

# Sistema de auxílio na locomoção de deficientes visuais

### Diego Antônio Amâncio Duarte

Escola Técnica Estadual Frederico Guilherme Schmidt, São Leopoldo - RS, Brasil  
E-mail: diego7042@gmail.com

### Gabriel Tamujo Meyrer

Escola Técnica Estadual Frederico Guilherme Schmidt, São Leopoldo - RS, Brasil  
E-mail: gabrieltamujo@outlook.com

### Thais de Oliveira Tonel

Escola Técnica Estadual Frederico Guilherme Schmidt, São Leopoldo - RS, Brasil  
E-mail: thaistonel@hotmail.com

### Marcos Freire Machado

Escola Técnica Estadual Frederico Guilherme Schmidt, São Leopoldo - RS, Brasil  
E-mail: freiremachado@ig.com.br

## RESUMO

O objetivo deste artigo é descrever a elaboração de um sistema inovador no ramo de mobilidade para deficientes visuais. A observação cotidiana evidencia sérias dificuldades no deslocamento urbano, principalmente em locais desconhecidos. O fluxo intenso de pessoas, ruídos, obstáculos físicos e inacessibilidade de informações sobre o local, são os principais problemas diariamente enfrentados pelos deficientes visuais. O sistema proposto baseia-se na utilização de um par de óculos providos de microfone, fones de ouvido e sensores ópticos. O software se encarrega da análise de dados coletados pelos periféricos, localização no ambiente, descrição de rotas, detecção de obstáculos e de pessoas, e reconhecimento de caracteres. A implementação de algoritmos de visão computacional, foi criada em linguagem Python, se alia à biblioteca OpenCV e Scikit-image. A partir da informação da planta do local, o software identifica a posição exata do dispositivo no ambiente assim como os obstáculos à sua volta, para então, guiar o deficiente visual via sintetização vocal.

**Palavras-chave:** Deficientes visuais. Visão computacional. Inteligência artificial. Dispositivo guia. Acessibilidade.

## Assistance system for locomotion of visually impaired

### ABSTRACT

The aim of this article is to describe the development of an innovative system in the field of mobility for the visually impaired. The daily observation evidences a series of difficulties in the urban displacement, mainly in unknown places. The intense flow of people, noise, physical obstacles and inaccessibility of local information are the main daily problems faced by the visually impaired. The proposed system is based on the use of glasses with a microphone, headphones and optical sensors. The software takes care of analyzing data collected by peripherals, location in the environment, route description, obstacle and people detection, and character recognition. The implementation of computer vision algorithms was created in Python language, allies with the OpenCV and Scikit-image library. From the site plan information, the software identifies the exact position of the device in the environment, as well as the obstacles around it, to guide the visually impaired via vocal synthesis.

**Keywords:** Visually impaired. Computer vision. Artificial intelligence. Guiding device. Accessibility.

## 1 INTRODUÇÃO

Não raro nos deparamos com as dificuldades enfrentadas pelos deficientes visuais. Tarefas relativamente simples, como a locomoção, muitas vezes tornam-se desafios a serem superados sem a dependência de terceiros, principalmente quando se trata de locais desconhecidos. Segundo o Instituto Brasileiro de Geografia e Estatística, essa é a realidade de aproximadamente 3,5% da população brasileira, ou seja, cerca de 6,5 milhões de indivíduos (IBGE, 2010).

A tecnologia, por sua vez, nunca esteve tão presente no cotidiano da humanidade. O constante processo de globalização busca, através de meios tecnológicos, soluções inovadoras para os problemas cotidianos. Smartphones cada vez mais velozes e inteligentes, tornaram-se ferramentas imprescindíveis na rotina diária. Entretanto, em meio a tantos gadgets dotados de alta tecnologia, percebe-se ainda a falta de dispositivos capazes de proporcionar aos deficientes visuais um certo grau de autonomia, de forma acessível e condizentes às suas necessidades.

Diante desta realidade, este trabalho busca descrever a elaboração de um novo sistema de orientação espacial para deficientes visuais: um dispositivo individual discreto, cientificamente confiável, financeiramente acessível e que possibilite a participação integral dos indivíduos na sociedade.

## 2 TRABALHOS RELACIONADOS

Durante às pesquisas, foram encontrados alguns dispositivos voltados à assistência de deficientes visuais. Um exemplo foi o OrcamMyEye2, de origem israelense. Trata-se de um dispositivo acoplável aos óculos que conta com recursos de leitura de texto, reconhecimento de cores, reconhecimento de produtos, reconhecimento de faces, reconhecimento de notas de dinheiro, além de ser capaz de dar informações sobre data e hora (ORCAM, 2017). O dispositivo não oferece suporte para orientação e mobilidade, apresenta pontos positivos, mas sua utilização no Brasil torna-se inviável para a grande maioria da população, devido ao custo de importação (GIZMODO, 2019).

Outro produto encontrado foi o WeWalk na área de mobilidade. O WeWalk é uma bengala que possui um sensor ultrassônico para avisar em caso de obstáculos aéreos (WEWALK, 2020). O produto conta com integração à internet para navegação por mapas virtuais e integração com assistentes comuns no mercado. Eficiente no que se propõe, tem uma lacuna na parte de detecção de objetos, leitura e tarefas que necessitam de um sensor óptico. Por fim, outro trabalho semelhante na área é o SoundSee, dispositivo desenvolvido pela Universidade de São Paulo (BERNARDES, 2016). O dispositivo também faz uso de uma localização por meio de ultrassom para o mapeamento do ambiente, tendo foco bem específico e necessitando de outros dispositivos para tarefas visuais. Este trabalho visa integrar o melhor das tecnologias pesquisadas, em um dispositivo que esteja ao alcance financeiro dos usuários.

### 2.1 Visão computacional, inteligência artificial e redes neurais

Neste trabalho, vamos utilizar do conceito de inteligência artificial, aliada à implementação de recursos de visão computacional para fornecer ao deficiente visual um assistente virtual capaz de integrar as melhores tecnologias na área de acessibilidade.

A visão é um dos processos mais complexos que o ser humano busca compreender, ou seja, como o cérebro processa a visão de modo que possamos conseguir identificar exatamente a posição de um objeto qualquer no espaço? De forma sucinta, segundo Coldevey (2016, p. 02),

o que acontece é que este objeto surge e golpeia a retina, que faz alguma análise e envia o resultado longitudinalmente ao cérebro, onde o córtex visual analisa mais profundamente a imagem”. Em seguida, ele envia para o resto do córtex, que compara a tudo o que já sabe, classifica os objetos e dimensões e, finalmente, decide sobre algo a fazer.

A Pesquisa não busca criar um equipamento que seja capaz de ver da mesma forma que o olho humano, pois está é uma tarefa realmente difícil, não só porque é difícil fazer computadores executarem um cálculo matemático que reproduza a visão humana, mas porque não estamos inteiramente certos de como o processo da visão realmente funciona.

Por meio deste trabalho, busca-se criar um equipamento que utiliza o conceito de visão computacional, segundo Coldewey (2016, p. 07),

A visão computacional é o processo de modelagem e replicação da visão humana usando software e hardware. Visão computacional e reconhecimento de imagem são termos frequentemente usados como sinônimos, mas o primeiro abrange mais do que apenas analisar imagens. Isso porque, mesmo para os seres humanos, “ver” também envolve a percepção em muitas outras frentes, juntamente com uma série de análises. [...] a visão computacional em seu sentido mais pleno está sendo usada na vida cotidiana e nos negócios para conduzir todos os tipos de tarefas, incluindo identificar doenças médicas em raios-x, identificar produtos e onde comprá-los, anúncios dentro de imagens editoriais, entre outros. A tecnologia é complexa e, assim como todas as tarefas acima mencionadas, requer mais do que apenas reconhecimento de imagem, mas também análise semântica de grandes conjuntos de dados.

As fronteiras da visão computacional e Inteligência Artificial convergem. O futuro da visão computacional é integrar sistemas poderosos e específicos, com os mais amplos que estão focados em conceitos que são um pouco mais difíceis de definir: contexto, atenção, intenção.

Apesar da dificuldade em se definir a inteligência artificial, para esta pesquisa é suficiente afirmar que, IA é o ramo da ciência que procura entender o comportamento humano, com o objetivo de reproduzi-lo em máquinas inteligentes. Este raciocínio baseia-se nas lições de McCarthy (2007, p. 01), ao afirmar que

inteligência artificial é a ciência e a engenharia de fabricação de máquinas inteligentes, especialmente programas de computador inteligentes. Está relacionado à tarefa semelhante de usar computadores para entender a inteligência humana, mas a IA não precisa se limitar a métodos que são biologicamente observáveis.

