



Fundação Universidade Federal do ABC

Pró reitoria de pesquisa

Av. dos Estados, 5001, Santa Terezinha, Santo André/SP, CEP 09210-580

Bloco L, 3º Andar, Fone (11) 3356-7617

iniciacao@ufabc.edu.br

Projeto de Iniciação Científica
submetido para avaliação no
Edital: **04/2022**

Título do projeto: Sistema de Reconhecimento de Semáforos para Pedestres.

Palavras-chave do projeto: Interface Humano-Computador. Visão Computacional.
Reconhecimento de Semáforos. Aprendizagem de Máquina. Tecnologia Assistiva

Área do conhecimento do projeto: Engenharia de Informação

Sumário

1	Resumo	3
2	Introdução	4
2.1	Motivação e Justificativa	4
3	Objetivos	4
4	Metodologia	5
4.1	Aquisição de imagens	6
4.2	Pré-processamento das imagens	7
4.3	Método de identificação de cores e objetos	7
4.4	Interface para pessoas com deficiência visual	7
4.5	Escolha do modelo de treinamento	8
4.6	Classificação	9
4.7	Alternativas de análise de cores para detecção no semáforo para pedestres . .	10
5	Viabilidade	10
6	Cronograma de atividades	10
	Referências	12

1 Resumo

Este trabalho propõe a aplicação de um sistema embarcado no reconhecimento em tempo real do estado dos semáforos para pedestres. Por meio da utilização de uma rede neural, identificaremos o estado dos semáforos em imagens capturadas por uma câmera. Uma vez realizada a análise por meio do algoritmo desenvolvido, sinais sonoros e vibratórios serão disparados de forma a informar o estado atual do semáforo. Nessa fase será realizada uma pesquisa nas áreas de rede neurais, o funcionamento, o treinamento e a implementação delas, junto com uma pesquisa sobre base de dados de imagens dinâmicas, que será usado para treinar a rede neural. Será definido também um sistema embarcado usando Raspberry Pi e seus módulos de câmeras, e também será elaborado um teste de desempenho e viabilidade para garantir o funcionamento perfeito do sistema embarcado com a rede neural. Com este trabalho esperamos desenvolver um dispositivo de baixo custo capaz de auxiliar pessoas com deficiência visual durante a travessia de rua informando o estado atual do semáforo, e portanto ampliando sua autonomia e segurança.

Palavras-chaves: Interface Humano-Computador. Visão Computacional. Reconhecimento de Semáforos. Aprendizagem de Máquina. Tecnologia Assistiva

2 Introdução

Visão computacional é um ramo multidisciplinar da inteligência artificial voltado para treinar computadores na interpretação de alto nível do mundo visual por meio de imagens em câmeras, vídeos ou outros dados visuais. O objetivo é a criação de sistemas que possibilitem ao computador processar, analisar, compreender e extrair conhecimentos relevantes a fim de tomar ações ou fazer recomendações com base no sentido apreendido [1].

As aplicações dessa tecnologia são vastas e podem abranger, entre outros, a detecção de objetos e a classificação de imagens. Para fins desta proposta, escolheu-se como alvo da pesquisa na área de visão computacional a classificação de imagens visando a identificação em tempo real do estado dos semáforos de pedestres, permitindo auxiliar deficientes visuais a realizarem uma travessia com mais autonomia e segurança.

