FLArobot - Interação Homem-Robô com drone Tello

Arthur Jung[1], **Daniel Ribeiro**[1], **Gabriel Schmidt**[1], **Hugo Fernandes**[1]

[1] Instituto de Informática
Universidade Federal de Goiás (UFG) – Goiânia, GO – Brazil

{jung, daniel.ribeiro, gabriel_schmidt, hugo.fernandes}@discente.ufg.br

Abstract. *This article describes a project for the subject of Robotic Perception and Action involving the use of a drone for human-robot interaction in order to instigate its control by means of human gestures. The drone used was DJI's Tello integrated with CNN to detect pre-established movements.*

Resumo. *Este artigo descreve um projeto para disciplina de Percepção e Ação Robótica envolvendo o uso de drone para interação humano-robô afim de instigar seu controle por meio de gestos do ser humano. O drone utilizado foi o Tello da DJI integrado com CNN para detecção de movimentos pré-estabelecidos.*

1. Introdução

Os fiscalizadores de aeronaves comerciais desempenham um papel indispensável na pista, servindo como ponte visual entre os pilotos e a tripulação de terra. Munidos com varinhas de sinalização – faróis portáteis iluminados – eles comunicam instruções vitais aos pilotos, desde desacelerar e virar, até parar ou desligar os motores. Esta sinalização visual, caracterizada pela sua intuitividade e redundância, garante clareza e minimiza erros. Este cenário resume o potencial de aproveitar os movimentos do corpo humano como entradas de comando dentro dos sistemas, estabelecendo as bases para a avançada Interação Humano-Robô (HRI). O HRI, como domínio, investiga as complexidades da incorporação de humanos no circuito de controle, influenciando assim o comportamento robótico. Uma extensão deste paradigma é a Interação Humano-Enxame (HSI), que amplifica o desafio ao integrar um enxame de robôs, garantindo que a sua autonomia inerente permaneça inalterada.

Tendo o desafio em vista, o projeto consistiu em utilizar o drone Tello da DJI integrado com técnicas de visão computacional para, por meio da câmera do próprio drone, inferir suas movimentações aéreas a partir de movimentos mapeados.

O dataset utilizado consiste em 1115 imagens de poses pré-definidas. Os testes seguiram duas abordagens, uma com 10% dos dados anotados e outra por meio de abordagens empíricas ao observar se o drone, com o script de interação humano-robô em execução, se movimentava de acordo com o esperado para cada comando gestual por diferentes pessoas.

2. Revisão Bibliográfica

Essa seção visa elencar um referencial teórico base para o desenvolvimento do projeto, os tópicos a seguir nos oferecem ferramentas, definições e aplicações para a Integração Humano-Robô com drones.

- **Human-Drone Interaction (HDI): Opportunities and Considerations in Construction** [1]: Esse capítulo fornece uma visão geral abrangente das diferentes aplicações da tecnologia de drones, também descreve elementos de interação humano-drone, áreas de pesquisa e oportunidades, bem como resume as considerações de interação enquanto propõe um roteiro de pesquisa futura com foco em canteiros de obras;
- **Human-Robot Interaction Based on Gaze Gestures for the Drone Teleoperation** [2]: Este artigo introduz gestos oculares como uma estratégia de seleção de objetos na Interação Humano-Robô para a teleoperação de drones;
- **Collocated Human-Drone Interaction: Methodology and Approach Strategy** [3]: Este artigo baseia-se em trabalhos anteriores e investiga como um robô voador deve se aproximar de uma pessoa. Também apresenta uma taxonomia de metodologias para estudos de interação humano-drone para guiar futuros pesquisadores na área;
- **The State-of-the-Art of Human–Drone Interaction: A Survey** [4]: Este trabalho começa com uma análise e comparação dos modelos de drones que são comumente usados por usuários finais e pesquisadores no campo da interação humano-drone. Em seguida, discute-se o estado atual do campo, incluindo os papéis dos humanos na interação humano-drone (HDI), métodos de controle inovadores, aspectos remanescentes da interação e novos protótipos e aplicações de drones;
- **Drone.io: A Gestural and Visual Interface for Human-Drone Interaction** [5]: Apresenta e descreve o processo de design do drone.io, uma interface gráfica de usuário centrada no corpo projetado para interação humano-drone. Usando dois gestos simples, os usuários podem interagir com um drone de maneira natural;
- **HRI in the sky: Creating and commanding teams of UAVs with a vision-mediated gestural interface** [6]: Apresenta um sistema de criação, modificação e comando de equipes de drones por um humano não instrumentalizado. Utilizam visão computacional para capturar o rosto do usuário e para controlar os drones.