Até recentemente, a criação e operação de redes neurais artificiais era impraticável por causa da imensa quantidade de computação que elas exigiam. Com os avanços na computação diminuíram essas barreiras, e nos últimos anos assistiu-se a uma explosão de investigação e utilização de sistemas que imitam – ainda muito aproximadamente – os do nosso cérebro. Para Zaffari (2015, p. 118),

redes neurais são sistemas de computação com nós interconectados que simulam matematicamente o funcionamento dos neurônios do cérebro humano. Utilizando algoritmos, as redes neurais podem, por exemplo, reconhecer padrões e correlações aplicadas ao reconhecimento de fala e imagem escondidos em dados brutos, agrupá-los e classificá-los, e – com o tempo – aprender e melhorar continuamente.

Para o desenvolvimento de uma ferramenta, capaz de tratar os dados capturados pelas câmeras do gadget, utilizou-se as bibliotecas multiplataformas na área de visão computacional e redes neurais, OpenCV e Scikit-image, totalmente livres ao uso acadêmico e comercial, para o desenvolvimento de aplicativos na área de visão computacional de código aberto, escrita em C e C++, projetadas para uma melhor eficiência computacional e aplicações em tempo real (BRADSKI, 2008). Com essas características e mais de 500 funções voltadas para computação, a biblioteca OpenCV se mostrou bastante versátil. Já a Scikit-image, biblioteca de processamento de imagem que implementa algoritmos e utilidades para o uso na área de pesquisa, educação e aplicações industriais (VAN DER WALT, 2014), forneceu métodos mais complexos de lidar com processamento de imagens em estéreo e pré-processamento.

### 3 MATERIAIS E MÉTODOS

Com o objetivo de compreender as principais dificuldades encontradas no cotidiano destes indivíduos e analisar como a tecnologia pode beneficiá-los, aplicou-se uma entrevista a 20 deficientes visuais na instituição ADEVIC (A Associação dos Deficientes Visuais de Canoas).

As principais dificuldades relatadas foram a de locomoção, localização, detecção de obstáculos, leitura e reconhecimento de objetos como notas de dinheiro e produtos. Quanto ao formato físico do dispositivo, os entrevistados, de forma unânime, citaram que deveria ser discreto, leve e ajustável. A busca pela compreensão da realidade dos indivíduos portadores de cegueira, através dos dados obtidos nas entrevistas, serve de base para a proposta de gadget<sup>1</sup> que os atenderia de forma mais efetiva.

A tradução do termo gadget para o português significa algo como bugiganga ou engenhoca. Gadget é definido como “um pequeno dispositivo ou ferramenta mecânica ou eletrônica; especialmente engenhosa ou nova” (GADGET, 2020).

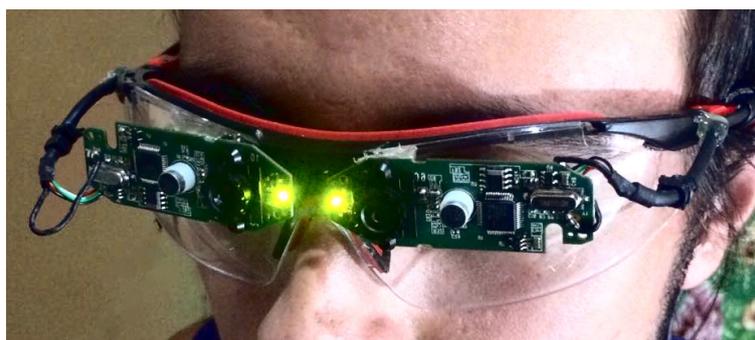
Os avanços da computação e a alta taxa de precisão em atividades de visão computacional e redes neurais aliado aos conceitos de inteligência artificial, trouxe à disposição uma série de ferramentas tecnológicas aplicáveis ao problema da pesquisa. Através de recursos como estes, frequentemente aplicados desde smartphones a carros autônomos, é que consiste a solução proposta: um gadget vestível, um par de óculos, dotado de recursos tecnológicos capazes de auxiliar nas principais dificuldades encontradas no cotidiano de quem sofre a perda de um sentido tão essencial quanto a visão. Isto é, um dispositivo capaz de manter uma comunicação ativa com o usuário via comandos e sintetização de voz em linguagem natural, através de um assistente virtual, atuando como guia em locais desconhecidos, identificando obstáculos, permitindo a leitura de caracteres, reconhecimento de objetos, e, até mesmo, a navegação na internet para acesso a informações como climatologia, notícias.

O protótipo do hardware desenvolvido (conf. fig. 01), conta com um par de câmeras Logitech C270 e um headset estéreo modelo Microfosft Lifechat lx-3000, o qual possui um sistema integrado de cancelamento de ruído. As câmeras, retiradas de sua carcaça original, foram acopladas em uma placa fabricada em acrílico e fixadas em um par de óculos. Como dispositivo de processamento de dados, foi utilizado um computador dentro das seguintes configurações: Sistema Operacional Linux® Ubuntu 16.04, processador Intel® Core™ i5-3330, CPU 3.00 GHz, RAM 4.00GB e um HDD como dispositivo de armazenamento padrão.

---

<sup>1</sup> A tradução do termo gadget para o português significa algo como bugiganga ou engenhoca. Gadget é definido como “um pequeno dispositivo ou ferramenta mecânica ou eletrônica; especialmente engenhosa ou nova” (GADGET, 2020).

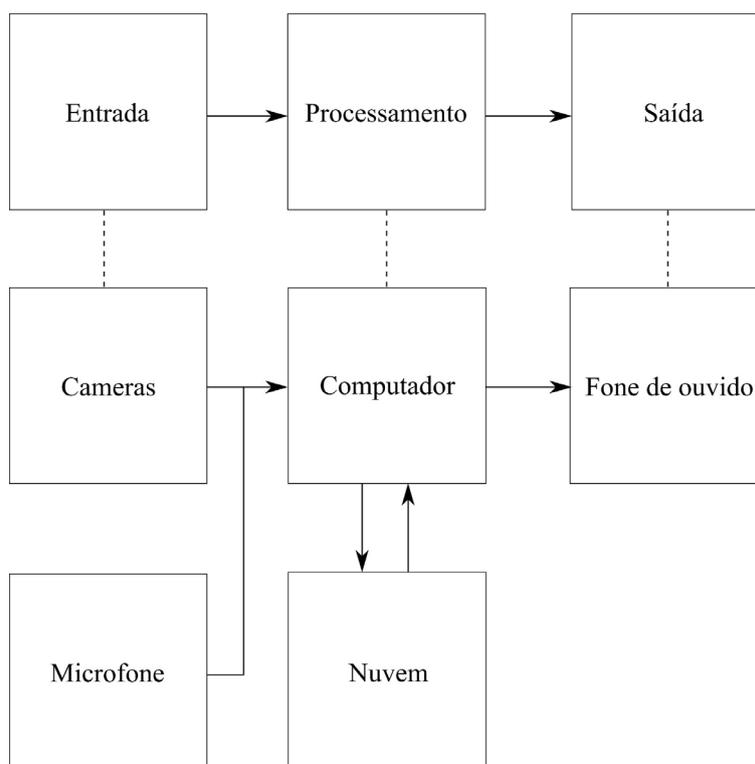
**Figura 1.** Protótipo.



A apresentação do fluxograma do sistema desenvolvido foi dividida em duas partes, (conf. fig. 2), a parte física do sistema se divide 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 através de um circuito eletrônico.

