



CINTERGEO

Congresso Internacional de Educação
e Geotecnologias

IV Congresso Internacional de Educação
e Geotecnologias

IX Encontro de Pesquisadores da Rádio

27 e 28 de Julho de 2023



APRENDIZADO DE MÁQUINA NO CAMPO DA VISÃO COMPUTACIONAL PARA USO NA ROBÓTICA INTELIGENTE

Davi Miguel Brito Barbosa¹

Tamir de Almeida Amorim²

Ana Patrícia Fontes Magalhães Mascarenhas³

Área Temática – Universidade pública: conhecimento científico e (geo)tecnológico
Agência Financiadora: não contou com financiamento

Resumo

Os robôs de serviço possuem uma ampla variedade de possibilidades de atuação, como em atividades cotidianas domésticas, cuidado de idosos, recepção e em atendimentos. Para essas tarefas é necessário que o robô consiga interagir com o meio externo se comunicando, por exemplo através da fala, assim como identificar pessoas e objetos ao seu redor e se locomover no ambiente. Este artigo apresenta o trabalho realizado pelos alunos de Iniciação Científica do Centro de Pesquisa em Arquitetura de Computadores, Sistemas Inteligentes e Robótica (ACSO) na Universidade do Estado da Bahia (UNEB) cujo objetivo geral é implementar a visão computacional para o robô de serviço autônomo chamado BILL (Robô Inteligente de Grande capacidade e Baixo custo). Como objetivo específico, o trabalho deve implementar, treinar e validar as funções de reconhecimento de pessoas e reconhecimento de objetos para BILL, possibilitando que este realize tarefas domésticas como, por exemplo, retirar o lixo de uma residência. A metodologia utilizada no desenvolvimento do trabalho se iniciou com uma pesquisa para seleção do algoritmo de visão computacional a ser adotado. Em seguida foi realizado o treinamento do algoritmo considerando as demandas de BILL. Atualmente estão sendo executados testes para preparar BILL para a Competição Brasileira de Robótica (CBR). Este artigo descreve as tecnologias utilizadas na implementação e detalha os resultados dos obtidos ao longo do treinamento e da validação. Os resultados atuais evidenciam uma boa acurácia do robô tanto no reconhecimento de objetos, acima de 80%, quanto de pessoas, acima de 75%. Novos treinamentos estão em andamento com um maior número de imagens, para melhorar ainda mais estes resultados.

Palavras-chave: BILL. Robótica de serviço. Machine learning. Reconhecimento de pessoas. Reconhecimento de objetos.

Introdução

Os robôs de serviço são construídos para realizar tarefas rotineiras ou que oferecem risco aos humanos, a exemplo de limpar a casa e resgatar pessoas em escombros. Para isso, o robô precisa interagir com o usuário, ver e reconhecer pessoas e objetos e se mover no ambiente.

¹Universidade do Estado da Bahia; Graduando; devildbar@gmail.com.

²Universidade do Estado da Bahia; Graduando; amorimlds@gmail.com

³Universidade do Estado da Bahia; Doutora em Ciência da Computação; apmagalhaes@uneb.br

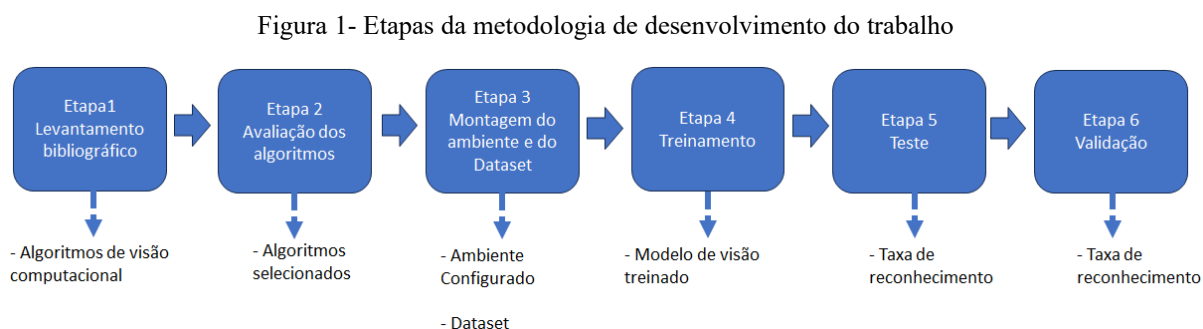
A visão computacional [1], tecnologia que possibilita que as máquinas enxerguem o mundo a sua volta, juntamente com o aprendizado de máquina [2], tecnologia utilizada para treinar robôs em conhecimentos específicos, são essenciais para cumprir essas tarefas, pois permitem que o robô identifique pessoas e objetos e realize ações de acordo com o que foi identificado.