3. Fundamentos Teóricos

Para integrar adequadamente os movimentos do drone com os movimentos humanos, foram utilizadas redes neurais, destacando-se dois tipos principais: Multi-Layer Perceptron (MLP) e Convolutional Neural Network (CNN) com camadas convolucionais de uma dimensão (conv1d).

No caso do Multi-Layer Perceptron, foi empregada a técnica Grid Search com validação cruzada para selecionar os melhores parâmetros. O MLP é uma rede neural utilizada para tarefas de aprendizado supervisionado, como classificação e regressão. Ele consiste em várias camadas de neurônios que transformam dados de entrada em saídas desejadas por meio de pesos aprendidos e funções de ativação.

A CNN com camadas convolucionais de uma dimensão (conv1d) é um modelo de aprendizado profundo projetado para processar dados sequenciais. Utiliza filtros convolucionais para extrair automaticamente características relevantes das sequências de dados de entrada. Esse tipo de CNN é particularmente eficaz para dados temporais ou sequências unidimensionais, como sinais de áudio ou séries temporais.

Para a detecção de landmarks, foi utilizado o algoritmo BlazePose do MediaPipe [7]. MediaPipe é um framework de código aberto para criar pipelines que realizam inferência de visão computacional em dados sensoriais, como vídeo ou áudio.

Utilizamos a função de ativação ReLU (Rectified Linear Unit) nas tentativas com CNN. A ReLU é importante para introduzir não-linearidade no treinamento dos modelos.

Essas técnicas representam a base teórica necessária para a interpretação adequada da interação humano-drone, que será detalhada na seção Metodologia.

4. Metodologia

O dataset foi construído pelos próprios integrantes da equipe, os quais mapearam poses para diferentes ações que o drone deveria tomar para ser capaz de agir em tasks de pouso, levantamento de voo, e se locomover no espaço 3D.

Um script foi construído para extração de landmarks de cada pose. Seu funcionamento consistiu em deixar a câmera ligada, e em tempo real fotos são tiradas, e a cada foto tirada, os landmarks são extraídos e armazenados num arquivo CSV.

O arquivo CSV gerado foi pós-processado para que apenas os pontos relevantes fossem considerados (afinal o algoritmo utilizado do MediaPipe também detecta landmarks faciais, pontos da mão, que não seriam necessários para execução do desafio). Não só isso, mas imagens consideradas não representantes de quaisquer poses foram excluídas, e os registros do CSV foram atualizados com base no restante relevante.

No total, 8 comandos foram mapeados com poses para deixar o drone parado, fazê-lo subir, descer, pousar, se afastar, se aproximar, ir para direita e ir para esquerda, com o adicional de 1 gesto bônus com o propósito de passar para outro drone a interpretabilidade de comandos (i.e., fazer com que o drone detector pare de interpretar os gestos e outro drone passe a interpretá-los, se tornando o detector). O motivo do gesto adicional foi um intuito de primeiro passo a trabalhos futuros os quais podem consistir em não se limitar a interação humano-robô, e possibilitar porta de entrada para interação humano-enxame.

O dataset com os landmarks anotados foi utilizado para treinamento da MLP e da CNN. A MLP foi treinada durante 1000 iterações para garantir convergência. Os valores de Alpha no Grid Search (5 folders) foram 0.001, 0.01 e 0.1, já a learning rate foi definida como 0.01, 0.1 e 0.2. Os melhores resultados foram 0.001 para alpha e 0.01 para learning rate, com 94% para acurácia, sendo que neste caso, o MinMaxScaler foi utilizado para normalizar input das features.