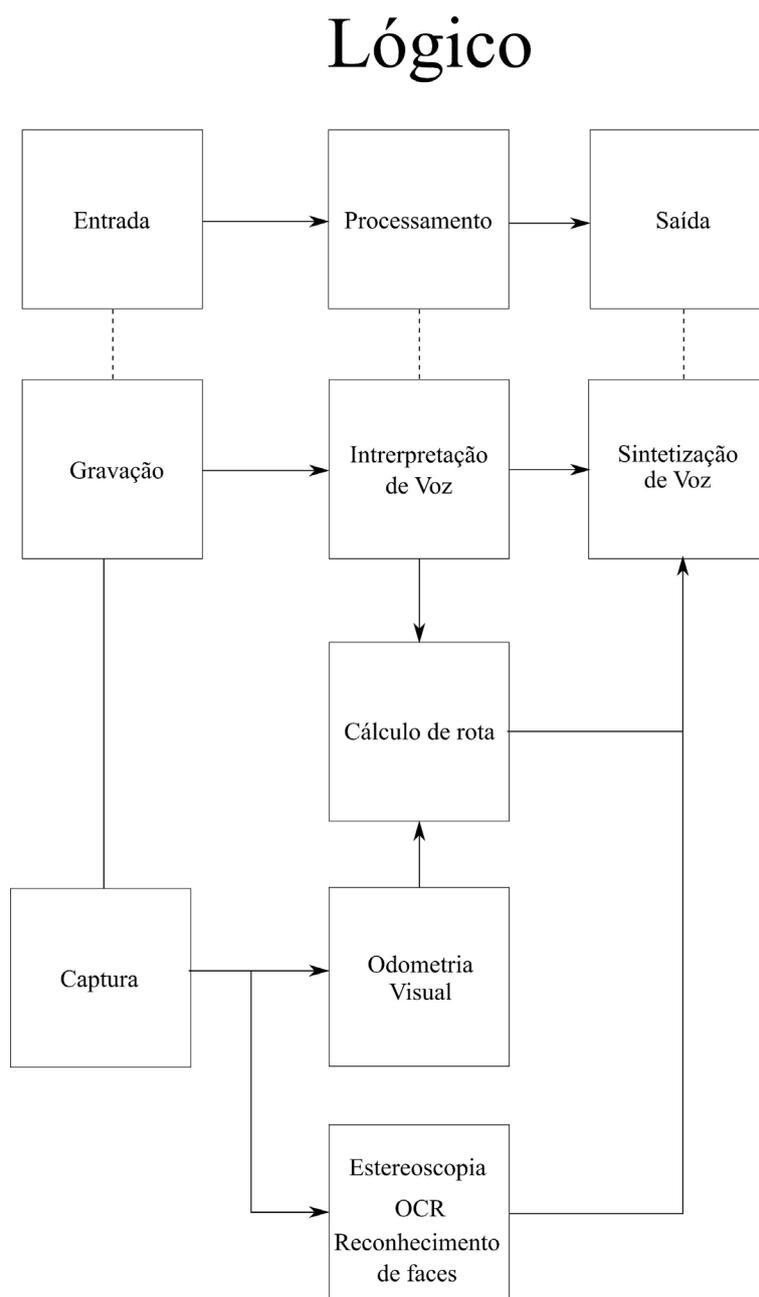
**Figura 2.** Parte física do sistema.

## Físico



A parte lógica do sistema (conf. fig.03) se encarrega de interpretar os sinais vindos do ambiente a ser reconhecido, realizando a conversão destes dados em sinais de áudio que fazem a interação com o usuário. É ele quem traduzirá e guiará o usuário de forma eficiente e segura. Nele é aplicado o conceito de inteligência artificial, sendo ele capaz de definir o melhor caminho, assim como interpretar a intenção de quem utilizará. Ele é toda a lógica envolvida para a realização do reconhecimento e comunicação.

Figura 3. Parte lógica do sistema.



### 3.1 “Isabel” a assistente virtual

A audição do cego é o principal meio de percepção do mundo exterior. Dessa maneira, se fez necessário desenvolver uma interface adequada a esta realidade. Portanto a criação de um assistente pessoal inteligente tornou-se a opção mais promissora devido à facilidade de interação com o sistema. Ou seja, desenvolvemos um sistema de comunicação por comando e sintetização de voz, através de linguagem natural, sendo capaz de atender solicitações e retornar informações como se de fato fosse uma pessoa e não uma máquina. Sendo assim, deu-se início no desenvolvimento da Isabel, isto é, a personificação assumida pelo sistema de assistência virtual.

O sistema de reconhecimento e sintetização de voz contaram com a integração das API's2 Google Text To Speech e Google Speech-To-Text (GSTT, 2020). Os trechos de áudio coletados pelo microfone são limitados por zonas de silêncio superiores a um limite pré-estabelecido. Em seguida, estes dados são enviados a nuvem para processamento e, como resultado, retorna um texto com a transcrição do áudio. Visando melhor desempenho de processamento, integrou-se no sistema o recurso Snowboy Hotword Detection. O Snowboy é um mecanismo de detecção de hotword, isto é, uma palavra-chave ou frase que um computador sempre escuta para acionar outras ações. Este recurso é comumente aplicado em sistemas de assistência pessoal como o do Google e Apple, tendo como objetivo a ativação de uma interface de interação de fala completa. Tal como as hotwords “Ok Google” e “Hey, Siri!”, a palavra-chave definida para a ativação deste sistema de assistência virtual foi “Isabel”.

Após a ativação do sistema e a conversão do áudio em texto, adentra-se a terceira etapa do processo: o processamento de linguagem natural. Para tal, utilizou-se um sistema de Chatterbot. As junções da palavra chatter, inglês para conversador, e da palavra robot, inglês para robô, originou o termo Chatterbot. Este termo refere-se a softwares que tem por objetivo simular o homem em conversações naturais, ou seja, sua finalidade é causar a sensação de que uma pessoa e não uma máquina está dialogando com o usuário. (LEONHARDT, 2005). Este tipo de sistema tem por base datasets<sup>3</sup> contendo diversas conversações naturais entre pessoas e, através deste, o sistema realiza o processo conhecido como treinamento, que consiste em uma série de cálculos estatísticos que permitem reconhecer padrões e estimar respostas. O algoritmo desenvolvido baseou-se na biblioteca Chatterbot do Python<sup>4</sup>. Esta biblioteca disponibiliza versáteis ferramentas para o processamento de linguagem natural, projetada para facilitar o desenvolvimento de softwares que possam envolver-se em diálogos. Esta biblioteca trabalha com uma seleção de algoritmos de aprendizado de máquina para produzir diferentes tipos de respostas, tornando fácil para desenvolvedores a automação de conversas. Além disso, a biblioteca também permite que o software melhore seu próprio conhecimento de possíveis respostas à medida que interage com seres humanos e outras fontes de dados informativos. (CHATTERBOT, 2018). Como base de treinamento, utilizou-se um dataset de conversações em português disponibilizado através do GitHub, ChatterBot Language Training Corpus (COX, 2020). Após o processo de treinamento, o algoritmo trabalha selecionando a frase conhecida mais próxima que corresponda à entrada e, em seguida, escolhe uma resposta da seleção de respostas conhecidas para determinada instrução.

### 3.2 A odometria visual como alternativa à localização

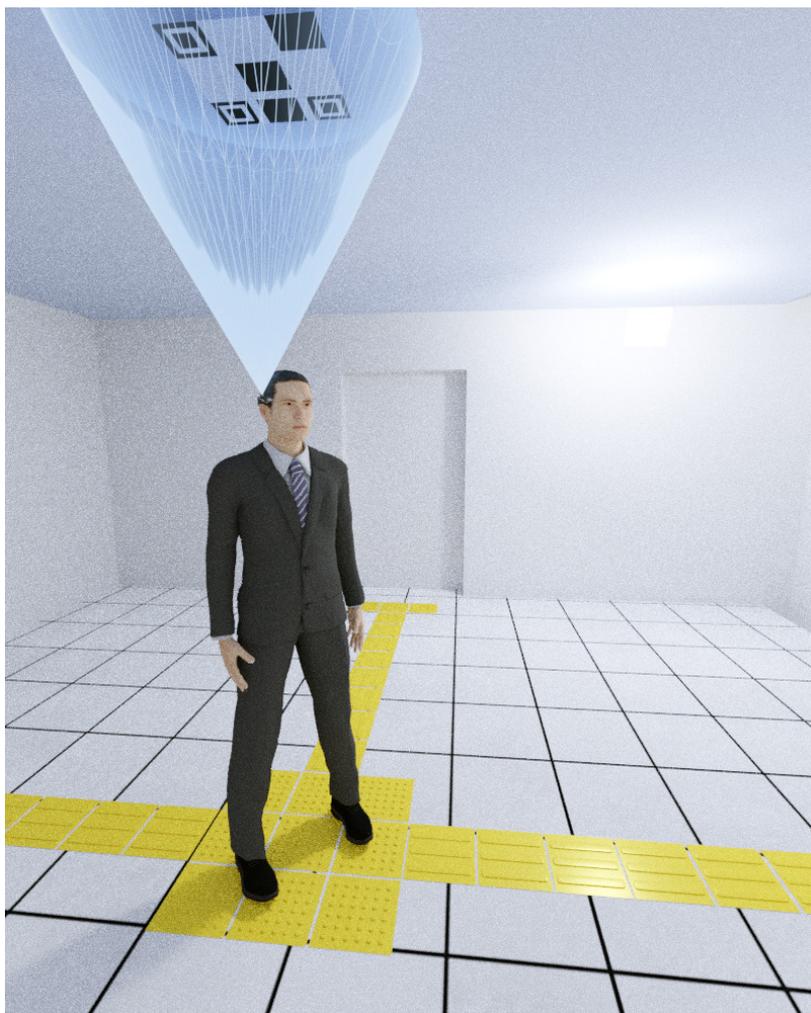
A primeira alternativa desenvolvida para obter a localização do dispositivo no ambiente baseou-se na utilização de códigos bidimensionais. O processo se dá pela utilização dos códigos para a compreensão do espaço, através da varredura de imagem para a averiguação de códigos visuais bidimensionais dispostos no teto dos ambientes para coleta de dados através de uma câmera. Uma trilha, com diversos pontos de referência alternadamente diferentes entre si, para que o dispositivo possa perceber o deslocamento do usuário no ambiente, enquanto códigos especiais, munidos de mais informação, dão detalhes do percurso, do local específico. Este sistema pode ser integrado com os novos pisos táteis já utilizados nesses ambientes para facilitação do deslocamento de deficientes.

