estipuladas por Alan Hevner et al, 2004, se faz necessário realizar um levantamento primário dos materiais e métodos, que devem ser empregados, para o bom desenvolvimento do projeto, criado com base nos conhecimentos adquiridos durante a criação deste trabalho de conclusão de curso, na **Tabela 07** a seguir, será descrito, estes elementos.

Tabela 07 – Recursos utilizados ou para futura integração.

Itens	Utilizado
Notebook samsung essentials e30 (Possui)	Sim
Fone de condução óssea (Possui)	Sim
SmartPhone LG k22 (Possui)	Sim
Webcam (Possui)	Sim
Raspberry pi (Não possui)	Não
Óculos com fone de condução óssea integrado e micro câmera acoplada (Não possui)	Não

Fonte: Elaboração própria

Descrição: A **Tabela 07**, lista uma série de itens utilizados no desenvolvimento do projeto “GuiaInteligente”.

3.2 Síntese e o Reconhecimento de voz

Nesta seção, serão demonstrados os principais pontos referentes ao funcionamento da biblioteca gTTs (Google Text-to-Speech) para conversão de texto

em fala TTs (Text-to-Speech), em seguida será abordado o funcionamento da biblioteca `speech_recognition` para o reconhecimento de fala. Neste sentido, a partir das detecções resultantes dos métodos de visão computacional, que serão abordados nas seções seguintes, resultaram textos com, pedestres, rostos, distância, obstáculos, linhas e cantos detectados, assim como textos por meio do reconhecimento óptico de caracteres.

Para utilizar a biblioteca gTTs, primeiro se faz necessário, instalar por meio do comando `pip` a referida biblioteca, utilizando o comando “`pip install gTTs`”, em seguida, deve-se importar a biblioteca da seguinte forma “`from gtts import gTTs`”, na maneira como o projeto foi construído, a partir do texto, atribuído para teste, o código salva o arquivo em formato “.mp3” e o executa em simultâneo. A sintetização de voz é própria do Google e no formato português (pt-br).

Figura 18 – Código que implementa a biblioteca gTTs.

```
1  # Import the required module for text
2  # to speech conversion
3  from gtts import gTTS
4
5  # This module is imported so that we can
6  # play the converted audio
7  import os
8
9  # The text that you want to convert to audio
10 mytext = 'Guia Inteligente'
11
12 # Language in which you want to convert
13 language = 'pt-br'
14
15 myobj = gTTS(text=mytext, lang=language, slow=False)
16
17 # Saving the converted audio in a mp3 file named
18 myobj.save("data\\audio\\output\\audiodescricao.mp3")
19
20 # Playing the converted file
21 os.system("data\\audio\\output\\audiodescricao.mp3")
```

Fonte: (Geeksforgeeks, 2023) (Adaptado).

Descrição: A **Figura 18**, contém um print screen do código “text_to_voice.py”. Na variável que recebe o texto do qual se deseja, a sintetização de voz, contém a seguinte frase: “Guia Inteligente”, é a variável que recebe, em qual linguagem se deseja, para com o resultado do sintetizador contém o Codex Multilíngue: “pt-br”.

3.3 Detecção e o Rastreamento de Pedestres

Nesta seção, serão demonstrados os principais pontos necessários para **Detecção e Rastreamento de pedestres** por meio da utilização do método Histogramas de gradientes orientados (HOG), e em seguida a implementação dos classificadores Haar Cascades. A **Figura 19** a seguir apresenta a sequência de códigos necessários para a implementação do método HOG.

Figura 19 – Código que implementa o método HOG na linguagem de programação Python.

```

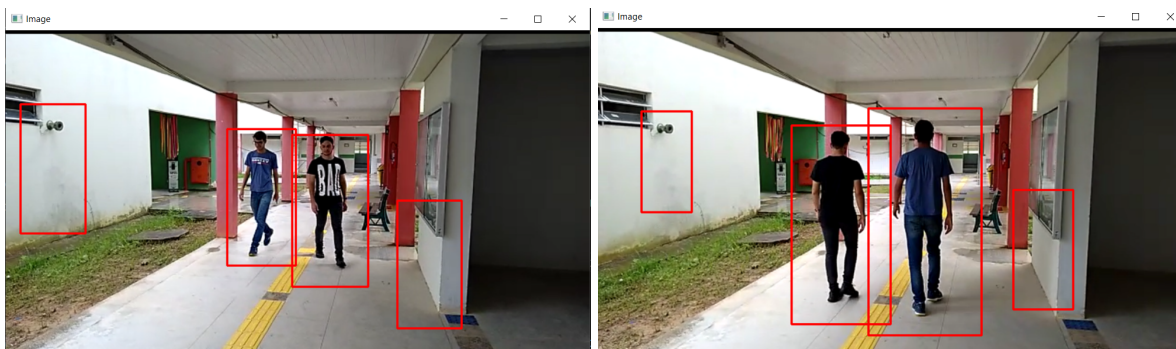
1  import cv2
2  import imutils
3
4  # Inicializando a pessoa HOG
5  # detector
6  hog = cv2.HOGDescriptor()
7  hog.setSVMDetector(cv2.HOGDescriptor_getDefaultPeopleDetector())
8
9  # Lendo a imagem
10 image = cv2.imread('data\\imagens\\input\\ifpe teste 01.1.png')
11
12 # Redimensionando a imagem
13 image = imutils.resize(image,
14                         width=min(800, image.shape[1]))
15
16 # Detectando todas as regiões do
17 # Imagem que tem um pedestre dentro dela
18 (regions, _) = hog.detectMultiScale(image,
19                                     winStride=(4, 4),
20                                     padding=(4, 4),
21                                     scale=1.05)
22
23 # Desenhando as regiões na imagem
24 for (x, y, w, h) in regions:
25     cv2.rectangle(image, (x, y),
26                  (x + w, y + h),
27                  (0, 0, 255), 2)
28
29 # Mostrando a imagem de saída
30 cv2.imshow("Image", image)
31 cv2.waitKey(0)
32
33 cv2.destroyAllWindows()

```

Fonte: (Geeksforgeeks, 2023) (Adaptado).

Descrição: A **Figura 19**, contém um print screen do código “image_pedestrian_detection_01.py”, que implementa o conjunto de métodos HOG para detecção e rastreamento de pedestres.

Figura 20 – Múltiplos Bounding box (caixas delimitadoras) geradas nas imagens.



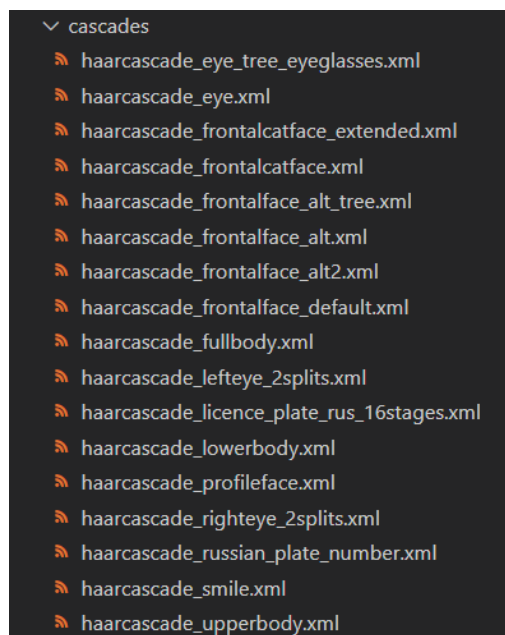
Fonte: Elaboração própria.

Descrição: A **Figura 20**, apresenta um print screen com a saída do método “cv2.imshow”, onde foram detectados dois pedestres/alunos, enquanto caminhavam juntos no corredor de acesso ao bloco D do Instituto Federal, campus Jaboatão dos Guararapes. Na primeira imagem os alunos encontram-se de frente para a câmara, já na segunda imagem, os dois alunos encontram-se de costas para a câmera.

