

Fusão de Dados LiDAR e Visual para Rastreamento 3D em Veículos Autônomos

Denise Silva Figueiredo / Marcelo Maia Juvencio

Abstract—This work explores the integration of visual and LiDAR sensors for 3D object detection and tracking in autonomous vehicle systems. A methodology leveraging data fusion from ZED stereo cameras and the Ouster OS1-64 LiDAR is implemented to improve perception accuracy in dynamic environments. The proposed system uses YOLO v3 for visual object classification and a probabilistic IMM-UKF-JPDAF tracker for motion estimation. The approach is validated using real-world data from the KITTI Dataset, demonstrating competitive performance in detection and tracking under various conditions. Results highlight the system's potential for embedded applications with real-time constraints, addressing challenges like occlusions and sensor limitations.

Index Terms—Autonomous Vehicles, 3D Object Detection, LiDAR, Visual Sensors, Data Fusion, IMM-UKF-JPDAF, Multiple Object Tracking, KITTI Dataset.

I. INTRODUÇÃO

AS REDES VEICULARES desempenham um papel central no avanço da mobilidade inteligente, possibilitando a comunicação eficiente entre veículos, infraestrutura e outros elementos do ecossistema urbano. A autonomia veicular de nível 5, definida pela Society of Automotive Engineers (SAE), representa o estágio mais avançado dessa evolução, no qual veículos operam de maneira completamente autônoma, sem necessidade de intervenção humana. No entanto, alcançar esse nível de automação apresenta desafios significativos que vão além da capacidade técnica dos sensores e algoritmos de decisão.

Entre os principais obstáculos estão as limitações em cenários complexos, como condições meteorológicas adversas, áreas urbanas densas e locais sem infraestrutura de comunicação robusta. Adicionalmente, a integração de tecnologias como LiDAR, câmeras de alta resolução e unidades referenciais inerciais (IMUs) demanda um alto nível de precisão e confiabilidade, enquanto os protocolos de comunicação enfrentam desafios relacionados à latência, segurança e interoperabilidade.

O trabalho busca desenvolver uma estrutura integrada para detecção e rastreamento de objetos 3D em sistemas embarcados, utilizando a fusão de dados de câmeras e sensores LiDAR. O objetivo principal é melhorar a eficiência e a precisão dessas tarefas por meio de uma abordagem conjunta, explorando a redundância de informações entre os sensores. A estrutura proposta combina a classificação visual de objetos com um rastreador baseado em modelos probabilísticos (IMM-UKF-JPDAF), oferecendo uma solução otimizada para aplicações em tempo real. A validação do sistema é realizada com base em métricas amplamente reconhecidas e dados reais do KITTI Dataset [1][2], demonstrando a aplicabilidade prática para veículos autônomos.

II. CONTEXTUALIZAÇÃO TEÓRICA

Redes veiculares (Vehicular Ad-hoc Networks – VANETs) são projetadas para que os dispositivos se conectem de maneira dinâmica e sem a necessidade de uma infraestrutura fixa (Ad-hoc) e assim, fornecer comunicação entre veículos (Vehicle-to-Vehicle, V2V) e entre veículos e infraestrutura (Vehicle-to-Infrastructure, V2I). Os objetivos destas redes é promover segurança no trânsito, eficiência no transporte e melhorias em serviços de mobilidade, tanto para veículos autônomos quanto convencionais. Enquanto veículos autônomos se aproveitam dessas redes para operar de forma mais inteligente, veículos convencionais também podem se integrar e se beneficiar das informações trocadas, melhorando a experiência geral de condução e contribuindo para um trânsito mais seguro e eficiente.

A arquitetura das redes veiculares é composta por unidades a bordo dos veículos (OBUs, On-Board Units) e unidades de infraestrutura (RSUs, Road-Side Units). Essas unidades se comunicam utilizando tecnologias de comunicação sem fio, como DSRC (Dedicated Short-Range Communications, que utiliza o padrão IEEE 802.11p que é uma extensão do padrão IEEE 802.11, para ambientes que requerem baixa latência e alta mobilidade), 4G, 5G e, mais recentemente, o C-V2X (Cellular Vehicle-to-Everything, que utiliza a infraestrutura do 4G LTE e 5G para melhorar a conectividade e a cobertura).

Diante do que foi apresentado sobre VANETs e suas tecnologias de comunicação sem fio, vamos introduzir conceitos relacionados a veículos autônomos.

Os veículos autônomos desempenham um papel crucial nas redes veiculares ad hoc (VANETs). A integração de veículos autônomos em VANETs tem o potencial de revolucionar a mobilidade, a segurança e a eficiência do tráfego nas cidades.

Os veículos autônomos são equipados com sensores avançados, sistemas de navegação e algoritmos de inteligência artificial que permitem a tomada de decisões em tempo real. O conjunto de tecnologias utilizadas em veículos autônomos vai definir o grau de automação destes veículos. Assim, foram definidos níveis para classificar o grau de automação dos veículos autônomos que vai do nível 0 (nenhuma automação) ao nível 5 (automação total), conforme tabela 1. É importante notar que apenas no nível 5 o veículo é totalmente autônomo e capaz de operar em qualquer condição, sem a necessidade de intervenção humana.