2 API é a abreviatura de Interface de Programação de Aplicação, em português. É um conjunto de instruções, rotinas e padrões de programação usadas para que se possa acessar um aplicativo baseado na internet. (RIBEIRO, 2016).

3 Datasets— conjunto de dados, em tradução livre —trata-se de um arquivo que contém centenas e até milhares de dados sobre um determinado assunto. Esse arquivo pode vir na forma de uma planilha no Excel (XLS), um arquivo CSV, TXT, JSON e até XML. (LEONHARDT, 2005).

4 Python é uma linguagem de programação criada por Guido van Rossum em 1991. Os objetivos do projeto da linguagem eram: produtividade e legibilidade. Em outras palavras, Python é uma linguagem que foi criada para produzir código bom e fácil de manter de maneira rápida. (Ibid, 2020).

**Figura 4.** Localização por códigos.

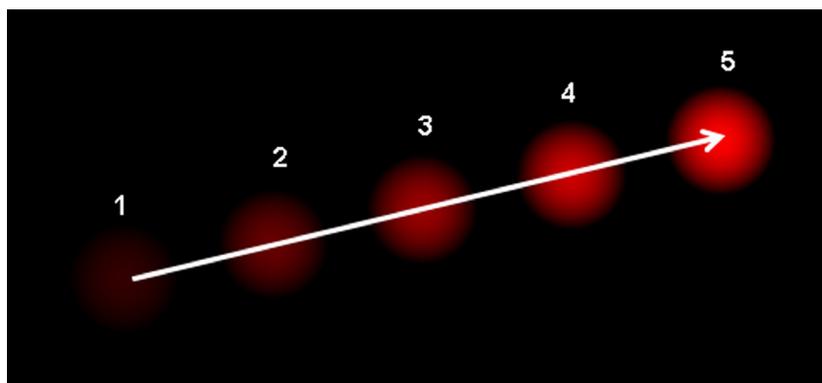


Embora este método seja muito eficaz, a sua implementação na prática demandaria adaptações nos ambientes, o que levou a pesquisa a continuar buscando métodos mais eficientes e de fácil aplicação.

Sete anos de sucesso e muitas descobertas na exploração da superfície terrestre de Marte classificam o Rover Curiosity, projeto da Mars Science Laboratory da NASA (National Aeronautics and Space Administration), como um dos maiores feitos já alcançados pela humanidade. O êxito desta missão contou com a ampliação e o desenvolvimento de muitas tecnologias, tal como a que promove um meio de localização alternativo ao GPS, devido à inviabilidade de sua utilização: a odometria visual (CORKE et al, 2004). Este método, ao contrário do GPS, que necessita pelo menos quatro satélites cruzando informações para obter a localização (PIRES et al, 2014), consegue utilizar apenas imagens capturadas por uma câmera para obter o resultado (NISTÉR, 2004). Dessa forma, este algoritmo tornou-se a opção mais favorável para o controle de localização por manter-se a condição de fácil aplicação à prática.

A odometria visual trabalha com o conceito de fluxo óptico, o qual consiste na observação da movimentação dos pixels de frame a frame, gerando, conseqüentemente, um vetor de movimento do determinado ponto. A Figura 05, a seguir, demonstra o vetor gerado pela trajetória de um ponto em cinco frames consecutivos.

Figura 5. Optical Flow.



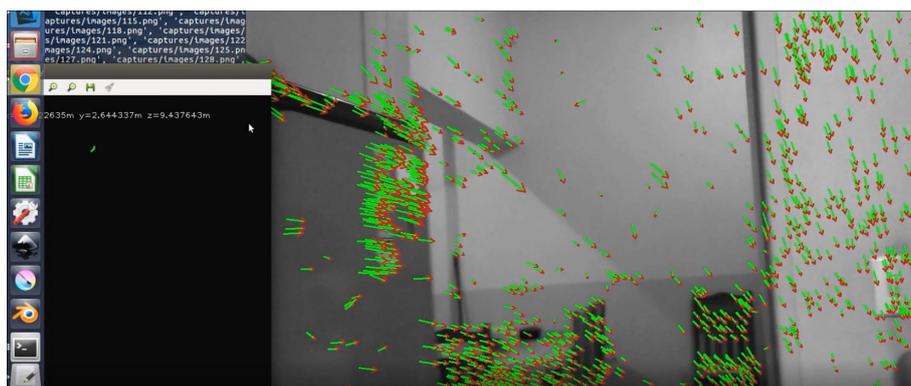
Fonte: OPENCV, 2018.

O algoritmo desenvolvido em linguagem Python para a odometria visual foi baseado no código open-source disponibilizado pelo MIT (Instituto de Tecnologia de Massachusetts), Py-MVO (CAO, 2020) através da plataforma Github. O software recebe a série de imagens em escalas de cinza. Tomamos por  $it$  e  $it+1$  pelas imagens capturadas nos momentos  $t$  e  $t+1$ , respectivamente. Conforme Avi (SINGH, 2015), abaixo a descrição do algoritmo:

- Estando as câmeras calibradas e todos os parâmetros intrínsecos configurados, as imagens passam pelo processo de correção de distorções;
- É necessário realizar a detecção de pontos em  $it$  e acompanhar os seus deslocamentos em  $it+1$ . Nesta etapa podem-se aplicar diversos métodos, variando em demanda de processamento e precisão. A técnica utilizada foi o SIFT (Scale-Invariant Feature Transform);
- Através da técnica de Nister's 5 points é calculada a matriz essencial da câmera, partindo de cinco pontos de interesse escolhidos aleatoriamente. O método RANSAC atua na verificação destes pontos aplicando a matriz obtida aos pontos de  $it$  e desprezando os que não coincidirem com  $it+1$ ;
- Os coeficientes de translação e rotação são estimados a partir da matriz essencial obtida na etapa anterior;
- A partir da translação e rotação calculada, a trajetória da câmera é plotada em um gráfico bidimensional.

A figura 06 a seguir, em seu lado inferior esquerdo apresenta o software em processo, ao lado direito, verifica-se a imagem da câmera e os pontos que estão sendo rastreados, bem como seus vetores de movimento. No lado superior esquerdo, apresenta o desenho da trajetória percorrida.

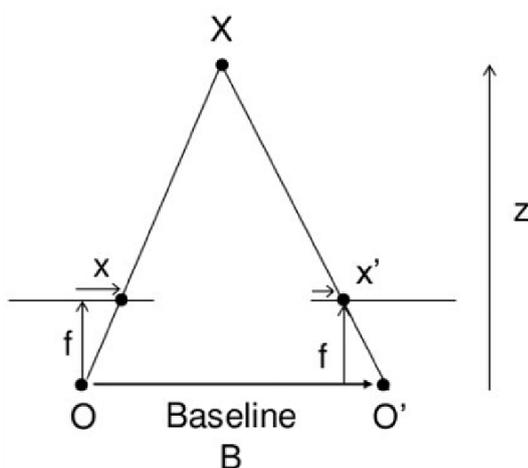
Figura 6. Odometria visual.



### 3.3 Estereoscopia

A tecnologia constantemente se baseia na própria natureza, e este foi mais um caso. O homem, assim como a maioria dos mamíferos, porta dois receptores ópticos. As imagens diferentes captadas por cada um dos olhos, ao serem processadas pelo cérebro, resultam na noção de profundidade. A aplicação desta técnica em dispositivos artificiais, conhecida como estereoscopia, consiste na utilização de duas câmeras, possibilitando o aproveitamento do fenômeno da paralaxe para computação da distância relativa entre os objetos observados (SISCOOTTO et al, 2012). Na prática desta técnica, dificilmente consegue-se obter um alinhamento físico perfeito de duas câmeras, o que é imprescindível para a técnica de estereoscopia. Nessas condições, além de corrigir as distorções das imagens, se faz necessária a utilização de métodos computacionais para corrigir o alinhamento, descritos em cv2.StereoRectify. (OpenCV, 2017a)

Figura 7. Triangulação.