2.1 Motivação e Justificativa

Um dos fatores consequentes da ausência da visão para as pessoas com deficiências visuais é a falta de independência para locomoção, sendo necessário um acompanhante para evitar possíveis acidentes. A independência e autonomia para se locomover têm diversos benefícios, como o aumento da autoestima e autoconfiança. Atualmente existem opções que permitem a movimentação, como um acompanhante, um cão-guia ou uma bengala. Entretanto, muitas vezes as pessoas não querem andar sempre acompanhadas pois afeta o seu senso de independência. A bengala é um auxiliar apenas físico e não possui a sensibilidade que é necessária na rua além do que é tátil. Desse modo, o cão-guia seria a melhor opção. Porém, segundo o Instituto IRIS [2], um dos pioneiros na doação de cães-guias no Brasil, o custo de um cão-guia é de aproximadamente 35 mil reais, com isso, a instituição, que existe desde 2002, afirma ter doado mais de 40 cães-guias e possui uma longa espera para a realização desse sonho de muitas pessoas. Esse preço está além das condições financeiras da maioria da população brasileira e mostra que ter um cão-guia é uma solução pouco acessível. Atualmente, mesmo após a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência) entrar em vigor em 2016, semáforos de pedestre com alertas sonoros são escassos em ruas brasileiras. Segundo a nota da CET de 2016, São Paulo possuía 6.204 semáforos, mas apenas 14 desses equipamentos eram sonoros. Eles estavam instalados em cinco cruzamentos - quatro na Rua Conselheiro Brotero, seis na Rua Líbero Badaró com o Viaduto do Chá e quatro nos acessos ao Aeroporto de Congonhas.), desse modo impossibilitando a travessia independente daqueles com deficiências visuais e podendo levar a acidentes de trânsito. Nos últimos anos a tecnologia tem sido uma importante aliada dessas, com um grande papel na reconquista da autonomia [3]. Um exemplo é o desenvolvimento de aplicativos de tecnologia assistiva para pessoas com deficiência [4]. Este projeto visa o desenvolvimento de um dispositivo de baixo custo que reconhece o semáforo de pedestre e sua cor, e avisa a pessoa com deficiência visual se ela pode ou não atravessar a rua por meio de avisos sonoros e vibrações. Evitando assim possíveis acidentes e não havendo a necessidade de um acompanhante.

3 Objetivos

O objetivo geral do projeto é investigar um sistema embarcado capaz de reconhecer o estado dos semáforos para pedestres voltado para aplicações como mobilidade urbana e autonomia para pessoas com deficiência visual, que demandam informações em tempo real.

O projeto contará com duas metas:

- Desenvolver e avaliar um sistema de reconhecimento dos estados do semáforo para pedestres para execução em tempo real.
- Desenvolver um hardware contendo uma câmera, computador embarcado (raspberry pi) e periféricos sonoros e vibratórios para informar o processamento do sistema em tempo real.

4 Metodologia

A Figura 1 apresenta o diagrama de blocos com a visão geral do projeto. A partir de uma câmera no dispositivo será feito o reconhecimento do semáforo de pedestre no seu campo de visão e também o estado do semáforo (se está aberto ou fechado para atravessar). Após esse reconhecimento o dispositivo, que terá um tipo de vibração para cada situação, avisará ao pedestre de acordo com a vibração se ele pode atravessar, se não pode ou se nenhum semáforo foi encontrado.

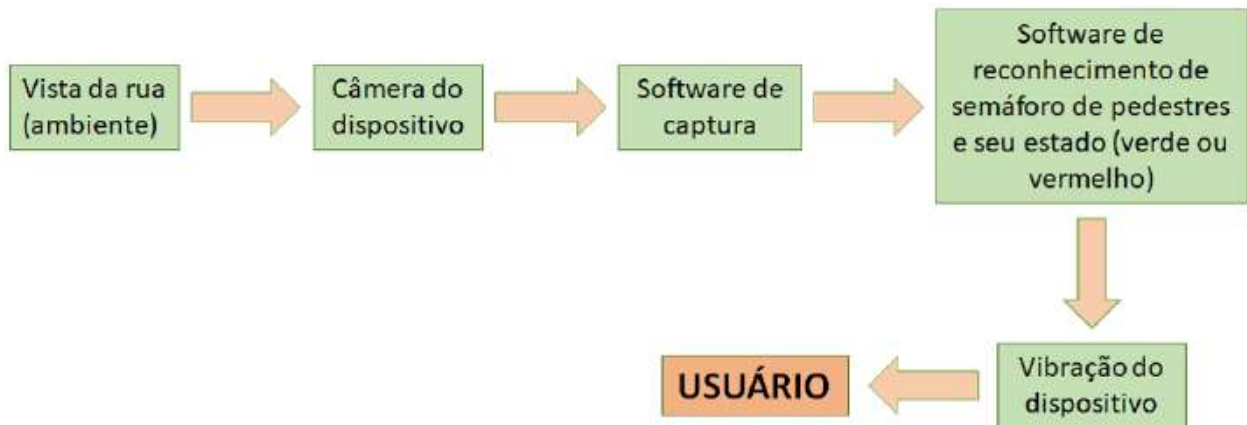


Figura 1: Diagrama de blocos do projeto.

