

## Resumo do artigo base

É proposto um algoritmo de mapeamento denso baseado em estéreo para ambientes urbanos em grande escala, comparando a projetos feitos anteriormente, os objetos estáticos foram reconstruídos simultaneamente enquanto os objetos dinâmicos e potencialmente dinâmicos foram construídos separadamente buscando alcançar maiores resultados. Para a avaliação da proposta foi utilizado os dados KITTI como base para o projeto.

A sequência de passos para produzir um mapa no artigo em questão pode ser entendido como: Inicialmente temos a etapa onde ocorre o calculo do mapa de profundidade denso, nesse momento ocorre todo o pré-processamento e estima-se um mapa de profundidade utilizando como base duas abordagens principais, uma geométrica e outra utilizando rede neural, logo em seguida é feita a segmentação de objetos 2D, nessa etapa é utilizado um algoritmo para reconhecer objetos dinâmicos e potencialmente dinâmicos, depois esses objetos rastreados são enviados para a etapa que gera o fluxo de cena. Com o fluxo de cena é possível então calcular a odometria visual e assim estimar o movimento 3D de objetos rastreados. Por fim é feito o rastreamento de objetos 3D, nesse momento o fluxo de cena associado a uma instância de objeto especifica é usado como entrada para que seja estimado o movimento da câmera em relação a essa instância e com isso o mapa estático e os objetos são reconstruídos, onde os dinâmicos e potencialmente dinâmicos são reconstruídos individualmente.

Foram feitos três experimentos na proposta:

- **Avaliação de precisão qualitativa**



(a) A fusão estática tende a corromper o mapa do ambiente com listras e outros artefatos produzidos por objetos que se movem independentemente.



(b) A fusão dinâmica, o principal modo de operação do nosso sistema, evita que trilhas de veículos e halos remanescentes sejam integrados ao mapa.

O principal resultado desse experimento foi mostrar que na reconstrução estática os veículos com movimento individual tendem a corromper o mapa com listras e “vultos” nas imagens, enquanto com a fusão dinâmica, isso tende a não acontecer pois os objetos dinâmicos são reconstruídos separadamente fazendo com que isso gere um mapa do ambiente livre de corrupção.

- **Avaliação de precisão quantitativa**

Buscou avaliar os mapas de profundidade da seguinte forma:

- 1) Os pontos do solo não associados a nenhum objeto potencialmente dinâmico são contados para as **estatísticas do mapa estático**.
- 2) Os pontos associados a objetos potencialmente dinâmicos que estão sendo reconstruídos são contados para as **estatísticas de objetos dinâmicos**.
- 3) Os pontos restantes que correspondem a objetos dinâmicos que não estão sendo reconstruídos (por exemplo, bicicletas, pedestres ou carros distantes

cujo movimento 3D não pode ser computado de forma confiável) são ignorados

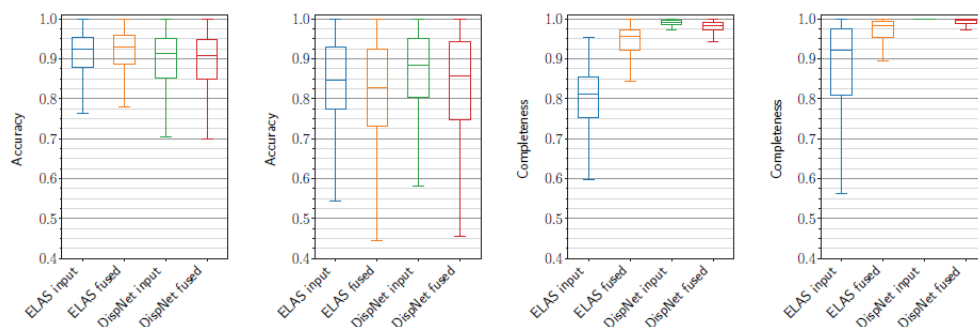
Apesar de ser um método imperfeito, pois depende da segmentação semântica computada, que nem sempre é 100% confiável, foi constatado que na prática funciona bem, permitindo obter vários insights sobre o desempenho do nosso sistema sob várias condições

Várias conclusões extras podem ser tiradas com esses resultados:

- 1) Usar mapas de profundidade ELAS (abordagem que leva em conta mais a geometria) leva a mapas estáticos mais precisos do que DispNet (abordagens com redes neurais), mas reconstruções de objetos menos precisas. Isso é explicado pela robustez superior do DispNet a desafios como superfícies reflexivas e transparentes, muito mais difundidas na reconstrução de carros.
- 2) O processo de fusão melhora a precisão da reconstrução ao usar mapas de profundidade ELAS, mas não ao usar DispNet. Isso se deve à natureza mais conservadora dos resultados produzidos pelo ELAS, que tendem a ter lacunas, mas são mais nítidos em geral.
- 3) A precisão da reconstrução dos objetos dinâmicos geralmente não é melhorada pela fusão quando comparada aos mapas de entrada, e a variação geral entre os quadros é muito maior. Isso decorre do fato de que os veículos são consideravelmente mais difíceis de reconstruir do que, por exemplo, superfícies de estradas, cercas e edifícios, devido às suas próprias propriedades.
- 4) DispNet leva a reconstruções muito mais densas, ou seja, mais completas do mapa estático e dos objetos do que ELAS, mas essa lacuna é reduzida pela fusão volumétrica em vários quadros. Isso decorre do fato de que os mapas de profundidade ELAS são mais esparsos, mas mais precisos, o que significa que eles se beneficiam mais da fusão do que o DispNet.
- 5) A variação da integridade do mapa de profundidade ELAS é alta porque geralmente produz resultados muito esparsos em condições de iluminação desafiadoras, como após sair de um túnel, antes que a exposição automática da câmera consiga se adaptar. No entanto, a magnitude desse efeito é reduzida significativamente pela fusão.

- **Eficácia do regularizador de mapa de atraso fixo**

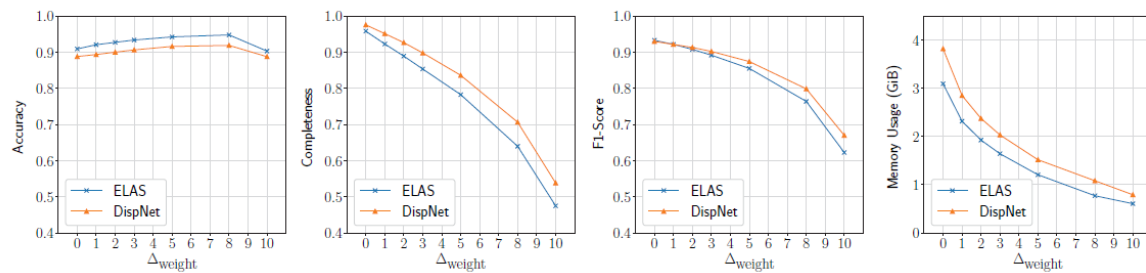
Nesse experimento foi avaliado o impacto da regularização do mapa no consumo de memória do sistema e na precisão da reconstrução



Os mapas produzidos usando DispNet são mais densos do que aqueles baseados em ELAS, o que também se reflete na integridade da reconstrução. Ao mesmo tempo, como nos experimentos anteriores, apesar de ser menos preciso que o DispNet, de acordo com o Benchmark KITTI, a profundidade dos mapas produzidos pelo ELAS levam a reconstruções mais precisas, embora menos completas.

Esta tendência é mantida mesmo com o aumento do  $\Delta\text{weight}$ .

Analisando o gráfico abaixo, fica claro que o impacto do ruído no consumo de memória das reconstruções é forte.



## Descrição do carro e sensores.

Automóvel Robótico Autônomo Inteligente (IARA) é um carro autônomo de pesquisa desenvolvido no LCAD/UFES (Laboratório de Computação de Alto Desempenho da Universidade Federal do Espírito Santo, Brasil). Foi o primeiro carro autônomo a percorrer 74 km de forma autônoma em vias urbanas e rodovias do Brasil. O projeto é baseado em um Ford Escape Hybrid 2011 adaptado para operação autônoma pela Torc Robotics e pela equipe do LCAD. Os sensores usados no IARA são um *Velodyne HDL-32E LiDAR*, uma câmera estéreo frontal *Point Gray Bumblebee XB3*, um *Xsens MTi IMU*, um *Trimble RTK GPS* além de um sensor de odometria.

O *Velodyne HDL-32E LiDAR* fornece dados de nuvem de pontos (*point cloud*), que consiste em um conjunto de coordenadas tridimensionais ( $x, y, z$ ) representando os pontos no ambiente escaneado pelo LiDAR, com ele é possível ter acesso a informações de profundidade mais precisas. Basicamente, é um sensor que possui 32 lasers montados verticalmente capazes de medir a distância até o objeto mais próximo e a refletividade de sua superfície. Essas 32 leituras são coletadas simultaneamente e chamadas de *shot*. O sensor é equipado com um motor giratório que permite a rotação dos lasers em torno do eixo vertical, proporcionando uma visão tridimensional do ambiente em forma de uma esfera centrada no carro. A cada *shot*, o ângulo horizontal do motor em rotação também é registrado.

