Um Método para Localização de Veículos Subaquáticos Baseado em Imagens Visíveis Aéreas e Acústicas Subaquáticas

Matheus Machado dos Santos[1], **Paulo Lilles Jorge Drews-Jr**[1], **Silvia Silva da Costa Botelho**[1]

[1]Universidade Federal do Rio Grande (FURG)
Centro de Ciências Computacionais (C3)
Av. Itália Km 8, 96203-900, Rio Grande, RS, Brazil

matheusmachado@furg.br, paulodrews@furg.br, silviacb@furg.br

Abstract. *This paper presents a cross-domain and cross-view framework for underwater robot localization, which does not require any Global Positioning System (GPS) information. The proposed localization method uses color aerial images and underwater acoustic images to estimate robot position. The method identifies the correlation among images from distinct domains, given by the matching of images acquired in partially structured environments with shared features. The validation of the proposed method is done using a real dataset, which was acquired by an underwater vehicle in a Marina. Besides, it was compared to Dead Reckoning and a new learning-based method. The experimental results present the feasibility of the proposed method and its advances in relation to the state-of-the-art algorithms.*

Resumo. *Este artigo apresenta um framework multi domínio e multi perspectiva para localização de robôs submarinos, que não requer nenhuma informação do Sistema de Posicionamento Global (GPS). O método de localização proposto usa imagens aéreas coloridas e imagens acústicas subaquáticas para estimar a posição do robô. O método identifica a correlação entre imagens de domínios distintos, dada pelo casamento de imagens adquiridas em ambientes parcialmente estruturados com características compartilhadas. A validação do método proposto é feita usando um conjunto de dados real, que foi adquirido por um veículo submarino em uma marina. Além disso, foi realizada a comparação com Dead Reckoning e um método baseado em aprendizado. Os resultados experimentais apresentam a viabilidade do método proposto e seus avanços em relação ao estado da arte.*

Student level PhD, concluded on August 3th, 2022

1. Introdução

A exploração de ambientes subaquáticos é uma tarefa desafiadora devido às características hostis do meio [Concha et al. 2015]. Especialmente para os humanos, a exploração dessas localidades exige cautela redobrada, e muitas vezes é necessário o uso de aparelhos especiais, como tanque de oxigênio e roupas de mergulho. Assim, o uso de Veículos Subaquáticos Não Tripulados (do Inglês *Unmanned Underwater Vehicle*,UUV), como os Veículos Operados Remotamente (do Inglês, *Remotely Operated Vehicle* ROV) e os Veículos Subaquáticos Autônomos (do Inglês *Underwater Autonomous Vehicle* AUV), surgem como uma solução viável, uma vez que afastam o homem das condições adversas do ambiente subaquático, sendo a maneira mais segura de realizar tarefas subaquáticas perigosas e de longo prazo.

O sucesso de um UUV realizar uma tarefa autônoma depende de sua localização. No entanto, o meio aquático impõe restrições que reduzem a capacidade de percepção dos sensores [Drews-Jr et al. 2015]. A principal dificuldade para a localização de UUVs está relacionada a rápida atenuação das ondas eletromagnéticas que impossibilita o uso do Sistema de Posição Global (GPS). Além disso, também reduz significativamente o alcance de sensores baseados em luz, como câmeras e lasers. Portanto, os sensores acústicos são os dispositivos mais utilizados para a percepção subaquática [C. S. Ribeiro et al. 2017], devido ao seu alcance e precisão que se destaca em relação a outros sensores. No entanto, seus dados apresentam níveis mais altos de ruído e são menos informativos do que outros sensores, como os ópticos [Hurtos et al. 2013].

A percepção de um ambiente não é limitada pela capacidade de um único veículo que trafega em um único meio [Wu 2019]. Destaca-se os métodos *Cross-view*, que realizam a correlação de dados a partir de perspectivas distintas, como por exemplo ambientes aéreos e terrestres, investigados em [Leung et al. 2008, Wolff et al. 2016, Hu et al. 2018]. Normalmente, imagens do *Street View* são localizadas realizando a correspondência de imagens aéreas georreferenciadas de drones ou de satélites [Hu et al. 2018].

Inspirados nos desafios da localização subaquática e considerando o sucesso dos métodos *cross-view*, que foram aplicados para casos de ambientes terrestres, propomos um método para lidar com o problema *cross-view* em ambientes subaquáticos. A metodologia proposta para localização de veículos subaquáticos utiliza imagens acústicas, adquiridas por um Sonar Multifeixe de Imageamento Frontal (do Inglês, Forward Looking Sonar FLS), e as compara com imagens de satélite georreferenciadas. Esta é uma solução viável para localização de robôs que navegam em ambientes parcialmente estruturados, como marinas e portos, pois possuem estruturas fixas (píeres, pedras e a linha da costa) que podem ser observadas em ambientes aéreos e subaquáticos.