Neste ponto podemos citar os veículos da Tesla, conhecidos como Autopilot. Estes veículos dependem principalmente de câmeras e software de visão computacional para a percepção do ambiente e dessa forma a intervenção humana pode ser

Nível	Descrição	Exemplo de Funcionalidade
0	Sem automação (Controle humano total)	O motorista controla todas as funções do veículo.
1	Assistência ao motorista (Automação mínima)	Controle de cruzeiro adaptativo (ACC)
2	Automação parcial (Controle simultâneo de várias funções)	Manutenção de faixa e controle de velocidade ao mesmo tempo
3	Automação condicional (Automação em condições limitadas)	Condução autônoma em rodovias, com motorista pronto para intervir
4	Alta automação (Condução autônoma em condições específicas)	Veículo autônomo que opera sem intervenção em áreas específicas
5	Automação total (Condução totalmente autônoma)	Veículo que pode dirigir em qualquer condição sem intervenção humana

TABLE I
NÍVEIS DE AUTOMAÇÃO PARA VEÍCULOS AUTÔNOMOS

necessária em cenários onde a visibilidade das câmeras é comprometida como, por exemplo, em uma tempestade ou uma nevasca. Essa abordagem levaria a uma classificação para os carros da Tesla de nível 2 ou 3, dependendo da situação, já que para nível 4, operando em áreas específicas como rodovias ou áreas urbanas com infraestrutura adequada, o cenário de pouca visibilidade ainda é possível.

As VANETs com as tecnologias, já citadas, C-V2X (Cellular Vehicle-to-Everything) e o DSRC (Dedicated Short-Range Communications) são essenciais para a operação eficaz dos veículos autônomos, pois proporcionam uma percepção aprimorada do ambiente, permitindo que esses veículos tomem decisões acertivas em situações dinâmicas. A integração de sensores avançados, como LIDAR, câmeras e radares, junto com a comunicação em tempo real oferecida pelas VANETs, cria um cenário onde os veículos autônomos não apenas reagem a eventos em seu ambiente, mas também antecipam e colaboram com outros veículos para uma condução mais segura e eficiente.

O exemplo dos veículos autônomos da Tesla mostra a necessidade da integração de sensores avançados para superar o desafio da visibilidade das câmeras para chegar ao nível 5. É claro que para superar um desafio outros desafios são criados, como por exemplo o poder computacional necessário que deverá ser embarcado em um veículo para processar informações que garantam uma detecção e rastreamento de objetos em cenário dinâmico como uma cidade.

Neste trabalho será apresentado um estudo de caso de veículo autônomo com detecção e rastreamento de objetos 3D [3], usando câmera e LiDAR 3D como sensores primários.

III. METODOLOGIA

O contexto principal é a aplicação em veículos autônomos, onde a percepção ambiental precisa ser robusta e eficiente. As tarefas de detecção e rastreamento são fundamentais para estimar a posição, orientação e movimento de objetos no ambiente dinâmico. O sistema foi projetado para operar em sistemas embarcados, garantindo processamento em tempo real e adaptabilidade a diferentes condições ambientais.

A. Componentes do Sistema

1) Sensores:

• LiDAR OS1-64 (Ouster):[4]

Fornece nuvens de pontos 3D para representar o ambiente ao redor do veículo. O sensor LIDAR é um dispositivo que utiliza pulsos de laser para determinar a distância entre o sensor e objetos ao seu redor. Ele é amplamente usado em aplicações de mapeamento, veículos autônomos, robótica, agricultura de precisão, e redes veiculares, devido a sua precisão e capacidade de medir distâncias com rapidez. O sensor LIDAR emite pulsos de luz laser infravermelha ou visível em direção ao ambiente, quando o pulso de laser atinge um objeto, ele é refletido de volta para o sensor e o tempo que o pulso leva para retornar ao LIDAR é registrado. Com este tempo (chamado de tempo de voo ou ToF - Time of Flight) e conhecendo a velocidade da luz é possível calcular a distância até o objeto como segue:

$$\text{Distância} = \frac{\text{Velocidade da luz} \times \text{Tempo de voo}}{2} \quad (1)$$

Emitindo milhões de pulsos por segundo em todas as direções com um sistema de feixes giratórios, conforme a luz reflete em diferentes pontos ao redor do sensor, é possível criar um mapa 3D preciso do ambiente com um campo de visão de 360 graus. Esse mapeamento tridimensional é muitas vezes chamado de nuvem de pontos.

O sensor LiDAR OS1-64, fabricado pela Ouster, é um dos dispositivos mais avançados para aplicações em detecção e percepção 3D e pode ser combinado com uma unidade inercial de medida (IMU) para fornecer dados de orientação e movimento, melhorando a estabilidade e precisão da nuvem de pontos em cenários dinâmicos ou em movimento. Ele é amplamente utilizado em veículos autônomos, mapeamento, robótica e sistemas de vigilância. Suas especificações e funcionalidades o tornam altamente eficiente para aplicações exigentes. Suas características principais são:

– Resolução e Canais:

Possui 64 canais de detecção, permitindo capturar nuvens de pontos com alta densidade e detalhamento.