A estimativa de distância baseia-se nos parâmetros intrínsecos e extrínsecos das câmeras. A figura 7 representa o triângulo formado pelo conjunto de pontos de vista (O e O') e o ponto observado (X). A projeção de X gera os pixels x e x' das imagens capturadas. As câmeras distam a B metros uma da outra na horizontal.

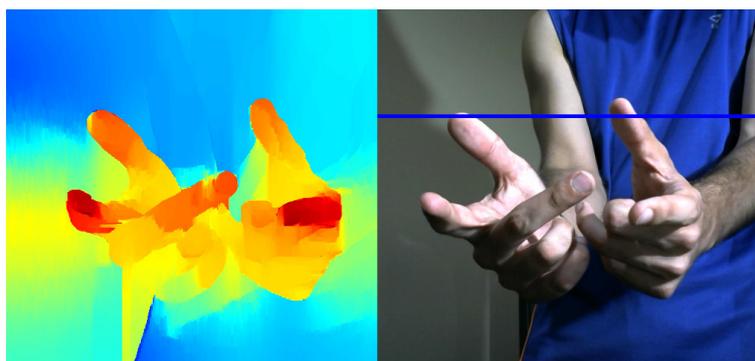
A distância é encontrada seguindo a relação matemática da figura 8, sendo Z a distância do ponto à câmera, f a distância focal, T a distância entre as câmeras e da disparidade (OpenCV, 2017a). A disparidade pode ser definida pela distância relativa entre dois pontos de imagens capturadas em diferentes perspectivas (STIVANELLO, 2008).

Figura 8. Relação entre distância e disparidade.

$$Z = \frac{fT}{d}$$

O mapa de disparidade (conf. fig. 09) é gerado denotando uma cor para representação da disparidade entre cada ponto da imagem, adequando a uma escala de cores que possa ser representada. Na figura 9 verifica-se que a escala está na proporção de quanto mais próximo do tom de vermelho, mais próximo da câmera está o ponto. O algoritmo desenvolvido para computar a disparidade demonstrada baseia-se na utilização do método cv2.StereoSGBM (OpenCV 2017a) e na utilização de filtros de normalização, ambos disponibilizados, respectivamente, pela OpenCV e Scikit-image.

**Figura 9.** Mapa de disparidade.



### 3.4 Cálculo da Rota

No sistema proposto, o registro da planta do local no sistema é feito pela passagem das coordenadas dos retângulos que formam os cômodos, assim como pontos de interesse, tais como banheiros e centros de informação. Tendo conhecimento destas informações é possível calcular rotas possíveis para alcançar qualquer ponto desejado no local. As coordenadas são convertidas para hexadecimal e adequadas a um padrão conveniente para o envio dos dados via Wireless, assim como para uma representação em um código bidimensional.

O dispositivo, ao receber a planta, converte-a em uma malha de nós. Através da odometria visual<sup>5</sup> ou o sistema de códigos bidimensionais se torna possível controlar sua localização na planta. Em caso de solicitação via comando de voz, o software traça o caminho até o ponto desejado. A cada nó é analisado as distâncias até o ponto de partida e a distância até o objetivo, escolhendo o menor caminho. Caso não haja nós vizinhos que não tenham sido visitados, o caminho é descartado, retornando ao nó inicial e escolhendo um caminho diferente. No fim da análise o menor caminho encontrado é escolhido, mantendo uma distância segura das paredes do local. As orientações sobre o caminho são passadas ao usuário via sintetização de voz.

### 3.5 Reconhecimento de Caracteres

Com a evolução dos carros autônomos, o reconhecimento de caracteres (OCR<sup>6</sup>) sofreu um grande avanço na última década, sendo utilizado para ler placas e sinalizações em geral. O processo se baseia na comparação entre as letras capturadas pelo sensor óptico com um banco de dados. Dessa forma, ele analisa a semelhança de cada letra com cada fonte registrada no sistema, retornando a que apresentar uma maior semelhança.

TesseractOCR é um software open-source para decodificação de imagens em texto (SMITH, 2007). Geralmente utilizada para conversão de livros escaneados em arquivos de texto. Tendo em vista a sua utilização padrão, dificilmente ela consegue identificar textos com distorções ou poluição visual. Este fato levou a utilização da OpenCV para adequar as imagens de entrada antes do reconhecimento pelo Tesseract. Algoritmos de Thresholding Binário Adaptativo, operações morfológicas foram utilizadas para remoção dos ruídos e oscilações de cores contidas nas imagens. A retirada da distorção causa da paralaxe foi realizada pelo cálculo da matriz de distorção apresentada pela diferença no alinhamento das arestas da folha e das linhas em que se encontram as letras. Os contornos são detectados e aproximados para formas geométricas, para simplificar a identificação.

<sup>5</sup> Odometria Visual consiste em estimar a posição relativa de um agente por meio de câmeras. (SCARAMUZZA, 2011).

<sup>6</sup> OCR é um acrônimo para optical character recognition (OCR, 2020). Sua tradução literal para o português significa “reconhecimento óptico de caracteres”.

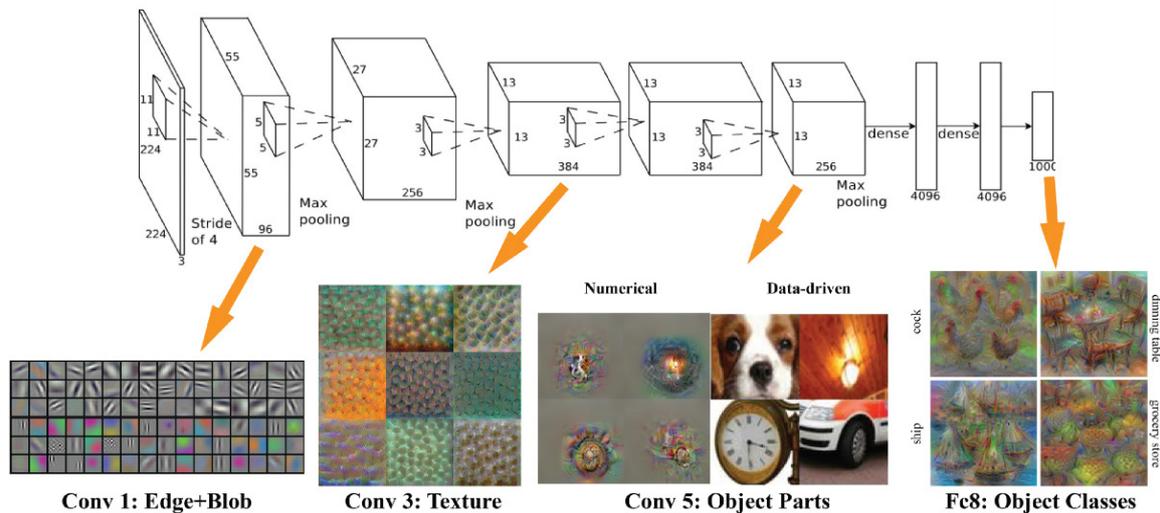
### 3.6 Detecção de Objetos

Nos últimos anos, as redes neurais foram as responsáveis por uma revolução em muitas áreas da inteligência artificial, principalmente as que exploram o reconhecimento de imagem. Porém, equacionar a precisão dos resultados com a capacidade de processamento tornou-se um desafio, principalmente quando se trata de dispositivos móveis.

Recentemente, a Google apresentou uma nova arquitetura de rede neural, a MobileNetV2 (SANDLER, 2018), que é especificamente projetada para ambientes restritos em capacidade computacional. Esta arquitetura possui grande desempenho no desenvolvimento de sistemas de detecção de múltiplos objetos em conjunto com redes como a SSD Multibox7 (LIU, 2016), tornando-se a opção mais favorável para a aplicação no dispositivo.

A eficiência de processamento desta rede pode ser ainda ampliada pela aplicação da técnica de TransferLearning. Conforme WEST et al (2007), esta técnica consiste em armazenar o conhecimento obtido na resolução de um problema e aplicá-lo em um novo problema diferente, porém, relacionado. Isto é, a transferência de aprendizado se dá pelo treinamento de uma rede básica com uma função base que, posteriormente, adapta e transfere o aprendizado obtido para uma segunda rede, de tarefa destino semelhante, que terá o seu treinamento simplificado pelo reaproveitamento destes dados. (SORIA et al, 2009). Conforme a figura 10, as camadas de uma rede neural trabalham com diferentes níveis de absorção de informação. A transferência de aprendizado deve-se ao fato de as informações obtidas nas camadas iniciais possuírem pequenas variações quando se trata de duas redes neurais com funções semelhantes, tornando possível, após o treinamento de uma rede base, economizar o processamento treinando apenas a última camada das redes seguintes. Dessa forma, o conhecimento obtido durante o treinamento de uma rede que reconhece carros poderia ser aplicado para identificar caminhões ou outros veículos semelhantes.