Além do problema de *cross-view*, também lidamos com o problema de imagem de domínios diferentes *cross-domain*, ou domínios cruzados, uma vez que imagens acústicas e ópticas são adotadas. A principal dificuldade para o casamento de imagens de domínio cruzado é a discrepância entre as imagens ópticas acústicas e aéreas. Observe na Fig. 1 que as imagens aéreas são ricas em informações de textura e cor, enquanto as imagens acústicas fornecem apenas as distâncias e formas dos objetos detectados em uma imagem em escala de cinza. Nesta figura, as imagens de satélite são semanticamente segmentadas em três classes: corpo d'água, estruturas estacionárias e estruturas móveis. Em seguida, um algoritmo de localização baseado em Filtro de Partículas é introduzido para fornecer

a localização subaquática.

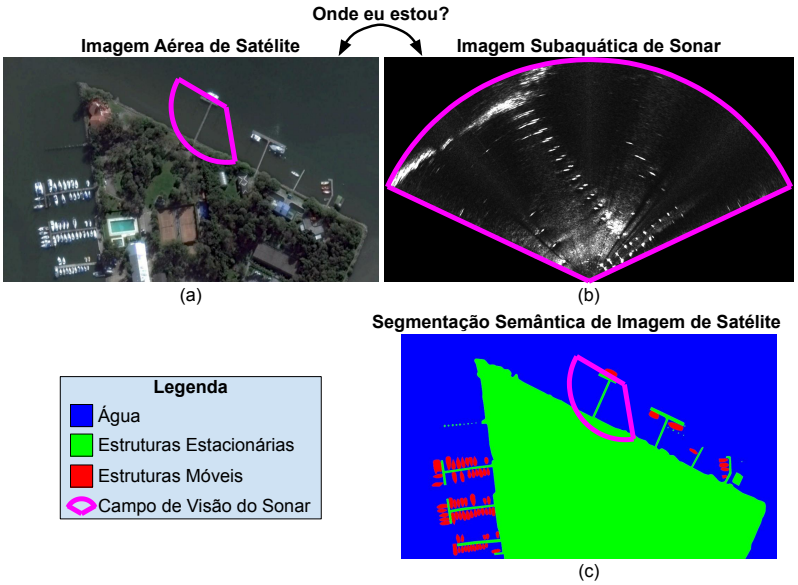

Figura 1. Localização de um robô submarino equipado com um sonar baseado na imagem aérea de satélite. (a) Imagem aérea de satélite, (b) Imagem acústica subaquática e (c) Imagem aérea de satélite semanticamente segmentada.

Assim, propusemos um framework de localização subaquática baseado em um mapa construído a partir de uma imagem aérea georreferenciada e o algoritmo de Filtro de Partículas [Thrun 2002]. Em [Dos Santos et al. 2021] adotamos a rede neural proposta em [De Giacomo et al. 2020] como modelo de observação do algoritmo de filtro de partículas. Propomos um modelo probabilístico ao invés de uma rede neural como modelo de observação do algoritmo de filtro de partículas. Adaptamos o *Likelihood Field Model* descrito em [Thrun 2002] para as imagens acústicas subaquáticas e comparamos com a abordagem de deep learning proposta por nós em [Dos Santos et al. 2021, Dos Santos et al. 2022].

O Filtro de Partículas, também conhecido como localização Monte Carlo, tem a capacidade de modelar uma função de distribuição probabilística multimodal não-Gaussiana espalhando hipóteses (partículas) em um mapa [Djuric et al. 2003]. As partículas nos permitem selecionar as regiões mais prováveis nas imagens de satélite (mapa) que correspondem à imagem acústica subaquática (percepção). Este método é escolhido por se adequar à associação de dados, permitindo recortar as imagens aéreas de forma que tenham um campo de visão compatível com as imagens acústicas. Observe, novamente na Fig. 1, que a imagem aérea cobre uma região maior do ambiente, o que a torna maior do que as imagens acústicas subaquáticas. Este problema é superado com o uso do Filtro de Partículas, pois apenas as regiões de posição das partículas são avaliadas.

O desempenho do método proposto é corroborado com sua aplicação em dados reais, que são coletados por um veículo submarino em uma marina, bem como em imagens aéreas ópticas de um satélite. Os resultados experimentais validaram o método apresentado, mostrando que ele pode se localizar no ambiente subaquático. Além disso, o método proposto alcança melhores resultados do que o dead reckoning do veículo, e do que tra-

balhos anteriores.