Avaliando os resultados para esta detecção em específico, percebe-se, a ocorrência de caixas delimitadoras em falsos positivos, isto se dá, provavelmente em decorrência de fatores externos ao método HOG. Os fatores para este resultado são diversos, a baixa luminosidade no ambiente, tendo em vista o horário em que foi gravado, a baixa qualidade da câmera utilizada para a captura dessas imagens, são exemplos de fatores que podem influenciar de forma significativa a obtenção dos resultados.

Com relação aos procedimentos necessários para Detecção e Rastreamento de Pedestres, por meio dos classificadores Haar Cascades. Na **Figura 21**, a seguir, será demonstrado os diferentes tipos de classificadores disponíveis, Cada classificador Haar cascade, corresponde a um tipo distinto e específico de detecção, o que significa por exemplo que, o arquivo “haarcascade_eye.xml”, foi treinado especificamente para detecção dos olhos, neste sentido a sua utilização para detecção de outros elementos do corpo humano, não resultaram em êxito.

Figura 21 – Diferentes tipos de Classificadores Haar Cascades disponíveis.



Fonte: Elaboração própria.

Descrição: A **Figura 21**, contém um print screen do projeto “GuiaInteligente”, com os classificadores Haar Cascades, importados para o projeto.

Figura 22 – Código que implementa os Classificadores Haar Cascades na linguagem de programação Python.

```

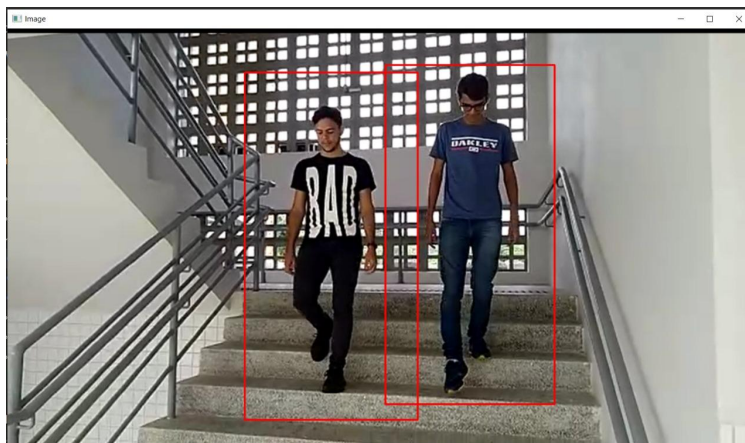
1  import cv2
2
3  print("Versão do OpenCV:", cv2.__version__)
4
5  # Utilização do Classificador para detecção de todo o corpo
6  classificador = cv2.CascadeClassifier("data\\cascades\\haarcascade_fullbody.xml")
7  # Video de referência
8  webCam = cv2.VideoCapture("data\\video\\input\\ifpe teste 01.2.mp4")
9
10 while(True):
11     conectou, imagem = webCam.read()
12     converteuCinza = cv2.cvtColor(imagem, cv2.COLOR_BGR2GRAY)
13
14     encontrarFases = classificador.detectMultiScale(converteuCinza,
15                                                     scaleFactor=1.5,
16                                                     minSize=(150,150),
17                                                     maxSize=(200,200))
18
19     cor = (0,0,255)
20     for(origemX, origemY, largura, altura) in encontrarFases:
21         cv2.rectangle(imagem,(origemX,origemY),
22                       (origemX + largura, origemY + altura),
23                       cor,2)
24         print("Largura", largura, "Altura", altura)
25     cv2.imshow("Rosto", imagem)
26
27     teclou = cv2.waitKey(20) & 0xFF
28     if teclou == ord('q') or teclou == 27: # se apertar q ou ESC
29         break
30
31 webCam.release()
32 cv2.destroyAllWindows()

```

Fonte: Elaboração própria.

Descrição: A **Figura 22**, contém um print screen do código “image_pedestrian_detection_02.py”, que implementa os classificadores Haar Cascades para detecção e rastreamento de pedestres.

Figura 23 – Bounding box contendo a detecção de dois alunos.



Fonte: Elaboração própria.

Descrição: A **Figura 23**, apresenta um print screen com a saída do método “cv2.imshow”, com a detecção de dois pedestres/alunos. Foram geradas duas caixas delimitadoras para cada aluno detectado no vídeo em questão, estes alunos encontram-se, descendo as escadarias do primeiro andar do bloco D, campus Jaboatão dos Guararapes. Na perspectiva em que a imagem foi gerada, os alunos estão de frente para a câmera.

Um fator que pode ter impacto significativo sobre os resultados obtido, diz respeito à quantidade limitada de imagens que foram utilizadas para gerar o classificador, tendo em vista que a base de arquivos “.xml”, utilizada nos testes fazem parte de uma base de modelos, pré-prontos, disponíveis de forma gratuita na internet.

Com relação ao funcionamento do YOLO para detecção e rastreamento de pedestres, é importante ressaltar que, devido a questões de domínio da ferramenta, os testes com a biblioteca em questão foram realizados exclusivamente com a utilização de imagens estatísticas, a partir dos vídeos produzidos no Instituto Federal, campus Jaboatão dos Guararapes.

A sequência com que foram dispostas as **Figuras 24, 25 e 26**, dizem respeito respectivamente a, imagem de teste gerada a partir de um vídeo, um print screen do código implementado na plataforma Google Colab e por fim, um print screen do console da ferramenta Google Colab com os resultados.

Figura 24 – Imagem carregada na plataforma Google Drive.



Fonte: Elaboração própria.

Descrição: A **Figura 24**, apresenta dois alunos/pedestres, descendo as escadarias do primeiro andar do bloco C, do Instituto Federal, campus Jaboatão dos Guararapes, na perspectiva em que a imagem foi gerada, os alunos estão de costas para quem está filmando.

Figura 25 – Código que implementa o YOLO na plataforma Google Colab.

```
import torch

# Model
model = torch.hub.load('ultralytics/yolov5', 'yolov5s') # or yolov5m, yolov5l, yolov5x, etc.
# model = torch.hub.load('ultralytics/yolov5', 'custom', 'path/to/best.pt') # custom trained model

# Images
im = '/content/drive/MyDrive/dados/ifpe teste 02.4.png' # or file, Path, URL, PIL, OpenCV, numpy, list

# Inference
results = model(im)

# Results
results.print() # or .show(), .save(), .crop(), .pandas(), etc.

results.xyxy[0] # im predictions (tensor)
results.pandas().xyxy[0] # im predictions (pandas)
#      xmin    ymin    xmax    ymax  confidence  class  name
# 0  749.50  43.50  1148.0  704.5    0.874023    0  person
# 2  114.75  195.75  1095.0  708.0    0.624512    0  person
# 3  986.00  304.00  1028.0  420.0    0.286865   27  tie
```

Fonte: Elaboração própria.

Descrição: A **Figura 25**, apresenta um print screen do código, implementado na ferramenta Google Colab.

A seguir na **Figura 26**, apresenta os resultados exibidos console da ferramenta Google Colab.

Figura 26 – Resultado para a detecção com YOLO na plataforma Google Colab.

```

100% 14.1M/14.1M [00:00<00:00, 24.6MB/s]
INFO:yolov5:
INFO:yolov5:Fusing layers...
Fusing layers...
INFO:yolov5:YOLOv5s summary: 213 layers, 7225885 parameters, 0 gradients
YOLOv5s summary: 213 layers, 7225885 parameters, 0 gradients
INFO:yolov5:Adding AutoShape...
Adding AutoShape...
image 1/1: 660x1200 2 persons
Speed: 98.1ms pre-process, 323.9ms inference, 18.7ms NMS per image at shape (1, 3, 352, 640)

```

	xmin	ymin	xmax	ymax	confidence	class	name
0	619.106934	158.268936	759.134888	538.148682	0.919808	0	person
1	351.085999	211.949783	490.588165	568.969849	0.907578	0	person

Fonte: Elaboração própria.

