

# Follow Me

Alexandre Costa Ferro Filho  
*Instituto de Informática*  
*Universidade Federal de Goiás*  
Goiânia, Brasil  
alexandre\_ferro@discente.ufg.br

Iago Alves Brito  
*Instituto de Informática*  
*Universidade Federal de Goiás*  
Goiânia, Brasil  
iagoalves@discente.ufg.br

Elisa Ayumi Masasi de Oliveira  
*Instituto de Informática*  
*Universidade Federal de Goiás*  
Goiânia, Brasil  
ayumi@discente.ufg.br

Murilo de Oliveira Guimarães  
*Instituto de Informática*  
*Universidade Federal de Goiás*  
Goiânia, Brasil  
muriloguimaraes@discente.ufg.br

**Abstract**—Este artigo apresenta uma abordagem abrangente para o desenvolvimento de um sistema robótico "Follow Me", que integra tecnologias de visão computacional, robótica e processamento de áudio e voz. O objetivo principal é permitir que o robô siga autonomamente uma pessoa designada em um ambiente dinâmico. Utilizando técnicas avançadas de visão computacional, o sistema identifica e rastreia o indivíduo alvo com base em pistas visuais. Simultaneamente, algoritmos de processamento de áudio e voz aprimoram a capacidade do robô de distinguir comandos e orientação do alvo em ambientes ruidosos. A integração dessas modalidades garante um rastreamento robusto e confiável, superando desafios impostos por oclusões, variações nas condições de iluminação e ruído ambiente. Os resultados experimentais demonstram a eficácia do sistema proposto, destacando seu potencial em aplicações de assistência pessoal, segurança e entretenimento interativo.

**Index Terms**—Computer Vision, Robotics, Audio and Voice Processing.

## I. INTRODUÇÃO

A task "Follow Me" em robótica refere-se ao desenvolvimento de sistemas autônomos capazes de seguir um usuário através de diferentes ambientes. Essa funcionalidade é especialmente útil em contextos domésticos ou de assistência, onde o robô pode acompanhar uma pessoa para prestar ajuda ou companhia. Implementar esta capacidade exige uma combinação avançada de visão computacional, processamento de áudio e algoritmos de navegação espacial que permitem ao robô detectar, identificar e seguir o usuário de forma contínua e segura. A task "Follow me" consiste em uma série de desafios que o robô deve seguir, sendo primeiro necessário que ele adentre um ambiente até determinado ponto, após isso ele deve retornar ao início sendo necessário desviar de objetos novos introduzidos no caminho, e por último ele deve seguir o usuário conforme o uso da WakeWord "Follow Me". A relevância desta task está na sua aplicabilidade direta ao melhoramento da qualidade de vida das pessoas, proporcionando suporte em atividades diárias e aumentando a autonomia de indivíduos com necessidades especiais.

No contexto brasileiro, a pesquisa e desenvolvimento em robótica têm ganhado espaço com iniciativas acadêmicas e colaborações tecnológicas. O Núcleo de Robótica do Pequeno Mecânico, da Universidade Federal de Goiás (UFG), é um exemplo proeminente dessa tendência. Especializado em engenharia robótica, o núcleo se dedica a adaptar e criar soluções que atendam às necessidades locais, abrangendo desde aplicações industriais até assistenciais. A solução para a task "Follow Me", desenvolvida por este projeto, visa complementar e expandir o projeto com aplicações adicionais.

## II. METODOLOGIA

### A. Navegação Autônoma

A navegação autônoma é uma componente essencial para o sucesso da task "Follow Me" em robótica, representando um dos maiores desafios técnicos do projeto. Na metodologia adotada, esta capacidade é alcançada através do uso de algoritmos avançados de localização e mapeamento, AMCL e GMapping, respectivamente. Para a coleta de dados, o projeto faz o uso do sensor LIDAR 2D, agregados ao robô TurtleBot 4. Esses equipamentos são fundamentais para a detecção de obstáculos, estimação de posição integrados a utilização do framework Nav2 para o planejamento de trajetórias seguras e eficientes. A implementação eficaz da navegação autônoma não só permite que o robô siga o usuário de maneira precisa, mas também adapta-se a mudanças no ambiente, garantindo uma interação segura e contínua entre o robô e o humano. Esta abordagem metodológica assegura que o robô possa operar de forma independente em diversos cenários domésticos, tornando-o uma ferramenta valiosa e versátil para assistência no lar.