A organização deste trabalho é dada da seguinte forma: a Seção 2 discute os trabalhos relacionados, seguido do método proposto na Seção 3. Em seguida, na Seção 4, são apresentados os resultados experimentais, e as conclusões são apresentadas na Seção 5.

2. Trabalhos Relacionados

Como mencionado anteriormente, nosso método de localização de robôs submarinos baseado em cross-domain e cross-view é uma maneira pioneira de lidar com esse problema desafiador. Portanto, nesta seção, é feita uma discussão sobre propostas semelhantes para outros ambientes, que também enfrentam questões análogas.

Um dos principais desafios no método proposto é realizar a correspondência de cross-view e cross-domain. Na literatura, existem algumas soluções para o casamento cross-view de imagens terrestres e aéreas, cujo objetivo é obter a geolocalização da imagem terrestre, como apresentado em [Gao et al. 2018]. Neste trabalho, os autores apresentam uma revisão dos métodos existentes para tratar deste problema, e os classificam em métodos baseados em imagem e baseados em estrutura. A partir dessa classificação, foi demonstrado que a maioria dos métodos de correspondência de imagens aéreas e terrestres utiliza recursos auto-similares, recursos semânticos ou abordagens de redes neurais profundas.

Métodos que utilizam recursos auto similares detectam as estruturas mais relevantes das imagens, como fachadas de arranha-céus, presentes em ambas as vistas, conforme mostrado em [Bansal et al. 2016, Wolff et al. 2016]. Por exemplo, para combinar imagens aéreas e imagens de vistas de rua, esses métodos procuram cores e texturas semelhantes na fachada de alguns prédios. Por outro lado, métodos baseados em estruturas visam identificar estruturas compartilhadas em ambas as vistas, aérea e terrestre, como mostrado em [Leung et al. 2008, Noda et al. 2010]. Esse tipo de abordagem é comumente aplicado quando há recursos reduzidos para comparar as imagens do ambiente. Além disso, métodos baseados em semântica combinam informações externas sobre a cena observada para realizar o casamento das imagens, como métodos apresentados em [Lin et al. 2013, Castaldo et al. 2015]. [Castaldo et al. 2015] que exploram mapas do Sistema de Informações Geográficas (SIG) para localização de imagens terrestres, onde as imagens ópticas terrestres foram segmentadas e transformadas em uma visão top-down. Além disso, os autores também propuseram descritores de layout de segmento semântico (SSL) para combinar os mapas SIG com as imagens de top-down e encontrar as correspondências com as imagens terrestres de rua. Alternativamente, [Lin et al. 2013] localizou as imagens ópticas terrestres usando uma abordagem baseada em tradução de recursos de visualização cruzada. Nesta metodologia, foram exploradas imagens aéreas e os mapas de atributos de cobertura do solo para realizar a geolocalização das imagens terrestres.

Diferentemente dos métodos discutidos anteriormente, os métodos de aprendizado profundo realizam a correspondência com base nos recursos aprendidos dos conjuntos de dados de imagem. Pioneiro neste tópico, [Lin et al. 2015] propôs a primeira rede de aprendizado profundo, chamada *Where-CNN*, que aprende a incorporação de recursos para correspondência de imagens. Além disso, [Workman and Jacobs 2015, Workman et al. 2015] obteve resultados em estado da arte para geolocalização de imagens

terrestres em área amplas, utilizando uma interessante estratégia de treinamento de visão cruzada para aprender uma representação de características semânticas conjuntas para imagens aéreas. De forma semelhante, focado em edifícios urbanos, [Tian et al. 2017] apresentou uma abordagem de correspondência de imagens terrestres e aéreas usando a rede neural Faster R-CNN [Ren et al. 2015], para detectar edifícios, e uma rede siamesa [Chopra et al. 2005], para aprender os recursos dos edifícios em ambas as vistas, aéreas e terrestres. As correspondências foram obtidas utilizando a os K casos mais similares, abordagem conhecida como k-vizinhos mais próximos (KNN).

Uma localização de imagens terrestres baseada em estruturas ortogonais, e usando imagens aéreas como referências, para áreas urbanas é proposta por [Leung et al. 2008]. Neste método, ortoimagens aéreas são usadas para gerar um mapa de características, e os limites do edifício foram identificados pela Transformada Probabilística Progressiva de Hough (PPHT) [Matas et al. 2000]. Assim, os pontos de fuga e as linhas encontradas são avaliados para definir as orientações das paredes que são usadas como observação no filtro de partículas, descrito em [Thrun 2002], que localiza as imagens do solo. Outro estudo interessante foi feito pelo [Noda et al. 2010], onde imagens aéreas foram utilizadas para geolocalização de veículos terrestres. Para isso, as imagens aéreas das estradas foram utilizadas para gerar um mapa. As imagens dos veículos foram transformadas em uma visão de cima para baixo, e a correspondência das imagens dos veículos e mapas de características foi realizada com o descritor SURF [Bay et al. 2008].