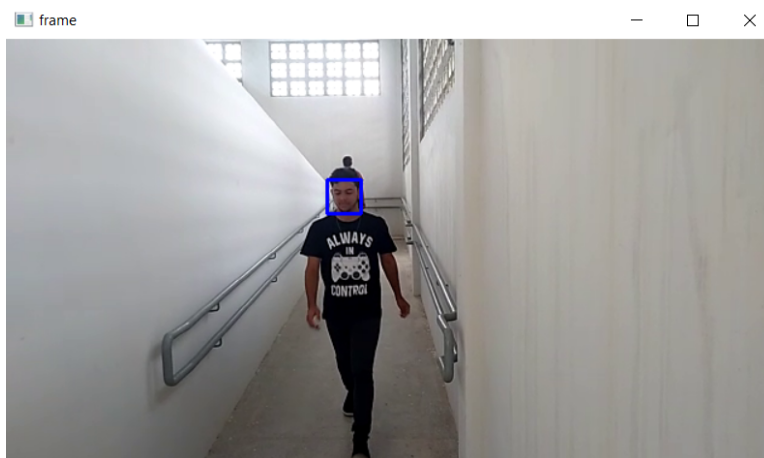
Descrição: A **Figura 26**, apresenta um print screen com o resultado da execução do código apresentado na **Figura 25** com a detecção de dois pedestres/alunos.

A **Figura 26**, indica que foram encontradas na imagem duas pessoas, e que o grau de confiança em porcentagem para estas detecções é de 91%, para o vetor [0], e 90% de confiança para o vetor [1].

3.4 Detecção e o Reconhecimento Facial

Nesta seção será demonstrado o funcionamento da biblioteca de visão computacional OpenCV, e os classificadores Haar Cascades implementados na linguagem de programação Python para a detecção e reconhecimento facial de alunos, funcionários e frequentadores do Instituto Federal, campus Jaboatão dos Guararares.

Figura 27 – Detecção da face de um aluno com a geração de uma caixa delimitadora.



Fonte: Elaboração Própria.

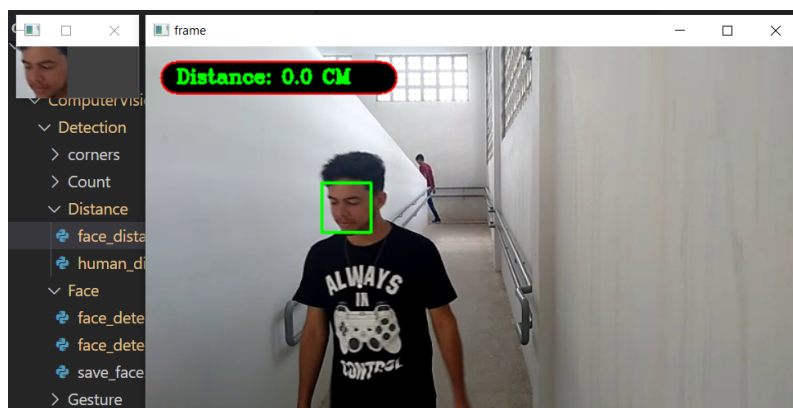
Descrição: A **Figura 27**, apresenta um print screen com a saída do método “cv2.imshow”, na imagem em questão o aluno/pedestre, encontra-se descendo o corredor a rampa de acesso ao primeiro andar do bloco D, do Instituto Federal, campus Jaboatão dos Guararapes.

3.5 Estimar Distância e a Antecipação de Obstáculos

Nesta seção será demonstrado o funcionamento da biblioteca de visão computacional OpenCV, e os classificadores Haar Cascades, implementados na linguagem de programação Python, a fim de, estimar a distância entre um objeto de referência pré-definido e um objeto alvo. A fim de simplificação, e com o intuito de gerar um número maior de casos de sucesso para a estimação de distância, foi utilizado o classificador “haarcascade_frontalface_default.xml”, com enfoque na detecção do rosto frontal de uma pessoa.

A escolha para se utilizar este classificador em específico, nasce em razão de que, como o usuário/pessoa com deficiência, utilizar o dispositivo (micro câmera), na altura dos olhos, facilitará a detecção dos rostos de outras pessoas, que estão dentro do campo de visão do dispositivo de captura.

Figura 28 – Imagem de teste utilizado para se estimar a distância.

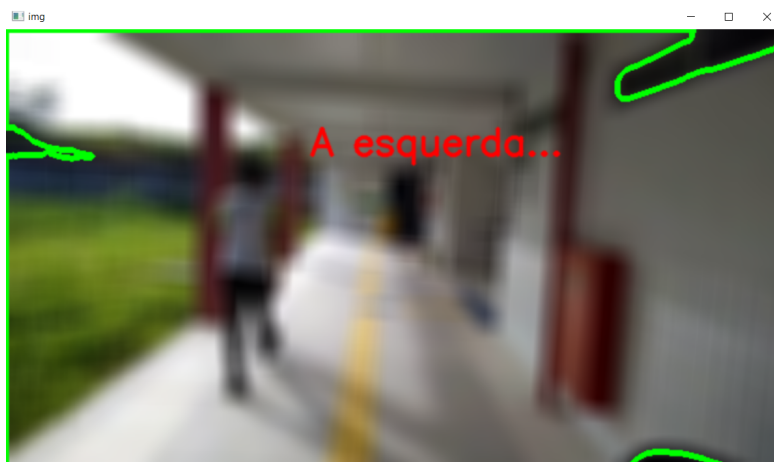


Fonte: Elaboração Própria.

Descrição: A **Figura 28**, apresenta um print screen com a saída do método “cv2.imshow”, onde foram geradas duas imagens, uma contendo uma imagem estática, com o rosto utilizado como referência para, e a outra, diz respeito a um momento específico do vídeo, onde , foi encontrando a imagem de referência utilizada, indicando a distância da câmera para o objeto em questão, que neste caso, é um rosto.

Com o objetivo de antecipar possíveis obstáculos durante o deslocamento, foi utilizado o método gaussian blur nativo do OpenCV, assim como a biblioteca de processamento de matrizes e funções matemáticas numpy, para estrutura do código, jura o resultado, com a base de dados de teste, produzida nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes, pode ser visualizado na **Figura 29** a seguir.

Figura 29 – Imagem de teste para antecipação de obstáculos.



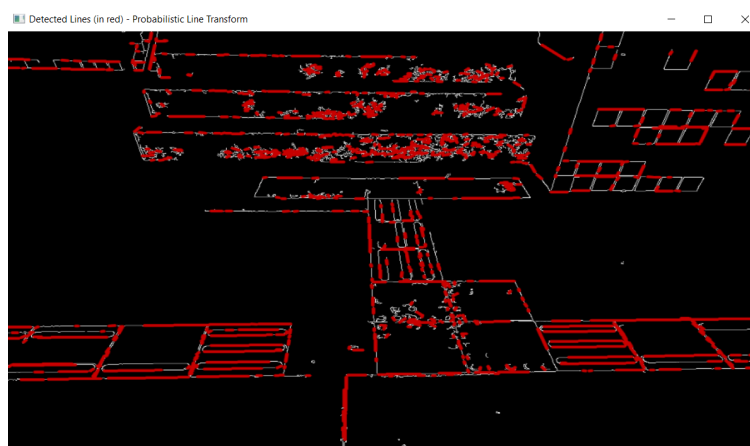
Fonte: Elaboração Própria.

Descrição: A **Figura 29**, apresenta um print screen com a saída do método “cv2.imshow”, devido a utilização do método gaussian blur, a imagem encontra se deslocada, na parte superior da imagem, a um texto escrito: “A esquerda”, e o aluno/pedestre presente na imagem encontra-se de fato nesta posição.

3.6 Detecção de Linhas e a Detecção de Cantos

Nesta seção será demonstrado o funcionamento da biblioteca de visão computacional OpenCV, e a técnica Hough transform, implementados na linguagem de programação Python. De forma simplificada, Hough transform é uma técnica matemática, especializada na detecção de formas geométricas em imagens digitais.

Figura 30 – Imagem de teste utilizada para detecção de linhas.



Fonte: Elaboração Própria.

Descrição: A **Figura 30**, apresenta um print screen com a saída do método “cv2.imread”, onde foram geradas diversas linhas em verde, onde o algoritmo considerou a incidência de linhas pré-definidas na imagem em questão.