– Frequência de Varredura:

Capaz de operar em frequências de até 20 Hz, garantindo uma taxa de atualização rápida para cenários dinâmicos.

– Alcance:

Detecta objetos a distâncias de até 120 metros, dependendo das condições de iluminação e refletividade do objeto.

– Campo de Visão (FoV):

87° vertical x 360° horizontal, cobrindo uma ampla área ao redor do sensor.

– Precisão e Consistência:

Mede distâncias com alta precisão, geralmente com uma margem de erro de poucos milímetros, ideal para cenários que exigem medições confiáveis.

– Tecnologia de Detecção:

Baseado em tecnologia de Time-of-Flight (ToF), onde pulsos de luz laser infravermelha são emitidos e

o tempo de retorno é medido para calcular distâncias.

– **Robustez:**

Projetado para suportar condições adversas, com classificação IP68 e IP69K, garantindo proteção contra poeira, água e vibrações.

– **Integração:**

Possui interfaces compatíveis com sistemas modernos, como Ethernet, e é facilmente integrável com frameworks como ROS (Robot Operating System).

• **Câmera ZED: [5]**

Usada para detecção visual de objetos. A câmera ZED, fabricada pela Stereolabs, é amplamente utilizada em aplicações de veículos autônomos devido às suas características avançadas que vão além das câmeras monoculares tradicionais, como:

– **Visão Estéreo para Detecção de Profundidade:**

A ZED utiliza duas lentes para capturar imagens estereoscópicas, permitindo calcular mapas de profundidade 3D em tempo real. Isso permite detectar a distância de objetos com precisão, essencial para evitar colisões, detectar pedestres e mapear o ambiente.

– **Campo de Visão Amplo:**

A ZED oferece um campo de visão amplo (até 120°), o que ajuda a capturar uma grande parte do ambiente ao redor do veículo, reduzindo pontos cegos.

– **Processamento em Tempo Real:**

A câmera processa informações diretamente no dispositivo ou com suporte de GPU externa, o que é vital para aplicações críticas em tempo real, como navegação autônoma.

– **Compatibilidade com SLAM:**

É compatível com algoritmos de Localização e Mapeamento Simultâneos (SLAM), usados para criar mapas 3D do ambiente e localizar o veículo dentro deles.

– **Alta Resolução e Taxa de Quadros:**

A ZED suporta resolução até 4K com taxas de quadros ajustáveis, permitindo capturar detalhes com alta qualidade em cenários dinâmicos.

– **Ambientes Dinâmicos e Iluminação Variável:**

Seu sistema é robusto em condições de iluminação variada (dia/noite) e pode lidar com ambientes dinâmicos, como trânsito intenso, oferecendo maior confiabilidade.

– **Integração com Redes Neurais:**

A ZED é compatível com frameworks como TensorFlow e PyTorch, permitindo integrar redes neurais para detecção visual avançada (e.g., classificação e segmentação semântica).

2) *Plataforma de Processamento:*

• **Jetson AGX Xavier (Nvidia): [6]**

Realiza o processamento da estrutura proposta, incluindo detecções visuais e rastreamento. O Jetson AGX Xavier, da Nvidia, é uma plataforma de computação avançada projetada especificamente para aplicações de inteligência artificial (IA) em sistemas embarcados, como veículos

autônomos, robótica e dispositivos IoT. Ele é ideal para o processamento de detecções visuais e rastreamento, especialmente quando combinado com sensores como câmeras e LiDAR. A seguir estão suas principais características:

– **Desempenho em IA:**

Possui até 32 TeraFLOPs de desempenho computacional para operações de IA, suficiente para rodar modelos complexos de detecção e rastreamento em tempo real.

– **GPU Integrada:**

Equipado com uma GPU Nvidia Volta com 512 núcleos CUDA e 64 Tensor Cores, otimizando tarefas de aprendizado profundo e operações matriciais para redes neurais como YOLO ou rastreadores baseados em IA.

– **Processamento em Tempo Real:**

Processa dados em tempo real, o que é essencial para detecção e rastreamento de objetos em ambientes dinâmicos, como trânsito urbano ou rodovias.

– **Módulos de Aceleração de Hardware:**

Inclui NVDLA (Nvidia Deep Learning Accelerator), que acelera inferências de redes neurais e libera a GPU para outras tarefas, como processamento de imagens ou geração de mapas tridimensionais.

– **Capacidade Multisensorial:**

Suporta entrada simultânea de múltiplos sensores, como a câmera ZED para visão estéreo e o LiDAR OS1-64 para geração de nuvens de pontos 3D. A integração dos sensores é facilitada pelo middleware ROS (Robot Operating System).

– **Eficiente em Energia:**

Consome entre 10W e 30W, sendo eficiente para aplicações embarcadas com restrições energéticas.

– **Interfaces Versáteis:**

Suporta diversas interfaces de entrada, incluindo USB, Ethernet e PCIe, permitindo conectar sensores LiDAR, câmeras e outras unidades periféricas.