Com o objetivo de reduzir a dimensão de imagens de um mesmo domínio, [De Giacomo et al. 2020] apresentou um método de treinamento cooperativo de múltiplas redes. Em geral, foi uma otimização colaborativa de duas redes neurais, que possuem a mesma arquitetura. No entanto, cada rede possui seus próprios pesos, atualizados por uma função objetivo de tripletes. As redes treinadas são usadas para extrair vetores que codificam as imagens alimentadas nelas. A partir dele, a distância entre os vetores de extração foi calculada pela distância euclidiana. O principal resultado desta rede é o matching e o ranking das imagens compradas, que são duas tarefas fundamentais para o presente trabalho.

Como pode ser observado, a localização de imagens acústicas subaquáticas usando imagens aéreas de satélite têm aspectos semelhantes à localização de imagens terrestres com imagens aéreas. Esses dois problemas envolvem correspondência de visões cruzadas (cross-view). No entanto, neste trabalho também estamos tratando imagens de multi-domínios (cross-domain), o que torna a tarefa de correspondência mais desafiadora. Além disso, os dados do sonar não fornecem informações de cor ou textura, como imagens aéreas, restringindo o uso de métodos baseados nos recursos das imagens. Nosso trabalho é inspirado em abordagens baseadas em estrutura, pois estruturas podem ser facilmente identificadas em imagens de sonar. Nossa abordagem utiliza essas estruturas para gerar um mapa a partir de imagens aéreas, em um framework baseado no algoritmo adaptativo de localização de Monte Carlo (do Inglês, Adaptive Monte Carlo Localization AMCL).

3. Metodologia

O framework proposto para localização subaquática baseada em imagens aéreas é mostrado na Fig. 2. Os dados de entrada consistem na imagem aérea I^a cobrindo o ambiente

navegável, imagens acústicas I_t^s observadas a cada instante t pelo veículo e o sinal de controle μ_t, acionado no instante t pelo sistema de controle do veículo. Alternativamente, o sinal μ_t pode ser uma estimativa de deslocamento do veículo. O método estima a localização mais provável x_t do veículo para o momento t.

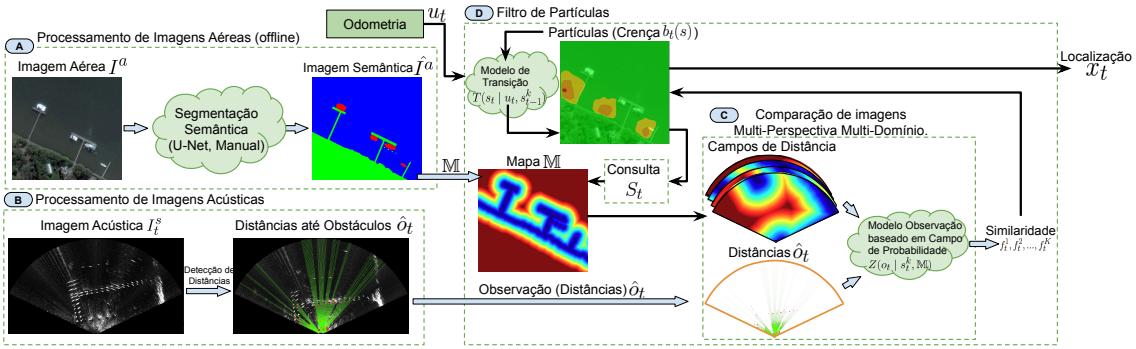

Figura 2. Proposta do framework multi-vista (cross-view), vista superior e frontal da cena, e multi-domínio (cross-domain), óptico e acústico, para localização subaquática de um veículo utilizando imagens acústicas de um sonar e imagens aéreas de satélite.

Assim como no algoritmo de filtro de partículas, o método mantém um mapa estático $\mathbb{M}$ do ambiente, um conjunto de K partículas $s_t^1, s_t^2, ..., s_t^K \in S_t$ para cada instante t, bem como um modelo de transição $T()$, e um modelo de observação $Z()$. A crença $b_t(s)$ de uma partícula s no instante t é dada pela distribuição de probabilidade,

$$b_t(s) \approx \langle s_t^k, w_t^k \rangle_{k=1:K}, \tag{1}$$

onde $\sum_k w_k = 1$; K é o número total de partículas; s_k é o estado da partícula; w_k é o peso (ou importância) da partícula; e t denota a tempo. O estado da partícula adotado neste trabalho consiste na posição x, y e orientação θ, ou seja, $s = [x\ y\ \theta]^T$.