Fonte: Os autores

A junção da tecnologia de identificação visual feita por máquinas, que tem sido estudada nos últimos anos, com a necessidade de uma ferramenta portátil para esse uso seria não apenas possível como também de suma importância para a locomoção desacompanhada dessas pessoas. No Brasil, a maioria dos semáforos de pedestres são apenas visuais. Desse modo, esse dispositivo poderia diminuir consideravelmente as chances de que acidentes de trânsito ocorram com essas pessoas. A maioria das pesquisas relacionadas à apresentação de cores para pessoas com deficiências visuais é por meio do tato, assim como o sistema de braille. Do mesmo modo que existe a possibilidade de leitura por essa estrutura, atualmente existem aplicativos que proporcionam a leitura de textos e livros em voz alta, como o Balabolka [5].

Porém, atualmente não há um dispositivo com a função de converter a identificação de cores para o modo auditivo ou vibrações, com prioridade para semáforos de pedestres. O dispositivo proposto é possível, considerando que já existem aplicativos que identificam cores e até mesmo diferenciam tonalidades de uma mesma cor [6]; e também a plataforma embarcada OpenCV AI Kit [7], que faz o processamento de imagens e consegue extrair diversas informações dessas.

A referência [8] apresenta uma abordagem para a identificação de semáforos de pedestres e principalmente disponibiliza a base de imagens deles, o que viabiliza o treinamento de modelos próprios de identificação. Adicionalmente, considerando a existência de recursos como a biblioteca OpenCV e aplicativos capazes de identificar cores por meio de uma câmera; e também ferramentas para expressão de informações de texto por meio de um sistema de voz e de vibração, a proposta se mostra viável de ser executada.

4.1 Aquisição de imagens

O banco de imagens utilizado foi extraído do artigo “Real-Time Detection of Pedestrian Traffic Lights for Visually-Impaired People” [8]. A base de imagens é composta por 4.399 imagens. Essas fotos foram criadas a partir de cortes em vídeos dos semáforos durante as trocas de estado. As imagens estão divididas em três categorias: treinamento (3.443), validação (478) e teste (478). A Figura 2 a seguir foi extraída do banco de imagens citado e mostra como são as imagens que o dispositivo faria durante seu uso.

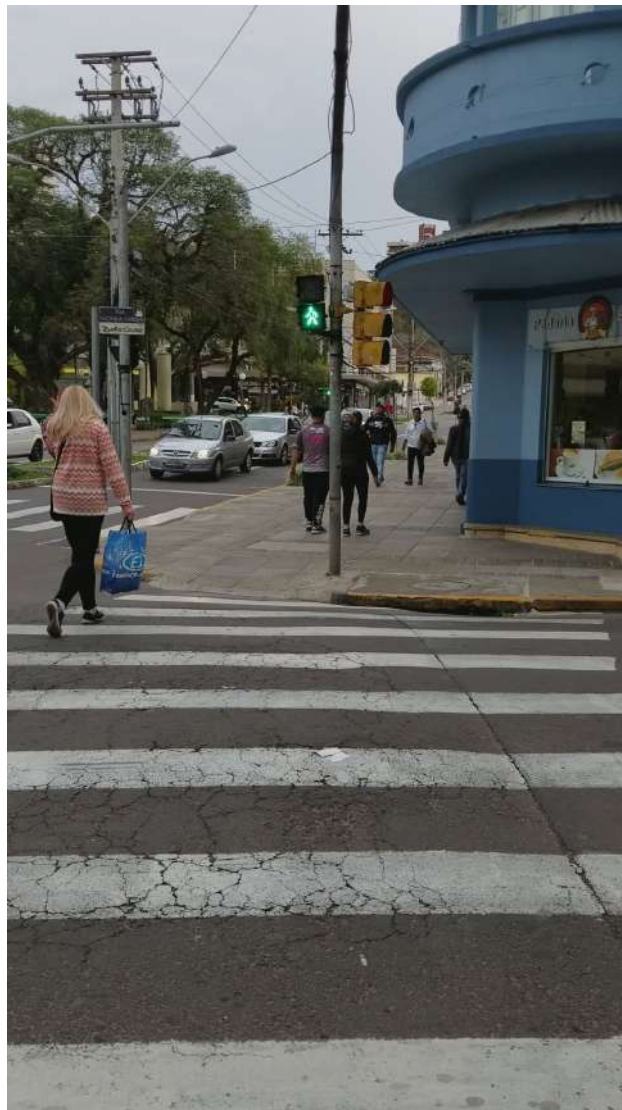


Figura 2: Fotografia do semáforo.