Um tipo de ruído que é comumente observado com esse tipo de sensor são os de movimento, causados pelo movimento do sensor durante a captura dos dados, o que pode levar a distorções na nuvem de pontos e à perda de informações importantes. Para a tarefa de mapeamento 3D o sensor é muito utilizado principalmente por possibilitar a captura de uma grande quantidade de pontos de nuvem em um curto período de tempo.



Figura 1: Ruído de movimento

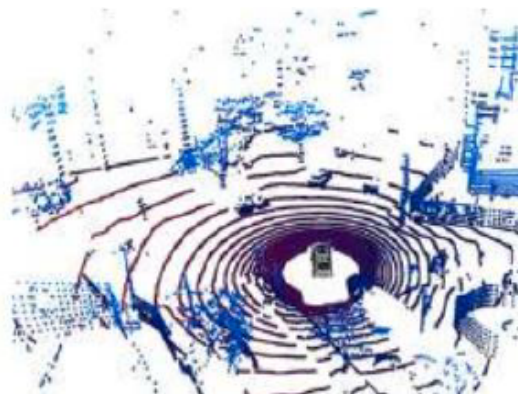


Figura 2: Imagem capturada e sua respectiva nuvem de pontos.

A câmera estéreo frontal *Point Gray Bumblebee XB3* fornece dados de imagem estéreo em tempo real. Mais especificamente, ela fornece um conjunto de duas imagens, uma para cada câmera, permitindo a captura de informações de profundidade através da disparidade entre as imagens. Além disso, são produzidos também os principais parâmetros para auxiliar na reconstrução da imagem. Os principais ruídos observados com esse tipo de sensor, seria os provenientes da própria iluminação do ambiente.



Figura 3: Câmera *Point Gray Bumblebee XB3*

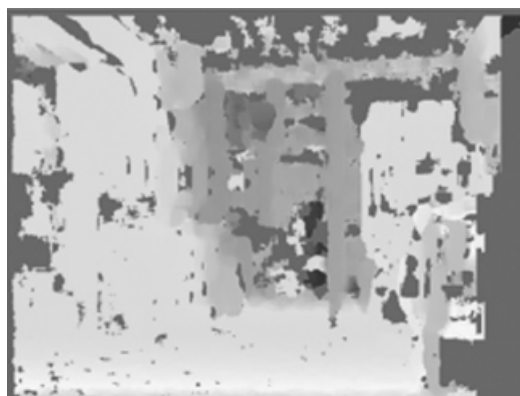


Figura 4: Imagem da câmera estéreo esquerda, direita e dado de profundidade.

O Xsens MTi IMU fornece dados de orientação e velocidade do veículo. Com ele obtemos fornece acelerações e velocidades angulares nas direções dos eixos x, y e z, bem como a orientação do sensor em relação a um sistema de coordenadas global na Terra.

O *Trimble RTK GPS* fornece dados de posição geográfica de alta precisão com correção em tempo real para compensar erros causados pelas próprias condições geográficas. No entanto, ainda podem ocorrer ruídos nos dados devido a condições ambientais adversas, podendo ser afetado também por erros sistemáticos do próprio sensor. Para minimizar esses impactos, normalmente esses sensores são combinados com outros, por exemplo com o IMU.

Por fim, o sensor de odometria fornece dados sobre a distância percorrida e a velocidade do veículo, que são adquiridos por meio da velocidade linear e da angulação do volante. O principal tipo de ruído observado nos dados da odometria é o erro acumulativo, que pode resultar em imprecisões na estimativa da posição e na velocidade do veículo ao longo do tempo.

### ***Descrição de como o modelo de movimento ackerman pode ser usado iterativamente para estimar a pose do carro ao longo do tempo***

Primeiramente, precisamos estabelecer uma posição e orientação inicial para definirmos a pose inicial do veículo, podemos obter esses dados utilizando algum sistema de navegação. A partir disso, tendo os dados de velocidade linear e ângulo do volante conseguimos estimar a próxima pose. Se fizermos isso iterativamente, considerando sempre que a pose estimada anteriormente será a nova pose do veículo e utilizando as informações de posição e orientação, conseguimos estimar toda a pose do carro ao longo do tempo, da forma que o modelo de movimento ackermann propõe.

Entretanto, cabe ressaltar que todo esse processo, pode sofrer erros cumulativos, já que em cada estimativa tem-se um erro, que será perpetuado em todas as outras estimativas.