As partículas mantidas pelo método são constantemente atualizadas de forma Bayesiana [Thrun 2002]. Além disso, as partículas são amostradas usando o modelo de transição,

$$s_t^k \sim T(s_t \mid u_t, s_{t-1}^k), \tag{2}$$

que considera o estado anterior da partícula s_{t-1}^k, e o último sinal de controle μ_t. Em seguida, o modelo de observação, dado por

$$f_t^k = Z(o_t \mid s_t^k, \mathbb{M}), \tag{3}$$

estima a probabilidade f_t^k da observação o_t estar no estado da partícula s_t^k, considerando o mapa de referência do ambiente $\mathbb{M}$. As observações são as leituras do sonar feitas pelo veículo durante a navegação, e o mapa é construído a partir das imagens aéreas que cobrem o ambiente.

O peso, ou importância, das partículas é atualizado considerando a probabilidade f_t^k para que

$$w_t^k = \eta f_t^k, \tag{4}$$

onde η é o fator de normalização dos pesos calculados, considerando $\eta^{-1} = \sum_{j=1:K} f_t^j$.

O framework proposto possui quatro etapas que permitem a execução do filtro de partículas no contexto do problema deste trabalho.

- **Passo A** realiza a segmentação semântica das imagens aéreas I^a. A segmentação é realizada *offline*, apenas uma vez para cada região em que o veículo irá navegar. As imagens são segmentadas em três classes: estruturas estacionárias, estruturas móveis e água. A segmentação permite a remoção de objetos em movimento, como barcos e navios, e utiliza estruturas estacionárias como mapa de referência para a localização do veículo debaixo d'água. Por ser uma etapa única e não depender dos dados coletados durante a navegação;
- **Passo B** é executado constantemente para cada imagem de som I_t^s capturada no instante t ao longo da missão. Consiste no pré-processamento das imagens. Ele lida com o problema do ruído da imagem, como apenas a forma dos objetos com uma resposta acústica clara, procedendo ao processo de correspondência da etapa C como uma observação o_t;
- **Etapa C** representa o modelo de observação do filtro de partículas, mostrado em (3). A observação atual $\hat{o_t}$, originada de uma imagem acústica, é comparada com o mapa estático $\mathbb{M}$ da cena, originado de imagens aéreas. A observação atual $\hat{o_t}$ é projetada no mapa estático $\mathbb{M}$ considerando o estado s_t^k da partícula k. Sobrepondo as informações, estima-se a similaridade f_t^k para a partícula k. Este processo é realizado para todas as partículas no tempo t. O resultado final desta etapa é a atualização dos pesos das partículas, conforme (4);
- **Passo D** é baseado no algoritmo de filtro de partículas. A localização mais provável do veículo é estimada pelo algoritmo de Monte Carlo adaptativo [Thrun 2002]. Várias hipóteses de localização do veículo estão espalhadas na forma de partículas no mapa $\mathbb{M}$. À medida que novas imagens acústicas são adquiridas, a confiança das partículas é atualizada através das correspondências calculadas na etapa C.

4. Resultados

Nosso método é avaliado em uma missão real submarina do conjunto de dados ARACATI 2017 usando o Robot Operating System (ROS) [Quigley et al. 2009].

4.1. Dataset ARACATI 2017

O conjunto de dados ARACATI 2017 [Dos Santos et al. 2022] foi registrado no Iate Clube da Cidade do Rio Grande, no Brasil, com um Seabotix Little Benthic Vehicle LBV 300-5 e um Forward-Looking Sonar (FLS) BlueView P900. O veículo foi fixado abaixo de uma prancha flutuante, de forma que permanece submerso durante toda a missão enquanto um Sistema de Posição Global Diferencial (DGPS) é mantido na parte superior da prancha, fora da água, coletando medições precisas de localização. O veículo percorreu 485 m a uma velocidade máxima de 0,6 m/s e avaliamos o desempenho do nosso método usando imagens acústicas, bússola e DGPS em uma missão de 29 minutos.

4.2. Resultados Experimentais

Nosso método é avaliado no conjunto de dados ARACATI 2017 adotando dois modelos de observação: o modelo de observação probabilística e uma abordagem baseada em aprendizado profundo [Dos Santos et al. 2022] que aprende o modelo de observação. Ambos os métodos são comparados com a odometria do veículo e com a referência de localização do DGPS.