### B. Identificação e Rastreamento de Pessoas

A identificação e rastreamento de pessoas é realizada por meio de técnicas avançadas de visão computacional. Câmeras RGB-D são utilizadas para capturar imagens em tempo real, das quais são extraídas características específicas da pessoa

a ser seguida. Para a detecção de pessoas, identificação de presença e posição utilizou-se o modelo YOLOv8, atingindo boa performance e resultado. Para a otimização e utilização em tempo real em GPUs desse algoritmo foi utilizado o framework TensorRT, que garante a baixa latência em sua inferência. Uma vez identificada, a pessoa é rastreada usando o algoritmo ByteTrack, garantindo que o robô possa acompanhar movimentos rápidos e mudanças de direção.

### C. Wake Word

A identificação de wakeword é essencial para que o robô possa reconhecer comandos verbais específicos da pessoa que deve seguir. O framework utilizado para realizar a detecção de wake word foi o OpenWakeWord, o que possibilitou detectar e identificar wakewords em tempo real. Técnicas de processamento de áudio são incorporadas para melhorar a qualidade do áudio capturado, permitindo que o robô seja capaz de reconhecer comandos em ambientes ruidosos. O sistema é treinado para responder a palavras-chave específicas, que atuam como gatilhos para ações subsequentes do robô, como iniciar ou parar o seguimento. A integração desta funcionalidade permite uma interação mais natural e intuitiva entre o usuário e o robô, aumentando a usabilidade e a eficácia do sistema "follow me" proposto.

## III. RESULTADOS

Para o escopo da aplicação, o sucesso da solução é verificado por meio do bom desempenho apresentado em inferência, que podem ser observados na referência [8] em anexo.

## IV. FUNDAMENTOS TEÓRICOS

Nesta seção, exploramos os fundamentos teóricos que sustentam a implementação da task "Follow Me" em robótica, cobrindo aspectos cruciais de visão computacional, navegação autônoma, processamento de áudio e avaliação de desempenho. Inicialmente, discutimos as técnicas avançadas de detecção e rastreamento de pessoas utilizando modelos como YOLOv8, TensorRT e ByteTrack. Em seguida, detalhamos os algoritmos de localização, mapeamento e navegação, como AMCL e GMapping, e o papel dos sensores LIDAR 2D e câmeras RGB-D na navegação autônoma. Além disso, abordamos o processamento de áudio e a detecção de wake words com o OpenWakeWord, essenciais para a interação verbal entre o robô e o usuário. Por fim, apresentamos o bom desempenho apresentado em inferência [8].

### A. Robótica

1) *Algoritmos de Localização e Mapeamento:* A navegação autônoma permite ao robô mover-se de forma independente e segura ao seguir uma pessoa, habilidade essencial para a task. Dois algoritmos principais são utilizados para localização e mapeamento: o Nav2 (Navigation 2) e o GMapping. O Nav2 é uma evolução dos sistemas de navegação autônoma, oferecendo uma arquitetura modular que facilita a implementação em diferentes plataformas robóticas. Ele utiliza técnicas de planejamento de caminho, controle de movimento e gerenciamento de custo de mapa para assegurar que o robô possa

navegar de forma eficiente. O GMapping, por sua vez, é um algoritmo de SLAM (Simultaneous Localization and Mapping) que cria mapas incrementais do ambiente enquanto rastreia a posição do robô. A integração do GMapping com o Nav2 permite ao robô adaptar-se a novos ambientes e atualizar seu mapa continuamente, garantindo uma navegação robusta.

A coleta de dados para navegação é realizada por meio de sensores, principalmente o LIDAR 2D e câmeras. O LIDAR 2D emite feixes de laser para medir distâncias até objetos ao redor, criando um mapa bidimensional do ambiente. Suas vantagens incluem alta precisão e capacidade de operar em diferentes condições de iluminação, embora tenha limitações em termos de alcance e resolução em ambientes complexos. As câmeras, especialmente as RGB-D, fornecem dados visuais e de profundidade que complementam as informações do LIDAR, ajudando na detecção de obstáculos e na estimativa de posição.

### B. Processamento de Áudio e Voz

1) *Automatic Speech Recognition:* O Reconhecimento Automático de Fala (ASR) é crucial para a task "Follow Me", permitindo que o robô responda a comandos verbais e se ajuste aos ambientes em que opera. As técnicas de captura são projetadas para garantir que os microfones do robô possam captar as instruções do usuário. Passando dessa maneira a captura do áudio em processamento contínuo para o robô.

2) *Detecção de WakeWord:* Para a detecção de wake words, o sistema utiliza o OpenWakeWord, uma plataforma de código aberto projetada para identificar palavras-chave específicas que acionam ações subsequentes do robô. O treinamento do OpenWakeWord envolve a coleta e análise de amostras de áudio para configurar o sistema a reconhecer comandos específicos com alta precisão. Uma vez treinado, o sistema é integrado ao robô, permitindo que ele responda imediatamente quando a wake word é detectada. Dessa forma, aumentando usabilidade e a segurança do sistema "Follow Me".

