

# ESTIMATIVA DE POSE HUMANA 3D EM AMBIENTE CIRÚRGICO A PARTIR DE IMAGENS RGB-D

Guilherme Soares Silvestre

Prof. Dr. Glauco A. P. Caurin, Paulo H. Polegato

Escola de Engenharia de São Carlos - Universidade de São Paulo

[gsoares.silvestre@usp.br](mailto:gsoares.silvestre@usp.br) [gcaurin@sc.usp.br](mailto:gcaurin@sc.usp.br) [paulopolegato@usp.br](mailto:paulopolegato@usp.br)

## Objetivos

O conceito de colaboração robótica almeja um método que tange o entendimento do ambiente, a comunicação e a sincronização das tarefas entre as partes [1]. Em ambientes semi-estruturados, como ambientes cirúrgicos, necessita-se do desenvolvimento de recursos que permitam ao sistema robótico “perceber” o ambiente e tomar decisões as quais não coloquem em risco os operantes.

Quando esse conceito é somado com determinado nível de autonomia, de forma que as tarefas possam ser automatizadas, há o aumento da complexidade do sistema [1]. Nessa linha, a motivação desse projeto é o melhoramento dos meios de entendimento do ambiente, visando a aplicação em ambiente cirúrgico para o robôs colaborativos utilizados em cirurgias.

O objetivo é buscar o entendimento da estrutura corporal dos operantes na sala de cirurgia a partir de imagens do ambiente. Os requisitos devem prezar pela velocidade de resposta e precisão adequada. O sistema também deve ser capaz de se comunicar com outras aplicações.

## Métodos e Procedimentos

Os métodos e procedimentos valem-se, primeiramente, da pesquisa bibliográfica sobre os trabalhos existentes. Também fazem parte do escopo do projeto o estudo das ferramentas que permitem o desenvolvimento do sistema.

O estado da arte trata esse campo como um problema de estimativa de pose humana tridimensional em ambientes com alta ocorrência de oclusão. Isso ocorre devido à proximidade dos operantes durante os procedimentos [2].

A literatura apresenta métodos inscritos no grande campo de aprendizado de máquina. Tal como outros desafios que envolvem visão computacional, o processamento de imagens utiliza majoritariamente ferramentas de *deep learning*. O objetivo das redes desenvolvidas é buscar a estrutura corporal dos humanos em imagens em lugares variados [3][4].

A aplicação em ambiente cirúrgico promove desafios para esses métodos. A homogeneidade das cores presentes nos jalecos e uniformes médicos é uma característica que dificulta a detecção dos membros corporais. Isso resulta em uma diminuição da precisão nessa situação nos modelos treinados e validados em datasets contendo imagens de ambientes gerais [2].

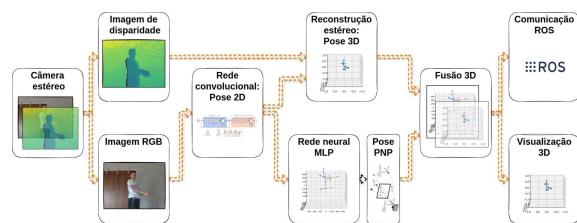


Figura 1. Processo arquitetônico para o método proposto.

Entende-se por estrutura corporal no ambiente as coordenadas tridimensionais dos membros e a relação com seus adjacentes, formando as poses corporais de todas as

pessoas presentes na imagem. O processo de extensão da informação projetada na imagem buscando a reconstrução da pose no espaço constitui um dos principais desafios do projeto.

A aplicação de câmeras RGB-D permite a utilização do canal de profundidade dos *pixels*. Ou seja, isso promove o maior entendimento espacial do ambiente, permitindo o aumento do desempenho do sistema [2].

Por fim, o estudo das ferramentas de comunicação entre sistemas e a visualização dos resultados completam o desenvolvimento da arquitetura proposta. Nesse campo, as ferramentas do *ROS* fizeram parte do estudo e desenvolvimento.

## Resultados

O método desenvolvido faz uso de trabalhos validados na literatura. Uma sequência de processamento foi arquitetada (Figura 1).

As imagens RGB e disparidade são coletadas da câmera *Intel Realsense D435i* utilizando o *RealSense SDK 2.0*. A imagem colorida é a entrada de uma rede convolucional (*OpenPose* [3] ou *Mask R-CNN* [4]) que prediz a pose dos operantes na imagem (pose 2D).

Essa informação é passada para dois métodos paralelos. A partir da imagem de disparidade, calcula-se a reprojeção dos *pixels* referentes a cada operante, alcançando as coordenadas espaciais.

Paralelamente, uma rede neural MLP (*SeffPose* [5]) recebe a pose 2D e infere a estrutura no espaço (pose 3D) em relação à base local de cada pessoa. Então, segue-se com estimativa da localização das bases locais, a partir do método PNP, obtendo as coordenadas em relação à câmera.

Com os dois resultados obtidos, busca-se a fusão das poses 3D, conservando as características de melhor desempenho em cada método. Por fim, a comunicação é feita com a estrutura do *ROS*. A visualização dos resultados é feita utilizando o *Matplotlib*.

## Conclusões

O projeto segue em andamento, buscando a validação em *datasets* montados para o ambiente cirúrgico, como o *MVOR*. Como também, busca-se o aprimoramento das partes que apresentarem desempenho insatisfatório como alta complexidade de tempo ou baixa precisão.



Figura 2. Aplicação do método em imagens de procedimentos médicos.

Com resultados favoráveis, buscamos a aplicação em robôs colaborativos. Nesse campo, a aplicação de robôs colaborativos é contra-intuitiva diante da excelência produtiva alcançada pelos robôs autônomos solitários. Entretanto, acreditamos que o futuro das aplicações robóticas necessitem do melhoramento das interfaces homem-robô. Essa ideia promove a motivação necessária para esse projeto.

## Referências Bibliográficas

- [1] Haidegger, T. (2019). Autonomy for Surgical Robots: Concepts and Paradigms. *IEEE Transactions on Medical Robotics and Bionics*, 1(2), 65–76. <https://doi.org/10.1109/tmrb.2019.2913282>
- [2] Kadkhodamohammadi, A., & Padoy, N. (2021). A generalizable approach for multi-view 3D human pose regression. *Machine Vision and Applications*, 32(1). <https://doi.org/10.1007/s00138-020-01120-2>
- [3] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172–186. <https://doi.org/10.1109/TPAMI.2019.2929257>
- [4] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386–397. <https://doi.org/10.1109/TPAMI.2018.2844175>
- [5] Martinez, J., Hossain, R., Romero, J., & Little, J. J. (2017). A Simple Yet Effective Baseline for 3d Human Pose Estimation. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob, 2659–2668. <https://doi.org/10.1109/ICCV.2017.288>