A tabela 1 mostra os parâmetros adotados no experimento[1]. Enquanto o método de aprendizado profundo espalhou no máximo 120 partículas devido a limitações de hardware, o método probabilístico simulou 5.000 partículas usando o mesmo hardware. A Fig. 3 mostra o caminho resultante, em metros, onde a linha laranja é o dado adotado como verdade (referência) em todas as subfiguras. Fig. 3-(a) a Fig. 3-(c) mostra o dead reckoning; método baseado em deep learning [Dos Santos et al. 2022]; nosso método baseado em probabilidade, respectivamente. A Fig. 4 mostra o erro de localização, em metros, do dead reckoning e os dois métodos avaliados em relação ao DGPS.

Parâmetro	Valor	Descrição
		Preparação de Dados Aéreos
d_{max}	20 metros	Distância Máxima ao Obstáculo
		Preparação de Dados Subaquáticos (Sonar Teledyne BlueView P900-130)
K	256	Num. de *beams* avaliados
l_{start}	180	Posição inicial do *beam* (pixels)
w_{sz}	2	Tamanho da janela de suavização (pixels)
i_{min}	180	Threshold para detectar objetos (valor do pixel)
θ_{max}	130	Campo de visão horizontal (graus)
p_{max}	50	Alcance máximo (metros)
		Correspondência
z_{hit}	0,5	Peso do erro de medição do sonar (valor normalizado)
z_{rand}	0,5	Peso da medição aleatória do sonar (valor normalizado)
σ_{hit}	0,2	Desvio padrão Gaussiano. (metros)

Tabela 1. Parameters value adopted on the experiment

Os resultados mostraram que o algoritmo do filtro de partículas melhorou a localização do veículo ao combinar imagens ópticas aéreas e acústicas submarinas. Conforme observado na Fig. 4, na maioria das vezes o erro de localização é menor do que o dead reckoning.

Quando o veículo se move para uma área aberta, o algoritmo baseia-se em puro dead reckoning. Isso acontece porque o método é baseado na correspondência de imagens aéreas e subaquáticas de estruturas estacionárias. Quando o veículo trafega em área aberta, as estruturas estão fora do alcance do sonar e, portanto, não são detectadas pelo nosso sistema. Assim, nenhuma correlação é detectada e o peso de todas as partículas é diminuído na mesma proporção. Neste momento, apenas o modelo de transação afeta a estimativa de pose, atualizando o estado das partículas usando o sinal de controle do veículo μ_t. As partículas começam a se espalhar ao longo da imagem aérea.

[1] Um vídeo do experimento 8x mais rápido está disponível em `https://youtu.be/-nC0DkgszmE`

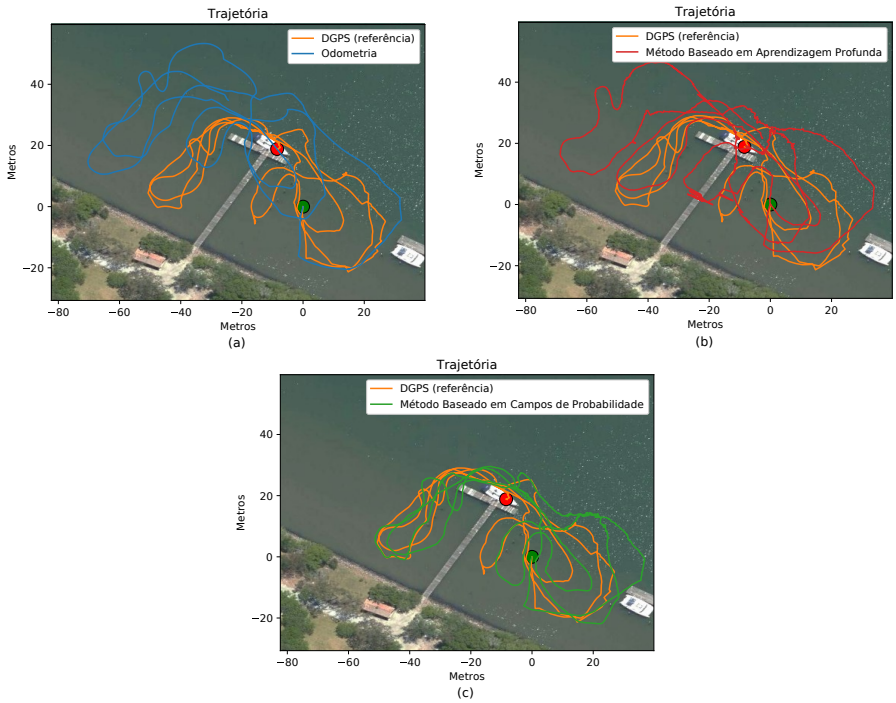

Figura 3. Trajeto do veículo, em metros, estimado por: (a) dead reckoning, (b) método utilizando modelo de observação baseado em deep learning estado da arte [Dos Santos et al. 2022], (c) o método de modelo de observação probabilístico proposto. A imagem de fundo mostra o ambiente, o norte aponta para cima e o leste aponta para a direita.