Fonte: [8]

4.2 Pré-processamento das imagens

O pré-processamento de imagens é um procedimento essencial para qualquer sistema de aprendizagem de máquina como o de reconhecimento de semáforos de pedestre. É necessário garantir a qualidade dos dados inseridos com o propósito de evitar resultados indesejados [9]. Para isso utilizamos a técnica ROI ("Region Of Interest" ou AOI - "Area Of Interest") a região definida automaticamente a partir de parâmetros obtidos na própria imagem (ou por um usuário) onde o processamento estará totalmente concentrado.

A ROI é de extrema importância para que o algoritmo compreenda quais partes da imagem ou vídeo utilizar para alimentar o treinamento nos estágios seguintes.

Um bom desempenho de um sistema de reconhecimento de semáforos de pedestres requer que as imagens utilizadas na classificação sejam geometricamente as mais semelhantes possíveis às imagens usadas durante o treinamento. Para esta normalização, utiliza-se as transformações geométricas nas imagens, que são a rotação, escala e translação [9].

4.3 Método de identificação de cores e objetos

Essa seção vai apresentar alguns softwares e métodos para a identificação de objetos, mais especificamente semáforos de pedestre, e de cores.

Utilização de um modelo para identificação de cores: RGB, onde todas as cores são representadas como variações de vermelho, azul e verde e é representado por um cubo com coordenadas cartesianas sendo cada eixo relacionado à intensidade das cores; HSV, derivado do RGB, tem seu espectro de cores representado em um cilindro reto e tem como parâmetros o matiz, o brilho e a saturação; HSI, apresentado na referência [10], é similar ao HSV mas sua representação é em um cone.

A referência [8] apresenta métodos de detecção de semáforos de pedestre, assim como uma base de imagens de semáforo de pedestres denominada *Pedestrian Traffic Light Dataset* - PTLTD. Esse artigo utilizou a abordagem de redes neurais convolucionais para a detecção de objetos denominados Faster R-CNN, SSD, YOLO Full e YOLO Tiny. Tendo apresentado ótimos resultados de aprendizado e de detecção de semáforos de pedestres. Além disso, no mesmo artigo foi feito um teste em semáforos da Alemanha mantendo como base os dados dos semáforos brasileiros, ao fim desse experimento foi indicado que o Faster R-CNN era a melhor opção, mesmo com suas limitações, como consumir mais tempo para desenvolvimento. Tal estudo também mostrou que ainda é preciso o desenvolvimento de uma estratégia de amostragem em tempo real.

A referência [11] apresenta uma metodologia de identificação de objetos por meio de redes neurais. Nessa alternativa cada pixel emite uma informação no momento em que é registrada. Assim, não é utilizada a noção de frames, mas uma concepção de eventos espalhados temporalmente e espacialmente. Essa metodologia permite identificar deslocamento dos objetos. A partir das informações recebidas é recriado a noção de frame. Por fim é necessário apresentar à máquina os objetos que serão identificados durante seu uso, sendo preciso gravar o objeto em diferentes ângulos.

4.4 Interface para pessoas com deficiência visual

Esta seção apresenta o conceito da "interface" de operação do dispositivo e a utilização esperada dos usuários com deficiência visual.

Após a câmera do dispositivo fazer o reconhecimento do semáforo de pedestres, esse emitirá um dos três tipos de vibrações nele, de acordo com a vibração o usuário saberá se o semáforo está aberto para ele, fechado ou se nenhum semáforo foi identificado.

De acordo com o Correio Braziliense [12], jornal com sede em Brasília, foi criado um aplicativo por estudantes após verem imagens de uma pessoa com deficiência visual não conseguindo atravessar a rua porque nenhuma pessoa a informava quando o semáforo estava fechado ou aberto. Esse dispositivo informa às pessoas com deficiência visual se o semáforo está fechado ou livre, em um tipo de relógio de pulso a pessoa recebe diferentes tipos de vibrações para indicar qual a cor do semáforo. Tal dispositivo tem objetivos similares que os desse projeto, porém com uma metodologia diferente, além disso, ele foi patenteado em 2012 mas não chegou ao mercado.

