



# MakeMeSee – Uma solução para melhorar a qualidade de vida de deficientes visuais por meio de visão computacional e inteligência artificial

Henrique R. Ricci<sup>1</sup>, Mário A. Pazoti<sup>2</sup>

1. Discente do Curso de Ciência da Computação – Faculdade de Informática de Presidente Prudente
  2. Docente – Faculdade de Informática de Presidente Prudente.
- E-mails: henrique.ricci@unoeste.edu.br, mario@unoeste.br

## Introdução

Problemas visuais atingem cerca de 1,3 bilhões de pessoas ao redor do mundo, levando algumas delas a desenvolverem problemas graves ou até mesmo a cegueira total. Pessoas com deficiência visual podem ter uma certa dificuldade de perceber o mundo, uma vez que tem que se adaptar aos outros sentidos. Tendo em vista isso, este trabalho apresenta uma solução composta por um hardware com duas câmeras e um software, em que o objetivo é reconhecer possíveis obstáculos, calcular suas distâncias e informar ao usuário, para que ele consiga desviar. A solução também realiza o reconhecimento facial de pessoas por meio de inteligência artificial.

## Metodologia

Para chegar a solução final, o projeto foi dividido em duas partes, sendo a primeira relacionada a Visão Computacional (VC) e a segunda relacionada a Inteligência Artificial (IA).

Em relação a VC, o principal conceito abordado nesse trabalho foi a estereoscopia. Para que pudessem ser realizados os cálculos de distâncias dos obstáculos, foi necessário utilizar duas câmeras estéreo, posicionadas a uma distância fixa uma da outra e fixadas em um suporte feito em madeira de forma paralela. Isso foi pensado para simular o processo de visão humana e ter a noção de profundidade da cena.

Por meio de busca de pontos correlatos e geometria epipolar, a partir de duas imagens estéreas é possível gerar um mapa de profundidade, que posteriormente pode ser convertido para distâncias reais por meio da medição real das distâncias e relacionamento com a cor do mapa (o mapa gerado está em tons de cinza).

Com relação a IA, foram divididas em três partes: reconhecimento dos obstáculos, reconhecimento de faces e processamento de linguagem natural. O reconhecimento de obstáculos foi implementado utilizando arquiteturas *Deep Learning*, por serem extremamente eficazes no reconhecimento de várias classes diferentes. Para o treinamento, foi utilizado uma técnica chamada *Fine-Tuning*, em que o modelo já está pré treinado porém realiza-se um breve treinamento de parte da rede com imagens próprias de alguma classe que seja diferente das originais. A base utilizada foi a *MS Coco*, excluindo algumas classes irrelevantes para o projeto como avião.

Para o reconhecimento facial, foi utilizado o algoritmo HOG para detecção das faces. Apesar de ser um pouco mais lento que outros algoritmos do mesmo segmento, ele apresenta resultados bem confiáveis. Após detectado um rosto, é realizado um processo de extração de características faciais. É por meio dessas características que novas faces podem ser classificadas. Para encontrar uma nova face, é utilizado o algoritmo KNN, sendo que se a menor distância estiver acima de um limiar estabelecido, a nova face é então classificada. Não foi utilizado um modelo de *Deep Learning* para o reconhecimento facial pois como a aplicação deve rodar em dispositivos móveis, fazer o treinamento completo da rede para cada nova face seria algo inviável. Com o KNN, o tempo de treinamento é reduzido, apesar de que quanto mais dados se tem treinados, mais lento é o processo de classificação. Por fim, o processamento de linguagem natural é utilizado para que o usuário se comunique com o sistema. Para que ele possa, por exemplo, salvar uma nova face para ser reconhecida, basta que ele diga “Salvar ‘Nome da pessoa’” que o sistema realiza o processamento dos sinais, converte a voz em texto e executa os comandos pré definidos.

## Resultados

Nos testes realizados até o momento foi possível notar um alto desempenho e precisão no reconhecimento e cálculo de distância dos obstáculos. Os testes foram realizados em um smartphone com 6 GB de memória RAM e um *Snapdragon 855*.

Com relação ao cálculo de distâncias, o sistema consegue calcular distâncias que estejam entre dois e onze metros. Existe uma margem de erro que pode chegar a 30 cm, mas que pode facilmente ser corrigido com o aumento do número de amostras reais utilizadas para a geração da função de cálculo.

Já em relação ao reconhecimento de obstáculos, o tempo de resposta é de aproximadamente 60 ms, desde a captação das imagens, detecção do obstáculo e cálculo da distância. A precisão da detecção está em torno de 96% e ainda não foram realizados testes no desvio desses obstáculos.

O reconhecimento facial precisa de 1000 a 2000 ms para conseguir efetivar o reconhecimento. Isso porque as características das novas faces de entrada necessitam ser comparadas a todas as características extraídas durante a fase de treinamento. Essa abordagem foi utilizada pois, um modelo de *Deep Learning* convencional necessitaria de um novo treinamento a cada nova pessoa que o usuário desejasse cadastrar, se tornando inviável de ser executado em um dispositivo móvel.

Na imagem 2 é possível ver o protótipo físico desenvolvido.

Imagem 2. Protótipo desenvolvido.



Fonte: Elaborada pelo autor.

## Conclusões

Visto o que foi apresentado, nota-se que a proposta já pode ser utilizada para algumas tarefas mais simples, pois seus resultados já são bem satisfatórios. Uma questão importante a ser analisada e, posteriormente implementada, é a navegação do usuário em ambientes externos. Além de saber o que existe à sua frente e a que distância está, ele também poderá saber para onde ir e assim melhorar sua qualidade de uso do sistema.

## Bibliografia

- CASTRO, M. D. **Alvisku: uso da visão computacional e sons 3D para auxílio a cegos**. 2018.
- GIANINI, S. et al. **JUVO – An Aid for the Visual Impaired**. 2018 International Conference on Smart City and Emerging Technology (ICSCET).
- NETO, L. S. B. **Sistemas wearable baseados em métodos de visão computacional para auxiliar pessoas com deficiência visual**. 2016. 1 recurso online (92 p.), Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Computação, Campinas, SP. Disponível em <<http://www.repositorio.unicamp.br/handle/REPOSIP/320884>>, acesso em 28/09/2020 às 21h
- POGGI, M.; MATTOCCIA, S. **A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning**. 2016 IEEE Symposium on Computers and Communication (ISCC).