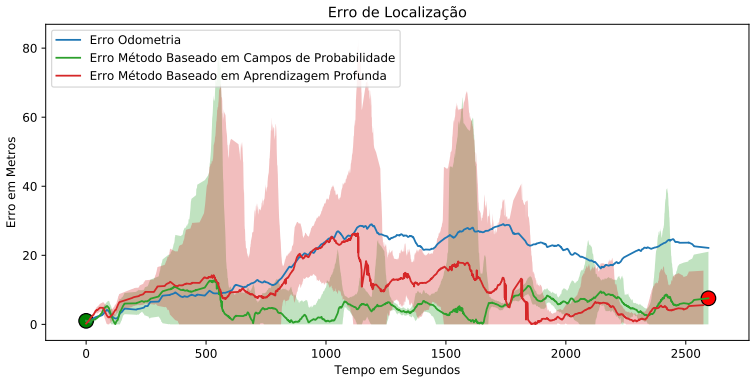

Figura 4. Erros de localização.

No entanto, quando o veículo retorna à região estruturada, graças às correlações das imagens, o modelo de observação aumenta o peso das partículas mais próximas das estruturas e diminui o peso das partículas longe das estruturas, resultando em uma melhor estimativa da localização do veículo . Este efeito é observado após 500 segundos na Fig. 4. Quando o veículo está viajando do cais leste para oeste na Fig. 3.

Os píeres são semelhantes e representam um desafio à nossa abordagem. Especi-

ficamente nos resultados da Fig. 3-(b), quando o veículo chega próximo ao píer, grupos de partículas são divididos entre o píer atual e o anterior, levando o sistema ao maior erro de localização devido a um problema de ambiguidade. Porém, assim que o veículo se movimenta e adquire novas imagens acústicas, as partículas do píer atual passam a ter um fator de maior importância fazendo com que as partículas do píer anterior sejam reamostradas para o píer correto. Este efeito gera os picos no gráfico de erro presente na Fig. 4. Os vales mostram que nosso método pode relocalizar após períodos sem ver nenhuma estrutura.

A Fig. 4 mostra que nosso método, que adota o *Likelihood Field Model* como modelo de observação, teve um desempenho melhor que os métodos anteriores.

Os experimentos foram executados em um computador com Processador Ryzen 7 2700x e GPU NVIDIA RTX 3070. O método baseado em rede neural [Dos Santos et al. 2022] executou o modelo de observação em GPU. O experimento com NVIDIA RTX 3070 mostrou que a rede pode avaliar 120 pares de imagens acústicas subaquáticas e aéreas ópticas por segundo (tempo de inferência de 8,3ms).

O modelo de observação proposto é executado na CPU em 0,003 ms. O processo de transformação da imagem acústica em distâncias definidos no **Passo C** na Seção 3 é executado em 2,85 ms. Assim, a abordagem probabilística proposta neste estudo é mais rápida do que trabalhos anteriores baseados na abordagem de aprendizado profundo [Dos Santos et al. 2022]. Nosso método obteve melhor localização e desempenho computacional do que os métodos baseados em redes neurais.

5. Conclusões

Este artigo apresenta um novo método baseado no *Likelihood Field Model* para o método de localização de veículos subaquáticos baseado em imagens cross-view e cross-domain. O método é capaz de localizar um veículo subaquático em um ambiente parcialmente estruturado com imagens acústicas de um sonar de imageamento frontal (FLS) e imagens aéreas ópticas do ambiente capturadas por satélite. Além disso, o método proposto é comparado com outros dois métodos da literatura e os superou. Em trabalhos futuros, planejamos avaliar nosso método em novos experimentos em diferentes lugares e realizar testes em plataformas embarcadas como NVIDIA Jetson Xavier ou mesmo Jetson Nano.

Referências

Bansal, M., Daniilidis, K., and Sawhney, H. (2016). Ultrawide baseline facade matching for geo-localization. In *Large-Scale Visual Geo-Localization*, pages 77–98. Springer.

Bay, H., Ess, A., Tuytelaars, T., and Gool, L. V. (2008). Speeded-up robust features (surf). *CVIU*, 110(3):346 – 359.