O reconhecimento de pessoas é um sub-ramo da visão computacional que realiza reconhecimento facial através das características faciais das pessoas. Consiste em analisar uma característica padrão facial e usar algoritmos para identificar uma pessoa específica. O reconhecimento de objetos, por sua vez, utiliza redes neurais [3], algoritmos de aprendizagem de máquina, para treinar o robô e permitir que ele reconheça e classifique um objeto.

O Centro de Pesquisa em Arquitetura de Computadores, Sistemas Inteligentes e Robótica (ACSO) da Universidade do Estado da Bahia (UNEB) desenvolveu o Robô Inteligente de Grande Capacidade e Baixo custo (BILL). BILL precisa realizar atividades em uma casa, por exemplo, pegar um refrigerante na geladeira. Para isso, deve reconhecer pessoas e objetos à sua volta. Este artigo apresenta a pesquisa realizada no ACSO para desenvolver o reconhecimento de pessoas e objetos para BILL. As próximas seções descrevem a metodologia utilizada para realizar o trabalho, os resultados e discussões e as considerações finais.

Metodologia

Este trabalho utiliza uma metodologia de pesquisa aplicada e quantitativa. Nesta direção, aplica conhecimento de visão computacional ao domínio de robô de serviço e avalia a taxa de assertividade no contexto das provas da Competição Brasileira de Robótica (CBR). Nesta direção a pesquisa compreendeu cinco etapas (Figura 1).



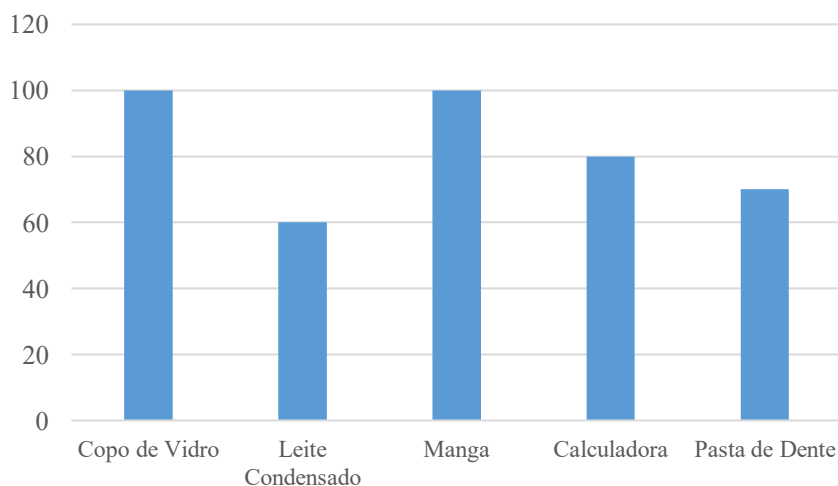
Fonte: os autores (2023)

Conforme ilustrado na Figura 1, o trabalho se iniciou com uma pesquisa bibliográfica sobre algoritmos de visão computacional. Como resultado foram identificados um conjunto de algoritmos aplicados à visão computacional, especificamente ao reconhecimento de objetos e ao reconhecimento de pessoas. Na segunda etapa, os algoritmos foram avaliados considerando a taxa de assertividade do reconhecimento, custo computacional (CPU, memória e placa de vídeo necessária) e tempo de treinamento. Como resultado foram escolhidos os algoritmos a serem utilizados. A partir disso, o ambiente foi instalado, configurado e definido o *dataset* (conjunto de imagens) a ser utilizado no treinamento (Etapa 3). Iniciou-se em seguida o treinamento do robô (Etapa 4). Para isso diversos parâmetros relacionados a brilho, cor, angulação, entre outros aspectos foram aplicados. Como resultado o modelo de visão treinado foi gerado. A etapa 5, em execução atualmente, consiste na realização de testes. Espera-se ter uma taxa de assertividade no reconhecimento de pelo menos 80%. Finalmente, na Etapa 6 o trabalho será validado com a participação do robô Bill nas provas da CBR, onde se espera um reconhecimento superior a meta de 80% para o qual o robô foi treinado. Adicionalmente, ao longo da pesquisa os resultados serão publicados em conferências.