A CNN foi treinada durante 1000 épocas com CrossEntropyLoss e Adam Optimizer, cuja arquitetura foi construída com Pytorch, a qual consistiu principalmente em 3 blocos (2 de conv1d e 1 para fully connected), com uma uma flatten aplicada entre o segundo e o terceiro bloco.

5. Resultados

Para avaliação dos resultados, foram utilizadas duas abordagens principais, a primeira foi a divisão dos dados anotados em treino e teste, sendo a base de teste correspondente a 10% do total das imagens adquiridas. Os pontos trackeados do usuário em cada imagem

foram passados para a MLP e a CNN, obtendo os resultados de acurácia de 0.94 e 0.97, respectivamente.

A princípio, ambos os resultados foram satisfatórios, porém era preciso uma avaliação em um cenário real. Para isso, realizamos testes de captura de imagens com o drone e inferência com os modelos em tempo real, também avaliamos a clareza da identificação dos gestos e posterior movimentação do drone. A partir disso, foi perceptível um desempenho superior da CNN nos cenários avaliados, uma vez que foram identificadas inconsistências nas inferências da MLP nas posições esquerda e direita.

Por fim, foi realizado um teste prático de controle gestual do drone, onde se tinha o objetivo de guiá-lo através de uma sala para pouso em uma plataforma, o teste ocorreu com sucesso e o resultado pode ser verificado em vídeo.

6. Conclusão

Portanto, com base nos resultados obtidos, entende-se que o trabalho aqui descrito é útil em uma ampla gama de aplicações, uma vez que demonstra o uso de redes neurais simples e visão computacional para controle de aeronaves e drones de forma gestual. Também, reforça-se o objetivo de uso de abordagens com baixo custo computacional, já que o pipeline aqui desenvolvido pode ser executado em processadores simples com baixa latência.

Por fim, considera-se que o trabalho atingiu os objetivos propostos e também cria oportunidades para implementações futuras, como o controle de um enxame de drones e o uso de tais equipamentos em cenários diversos, como agricultura, engenharia civil, segurança pública, dentre outros.

7. Referências

[1]: ALBEAINO, G. et al. **Human-Drone Interaction (HDI): Opportunities and Considerations in Construction**. 2022. Automation and Robotics in the Architecture, Engineering, and Construction Industry. Springer, Cham. Disponível em: https://doi.org/10.1007/978-3-030-77163-8_6. Acesso em 09 jul. 24.

[2]: YU, M. et al. **Human-Robot Interaction Based on Gaze Gestures for the Drone Teleoperation**. 2014. School of Automation, Beijing Institute of Technology, Beijing, China. Disponível em: https://core.ac.uk/download/pdf/158976476.pdf. Acesso em 09 jul. 24.

[3]: WOJCIECHOWSKA, A. et al. **Collocated Human-Drone Interaction: Methodology and Approach Strategy**. 2019. 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), pp. 172-181, Disponível em: https://ieeexplore.ieee.org/abstract/document/8673127. Acesso em 09 jul. 24.

[4]: TEZZA, D. et al. **The State-of-the-Art of Human–Drone Interaction: A Survey**. 2019. IEEE Access, vol. 7, pp. 167438-167454. Disponível em: https://ieeexplore.ieee.org/abstract/document/8903295. Acesso em 09 jul. 24.

[5]: CAUCHARD, J. R. et al. **Drone.io: A Gestural and Visual Interface for Human-Drone Interaction**. 2019. 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), pp. 153-162. Disponível em: https://ieeexplore.ieee.org/abstract/document/8673011. Acesso em 09 jul. 24.

[6]: MONAJJEMI, V. M. et al. **HRI in the sky: Creating and commanding teams of UAVs with a vision-mediated gestural interface**. 2013. IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 2013, pp. 617-623. Disponível em: https://ieeexplore.ieee.org/abstract/document/6696415. Acesso em: 09 jul. 24.

[7]: Google AI Edge. **MediaPipe**. GitHub. Disponível em: https://github.com/google-ai-edge/mediapipe. Acesso em 09 jul. 24.