O Jetson AGX Xavier é uma solução poderosa e eficiente para sistemas de detecção e rastreamento em veículos autônomos. Sua combinação de alto desempenho em IA, capacidade de processamento de múltiplos sensores e eficiência energética o torna ideal para integrar câmeras, LiDAR e algoritmos avançados de percepção. Ele desempenha um papel crucial na implementação de sistemas embarcados que requerem processamento em tempo real e alta confiabilidade.

- Middleware ROS “Melodic Morenia” em Ubuntu Linux 18.04.1.

- GPU configurada com bibliotecas CUDA 10.0 para tarefas visuais.

A estrutura proposta é implementada e testada no Hyundai i30 (Hyundai Motor Company, Seul, Coreia do Sul), mostrado na Fig. 1 com os sensores e plataforma de processamento já apresentados e componentes auxiliares.

• **Componentes auxiliares:**

- **CAN (Controller Area Network):**

É um protocolo de comunicação robusto e de baixa latência amplamente utilizado em veículos automotivos. Ele permite que os diferentes sistemas eletrônicos de um veículo (como sensores, atuadores e controladores) se comuniquem entre si sem a necessidade de um computador central.

– **V2X Modem (Vehicle-to-Everything):**

O modem V2X é responsável pela comunicação entre o veículo e outros elementos no ambiente, como outros veículos (V2V), infraestrutura rodoviária (V2I), pedestres (V2P) e redes de comunicação (V2N). Esse tipo de comunicação melhora a percepção do veículo além do que os sensores embarcados conseguem alcançar, permitindo, por exemplo, que o veículo receba informações sobre condições de tráfego, sinais de trânsito ou possíveis colisões com antecedência.

– **Display:**

Fornece uma interface visual para os ocupantes do veículo ou os operadores, exibindo informações como o status do sistema autônomo, os dados do trajeto, e alertas. Pode ser usado para monitorar o comportamento do sistema, exibir a visão do que o carro "vê" (dados da câmera ou LiDAR), ou até para interações manuais em situações específicas.

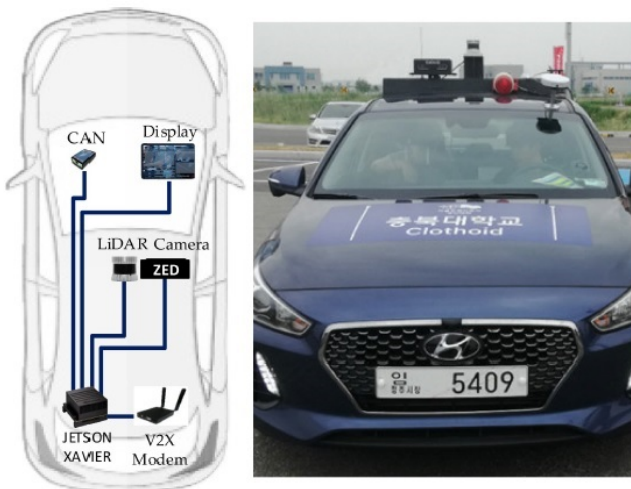


Fig. 1. Sensor setup and implementation platform. [3]

3) *Bibliotecas e Algoritmos:*

- YOLO v3 para detecção visual de objetos.
- Redes neurais leves para processamento visual de objetos em tempo real.

Uma rede neural leve (ou "lightweight neural network") é um tipo de rede neural artificial projetada para ser menos complexa e mais eficiente em termos de recursos computacionais, como tempo de processamento, uso de memória e consumo de energia. Esse tipo de rede é especialmente adequado para ser executado em dispositivos com capacidades limitadas, como sistemas embarcados ou dispositivos móveis.

B. *Métodos utilizados*

1) *Fusão de Sensores:*

– **Fusão de Câmera e LIDAR:**

O sistema combina informações de cor, textura e contornos dos objetos, presentes nas imagens ou vídeos do ambiente captados pelas câmeras, com dados geométricos precisos sobre a posição, forma e profundidade, presentes na nuvem de pontos 3D geradas pelo LIDAR. Esta combinação de dados da câmera com dados do LIDAR em sistemas Visual-LiDAR é denominada fusão de dados, que é a forma como as informações são combinadas para melhorar a detecção dos objetos. Existem três principais tipos de fusão: Fusão Precoce (Early Fusion), Fusão Tardia (Late Fusion) e Fusão Profunda (Deep Fusion). Esses esquemas determinam em que estágio da captação os dados dos dois sensores (câmera e LIDAR) são combinados para processar a informação e detectar objetos. Em resumo temos:

* **Fusão Precoce:**

Combina os dados logo no início, aproveitando ao máximo a complementaridade dos dois sensores desde o começo. Pode ser mais preciso, porém mais complexo.

* **Fusão Tardia:**

Os dados são processados separadamente e combinados no final. É mais simples de implementar, mas pode perder detalhes cruciais.

* **Fusão Profunda:**

Integra os dados ao longo de todo o processo usando aprendizado profundo. Oferece alta precisão, mas é mais exigente em termos de recursos computacionais e dados de treinamento.