Figura 10. Camadas de uma rede neural.

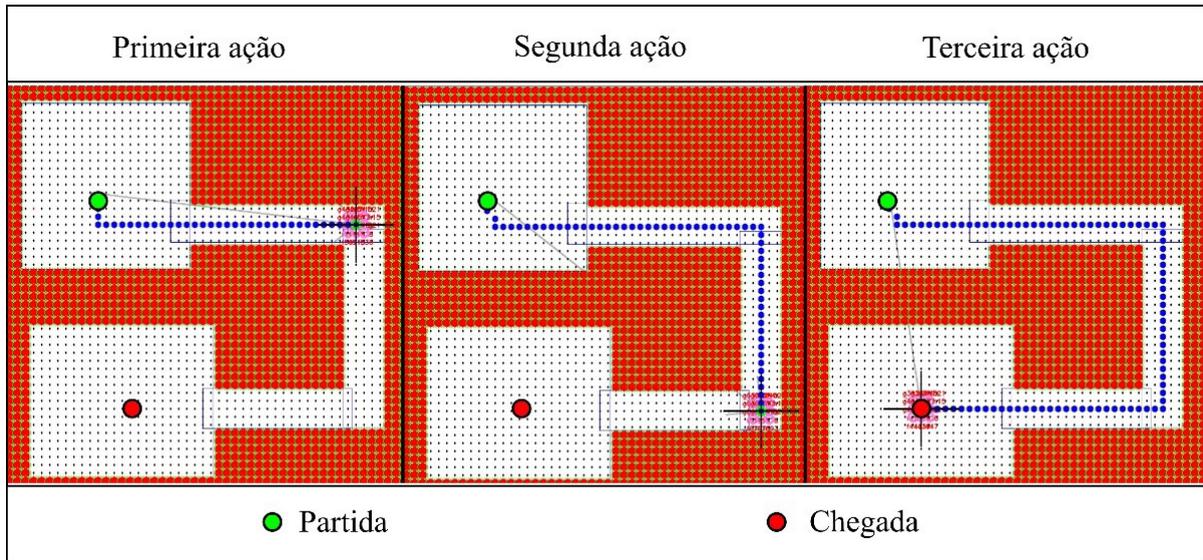


7 SSD é uma forma abreviada de Single Shot Detector, se referindo ao fato de que o processamento ser feito todo de uma só vez, em um único passe na rede neural, gerando respostas mais rápidas.

## 4 RESULTADOS E DISCUSSÕES

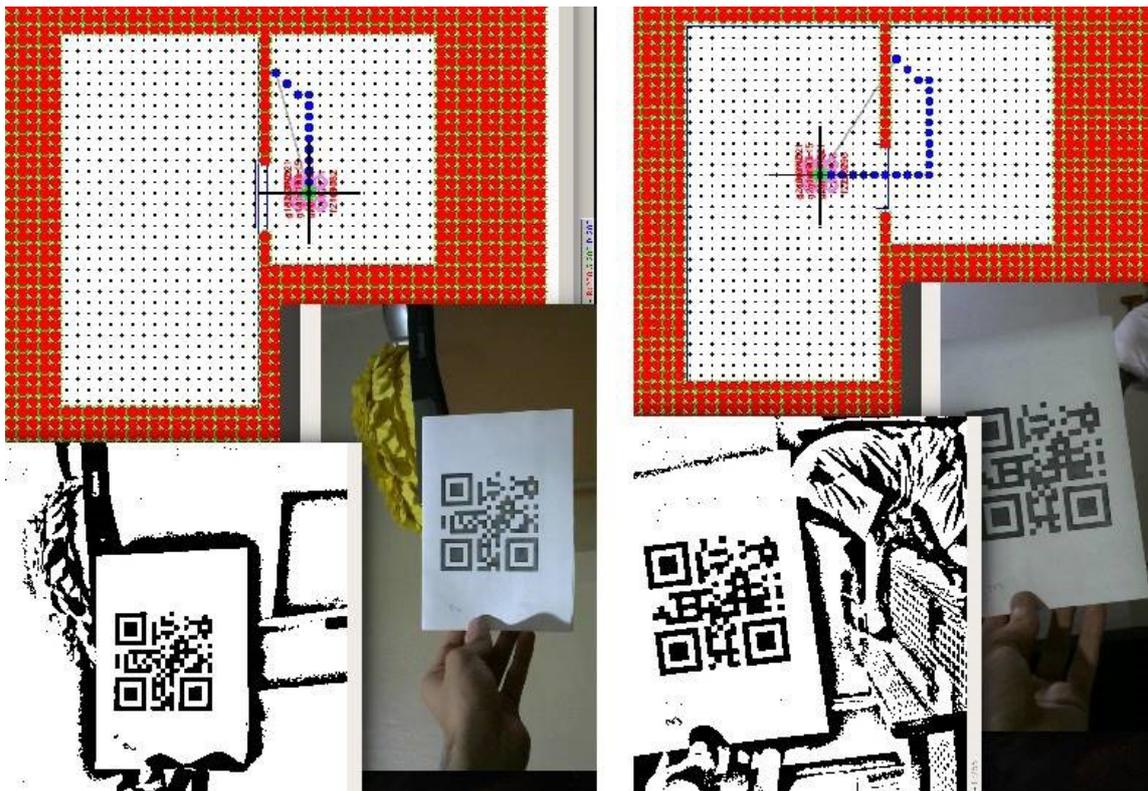
O algoritmo desenvolvido para o cálculo de rotas, demonstrado na figura 12, funcionou como o esperado em testes iniciais com plantas simples, de dois a três cômodos. Na figura 11, mostra-se em verde, no primeiro cômodo, o ponto de partida e em vermelho, no segundo cômodo, representa-se o ponto de chegada. A trajetória traçada em azul pelo software representa a rota que o usuário necessita seguir para atingir o objetivo.

Figura 11. Algoritmo Pathfinder.



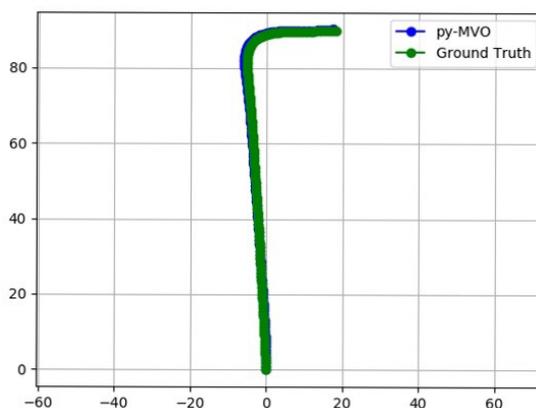
Obteve-se sucesso nos testes realizados com a localização via códigos bidimensionais. A cada novo código detectado a localização do dispositivo foi atualizada com êxito para o respectivo ponto, como se verifica no teste representado na figura 12.

Figura 12. Teste código bidimensional.



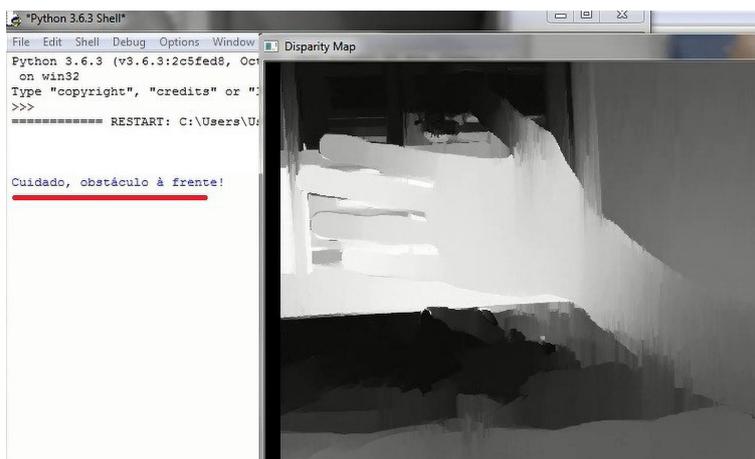
Porém, a odometria, assim como o esperado, revelou-se um método mais eficiente para localização, mostrando um controle constante e bem próximo da realidade. Um dos testes realizados utilizou como base de imagens o dataset KITTI (KITTI, 2020) para testes, o que permitiu comparar a trajetória estimada pelo software com a trajetória real. No gráfico da figura 13 pode-se observar, em azul, a trajetória calculada e, em verde, o caminho percorrido de fato. Em testes realizados com funcionamento em tempo real, detectaram-se problemas quando em situações de baixa luminosidade, por não conseguir obter pontos de rastreamento nas imagens, ocasionando em desvios de rota. Estes problemas estão diretamente relacionados com a qualidade da câmera utilizada. Uma câmera com um sensor mais sensível a luz pode reduzir notavelmente as falhas desta natureza.