Resultados e discussões

A detecção de objetos em BILL utiliza o algoritmo Yolo [4] versão 5. Trata-se de um algoritmo de rede neural convolucional [4], um tipo de rede que foi desenvolvida para trabalho com imagens, que antes de passar os dados para a rede neural densa faz um pré-processamento para filtrar características das imagens que indicam a existência de um determinado objeto. O treinamento de BILL foi realizado considerando cinco objetos e dez fotos para cada objeto, usando diferentes locais e posições. Este número foi definido para realizar um teste pequeno, mas diversificado. Usamos diferentes quantidades de objetos em cada imagem, sempre visando analisar o impacto dos parâmetros e o número de fotos passadas no treinamento de reconhecimento. A Figura 2 apresenta a porcentagem de acertos dos objetos detectados. Falsos positivos também foram analisados durante os testes, quando o algoritmo não apresenta o resultado esperado. O eixo x apresenta os objetos treinados e o eixo y o percentual de acerto após o treinamento. O menor resultado foi o do objeto “leite condensado” com 60% de acerto. Todos os demais tiveram pelo menos 70% de acerto, um resultado satisfatório.

Figura 2 - Percentual de acerto para cada objeto treinado



Fonte: os autores (2023)

O reconhecimento de pessoas utiliza uma rede neural convolucional (CNN) chamada VGG16-Mixed-Finetuning [5], que já vem pré-treinada por um banco de dados de imagens chamado ImageNet [6]. Para validar o treinamento utilizamos um banco de imagens que consiste em 19.370 fotos; no entanto, para facilitar os testes, selecionamos aleatoriamente apenas 200 fotos com diferentes pessoas, uma pessoa por foto, variando em diferentes faixas etárias de 0-2 anos até 60-100 anos. Além disso, as fotos não são controladas, possuindo diferentes tamanhos, ângulos, distâncias, e as pessoas usam acessórios, como óculos, chapéus e maquiagem. O experimento consistiu em submeter essas fotos ao algoritmo apenas uma vez, pois os resultados são sempre os mesmos, e verificar se ele reconhece corretamente a pessoa para a classificação do grupo etário. Observa-se um aumento de 23% nas respostas corretas ao variar os valores de 0 a 64 pixels. Atualmente estão sendo realizados mais testes utilizando valores maiores de preenchimento para avaliar se a taxa de acerto melhora, piora ou se mantém sem alterações significativas.

Considerações finais

Este artigo apresentou a implementação da visão computacional para o robô BILL. Os resultados dos testes realizados até o momento mostram que o robô BILL consegue reconhecer pessoas e objetos com taxa de acerto superior a 80% no reconhecimento de objetos e acima de 75% no reconhecimento de pessoas. Atualmente estamos realizando novos treinamentos para aumentar a quantidade de objetos personalizados e de pessoas que podem ser reconhecidos pela

máquina. Os alunos que participam desta pesquisa são bolsistas de Iniciação Científica do CNPq de 2022.

REFERÊNCIAS

- [1] BARELLI, Felipe. **Visão Computacional: uma Abordagem Prática com Python e OpenCV**. Casa do Código, 2018.
- [2] WICKHAN, Mark. **Practical Java Machine Learning**. Springer Nature B.V., 2018.
- [3] REDMON, Joseph; FARHADI, Ali. **Yolo900: Better, Faster, Stronger**. Disponível em: https://openaccess.thecvf.com/content_cvpr_2017/papers/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.pdf, acessado: 2023-06-27.
- [4] REDMON, Joseph. **YOLO: Real time object detection**. Disponível em: <https://pjreddie.com/darknet/yolo/>, acessado 2023-06-25.
- [5] [KHANDELWAL](#), Vaibhav. **The architecture and implementation of vgg16**. Disponível em: <https://pub.towardsai.net/the-architecture-and-implementation-of-vgg-16-b050e5a5920b>, acessado em: 2023-06-27
- [6] STANFORD VISION LAB. ImageNet. **Base de Dados de Imagens para propósitos de Visão Computacional**. Disponível em <http://www.image-net.org/>, acessado 2023-06-27.