2) *Detecção e Rastreamento:*

– **Detecção 3D:**

A detecção de objetos refere-se à capacidade do sistema de identificar e reconhecer a presença de objetos no ambiente ao redor do veículo. Ela envolve a localização inicial e a classificação desses objetos (por exemplo, carros, pedestres, ciclistas, sinais de trânsito, etc.). Sensores como LIDAR, câmeras e radares são amplamente utilizados para esta finalidade.

A detecção é feita com a configuração Visual-LIDAR. Visual-LiDAR são algoritmos que utilizam tanto informações visuais (imagens de câmeras) quanto dados de sensores LIDAR para detectar e identificar objetos em um ambiente tridimensional. A nuvem de pontos LiDAR é processada para remover informações do solo, segmentar objetos e ajustar dimensões e poses. As detecções visuais são realizadas em paralelo, usando YOLO v3, fornecendo informações de classe e posição.

* **YOLO v3:**

YOLO (You Only Look Once)[7] é uma das arquiteturas mais populares para a detecção de objetos em imagens e vídeos. A versão 3, con-

hecida como YOLOv3, foi lançada em 2018 e trouxe várias melhorias em relação às versões anteriores. O YOLOv3 é um modelo de aprendizado profundo que realiza a detecção de objetos em tempo real, utilizando uma única rede neural convolucional (CNN) para prever as classes e as localizações dos objetos simultaneamente. O YOLOv3 é pré-treinado em conjuntos de dados Microsoft COCO [8]. O Microsoft COCO (Common Objects in Context) é um grande conjunto de dados projetado para a detecção de objetos, segmentação de imagens e legendagem de imagens. Embora o YOLOv3 seja primariamente utilizado para detecção de objetos em 2D, existem extensões e modificações que permitem a detecção de objetos em 3D usando técnicas adicionais como Sensores 3D, como o LIDAR neste estudo de caso, Modelos de Aprendizado Profundo e Fusão de Dados. Modelos de Aprendizado Profundo são adaptações do YOLO para lidar com dados 3D, prevendo não apenas a localização em 2D, mas também a profundidade ou a posição no espaço tridimensional.

Fusão de Dados é a combinação de dados de diferentes sensores (LiDAR OS1-64 e Câmera ZED, neste estudo de caso) para melhorar a detecção e a classificação de objetos em 3D.

Em resumo:

Detecção é o processo de localizar e identificar objetos em uma cena, sem necessariamente se preocupar com o que acontece com esses objetos ao longo do tempo.

A figura 2 mostra a identificação e classificação dos objetos "person" e "car" feita pelo YOLO que foi treinado utilizando o dataset COCO. Esses objetos são então associados aos dados do LiDAR para melhorar a precisão e a segmentação.

– Rastreamento de Múltiplos Objetos (MODT):

Um rastreador baseado em IMM-UKF-JPDAF é implementado, que combina modelos de movimento interativos com filtragem probabilística para lidar com incertezas no movimento e nas detecções. O rastreamento de objetos vai além da simples detecção. Depois que o objeto é identificado, o sistema de rastreamento acompanha o movimento do objeto ao longo do tempo, monitorando sua trajetória, velocidade e direção. O rastreamento utiliza os dados de detecção em sequência para prever a futura posição do objeto, o que é essencial para a tomada de decisões e planejamento de rotas seguras.

O rastreamento é feito por um rastreador de objetos baseado em IMM-UKF-JPDAF, um sistema avançado de rastreamento de múltiplos objetos que combina várias técnicas para lidar com a complexidade de movimentos e incertezas em ambientes com muitos alvos. Cada uma dessas siglas, como será mostrado, refere-se a um componente específico do



Fig. 2. Visual object detector to classify tracked objects using class [3]

rastreador, IMM (Interacting Multiple Model), UKF (Unscented Kalman Filter) e JPDAF (Joint Probabilistic Data Association Filter) que juntos formam um algoritmo robusto e preciso.

* MOT (Multiple Object Tracking):

Refere-se ao rastreamento de múltiplos objetos. Em um sistema MOT, o foco está em seguir o movimento de vários objetos ao longo do tempo, depois que eles já foram detectados. A tarefa é identificar a trajetória e prever o movimento futuro desses objetos, atualizando continuamente sua posição. No caso de veículos autônomos, por exemplo, um sistema MOT poderia rastrear a posição de pedestres, veículos e ciclistas simultaneamente enquanto eles se movem em torno do veículo. MOT pressupõe que a detecção inicial já ocorreu, e o foco agora é no rastreamento consistente desses objetos ao longo do tempo.

* MODT (Multiple Object Detection and Tracking):[9][10]

A Detecção e Rastreamento de Múltiplos Objetos (MODT) é usado para associar detecções consecutivas, criando trajetórias coerentes mesmo em cenários com oclusão. MODT engloba tanto a detecção quanto o rastreamento de múltiplos objetos. Neste tipo de sistema, os objetos são detectados continuamente, e o sistema simultaneamente realiza o rastreamento desses objetos. Isso pode

incluir a detecção de novos objetos que entram no campo de visão e o rastreamento contínuo de objetos já detectados. É uma abordagem mais complexa, já que exige que o sistema execute a detecção e o rastreamento de forma integrada e em tempo real, podendo lidar com a entrada e saída de objetos do campo de visão do sensor.