### C. Visão Computacional

1) *Detecção de Objetos:* Uma das técnicas centrais utilizadas é a detecção de objetos com o modelo YOLOv8 (You Only Look Once, versão 8). A série YOLO é conhecida pela sua capacidade de realizar detecções rápidas e precisas em tempo real. O YOLOv8, a versão mais recente, introduz melhorias significativas em termos de arquitetura e desempenho, utilizando uma Rede Neural Convolucional (CNN) que divide a imagem em regiões e prevê bounding boxes e probabilidades para cada classe de objeto. Essa abordagem permite ao robô detectar rapidamente a presença de pessoas em seu campo de visão, crucial para iniciar o processo da task.

2) *Otimização da Inferência:* Uma vez que a pessoa é detectada, a extração de características e a identificação são realizadas utilizando o YOLOv8, onde esse processo é otimizado pelo TensorRT. O TensorRT é uma plataforma de otimização e inferência de alta performance para redes neurais profundas desenvolvida pela NVIDIA. Ele permite a aceleração do processamento de imagens e a execução eficiente dos

modelos de visão computacional em hardware de GPU. No contexto do "Follow Me", o TensorRT otimiza o YoLov8 que processa as imagens capturadas pelas câmeras RGB-D do robô, extraindo características específicas da pessoa a ser seguida, como forma, cor e posição no espaço tridimensional.

3) *Rastreamento de Pessoas*: Para garantir que o robô consiga acompanhar movimentos rápidos e mudanças de direção, foi utilizado o algoritmo ByteTrack para o rastreamento de pessoas. O ByteTrack é um algoritmo de rastreamento que combina detecções sucessivas em frames de vídeo, mantendo uma associação contínua entre a posição da pessoa e o movimento capturado. Ele funciona ao empregar uma combinação de técnicas de associação de objetos e predição de movimento, garantindo que o robô mantenha o foco na pessoa mesmo em ambientes dinâmicos e complexos. A adaptação do ByteTrack para robôs seguidores envolve ajustar seus parâmetros e heurísticas para lidar com as especificidades dos cenários domésticos, onde o robô deve navegar em torno de móveis e outros obstáculos enquanto mantém a pessoa no campo de visão.

A integração dessas tecnologias (YOLOv8 para detecção e identificação, TensorRT para otimização e ByteTrack para rastreamento) forma a base da solução de visão computacional para a task "Follow Me". Esse sistema permite ao robô não apenas detectar e identificar a pessoa a ser seguida, mas também acompanhar seus movimentos de maneira precisa.

## V. CONCLUSÕES

A performance em aplicações em tempo real do sistema robótico "Follow Me" desenvolvido neste estudo demonstra um avanço na integração de tecnologias de visão computacional, robótica e processamento de áudio e voz. A capacidade do robô de identificar, rastrear e seguir uma pessoa designada em ambientes dinâmicos, mesmo sob condições desafiadoras de oclusões, variações de iluminação e ruído ambiente, destaca a eficácia das soluções implementadas.

Os experimentos realizados comprovam que a combinação de algoritmos avançados, como YOLOv8, TensorRT e ByteTrack, juntamente com técnicas de processamento de áudio e navegação autônoma (AMCL, GMapping e Nav2), proporciona um desempenho robusto e confiável. A implementação do sistema em tempo real é particularmente relevante em cenários de assistência pessoal, segurança e entretenimento interativo, onde a resposta imediata e a interação contínua são cruciais, assim como a aplicação do estudo.

## REFERENCES

- [1] NVIDIA. TensorRT SDK. Disponível em: <https://developer.nvidia.com/tensorrt>.
- [2] DScripka. OpenWakeWord. GitHub, 2024. Disponível em: <https://github.com/dscripka/openWakeWord>.
- [3] IFZHANG. ByteTrack. GitHub, 2024. Disponível em: <https://github.com/ifzhang/ByteTrack>.
- [4] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics>
- [5] ROS Navigation. Navigation2. GitHub, 2024. Disponível em: <https://github.com/ros-navigation/navigation2>.
- [6] ROS Wiki. gmapping. ROS Wiki, 2024. Disponível em: <https://wiki.ros.org/gmapping>.
- [7] ROS Wiki. AMCL. ROS Wiki, 2024. Disponível em: <https://wiki.ros.org/amcl>
- [8] Google Drive. Follow Me. Disponível em: <https://drive.google.com/drive/folders/11pqpHYifnsumJ7EDl95nuUyFzjFnoyyp>