A imagem, representada na **Figura 30**, é um resultado da aplicação de um método conhecido como “Transformação de linha probabilística”, que foi aplicado na mesma imagem apresentada na **Figura 01**, inserida na introdução deste documento.

Com o objetivo de detectar regiões nas imagens, com uma grande variação de intensidade em todas as direções, foi desenvolvido por Chris Harris e Mike Stephens em seu artigo A Combined Corner and Edge Detector em 1988, uma método para detecção de cantos, que este método é conhecido atualmente como Harris Corner Detector.

Figura 31 – Aplicação prática do Harris Corner Detector.



Fonte: Elaboração Própria.

Descrição: A **Figura 31**, apresenta um print screen com a saída do método “cv2.imshow”, onde foram gerados na imagem, da fachada da biblioteca do Instituto Federal, campus Jaboatão dos Guararapes, diversos pontos em vermelho, onde o método Harris Corner, identificou grande variações de intensidade.

3.7 Flickr 8k

Nesta seção será demonstrado o funcionamento do conjunto de dados Flickr 8k para geração de legendas em imagens. Os resultados apresentados nesta seção,

são resultado, do trabalho desenvolvido por Raman Shinde em seu artigo intitulado, “Image Captioning With Flickr 8k Dataset & BLEU”, disponível de forma integral na plataforma Medium, cujo o link para este artigo, encontra-se anexado a este documento. Os elementos abordados no já citado artigo, de forma resumida, elenca alguns passos elementares para se utilizar a base de dados de forma correta.

- Caracterização das imagens.
- Pré-processamento das legendas.

A Caracterização utiliza um modelo pré-treinado do conjunto de dados Imagenet fornecido em keras, contendo 14 milhões imagens no conjunto de dados, é mais de 21 mil grupos de classes distintas. no já mencionado trabalho, foi utilizado o InceptionV3 disponibilizado pelo Google, proporciona um arquivo menor de 96 MB é sendo mais rápido para treinar. Vale ressaltar que é necessário redimensionar todas as imagens para o tamanho (299 x299).

Durante o pré-processamento, as legendas são lidas diretamente do arquivo “Flickr8k.token.txt”, é adicionar as legendas no dicionário k:v, sendo k = identificador da imagem (id) e v = valor (lista de legenda). A decodificação para o dicionário acontece da seguinte forma: “startseq “ + legenda + “ endseq”.

- startseq : Atuará como nossa primeira palavra quando o vetor de imagem extraído do recurso for alimentado ao decodificador. Isso iniciará o processo de geração de legendas.
- enseq : Isso dirá ao decodificador quando parar. Pararemos de prever palavras assim que endseq aparecer ou tivermos previsto todas as palavras do dicionário de trem, o que vier primeiro.

Figura 32 – Pré-processamento das legendas no Flickr 8k.

```

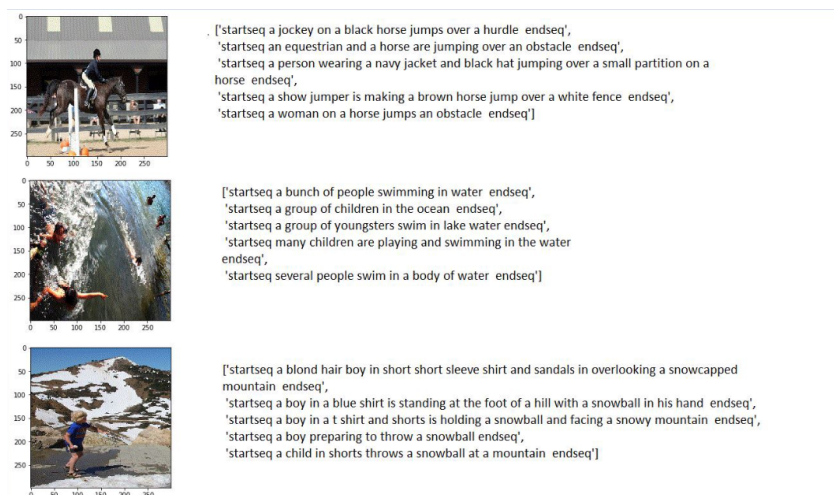
1 #abrindo arquivo texto
2 PATH = "data\\flickr8k\\input"
3
4 with open(PATH+"flickr8k.token.txt") as f:
5     data = f.read()
6
7 # chave contínua do dicionário como image_id e valor como lista de legendas.
8 descriptions = dict()
9
10 try:
11     for el in data.split("\n"):
12         tokens = el.split()
13         image_id , image_desc = tokens[0],tokens[1:]
14
15         # removendo .jpg do id da imagem
16         image_id = image_id.split(".")[0]
17
18         image_desc = " ".join(image_desc)
19
20         # verifica se image_id já está presente ou não
21         if image_id in descriptions:
22             descriptions[image_id].append(image_desc)
23         else:
24             descriptions[image_id] = list()
25             descriptions[image_id].append(image_desc)
26 except Exception as e:
27     print("Exception got :- \n",e)
28
29 descriptions["1000268201_693b08cb0e"]

```

Fonte: (Raman Shinde, 2019) (adaptado).

Descrição: A **Figura 32**, apresenta um print screen do código utilizado para o pré-processamento das legendas atribuídas a cada uma das 8 mil imagens do conjunto de dados Flickr 8k.

Figura 33 – Visualização das legendas atribuídas às imagens de referência.



Fonte: Raman Shinde (adaptado).

Descrição: A **Figura 33**, apresenta um print screen da saída de execução do notebook criado no Google Colab, contendo três imagens do conjunto de referência e cinco legendas associadas a cada uma destas imagens.

3.8 Tesseract (OCR)

Nesta seção será demonstrado o funcionamento da Tesseract para o reconhecimento óptico de caracteres, implementados na linguagem de programação Python. No Python especificamente, a utilização do Tesseract se dá por meio da utilização da biblioteca PyTesseract. no exemplo demonstrado na **Figura 34**, a seguir o método utilizado foi o “pytesseract.image_to_data”, para identificação de regiões de interesse, para a geração das caixas delimitadoras foi utilizado o método “cv2.rectangle”, em conjunto com um algoritmo de ordenação.

Figura 34 – Múltiplas caixas delimitadoras geradas na imagem de referência.



Fonte: Elaboração Própria.

Descrição: A **Figura 34**, apresenta um print screen com a saída do método “cv2.imshow”, na imagem em questão, foram geradas múltiplas caixas delimitadoras, havendo a ocorrência de verdadeiros positivos e falsos positivos.

Este resultado demonstrado na **Figura 34**, foi obtido, por meio da execução do arquivo “tesseract_02.py”, onde também foi utilizado, o método “pytesseract.image_to_string”, resultado na detecção de partes dispersas das palavras. Palavras como: “ARIANO SUASSUN”, “INSTITUTO FEDERAL”, “Campus Jaboatão dos”, entre outras letras avulsas.

4. ANÁLISE DOS RESULTADOS OBTIDOS

Neste capítulo, serão apresentados os resultados da avaliação realizada através da proposta de construção de uma ferramenta de acessibilidade utilizando as bibliotecas de processamento de linguagem natural, Speech_recognition, assim como as bibliotecas de visão computacional OpenCV, YOLO para detecção e rastreamento de pedestres, entre outros tópicos importantes, a fim de construir uma ferramenta de acessibilidade, para pessoas com deficiência visual, observando os fundamentos da orientação e mobilidade, da audiodescrição simultânea por meio de um sintetizador de voz.

Também serão abordados o funcionamento da base de dados Flickr 8k para geração automática de legendas em imagens, devendo-se ressaltar que os resultados apresentando, exclusivamente com relação a esta ferramenta, são resultados referentes à execução do trabalho desenvolvido por Raman Shinde em seu artigo intitulado, "Image Captioning With Flickr 8k Dataset & BLEU". (Raman Shinde, 2019)

4.1 Speech_recognition e o Google Text-to-Speech (gTTS)

Nesta seção será apresentado o resultados dos testes realizados com a utilização da biblioteca de processamento de linguagem natural, gtts, para conversão de textos digitais em áudio. Durante os testes com a já mencionada biblioteca, foi constatado, um ótimo resultado para a sintetização da voz humana, optando-se pela geração dos dados, na língua portuguesa 'pt-br'.