Em resumo:

Rastreamento é o processo de acompanhar os objetos ao longo do tempo, prevendo suas trajetórias futuras e ajustando as ações do veículo com base nesses dados.

MOT se concentra no rastreamento de objetos previamente detectados e foca na manutenção das identidades e trajetórias desses objetos ao longo do tempo.

MODT envolve tanto a detecção quanto o rastreamento simultâneo de múltiplos objetos, oferecendo uma solução completa para ambientes dinâmicos onde novos objetos podem surgir a qualquer momento.

– Algoritmos de rastreamento IMM-UKF-JPDAF:

O rastreador de objetos baseado em IMM-UKF-JPDAF é um sistema avançado de rastreamento de múltiplos objetos que combina várias técnicas para lidar com a complexidade de movimentos e incertezas em ambientes com muitos alvos. Cada uma dessas siglas refere-se a um componente específico do rastreador, e juntas formam um algoritmo robusto e preciso. Vamos entender cada parte:

* IMM (Interacting Multiple Model):

Interacting Multiple Model é um método usado para rastreamento de objetos que se movem de maneira complexa, seguindo diferentes dinâmicas de movimento. O IMM assume que o objeto pode alternar entre diferentes comportamentos ou modos de movimento (por exemplo, um carro pode acelerar, frear, virar, ou manter velocidade constante). O rastreador usa múltiplos modelos (cada um representando um comportamento) simultaneamente e calcula a probabilidade de o objeto estar seguindo cada modelo em determinado momento. O algoritmo combina as previsões desses diferentes modelos para obter uma estimativa precisa da posição e velocidade do objeto.

* UKF (Unscented Kalman Filter):

O Unscented Kalman Filter (UKF) é uma variante do Filtro de Kalman, que é amplamente utilizado para rastreamento de objetos em sistemas dinâmicos. O UKF é especialmente útil quando o sistema a ser rastreado é não linear (por exemplo, um carro fazendo curvas ou mudanças bruscas de direção). Ele usa uma técnica chamada transformação unscented para lidar melhor com a não linearidade ao prever o estado futuro do objeto (posição, velocidade, etc.) a partir de observações ruidosas, como medições de sensores LIDAR ou câmeras. Comparado ao Filtro de Kalman padrão,

o UKF oferece melhor desempenho em situações onde o movimento do objeto não segue um padrão linear.

* JPDAF (Joint Probabilistic Data Association Filter):

JPDAF é um método que resolve o problema de associação de dados em cenários com múltiplos objetos. Em um ambiente com vários objetos, pode ser difícil determinar qual detecção (por exemplo, de um sensor LIDAR) pertence a qual objeto. O JPDAF trata isso de maneira probabilística, associando as detecções aos objetos com base em probabilidades, levando em conta a incerteza nas medições. Ele considera múltiplas hipóteses simultaneamente, ou seja, ele calcula probabilidades de associação para várias detecções e distribui essas probabilidades para determinar o melhor ajuste de cada objeto às detecções observadas.

Em resumo:

IMM lida com a incerteza do modelo de movimento, permitindo que o rastreador considere diferentes dinâmicas para prever o movimento de um objeto.

UKF trata de prever e rastrear objetos que se movem de maneira não linear (como em curvas ou acelerações/desacelerações complexas).

JPDAF resolve o problema de atribuir corretamente cada detecção de sensor a um objeto específico em cenários onde há múltiplos alvos presentes (por exemplo, rastrear vários veículos ou pedestres ao mesmo tempo em uma rua movimentada).

A figura 3 mostra a aplicação dos algoritmos de rastreamento **IMM-UKF-JPDAF** na forma de um gráfico de visualização de rastreamento sensorial.

Na imagem, o **IMM** rastreia o objeto identificado como "person" (pessoa), ajustando os parâmetros (Yaw e Vel) conforme a movimentação detectada.

O **UKF** estima a posição, velocidade e orientação do objeto (pessoa), levando em conta as incertezas nas medições provenientes do sensor LiDAR.

O **JPDAF** associa as detecções do sensor LiDAR (os pontos em vermelho na figura) com alvos rastreados (representados pelo contorno verde da pessoa). Ele ajuda a resolver ambiguidades quando múltiplos objetos estão presentes no campo de visão.

3) Processamento em Tempo Real:

- Otimizações são realizadas no módulo de rastreamento e na classificação de solo para garantir que todo o sistema funcione dentro dos limites de tempo definidos pelo ciclo de amostragem do LiDAR.

