

Tema:

**Deep learning analysis of RGBD image sequences for pose detection and tracking**

#### MOTIVAÇÃO

No domínio da visão computacional, a detecção de pose e o rastreamento são dois desafios de alta relevância. Suas aplicações variam desde a indústria de jogos (AR/VR), até técnicas de *data analytics* visando análise de intenção.

#### OBJETIVOS

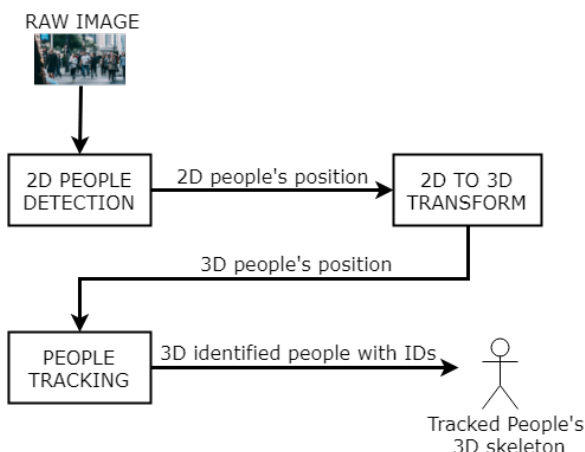
Nesse contexto, os objetivos do projeto são:

- O estudo de técnicas do estado da arte em detecção de pose e rastreamento temporal;
- A construção de um sistema que, a partir de imagens RGBD, permita realizar a detecção e o rastreamento da pose das pessoas em 3D, respeitando requisitos de processamento em tempo real.

#### FRAMEWORK

O framework proposto é composto por três módulos principais e um módulo auxiliar de visualização 3D. A entrada global do sistema é uma imagem RGBD, proveniente de uma câmera Kinect v1. Os três módulos principais são:

1. **Módulo de detecção 2D** (*2D People Detection*): obtém poses 2D a partir de imagens RGB;
2. **Módulo de extração da profundidade** (*2D to 3D Transform*): projeta as poses no plano de coordenadas do mundo real ( $x, y, z$ );
3. **Módulo de rastreamento 3D** (*People Tracking*): rastreia as pessoas a partir das poses 3D.



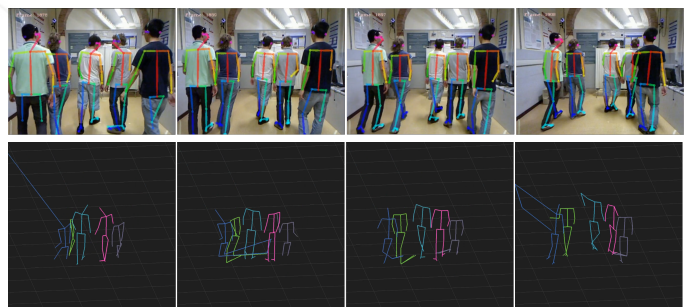
#### TÉCNICAS EMPREGADAS

Para cada um dos três módulos, diversas técnicas foram estudadas e testadas, objetivando a escolha da técnica com menor custo computacional e maior precisão de detecção e rastreamento.

- **O módulo de detecção 2D** foi construído usando a rede neural convolucional proposta por Zhe Cao et al. [1], que se baseia no modelo de Part Affinity Fields (PAF) descrito no mesmo artigo;
- **O módulo de extração da profundidade** realiza a projeção dos pontos do sistema de coordenadas da imagem  $(i, j)$  para pontos no sistema de coordenadas do mundo real  $(x, y, z)$  a partir dos parâmetros intrínsecos da câmera Kinect v1;
- **O módulo de rastreamento 3D** utiliza a filtragem de Kalman Unscented (UKF, em inglês), como proposto por Julier e Uhlmann [2] para realizar o rastreamento temporal;
- **O módulo auxiliar de visualização 3D** foi construído em OpenGL, e permite a visualização da cena 3D em tempo real a partir de qualquer ponto de vista.

#### RESULTADOS

Com a construção e validação de cada módulo, o framework completo foi construído e testado em datasets de imagens RGBD da câmera Kinect v1. O resultado abaixo foi obtido a partir do dataset KTP [3].



#### REFERÊNCIAS

- [1] Cao Z. et al. "Realtime multi-person 2d pose estimation using part affinity fields" (2016).
- [2] Julier S. J. e Uhlmann J. K. "New extension of the kalman filter for nonlinear systems". International Society for Optics and Photonics (1997), p.182-184.
- [3] Padova University "KTP Dataset". Acessado em: 15/04/2019. Disponível em <<http://www.dei.unipd.it/munaro/KTP-dataset.html>>

Integrantes: [Alexsander Henrique Baldin](#)  
[Guilherme Kurike Matsumoto](#)

Orientador: Prof. Dr. Jaime Simão Sichman