Figura 13. Trajetória.



O dispositivo facilmente detecta obstáculos próximos via estereoscopia. A figura 14 apresenta o algoritmo emitindo um alerta ao detectar um corpo demasiadamente próximo do usuário.

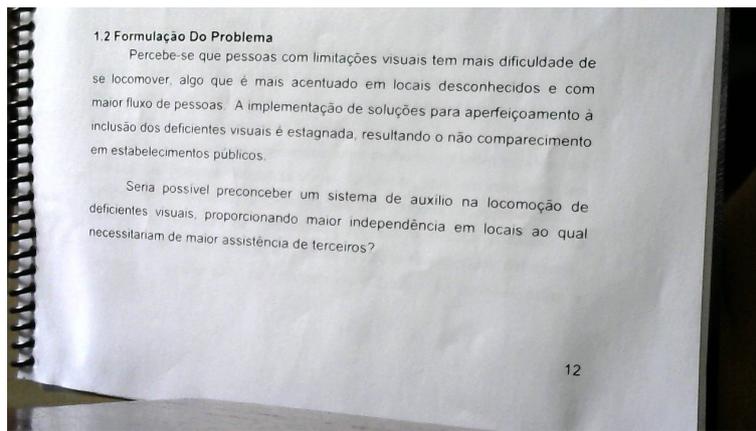
Figura 14. Detecção de obstáculos.



A estereoscopia possui algumas limitações que podem causar erros nas medições. Superfícies reflexivas, como espelhos, confundem o algoritmo, pois causam uma ilusão de ótica, mudando, dessa forma, a disparidade analisada.

Os testes realizados com o algoritmo de OCR consistiram na análise da taxa de acerto obtida por diferentes configurações de resolução e pré-processamento da imagem.

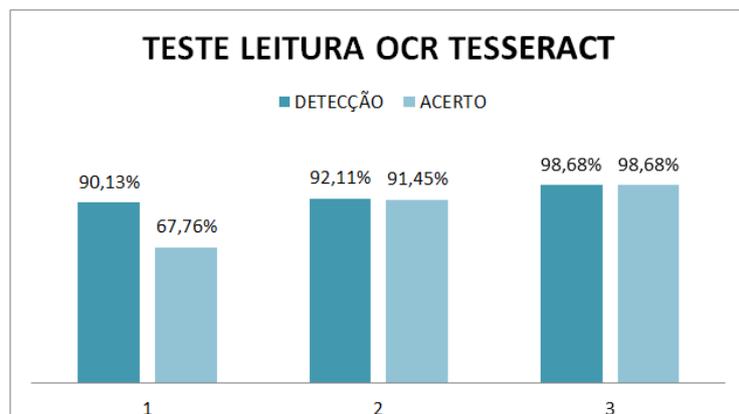
**Figura 15.** Imagem teste de OCR.



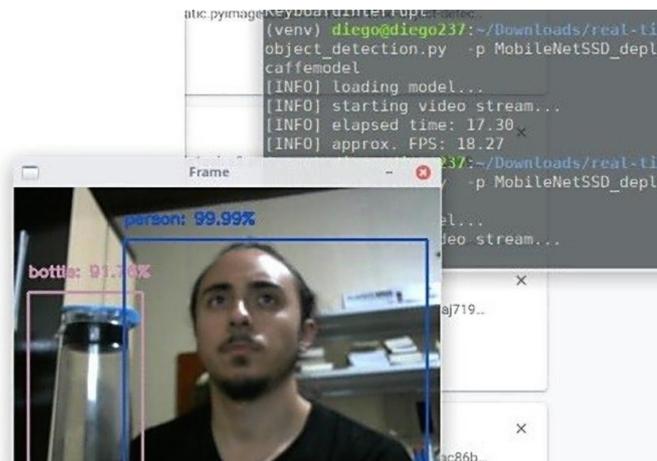
A figura 15 apresenta a imagem do texto utilizado para o teste, contendo um total de 76 palavras. Conforme os resultados evidenciados no gráfico da figura 16, atingiu-se o melhor desempenho no terceiro teste, onde se aplicou a resolução de 1280x780 e o pré-processamento da imagem, como alternativa para limpeza da poluição visual, o maior causador de erros na detecção.

**Figura 16.** Teste de leitura de OCR.

TESTE	DEVICE	RESOLUTION	PRÉ-PROCESSAMENTO	Nº PALAVRAS
1	C270	800x600	-	76
2	C270	1280x720	-	76
3	C270	1280x720	Warp+Crop+Normalization	76



**Figura 17.** Detecção de Objetos.



A detecção de objetos reconheceu múltiplos corpos em diferentes distâncias de forma rápida e eficiente em capacidade computacional. A figura 17 mostra um teste do algoritmo, que dentro da média observada conseguiu entregar a taxa de atualização de 18 quadros por segundo com o hardware utilizado no projeto.

## 5 CONCLUSÃO

O presente trabalho apresentou Isabel, um gadget com sistema de comunicação por comando e sintetização de voz, através de linguagem natural, capaz de atender solicitações e retornar informações como se de fato fosse uma pessoa e não uma máquina. A possibilidade e habilidade de se mover livremente possui grande impacto na vida das pessoas, as tecnologias devem ser utilizadas para facilitar a vivência dessas pessoas que apresentam deficiência visual. Pessoas com deficiência visual, ao serem capazes de se movimentar e se orientar no espaço de maneira mais eficiente, apresentam comportamentos de alta autoestima e se tornam mais dispostas a realizar mais atividades. A partir da integração das tecnologias de visão computacional, inteligência artificial, redes neurais e suas ferramentas e algoritmos, este trabalho pode cada vez mais com o avanço da tecnologia, proporcionar uma melhora gradativa na qualidade de vida das pessoas que possuem deficiência visual, que segundo o IBGE somam cerca de 3,5% da população brasileira.

Assim o trabalho alcançou efetivamente todos os objetivos inicialmente traçados. Resultados promissores revelaram tecnologias confiáveis e aplicáveis ao problema da pesquisa, corroborando a ideia de que este projeto se encaminha para uma solução tecnológica completa e, principalmente, viável para ampliar a necessitada independência dos deficientes visuais no futuro. Além disso, foi de suma importância para aquisição, consolidação e assimilação de conhecimentos adquiridos durante todo o curso técnico, bem como, foi determinante na escolha e inserção dos integrantes no curso de graduação na área da computação.

Dessa maneira, o desenvolvimento deste sistema é de grande utilidade prática. O projeto de pesquisa foi realizado com o objetivo de abranger a maior quantidade de tecnologias integradas para auxiliar o cotidiano do usuário. Isto torna o projeto uma alternativa viável e de grande utilidade.

Esperamos que este trabalho venha contribuir para o aumento do interesse em pesquisa e mais investimentos nas áreas da ciência, para que cada vez mais possamos resolver problemas através do desenvolvimento tecnológico.

## REFERÊNCIAS

BERNARDES, Júlio et al. **Sistema ajuda portador de deficiência visual a se locomover**. 2016. Disponível em: <<https://www.revistaespacios.com/a16v37n09/16370920.html>> Acesso em: out. 2018.

BRADSKI, Gary; KAEHLER, Adrian. **Learning OpenCV: Computer vision with the OpenCV library**. Editora O'Reilly Media, Inc. 2008.

CAO, Mingwei. **Py-MVO: Monocular Visual Odometry using Python**. Disponível em: <[https://github.com/caomw/visual\\_odometry-1](https://github.com/caomw/visual_odometry-1)> Acesso em fev. 2020.

COLDEWEY, Devin. **WTF is computer vision?** – Disponível em: <<https://techcrunch.com/2016/11/13/wtf-is-computer-vision/>> Acesso em mar. 2020.

CHATTERBOT; **About ChatterBot**. 2018. Disponível em: <<https://chatterbot.readthedocs.io/en/stable/index.html>> Acesso em: out. 2018

CORKE, P; STRELOW, D; SINGH, S. **Omnidirectional Visual Odometry for a Planetary Rover**. Pittsburgh: Carnegie Mellon University, 2004.

COX, Gunther. **ChatterBot Language Training Corpus**. 2020. Repositório GitHub. Disponível em: <<https://github.com/gunthercox/chatterbot-corpus>> Acesso em fev. 2020.

GADGET. **Dicionário online Oxford**, 09 fev. 2020. Disponível em <<https://www.oxfordlearnersdictionaries.com/definition/english/gadget>> Acesso em: fev. 2020.