Foram elaboradas, uma série de frases, aplicadas ao contexto que este trabalho de conclusão de curso está delimitado, sendo esta delimitação, as dependências internas do Instituto Federal, campus Jaboatão dos Guararapes. Exemplos de frases como:

- "Uma pessoa está a 6 metros de distância."
- "2 alunos estão logo à frente."
- "Continue sobre o piso tátil."
- "Obstáculo à direita."
- "Obstáculo à esquerda."

Estas frases, entram no contexto do que se pretende como resultado final para com o uso destas duas bibliotecas, em um contexto no qual, todos os conhecimentos abordados até então, trabalhando de forma 100% integrada, sendo elas, a biblioteca `Speech_recognition` e o Google Text-to-Speech (gTTs).

Especificamente, nos testes com a biblioteca de `Speech_recognition`, foi utilizado o notebook mencionado nos materiais e métodos deste trabalho de conclusão de curso, e detalhado no capítulo 03 seção 3.1. De forma resumida, foi possível observar que o uso da ferramenta resultou em excelentes resultados, algo próximo a 100% (cem por cento) de assertividade para as detecções das palavras/frases de teste faladas.

Com relação a biblioteca Google Text-to-Speech (gTTs), para sintetização de vocês, as métricas, assim como os resultados manteve-se praticamente os mesmos, com a voz gerada pelo sintetizador, sendo, quase em sua totalidade, 100% (cem por cento) audível, com clareza e precisão.

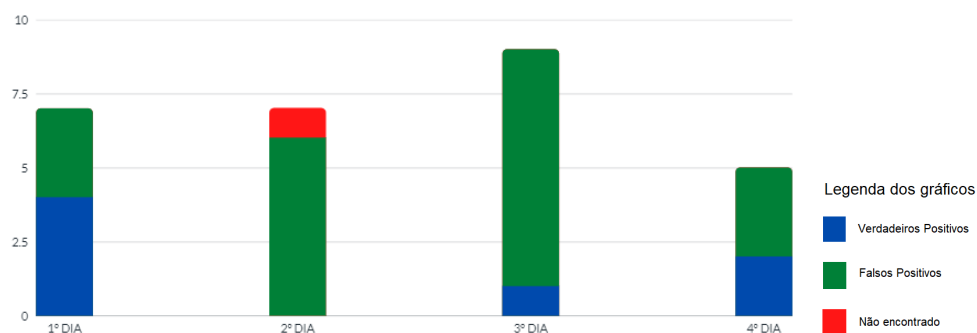
4.2 OpenCV

Na subseção a seguir será apresentado o resultados dos testes realizados com a biblioteca de visão computacional OpenCV por meio da utilização dos métodos HOG (Histogramas de gradientes orientados), assim como, os resultados obtidos com a utilização dos classificadores Haar Cascades, entre outros métodos, comuns ao bom funcionamento destas duas abordagens.

4.2.1 Histogram of Oriented Gradients (HOG)

Os resultados apresentados a seguir são referentes à execução do código “`video_depestrian_detection_01.py`”, presente no projeto “GuiaInteligente”. A **Figura 35**, apresenta um gráfico com os resultados para as detecções realizadas com a base de dados simulada, nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes.

Figura 35 – Representação gráfica dos resultados com o método HOG.



Fonte: Elaboração própria.

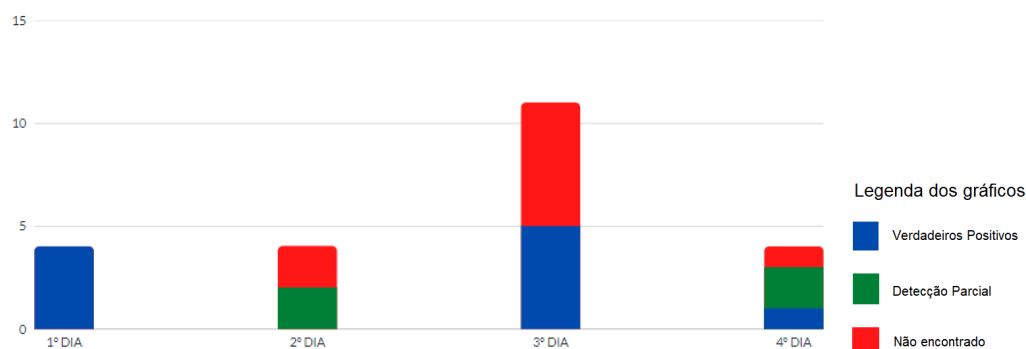
Descrição: A **Figura 35**, contém um gráfico, representando o somatório dos casos de incidência de verdadeiros positivos e falsos positivos para as detecções, assim como os casos em que não foram encontrados pedestres, em um período de quatro dias.

O objetivo geral com os testes realizados, foi avaliar o quão significativo, os elementos do ambiente, podem interferir na assertividade das detecções e rastreamentos dos pedestres. Questões como a luminosidade (luz solar ou artificial), sombra nos corredores e escadarias de acesso, assim como a qualidade do equipamento/câmera utilizada para as gravações, interferem no resultado final.

4.2.2 Haar Cascades

Os resultados apresentados a seguir são referentes à execução do código “video_depedestrian_detection_02.py” presente no projeto “GuiaInteligente”. A **Figura 36**, apresenta um gráfico com os resultados destas detecções, realizadas nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes.

Figura 36 – Representação gráfica dos resultados com o Haar Cascades.



Fonte: Elaboração própria.

Descrição: A **Figura 36**, contém um gráfico, representando o somatório dos casos de incidência de verdadeiros positivos e falsos positivos para as detecções, assim como os casos em que não foram encontrados pedestres, em um período de quatro dias.

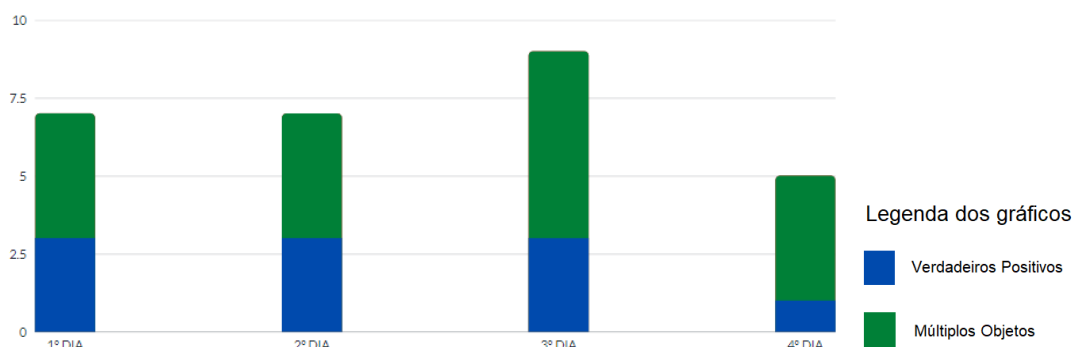
O objetivo geral para com estes testes, foi avaliar a incidência de casos, positivos e negativos para com as detecções, assim como a ocorrência de casos em que não foi possível detectar nenhum pedestre/aluno.

Comparando os dois gráficos, apresentados nas **Figura 35 e 36**, percebe-se que no gráfico apresentado na **Figura 36**, ocorrem mais casos de não detecções, o que num primeiro momento, pode indicar que o classificador Haar Cascade, é inferior ao método HOG, o que não necessariamente corresponde à realidade, tendo em vista que, como já foi citado anteriormente, elementos do ambiente podem ter representar grande influência sobre o resultado final.

4.3 You Only Look Once (YOLO)

Os resultados apresentados a seguir são referentes à execução do código adaptado na plataforma Google Colab com o objetivo de implementar a utilização do modelo de rede neural do YOLO para detecção e rastreamento de pedestres/alunos. A **Figura 37**, apresenta um gráfico com os resultados para estas detecções, realizados nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes.