4) Avaliação de Desempenho:

– Métricas MOT16 e o KITTI Datasets:

MOT16 e o KITTI Datasets são padrões amplamente usados para avaliação de algoritmos de rastreamento de objetos. O MOT16 é usado para avaliar o desempenho do sistema de rastreamento de objetos, enquanto o KITTI Datasets fornece os dados de nuvem de pontos 3D e verdades básicas necessárias

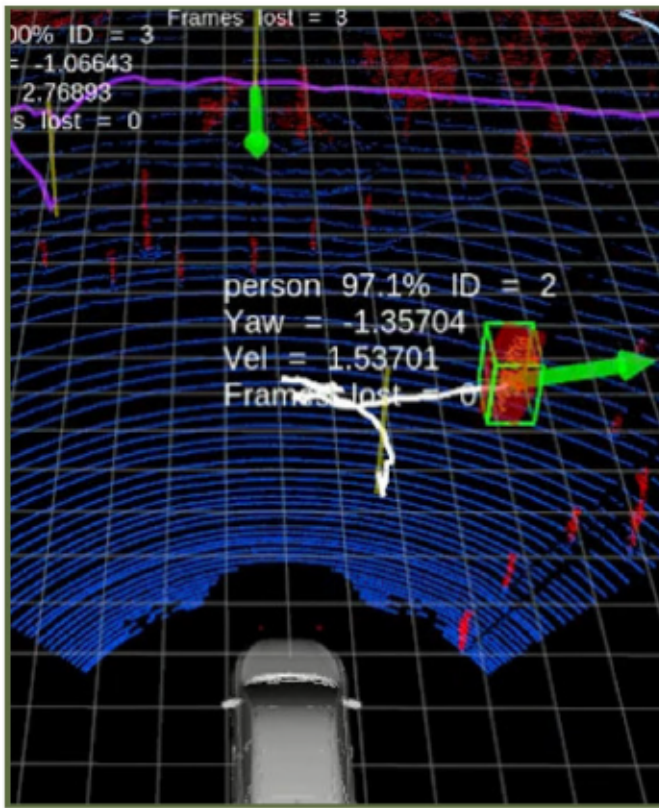


Fig. 3. IMM-UKF-JPDAF tracker with clustered associations. [3]

para treinar e testar o algoritmo de detecção e rastreamento de objetos 3D descrito na citação. O sistema pode ser avaliado em termos de precisão de detecção e qualidade do rastreamento usando esses dois recursos combinados.

* **MOT16:**

Multiple Object Tracking 2016 (MOT16) é um benchmark usado para avaliar algoritmos de rastreamento de múltiplos objetos (MOT). Ele faz parte de uma série de benchmarks da série MOTChallenge, que fornece conjuntos de dados anotados para comparar o desempenho de diferentes algoritmos de rastreamento. As métricas MOT16 avaliam a capacidade de um algoritmo em rastrear corretamente diversos objetos em vídeos. As principais métricas incluem MOTA e MOTP.

* **MOTA (Multiple Object Tracking Accuracy):** Uma métrica que combina a precisão de detecção, o número de falsas detecções e objetos perdidos.

* **MOTP (Multiple Object Tracking Precision):** Mede a precisão do rastreamento em termos de distância entre a predição e a posição real dos objetos.

* **KITTI Datasets:**

O KITTI é uma coleção de conjuntos de dados para a pesquisa de visão computacional, frequentemente usada em sistemas de percepção para veículos autônomos. Ele foi criado por pesquisadores do Karlsruhe Institute of Technol-

ogy (KIT) e do Toyota Technological Institute at Chicago (TTIC). O KITTI Object Detection and Tracking Dataset inclui dados de LIDAR, câmeras estéreo e GPS, permitindo a avaliação de algoritmos para detecção e rastreamento de objetos em 3D. As anotações fornecem as chamadas "verdades básicas" (ground truths) que especificam a posição e o tipo de objetos, como carros e pedestres, em cada frame. O KITTI fornece um ambiente realista para testar algoritmos, incluindo cenas de tráfego urbano com diferentes níveis de complexidade.

Em resumo, o MOT16 refere-se ao conjunto de métricas e benchmarks usados para avaliar o rastreamento de múltiplos objetos, e o KITTI Datasets fornece os dados de sensor necessários para esses testes, incluindo as "verdades básicas" que são usadas como referência para medir o desempenho dos algoritmos de rastreamento e detecção.

IV. RESULTADOS E DISCUSSÕES

Este estudo de caso analisa uma estrutura para detecção e rastreamento de objetos 3D baseada na fusão de sensores Visual-LiDAR, projetada para operar em sistemas embarcados com limitações de recursos computacionais. O objetivo é abordar desafios relacionados à detecção e rastreamento simultâneo de múltiplos objetos em tempo real, especialmente em cenários complexos de veículos autônomos.

O algoritmo proposto é colocado em teste sob o cenário de um veículo de emergência rápido passando e um manequim modelo sendo rapidamente arrastado pela frente da plataforma como um pedestre. A Fig. 4 mostra que a estrutura rastreia e relata eficientemente os parâmetros de rastreamento em ambos os cenários. O veículo de emergência é identificado como um carro com certeza de 100%, o que implica que o objeto rastreado em cada período de tempo recebeu a mesma classe, mantendo uma ID exclusiva e obtendo uma medição associada no último passo de tempo.