A Figura 3 é um esboço preliminar que foi inicialmente idealizado de qual seria a aparência do dispositivo quando pronto. Este teria, em um extremo, a entrada para conectar um computador embarcado como o Raspberry Pi 3, uma câmera na outra extremidade, e um botão de ativação com um LED na parte superior do dispositivo mais próximo da câmera para que o usuário possa identificar qual lado apontar para o semáforo com mais facilidade.

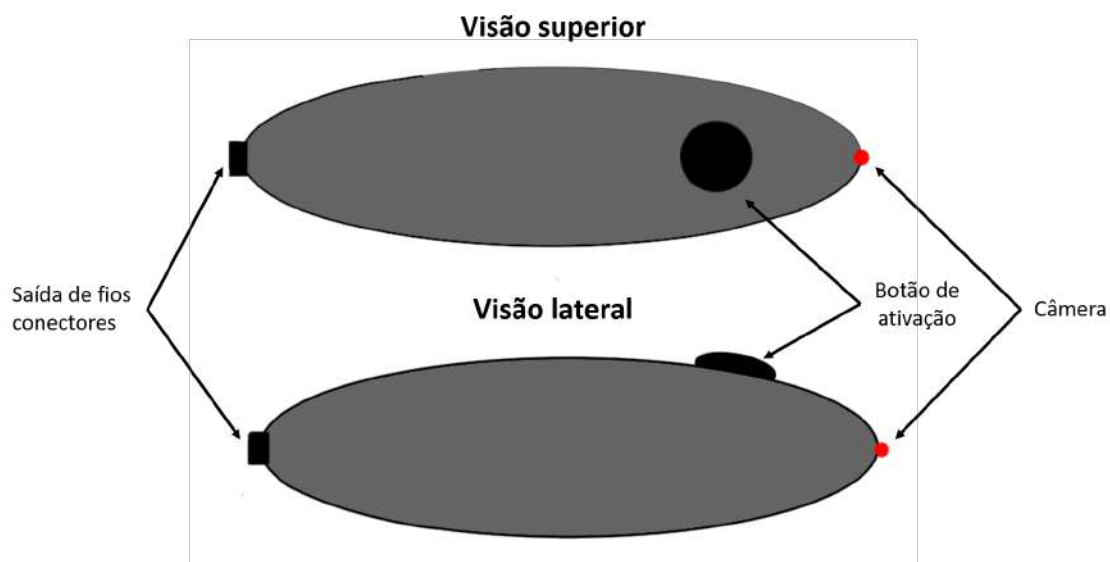


Figura 3: Representação do dispositivo.

Fonte: o autor.

O usuário carregará o dispositivo com um carregador. Quando ele estiver na rua e quiser acionar o dispositivo apertará o botão de ativação na parte superior dele e apontará a câmera, que está na parte frontal, para onde tiver um semáforo de pedestre.

4.5 Escolha do modelo de treinamento

Há alguns modelos de treinamento que podem ser utilizados para sistemas de reconhecimento de semáforos, sendo o primeiro o mais utilizado nessa pesquisa. [9]:

- *Convolutional Neural Networks (CNNs)* - HO [13].

- *Long Short Term Memory Networks (LSTMs)* [13].
- *Recurrent Neural Networks (RNNs)* [13].

4.6 Classificação

Classificar é prever uma classe, em valores discretos, a partir de uma imagem de entrada ou de parâmetros de características do conteúdo. Um modelo de classificação bastante utilizado para sistemas de reconhecimento de expressão facial é o CNN (*Convolutional Neural Network*), pois este apresenta bom desempenho em aplicações de processamento de imagens [9].