Figura 37 – Representação gráfica dos resultados com o YOLO.



Fonte: Elaboração própria.

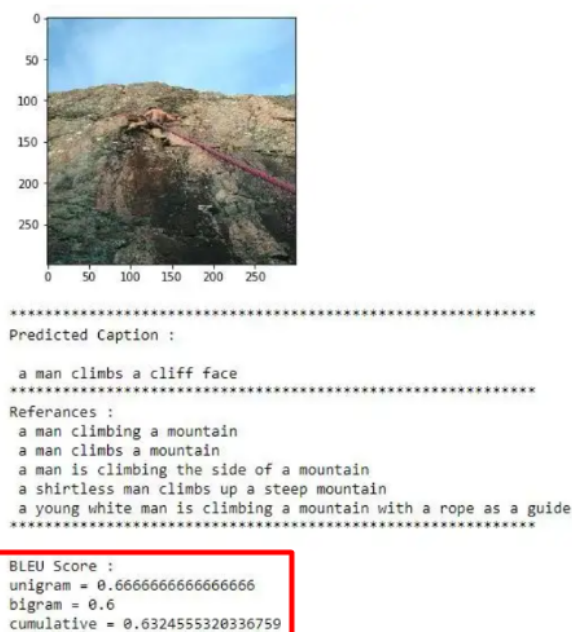
Descrição: A **Figura 37**, apresenta um gráfico, representando o somatório dos casos de verdadeiros positivos (em azul) para com as detecções dos pedestres especificamente, em quando em verde, destaca os casos em foram detectados múltiplos objetos (não pedestres) nas imagens.

O objetivo geral com os testes realizados utilizando o YOLO, foi avaliar de forma introdutória, as múltiplas detecções nas imagens, produzidas, nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes.

4.4 Flickr 8k

Os resultados apresentados a seguir são referentes à execução do trabalho desenvolvido por Raman Shinde em seu artigo intitulado, “Image Captioning With Flickr 8k Dataset & BLEU”, este mesmo projeto, foi abordado no capítulo anterior, em razão disto, nesta seção, será analisado os resultados dos testes realizados neste documento. (Raman Shinde, 2019)

Figura 38 – Exemplo de output para a geração de legendas.



Fonte: (Raman Shinde, 2019) (Adaptado).

Descrição: A **Figura 38** apresenta um print screen, um exemplo de saída para a geração das legendas em um imagem.

De um modo geral, o desempenho do modelo utilizado para obtenção dos resultados, poderiam ser melhorados, a partir do treinamento de um conjunto de dados maior, assim como, o ajuste de hiperparâmetros mais precisos. Acima, pode-se observar, destacado em vermelho, que as pontuações de unigram BLEU favorecem as previsões curtas.

4.5 Tesseract (OCR)

Os resultados apresentados a seguir são referentes à execução do código, “tesseract_02.py”, onde foi avaliado, o desempenho da biblioteca de conversão de imagens digitais para textos digitais, conhecido como, reconhecimento óptico de caracteres ou (OCR), aplicado às dependências internas do Instituto Federal, campus Jaboatão dos Guararapes.

Afim de visualizar de forma clara, a integração dos diversos conhecimentos desenvolvidos neste trabalho de conclusão de curso, os testes para o reconhecimento óptico de caracteres, foram realizados com a biblioteca Tesseract

em consonância com a biblioteca gTTS, para o reconhecimento de textos e sintetização destes textos em áudio, de forma simultânea, em tempo real.

Sendo assim, a sintetização de voz, assim como sua compreensão, está condicionada a qualidade da leitura dos caracteres/palavras detectadas nas imagens. Como resultado, a porcentagem para estas detecções é incerta pois, na base de dados simulada de teste, não foram registrados no Instituto Federal, campus Jaboatão dos Guararapes, placas e banners com textos, apenas a fachada da biblioteca do referido campus.

5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Neste último capítulo, serão apresentadas as considerações finais sobre o uso das bibliotecas de processamento de linguagem natural, Speech_recognition e o Google Text-to-Speech (gTTs), assim como a utilização da biblioteca de visão computacional OpenCV, especificamente, por meio da utilização do método HOG (Histogramas de gradientes orientados) e dos classificadores Haar Cascades, aplicados a base de dados de teste, produzida em caráter de simulação, nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes. Também será discutido, o funcionamento da ferramenta de visão computacional YOLO.

O funcionamento do Flickr 8k, utilizado para a geração de legendas em imagens, com base nos os resultados apresentados a partir da execução do trabalho desenvolvido por Raman Shinde em seu artigo intitulado, “Image Captioning With Flickr 8k Dataset & BLEU”. E por fim, o funcionamento da biblioteca Tesseract para o reconhecimento óptico de caracteres (OCR). (Raman Shinde, 2019)

5.1 Considerações Finais

Neste trabalho, foram abordados os principais conceitos referente à deficiência visual, assim como os desafios que as deficiências visuais impõem para a vida das pessoas. Também foram abordados os conceitos norteadores, referentes a Orientação e mobilidade (OM). Em seguida, foi discutido, as tecnologias assistivas, a fim de aumentar as capacidades funcionais destas pessoas, em seguida, foi abordado, referentes a audiodescrição de cenários Mais adiante foram abordados dois temas comuns ao processamento de linguagem natural, e por fim, discutido de forma mais aprofundada, elementos norteadores do processamento de linguagem natural, vale ressaltar que durante a o desenvolvimento deste trabalho de conclusão de curso, pretendeu se seguir este ordem, descrita acima.

A partir desses conceitos, foi proposta uma metodologia de desenvolvimento que tinha como premissa, a integração completa destes recursos, a fim de promover a acessibilidade para pessoas com deficiência visual, por meio do desenvolvimento de uma ferramenta específica para este fim. Sendo importante ressaltar que, em razão de limitações financeiras, todo o trabalho desenvolvido neste trabalho de

conclusão de curso, não se materializou em um dispositivo de hardware, apenas e exclusivamente software, estando os códigos desenvolvidos até então, disponíveis de forma integral e publica, na plataforma github em anexo a este documento, possibilitando desta forma, com que, se possa mensurar, a viabilidade da construção de uma ferramenta de acessibilidade para pessoas com deficiência visual.

5.2 Limitações

A proposta de desenvolvimento de uma ferramenta de acessibilidade para pessoas com deficiência visual, discutida neste trabalho de conclusão de curso, limita-se, em alguns aspectos, a questões financeiras e orçamentárias, demasiadamente custosas, para com o seu autor.

Outro aspecto, que deve ser observado, diz respeito às métricas utilizada para a avaliação do grau de assertividade para com os algoritmos, métodos e ferramentas utilizados durante, os testes, com a base de dados simulada, nas dependências internas do Instituto Federal, campus Jaboatão dos Guararapes, neste sentido é importante pontuar que as gravações ocorreram exclusivamente no período da manhã, tendo em vista a compatibilidade de horários entre o autor deste documento e os alunos devidamente citados nos agradecimentos deste documento.

5.3 Trabalhos Futuros

Alguns pontos importantes podem vir a serem desenvolvimentos em propostas futuras, a partir dos temas, abordados neste documento, possibilitando assim, o aprimoramento para com as detecções e rastreamentos de pedestres, por meio das bibliotecas de visão computacional, OpenCV, YOLO, assim como a geração de legendas em imagens por meio da base de dados Flickr 8k. Especificamente com relação ao OpenCV, os classificadores Haar Cascades, pode ser explorado um número maior de classificadores.

Estudar maneiras de contar objetos/pessoas em uma imagem, assim como o reconhecimento parcial de gestos humanos, e o reconhecimento de feições humanas. Estudar os avanços que a robótica aliada à visão computacional, poderiam proporcionar aos temas abordados neste documento, como por exemplo,

a utilização de geolocalização e a delimitação de um trajeto, integrado com ferramentas como Google Maps.