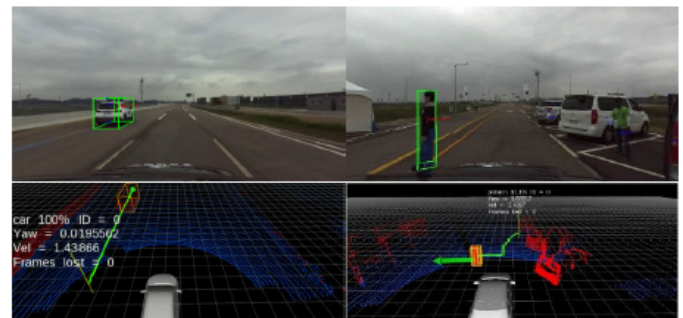


Fig. 4. Detection and tracking of vehicle and pedestrian. [3]

A. Resultados gerais

* **Eficiência em Sistemas Embarcados:**

A estrutura implementada no hardware Jetson AGX Xavier apresentou tempos de processamento que atendem aos limites do ciclo de amostragem do sensor LiDAR. Isso confirma a viabilidade do sistema para aplicações em tempo real.

* **Precisão e Rastreabilidade:**

Nos experimentos conduzidos com o KITTI Dataset, dois terços das trilhas foram categorizadas como "Principalmente Rastreadas" (MT), indicando um bom desempenho de rastreamento. O sistema conseguiu manter baixos os índices de troca de identidade (IDSW), um fator importante para a confiabilidade do rastreamento.

* **Limitações de Alcance e Condições de Oclusão:**

As métricas de precisão degradam significativamente em distâncias maiores que 40 metros, devido à escassez de medições do LiDAR. Além disso, objetos parcialmente oclusos ao longo de sua trajetória apresentam dimensões incorretas e erros de centroides.

B. Métricas de avaliação

* **MOTA (Multiple Object Tracking Accuracy):**

A precisão do rastreamento foi consistente para objetos próximos, com erro médio de centroides em torno de 1 metro. Esse valor reflete medições adequadas mesmo em condições de oclusão parcial.

* **MOTP (Multiple Object Tracking Precision):**

Indicou alta precisão para a associação temporal das trilhas, especialmente em cenários urbanos densos.

* **Qualidade do Rastreamento:**

A maioria dos objetos foi rastreada por mais de 80% de sua trajetória. No entanto, conjuntos de dados com pedestres aglomerados mostraram maior fragmentação das trilhas.

C. Comparações

- * Comparado com abordagens de última geração (e.g., AB3DMOT, FANTrack), o sistema apresentou desempenho competitivo em termos de precisão de rastreamento (AMOTA) e troca de identidade, enquanto se destacou pela alta velocidade de processamento, alcançando 236 FPS.
- * A vantagem de operar em sistemas embarcados sem a necessidade de GPUs dedicadas foi destacada como um diferencial, embora o sistema perca desempenho em cenários mais complexos.

D. Limitações do sistema

* **Velocidade Relativa dos Objetos:**

Objetos com velocidade superior a 80 km/h não foram rastreados adequadamente, resultando em trocas de ID.

* **Dependência de Detecções Visuais:**

O sistema apresentou quedas de desempenho em condições de baixa luminosidade, onde as detecções visuais são insuficientes.

* **Oclusões Persistentes:**

O rastreamento de objetos parcialmente oclusos ao longo de toda sua trajetória compromete a precisão das dimensões e da pose estimadas.

V. CONCLUSÃO

Os resultados indicam que a estrutura proposta é uma solução eficiente e robusta para detecção e rastreamento de objetos 3D em sistemas embarcados, adequada para veículos autônomos. Apesar das limitações, o sistema demonstrou ser competitivo com abordagens de última geração, especialmente na relação entre desempenho e eficiência computacional. As extensões futuras podem fortalecer ainda mais a aplicabilidade da abordagem em cenários complexos.

REFERENCES

- [1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2012, pp. 3354–3361.
- [2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, September 2013.
- [3] M. Sualeh and G.-W. Kim, "Visual-lidar based 3d object detection and tracking for embedded systems," *IEEE Access*, vol. 8, pp. 156 285–156 298, 2020.
- [4] O. Inc., *OS1-64 Product Documentation*, Online, 2024, available: <https://ouster.com/products/hardware/os1-lidar-sensor>.
- [5] S. Inc., *ZED Stereo Camera Product Documentation*, Online, 2024, available: <https://www.stereolabs.com/zed/>.
- [6] NVIDIA Corporation, "Jetson AGX Xavier Developer Kit," <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>, NVIDIA Corporation, Santa Clara, CA, USA, 2024, accessed: 2024-12-02.
- [7] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018. [Online]. Available: <https://arxiv.org/abs/1804.02767>
- [8] T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," *CoRR*, vol. abs/1405.0312, 2014. [Online]. Available: <http://arxiv.org/abs/1405.0312>
- [9] S. Kim, J. Ha, and K. Jo, "Semantic point cloud-based adaptive multiple object detection and tracking for autonomous vehicles," *IEEE Access*, vol. 9, pp. 157 550–157 562, 2021.
- [10] M. Sualeh and G.-W. Kim, "Dynamic multi-lidar based multiple object detection and tracking," *Sensors*, vol. 19, no. 6, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/6/1474>