A arquitetura básica de uma CNN geralmente consiste em três camadas: camada de convolução, camada de *pooling* e camada de conexão completa:

- Camada de convolução: o objetivo principal é extrair os recursos de entrada, que são fornecidos a um grupo de neurônios de aprendizagem. A camada seguinte, a qual inclui neurônios com pesos treinados e distorções, é abastecida desses recursos. Os pesos podem ser alterados em uma mapa de ativação quando na necessidade de inserir uma nova característica. Essa camada atua conjuntamente com kernels - filtros de extração de características tais como linhas verticais ou horizontais. Quanto melhores os kernels, melhores a função de aprendizado e maior a acurácia do sistema [14].
- Camada de *pooling*: é responsável por discriminar características visuais primárias, arranjando-as em características visuais abstratas [14].
- Camada de conexão completa: as últimas camadas são camadas totalmente conectadas que compilam os dados extraídos pelas camadas anteriores para formar a saída final [14]. A Figura 4 exemplifica uma rede neural com camada de conexão completa.

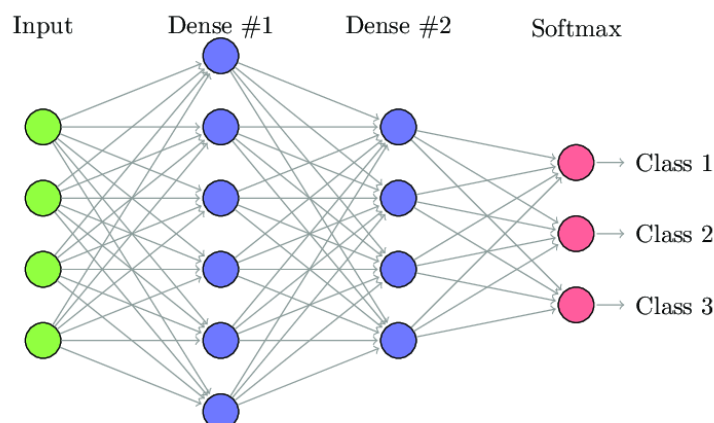


Figura 4: Diagrama representando uma rede neural com conexão completa.

Fonte: [15]

4.7 Alternativas de análise de cores para detecção no semáforo para pedestres

Os principais sistemas de cores são RGB, HSV e YUV [16]. O sistema de cores RGB é formado por vermelho, verde e azul e é comumente utilizado em computação gráfica, mas não é muito eficiente para imagens reais. O sistema é representado em um cubo com coordenadas cartesianas sendo cada coordenada uma das três cores e a diagonal representa a variação de cinza (de branco a preto). O problema com imagens reais é causado pelo nivelamento de profundidade e resolução de cada pixel.

Uma das versões do RGB é o sRGB - “Studio RGB” - que possui maior compatibilidade com vídeo. Porém, a limitação desse sistema é que os valores normalizados ficam estritamente entre 0 e 1 e cores fora desse espaço não podem ser usadas.

O HSV e seu derivado HSI são sistemas que foram desenvolvidos para ser mais intuitivo na manipulação de cores e foi criado para aproximar o modo que as pessoas percebem e interpretam cores. O HSI, com representação das cores em um cone, é melhor para imagens tradicionais, o HSV, com representação das cores em um cilindro, é melhor para manipulação de matiz e saturação.

O YUV é um sistema preto e branco que usa somente a informação do brilho, a informação da cor é recebida de como que os receptores preto e branco apresentam uma imagem em preto e branco e os receptores de cores decodificam as informações das cores adicionais para apresentar uma imagem colorida.

5 Viabilidade

O estudante e o orientador proponentes já realizaram uma pesquisa anterior no período de 2020 e 2021, sem fins acadêmicos, na qual procederam análises preliminares com um modelo treinado disponível para reconhecimento de objetos e pessoas.

O projeto proposto será realizado em modo remoto (*home office*), coordenado pelo orientador, na residência do aluno, tendo disponível espaço físico e infraestrutura adequados para a sua realização.

6 Cronograma de atividades

As atividades do projeto foram divididas em seis etapas, conforme descritas a seguir:

1. Etapa 1: Pesquisa bibliográfica

Nesta fase será realizada a pesquisa bibliográfica relacionada ao pré-processamento de imagens referente, especialmente, ao enquadramento e à luminosidade das imagens, bem como as técnicas de aprendizado de máquina necessárias para o reconhecimento de expressão facial.

2. Etapa 2: Elaboração da base de dados de treinamento e de testes

Será realizado o levantamento de imagens dos diferentes tipos de semáforos de pedestres existentes no Brasil. Buscar-se-á não somente bancos de dados já existentes como também a ampliação dos mesmos por meio de técnicas de aumento de dados.