Outro ponto que pode ser considerado em trabalhos futuros está na utilização de elementos abordados em outras propostas de ferramentas de acessibilidade para pessoas com deficiência visual, propostas que apresentem resultados em cenários e situações diversas e diferentes das que foram abordadas neste trabalho de conclusão de curso.

Em razão do grande conjunto de dados que a integração completa de todas as tecnologias citadas neste documento, pode acabar resultando em um cenário onde, uma enorme quantidade de informações como, obstáculos a frente, distância entre o usuário do dispositivo e um pedestre/aluno, placas informativas, pisos táteis, entre outros elementos, detectados em simultâneo, deve-se observar, a criação de elementos com base em aprendizagem de máquina e aprendizagem profunda, a fim de ABSTRAIR esta enorme quantidade de informação, afim de criar uma “linha lógica, e situacional”, a fim de orientar a pessoa com deficiência de forma segura e determinada.

Outro ponto importante, diz respeito a base de dados simulada, utilizada nos testes, devendo-se gerar novas imagens/vídeos das salas e laboratórios do campus, assim como a detecção de, cartazes e placas e banners, expostos no Instituto Federal.

REFERÊNCIAS

- ALVES, MARIA EDUARDA SOFFIATI, LIMA, FERNANDO PAIM. **Detecção e reconhecimento facial a partir de uma base de dados**, IFMG - Campus Formiga - 2019.
- Alexssandro Ferreira Cordeiro, Cristian A. U. Ojeda, Pedro L. P. Filho, Gustavo R. Valiati, **Rastreamento e contagem de pedestre em tempo real por meio de imagens digitais**, LATINOWARE 2019, XVI Congresso Latino-americano de software livre e tecnologias abertas, 2019.
- ALMEIDA, M.G.S. **Instituto Benjamin Constant: 160 anos de inclusão**. Revista Benjamin Constant, 2014.
- BERNARDES, Júnior et al. **Sistema ajuda portador de deficiência visual a se locomover**. 2016.
- BRASIL. **Constituição** (1988). **Constituição** da República Federativa do Brasil. Brasília, DF: Senado **Federal**: Centro Gráfico, 1988.
- Bradski, Gary; Kaehler, Adrian. **Learning OpenCV: Computer Vision with the OpenCV Library**, 2008.
- BITTENCOURT, Z. Z.; HOEHNE, E. L. **Qualidade de vida de deficientes visuais**. Medicina (Ribeirão Preto. Online), [S.l.], v39, n.2, p.260, 2006.
- Boland, R. Design in the punctuation of management action. Managing as designing: **Creating a vocabulary for management education and research** (2002), 106-112.
- Dr. Vipulsangram.K.Kadam, Deepali G. Ganakwar, Face Detection: A Literature Review, Maharashtra, India, IJIRSET, July 2017.
- D. Jay Gense, En.S., Marilyn Gense, M.A., **The Importance of Orientation and Mobility Skills for Students who are Deaf-Blind**, Helen Keller National Center Perkins School for the Blind Teaching Research, 2004.
- ELIAS AUGUSTO FANK, DENIO DUARTE. **Utilização de visão computacional e detecção de caracteres para auxiliar na navegação de pessoas com deficiência visual em ambientes internos**, UFFS - Campus Chapecó, 2016.

Flavio Augusto Schiave Germano et al, **Estudo das causas de cegueira e baixa de visão em uma escola para deficientes visuais na cidade de bauru**, 2019

FELIPPE, J.A. de M.; FELIPPE, V. L. R.. **Orientação e mobilidade**. São Paulo: Laramara - Associação Brasileira de Assistência ao Deficiencia Visual, 1997

FELIPPE, J.A. de M. Caminhando juntos: manual das habilidades básicas de orientação e mobilidade. São Paulo: **Laramara**, 2001.

FRANCO, Eliana Paes Cardoso; SILVA, Manoela Cris na Correia Carvalho. Audiodescrição: Breve passeio histórico. In: MOTTA, Livia Maria Villela de Melo; ROMEU FILHO, Paulo. (Orgs.). **Audiodescrição: transformando imagens em palavras**. São Paulo: Secretaria dos Direitos da Pessoa com Deficiência do Estado de São Paulo, 2010.

GOMES, D. dos S. **Inteligência Artificial: conceitos e aplicações**. Olhar Científico. v1, n. 2, p. 234-246. 2010.

GABRIELA P, MARIANTO, JOYCE A. P. SOARES, EDUARDO M. A. AMARAL, **Sistema de detecção e resgate de vítima para um robô autônomo seguidor de linha baseado em visão computacional**. Mostra Nacional de Robótica (MNR), Instituto Federal, campus Serra.

Hyun-Tae Kim, Sang-Hyun Lee, **A study on object distance measurement using OpenCV-based Yolo 5**, Computer Engineering, Homan University, Korea, 2021.

Ing-Jr Ding, Chih-Ta Yen, and Yen-Ming Hsu, **Developments of Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition**, Hindawi Research Article, National Formosa University - Huwei, Taiwan, 2013.

João Álvaro de Moraes Felipe. **Caminhando juntos: manual das orientações básicas de orientação e mobilidade**. governo do estado de são paulo. secretaria de estado dos direitos da pessoa com deficiência. são paulo, 2009. Disponível em: <http://visaosubnormal.org.br/downloads/serie_deficiencia_visual_vol4_cbo_bq.pdf>

Souza C.A. (2003) “**Desenvolvimento de um Sistema de Visão Computacional para Aplicações de Controle Servo-Visual de Robôs Moveis**”. Monografia. Universidade SBC, Salvador.

MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento digital de imagens**. Rio de Janeiro: Brasport, 1999.

SILVA, J. B. **Explorando o algoritmo de Viola-Jones na detecção e reconhecimento facial**. 2018. 91 f. Trabalho de Conclusão de Curso (Engenharia da Computação) - Universidade Federal de São Carlos, São Carlos, 2018.

Scientia Prima, v. 6, n. 1, p. 98-116, **Sistema de auxílio na locomoção de deficientes visuais**. maio de 2020.

MOTA, Maria Glória B. **Orientação e mobilidade: conhecimentos básicos para a inclusão da pessoa com deficiência visual**. Brasília: MEC, 2003.

WILLIAM MARRION CAVENAGLI, CAROLINE MAZETTO MENDES, **Deteccção de vagas em estacionamento usando visão computacional e redes neurais convolucionais**. Universidade Positivo, 2020.

Lucas K. Uezima, Henrique Y. Y. Nishimoto, Rafael Barbosa, Prof. Dr. Antonio Luiz Basile, **Utilizando o Tesseract OCR para reconhecimento de textos a mão**, Universidade Presbiteriana Mackenzie (UPM) - São Paulo, SP, Faculdade de computação e informática, 2021

Robótica inteligente: **Uso de visão computacional voltada para o controle e navegação de robôs móveis autônomos**. Instituto de Ciências Matemáticas e de Computação (ICMC), USP/São Carlos - MNR Mostra Nacional de Robótica, 2012.

Milena L. dos Santos, Claudio R. M. Mauricio, Valéria N. dos Santos, Fabiana F. F. Peres, **Rede neural convolucional para detecção de pedestres realizando travessias de risco**, Universidade Estadual do Oeste do Paraná - Centro de Engenharias e Ciências Exatas, Foz do Iguaçu.

Paul Viola, Michael Jones, **Rapid Object Detection using a Boosted Cascade os Simple Features**, Accepted Conference on Computer Vision and Pattern Recognition, 2001.

Li Yi-bo, LI Jun-jun a, **Harris Corner Detection Algorithm Based on Improved Contourlet Transform**, Shenyang University of aeronautics and Astronautics- China, 2011.

Renato Martins Redovalio Ferreira, Ruth Maria Mariani Braz, **A Orientação e Mobilidade como Prática para Inclusão de Pessoas com Deficiência Visual**, , Conedu VII Congresso Nacional de Educação, Outubro 2020.

Loiane Maria Zengo, et al, **Relatos de Alunos Cegos Sobre o uso do Guia Vidente como Estratégia de Locomoção nos Ambientes Escolares**, Faculdade de Filosofia e Ciências - UNESP, Campus de Marília, Faculdade de Ensino Superior do Interior Paulista – FIAP, 2017.