GIZMODO. **Uma câmera que pode ser colocada em quaisquer óculos é uma baita solução para deficientes visuais**, jan. 2019. Disponível em: <<https://gizmodo.uol.com.br/orcam-myeye-2-ces-2019>> Acesso em: jan. 2019

GSTT. **Cloud Speech-To-Text**. Disponível em: <<https://cloud.google.com/speech-to-text>> Acesso em: fev. 2020.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Características gerais da população, religião e pessoas com deficiência. Disponível em: <[https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd\\_2010\\_religiao\\_deficiencia.pdf](https://biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd_2010_religiao_deficiencia.pdf)> Acesso em fev. 2020.

KITTI. Karlsruhe Institute of Technology and Toyota institute of Technology in Chicago. **The KITTI Vision Benchmark Suite**. Disponível em: <<http://www.cvlibs.net/datasets/kitti/>> Acesso em: fev. 2020.

LEONHARDT, M. D. **Doroty**: um Chatterbot para Treinamento de Profissionais Atuantes no Gerenciamento de Redes de Computadores. Dissertação (Mestrado em Ciência da Computação) Pós-Graduação em Computação do Instituto de Informática da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005. Disponível em: <<https://lume.ufrgs.br/handle/10183/5659>> Acesso em: out. 2018.

LIU, W. et al. **SSD: Single shot multibox detector**. arXiv 2016. Disponível em: <<https://arxiv.org/abs/1512.02325>> Acesso em: fev. 2020.

MCCARTHY, J. **What is artificial intelligence?** - Stanford University. Computer Science Department, nov./2007. Disponível em: <<http://www-formal.stanford.edu/jmc/whatisai/>> Acesso em: fev. 2020.

NISTÉR, David; NARODITSKY, Oleg; BERGEN, James. "Visual odometry". **Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 2004. CVPR 2004. Ieee, 2004. p. I-I. Disponível em: <<https://ieeexplore.ieee.org/document/1315094/references#references>>. Acesso em: out. 2018.

OCR. **Dicionário online Oxford**, 09 fev. 2020. Disponível em <<https://www.oxfordlearnersdictionaries.com/definition/english/ocr>> Acesso em: fev. 2020.

OPENCV. **Object Tracking**: 2017a. Disponível em: <[https://docs.opencv.org/3.3.1/dc/d6b/group\\_\\_video\\_\\_track.html#ga473e4b886d0bcc6b65831eb88ed93323](https://docs.opencv.org/3.3.1/dc/d6b/group__video__track.html#ga473e4b886d0bcc6b65831eb88ed93323)> Acesso em fev. 2018.

OPENCV. **Optical Flow**. 2017b. Disponível em <[https://docs.opencv.org/3.3.1/d7/d8b/tutorial\\_py\\_lucas\\_kanade.html](https://docs.opencv.org/3.3.1/d7/d8b/tutorial_py_lucas_kanade.html)> Acesso em fev. 2018.

ORCAM. **OrCam MyEye 2** - uma tecnologia de visão revolucionária. 2017. Disponível em: <<https://www.orcam.com/pt/myeye2/>> Acesso em: fev. 2018.

PIRES, João. et. al. **Sistema GPS, 2013/2014**. Projeto FEUP. Porto: Universidade do Porto. Disponível em: <[https://paginas.fe.up.pt/~projfeup/bestof/13\\_14/index.html](https://paginas.fe.up.pt/~projfeup/bestof/13_14/index.html)> Acesso em: fev. 2018

RIBEIRO, Marcos. **O que é API e como ela aumenta a produtividade nas empresas.** Disponível: <<https://pluga.co/blog/api/o-que-e-api/>> Acesso em: mar. 2020. ANDLER, Mark et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: **Proceedings of the IEEE conference on computer vision and pattern recognition.** 2018. p. 4510-4520.

SCARAMUZZA, Davide; FRAUNDORFER, Friedrich. **Visual odometry [tutorial].** IEEE robotics & automation magazine, v. 18, n. 4, p. 80-92, 2011. Disponível em: <[https://www.ifi.uzh.ch/dam/jcr:5759a719-55db-4930-8051-4cc534f812b1/VO\\_Part\\_I\\_Scaramuzza.pdf](https://www.ifi.uzh.ch/dam/jcr:5759a719-55db-4930-8051-4cc534f812b1/VO_Part_I_Scaramuzza.pdf)> Acesso em fev. 2020.

SINGH, Avi. **Visual Odometry from scratch**, 2015. Disponível em: <<https://avisingh599.github.io/vision/visual-odometry-full/>> Acesso em: mar. 2019.

KILNER, Claudio; TORI, Romero (Eds.). **Realidade Virtual: conceitos e tendências.** São Paulo: Editora Mania de Livro, 2004.

SMITH, Ray. An overview of the Tesseract OCR engine. In: **Ninth International Conference on Document Analysis and Recognition (ICDAR 2007).** IEEE, 2007. p. 629-633.

SORIA, E; MARTIN, J; MAGDALENA, R; MARTINEZ, M; SERRANO, A. **Handbook of Research on Machine Learning Application.** IGI GLOBAL, 2009.

STIVANELLO, Mauricio. **Desenvolvimento de uma biblioteca para sistemas de visão estereoscópica para robótica móvel.** Dissertação (Mestrado em Engenharia Elétrica) Programa de Pós-Graduação em Engenharia Elétrica. 2008. Disponível em: <<http://repositorio.ufsc.br/xmlui/handle/123456789/92029>> Acesso em mar. 2019.

TESSERACT. **Tesseract OCR.** Disponível em <<https://github.com/tesseract-ocr/tesseract>> Acesso em fev. 2020.

VAN DER WALT, Stefan et al. Scikit-image: image processing in Python. **PeerJ**, v. 2, p. e453, 2014.

WEST, J; VENTURA, D; WARNICK, S. **Spring Research Presentation: A Theoretical Foundation for Inductive Transfer.** Brigham Young University, College of Physical and Mathematical Sciences, 2007.

WEWALK. WeWALK Smart Cane. Disponível em: <<https://wewalk.io>> Acesso em: fev. 2020.

ZAFFARI, Felipe Pozueco; ESPÍNDOLA, Jean Carlo de Borba. Conceitos – O que é a inteligência artificial. In: BARONE, Dante Augusto Couto; BOESING, Ivan Jorge. **Inteligência Artificial: diálogo entre mentes e máquinas.** Porto Alegre: AGE / Evangraf, 2015, p.119.

**Diego Antonio Amancio Duarte** cursou Técnico em Eletromecânica Integrado ao Ensino Médio na Escola Técnica Estadual Frederico Guilherme Schmidt (2014 - 2019). Atualmente, cursa Engenharia da computação na Universidade do Vale do Rio dos Sinos (UNISINOS).

**Gabriel Tamujo Meyrer** cursou Técnico em Eletromecânica Integrado ao Ensino Médio na Escola Técnica Estadual Frederico Guilherme Schmidt (2014 - 2019). Atualmente, cursa Ciências da Computação na Universidade do Vale do Rio dos Sinos (UNISINOS).

**Thaís De Oliveira Tonel** cursou Técnico em Eletromecânica Integrado ao Ensino Médio na Escola Técnica Estadual Frederico Guilherme Schmidt (2014 - 2019). Atualmente, cursa Engenharia Mecânica na Universidade Do Vale Do Rio Dos Sinos (UNISINOS).

**Marcos Freire Machado** possui Especialização em Engenharia Biomédica com ênfase em Engenharia Clínica pela Universidade Estácio de Sá (2020); Especialização em Ensino de Sociologia pela Universidade Federal do Rio Grande - FURG (2018); Graduação em História pela Universidade La Salle (2012). Atualmente é Técnico de Laboratório/Área Industrial da Faculdade de Odontologia da Universidade Federal do Rio Grande do Sul (UFRGS) onde atua como Coordenador do Núcleo de Infraestrutura. Atua como Professor do Ensino Médio Integrado ao Técnico em Eletromecânica, da Escola Técnica Estadual Frederico Guilherme Schmidt. Tem experiência na área de História e Eletromecânica, bem como na orientação de projetos de iniciação científica.

Essa pesquisa foi desenvolvida na Escola Técnica Estadual Frederico Guilherme Schmidt.

**Contribuição de autoria.** Diego Antonio Amancio Duarte, Gabriel Tamujo Meyrer, Thaís De Oliveira Tonel: concepção, coleta e análise de dados, elaboração, revisão e redação do manuscrito. Marcos Freire Machado: orientação, redação e revisão do manuscrito.



**COMO CITAR ESSE ARTIGO (ABNT NBR 60230):**

DUARTE, D. A. A.; MEYRER, G. T.; TONEL, T. O.; MACHADO, M. F. Sistema de auxílio na locomoção de deficientes visuais. **Scientia Prima**, São Leopoldo, v. 6, n. 1, p. 98-116, maio 2020.