C. S. Ribeiro, P. O., M. dos Santos, M., L. J. Drews-Jr, P., and S. C. Botelho, S. (2017). Forward looking sonar scene matching using deep learning. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 574–579.

Castaldo, F., Zamir, A., Angst, R., Palmieri, F., and Savarese, S. (2015). Semantic cross-view matching. In *IEEE ICCVw*, pages 9–17.

Chopra, S., Hadsell, R., and LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *IEEE CVPR*, volume 1, pages 539–546.

Concha, A., Drews-Jr, P., Campos, M., and Civera, J. (2015). Real-time localization and dense mapping in underwater environments from a monocular sequence. In *OCEANS 2015 - Genova*, pages 1–5.

De Giacomo, G. G., dos Santos, M. M., Drews-Jr, P. L., and Botelho, S. S. (2020). Cooperative training of triplet networks for cross-domain matching. In *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*, pages 1–6. IEEE.

Djuric, P. M., Kotecha, J. H., Zhang, J., Huang, Y., Ghirmai, T., Bugallo, M. F., and Miguez, J. (2003). Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38.

Dos Santos, M. M., De Giacomo, G. G., Drews-Jr, P. L., and Botelho, S. S. (2022). Cross-view and cross-domain underwater localization based on optical aerial and acoustic underwater images. *IEEE Robotics and Automation Letters*, 7(2):4969–4974.

Dos Santos, M. M., De Giacomo, G. G., Drews-Jr, P. L., Botelho, S. S., and Mello, C. D. (2021). A framework for underwater vehicle localization based on cross-view and cross-domain acoustic and aerial images. In *2021 Latin American Robotics Symposium (LARS), 2021 Brazilian Symposium on Robotics (SBR), and 2021 Workshop on Robotics in Education (WRE)*, pages 204–209. IEEE.

Drews-Jr, P., Nascimento, E. R., Campos, M. F. M., and Elfes, A. (2015). Automatic restoration of underwater monocular sequences of images. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1058–1064.

Gao, X., Shen, S., Hu, Z., and Wang, Z. (2018). Ground and aerial meta-data integration for localization and reconstruction: A review. *Pattern Recognition Letters*, 127:202–214.

Hu, S., Feng, M., Nguyen, R. M. H., and Hee Lee, G. (2018). CVM-Net: Cross-view matching network for image-based ground-to-aerial geo-localization. In *IEEE CVPR*, pages 7258–7267.

Hurtos, N., Nagappa, S., Cufi, X., Petillot, Y., and Salvi, J. (2013). Evaluation of registration methods on two-dimensional forward-looking sonar imagery. In *OCEANS - Bergen, 2013 MTS/IEEE*, pages 1–8.

Leung, K. Y. K., Clark, C. M., and Huissoon, J. P. (2008). Localization in urban environments by matching ground level video images with an aerial image. In *IEEE ICRA 2008*, pages 551–556.

Lin, T.-Y., Belongie, S., and Hays, J. (2013). Cross-view image geolocalization. In *IEEE CVPR*, pages 891–898.

Lin, T.-Y., Cui, Y., Belongie, S., and Hays, J. (2015). Learning deep representations for ground-to-aerial geolocalization. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Matas, J., Galambos, C., and Kittler, J. (2000). Robust detection of lines using the progressive probabilistic hough transform. *CVIU*, 78(1):119–137.

Noda, M., Takahashi, T., Deguchi, D., Ide, I., Murase, H., Kojima, Y., and Naito, T. (2010). Vehicle ego-localization by matching in-vehicle camera images to an aerial image. In *ACCV*, pages 163–173.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5. Kobe, Japan.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc.

Thrun, S. (2002). Probabilistic robotics. *Communications of the ACM*, 45(3):52–57.

Tian, Y., Chen, C., and Shah, M. (2017). Cross-view image matching for geo-localization in urban environments. In *IEEE CVPR*, pages 3608–3616.

Wolff, M., Collins, R. T., and Liu, Y. (2016). Regularity-driven building facade matching between aerial and street views. In *IEEE CVPR*, pages 1591–1600.

Workman, S. and Jacobs, N. (2015). On the location dependence of convolutional neural network features. In *IEEE CVPRw*, pages 70–78.

Workman, S., Souvenir, R., and Jacobs, N. (2015). Wide-area image geolocalization with aerial reference imagery. In *The IEEE International Conference on Computer Vision (ICCV)*.

Wu, Y. (2019). Coordinated path planning for an unmanned aerial-aquatic vehicle (UAAV) and an autonomous underwater vehicle (AUV) in an underwater target strike mission. *Ocean Engineering*, 182:162 – 173.