Joseph Redmon, Ali Farhadi, et al, **You Only Look Once: Unified, Real-Time Object Detection**, University of Washington, 12 Nov 2015

Mai Thanh Nhat Truong, Sanghoon Kim, **A Review on Image Feature Detection and Description**, Hankyong National University, 2016

Kai Zhao et al, **Deep Hough Transform for Semantic Line Detection**, IEEE Transactions on Pattern Analysis and Machines Intelligence, May 2021.

Navneet Dalal, Bill Triggs, HAK archives-ouvertes, **Histograms of Oriented Gradients for Human Detection**, Dec 2010

Itunu A. Bamidele Ph.D, **Information Needs of Blind and Visually Impaired People**, Babcock University, Ilisan – Remo, gun State, Nigeria

MOTTA, Livia Maria Vilella de Mello. **A AUDIODESCRIÇÃO NA ESCOLA: ABRINDO CAMINHOS PARA LEITURA DE MUNDO**. 2016. Disponível em: <<http://www.vercompalavras.com.br/pdf/a-audiodescricao-na-escola.pdf>>

MOTTA, Livia Maria Vilella de Mello; ROMEU FILHO, Paulo (org). **Audiodescrição transformando imagens em Palavras**. São Paulo: Secretaria dos Direitos das Pessoas com Deficiência do Estado de São Paulo. 2010.

Nilton Pinto Ribeiro Filho, **VISÃO COMPUTACIONAL: UM NOVO CAMPO DE PESQUISA EM COGNIÇÃO VISUAL**, 2022. Disponível em: <<https://periodicos.unb.br/index.php/revistapt/article/download/17018/15504/28764>>

Raman Shinde, **Image Captioning With Flickr 8k Dataset & BLEU**, 2019. Disponível em:

<<https://medium.com/@raman.shinde15/image-captioning-with-flickr8k-dataset-bleu-4bcba0b52926>>

Instituto dos Cegos, deficiência visual, 2023. Disponível em:
<<https://www.institutodoscegos.org.br/deficiencia-visual>>

OMS, Primeiro relatório global sobre cegueira, OMS diz que mundo poderia evitar metade dos casos, 2019, Disponível em:
<<https://news.un.org/pt/story/2019/10/1690122>>

CIF, Classificação Internacional da Funcionalidade, Incapacidade e Saúde, 2001. Disponível em: <https://www.periciamedicadf.com.br/cif2/cif_portugues.pdf>

Blog da Audiodescrição, Audiodescrição, 2023. Disponível em:
<<https://www.blogdaaudiodescricao.com.br/audiodescricao>>

eveo, Saiba tudo sobre a inteligência artificial, 2023, Disponível em:
<<https://blog.eveo.com.br/tudo-sobre-inteligencia-artificial>>

Calebe Augusto dos Santos, Gino Dornelles, Text-To-Speech: Sintetizador de Voz para Documentos como Ferramenta de Apoio a Deficientes Visuais, 2008, Disponível em:
<https://repositorio.ufsc.br/bitstream/handle/123456789/184450/TCC_Calebe_Gino.pdf?sequence=-1&isAllowed=y>

IBGE, SNIG - Pesquisa Nacional de Informação de Gênero, 2012.
<<https://cidades.ibge.gov.br/brasil/pe/recife/pesquisa/11/0>>

WHO, World Report on Disability 2011, 2021. Disponível em:
<<https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/world-report-on-disability>>

Saúde Plena, Conselho Brasileiro de Oftalmologia alerta para casos de cegueira evitáveis, 2019, Disponível em:
<<https://www.uai.com.br/app/noticia/saude/2019/07/06/noticias-saude.248471/conselho-brasileiro-de-oftalmologia-alerta-para-casos-de-cegueira-evit.shtml>>

Assistiva Tecnologia e Educação, O que é Tecnologia Assistiva, 2023, Disponível em:
<<https://www.assistiva.com.br/tassistiva.html#:~:text=Tecnologia%20Assistiva%20%C3%A9%20o%20termo,promover%20Vida%20Independente%20e%20Inclus%C3%A3o>>

ADA, Americans with Disabilities Act, 2023. Disponível em: <<https://www.ada.gov/>>

Colégio gtm, Linguagem em Braille: saiba mais sobre a criação desse sistema de leitura, 2023. Disponível em:

<<https://www.colegiotgm.com.br/2021/04/08/linguagem-braille-criacao/>>

Enap, Escola Nacional de Administração Pública, 2023. Disponível em:

<<https://enap.gov.br/pt/>>

Jéssica Rodrigues, O que é o Processamento de Linguagem Natural?, 2017,

Disponível em:

<<https://medium.com/botsbrasil/o-que-%C3%A9-o-processamento-de-linguagem-natural-49ece9371cff>>

Jones Granatyr, Introdução a Processamento de Linguagem Natural, 2020.

Disponível em:

<https://iaexpert.academy/2020/12/15/introducao-a-processamento-de-linguagem-natural/?utm_source=rss&utm_medium=rss&utm_campaign=introducao-a-processamento-de-linguagem-natural&doing_wp_cron=1677801456.9338579177856445312500>

Leandro Kovacs, O que é processamento de linguagem natural? [NLP], 2021.

Disponível em:

<<https://tecnoblog.net/responde/o-que-e-processamento-de-linguagem-natural-nlp/>>

PrograMaria, Carreira em Visão Computacional: como se preparar para trabalhar na área, 2020. Disponível em:

<<https://www.programaria.org/carreira-em-visao-computacional-como-se-preparar-para-trabalhar-na-area/>>

OpenCV, documentation, 2023, Disponível em: <<https://docs.opencv.org/4.x/>>

Adrian Rosebrock, OpenCV Thresholding (cv2.threshold), 2021. Disponível em:

<<https://pyimagesearch.com/2021/04/28/opencv-thresholding-cv2-threshold/>>

Arnaldo Gualberto, Detectando objetos com métodos clássicos — OpenCV (cascades), 2018. Disponível em:

<<https://medium.com/ensina-ai/detectando-objetos-com-m%C3%A9todos-cl%C3%A1ssicos-opencv-cascades-440e29913b1b>>

Adrian Rosebrock, pyimagesearch, OpenCV Face detection with Haar cascades, 2021. Disponível em:

<<https://pyimagesearch.com/2021/04/05/opencv-face-detection-with-haar-cascades/>>

Jobu, YOLOv5 Como treinar um modelo de detecção de objeto personalizado, 2023. Disponível em: <<https://jobu.com.br/2021/06/28/como-treinar-um-yolov5/>>

HSANKESARA, Flickr Image dataset, 2018. Disponível em:

<<https://www.kaggle.com/datasets/hsankesara/flickr-image-dataset>>

edureka, How to Implement Optical Character Recognition in Python, 2021. Disponível em:

<<https://www.edureka.co/blog/optical-character-recognition-in-python/>>

mordorintelligence, TENDÊNCIAS, IMPACTO DO COVID-19 E PREVISÕES (2023-2028), 2023, Disponível em:

<<https://www.mordorintelligence.com/pt/industry-reports/computer-vision-market>>

Alan Hevner et al, Design Science in Information Systems Research, 2004. Disponível em:

<https://www.researchgate.net/publication/201168946_Design_Science_in_Information_Systems_Research>

geeksforgeeks, Convert Text to Speech in Python, 2023. Disponível em:

<<https://www.geeksforgeeks.org/convert-text-speech-python/>>

geeksforgeeks, Pedestrian Detection using OpenCV-Python, 2023. Disponível em:

<<https://www.geeksforgeeks.org/pedestrian-detection-using-opencv-python/>>

safe park sinalizacao, 2023. Disponível em: <<https://safeparksinalizacao.com/>>

freepik, 2023. Disponível em: <<https://br.freepik.com/fotos-vetores-gratis/braille>>

APÊNDICE

Projeto “GuiaInteligente”, disponível no repositório de acesso público no seguinte link: <<https://github.com/williams181/GuiaInteligente>>