3. Etapa 3: Reconhecimento do semáforo

Elaboração do sistema de reconhecimento de semáforos com execução em tempo real, utilizando-se as técnicas de aprendizagem de máquina cabíveis.

4. Etapa 4: Análise dos resultados

O sistema elaborado e seu desempenho serão analisados quanto às funcionalidades, bem como à capacidade de reconhecimento em tempo real dos estados de semáforos para pedestres.

5. Etapa 5: Refinamento do sistema

Serão realizados ajustes e refinamentos do sistema além de possíveis testes com outras técnicas.

6. Etapa 6: Relatório final

Elaboração da análise final do desempenho do sistema e execução do relatório final.

O cronograma destas etapas previstas está apresentada na Tabela 1.

Tabela 1: Cronograma de atividades previstas. Fonte: Os autores

Etapa:	MÊS											
	01	02	03	04	05	06	07	08	09	10	11	12
1.Pesquisa bibliográfica	X	X	X	X	X	X	X	X	X	X	X	
2.Base de dados para treino e teste	X	X	X									
3.Reconhecimento		X	X	X	X	X	X					
4.Análise dos resultados							X	X	X	X	X	
5.Refinoamento do sistema								X	X	X	X	X
6.Relatório final										X	X	X

Referências

- [1] IBM. *What is computer vision?* 2021. <https://www.ibm.com/topics/computer-vision>. Acesso em: 15.maio.2021.
- [2] IRIS, I. *Perguntas Comuns*. 2016. <http://www.iris.org.br/faq>. Acesso em: 26.Outubro.2020.
- [3] SANTOS, A. J. de O.; CASTRO, S. A. de. Autoestima a partir do caminhar: orientação e mobilidade da pessoa com deficiência visual. *Revista Benjamin Constant*, Instituto Benjamin Constant, v. 54, n. 3, 2013.
- [4] PORTO, L. E. *CittaMobi Acessibilidade*. 2020. <https://cittamobi.com.br/home/cittamobi-acessibilidade/>. Acesso em: 08.Novembro.2020.
- [5] LÓPEZ, A.; GOMES, T. *Balabolka*. 2020. <https://balabolka.br.uptodown.com/windows>. Acesso em: 10.Fevereiro.2021.
- [6] HIDALGO, G. D. Aplicativo para smartphone: Identificador de cores para pessoas com deficiência visual. *Monografia de trabalho de graduação em Engenharia de Informação. Universidade Federal do ABC*, Universidade Federal do ABC, Santo André, 2019.
- [7] NELSON, P. *Introducing OAK: Spatial AI Powered by OpenCV*. 2020. <https://opencv.org/introducing-oak-spatial-ai-powered-by-opencv/>. Acesso em: 02.Outubro.2020.
- [8] GHILARDI, M. C. et al. Real-time detection of pedestrian traffic lights for visually-impaired people. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2018. p. 1–8.
- [9] CANEDO, D.; NEVES, A. J. R. Facial expression recognition using computer vision: A systematic review. *Applied Sciences*, v. 9, n. 21, p. 4678, 2019. ISSN 2076-3417. Disponível em: <<https://www.mdpi.com/2076-3417/9/21/4678>>.
- [10] SHALEV, A.; LAUTERBACH, B. Pedestrian traffic light recognition for the visually impaired. 2012. Acesso em: 22.Novembro.2020.
- [11] GOUVEIA, E. B. Identificação de objetos utilizando visão neuromórfica e redes neurais convolutivas. 2019. Acesso em: 29.Novembro.2020.
- [12] AYER, F. Brasileiros criam app que informa a cegos se semáforo está fechado ou livre. *Correio Braziliense*, Correio Braziliense, 2016. Acesso em: 26.Outubro.2020.
- [13] RASCHKA, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, 2018.
- [14] ZHANG, T. Facial expression recognition based on deep learning: A survey. In: . [S.l.: s.n.], 2018. p. 345–352. ISBN 978-3-319-69095-7.
- [15] PELLETIER, C.; WEBB, G.; PETITJEAN, F. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, v. 11, p. 523, 03 2019.
- [16] HASTING, G.; RUBIN, A. Colour spaces - a review of historic and modern colour models*. *African Vision and Eye Health*, v. 71, 12 2012.