Marker-less Pose Estimation

Andy Gilbert, Simon Kalouche, Patrick Slade

Introduction

- Capturing human motion is beneficial for biomechanics and assistive devices
- Motion capture systems accurate but expensive and uses a preset space
  - Easily occluded by objects in way of camera
- IMU's offer general information but too noisy
- EgoCap uses cameras worn on a helmet[1]
  - Less noisy, portable, and avoids occlusion
-

Problem Statement

- We will develop a marker-less method for 3D pose estimation
- Use two networks for 2D joint predictions and 2D to 3D body pose estimation
- Evaluate based on estimation accuracy

Datasets

- MPII dataset of 20k images with body joint annotations used to learn heatmaps[2]
- EgoCap dataset of 70k+ images from a pair of fish-eye cameras in a first person style[1]

Fig. 1: EgoCap image capture setup

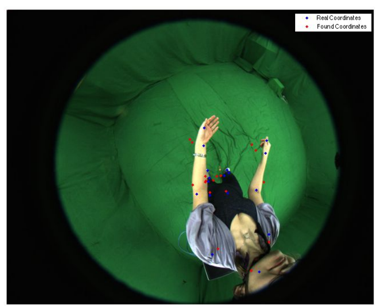

Fig. 2: EgoCap image with joints

Methods

- Resnet trained on MPII
- FC layers zeroed and retrained on EgoCap
- Hyperparameter tuning performed on EgoCap training
- K-nearest neighbor (KNN) and neural networks (NN) used to map 2D joint predictions to 3D pose estimations

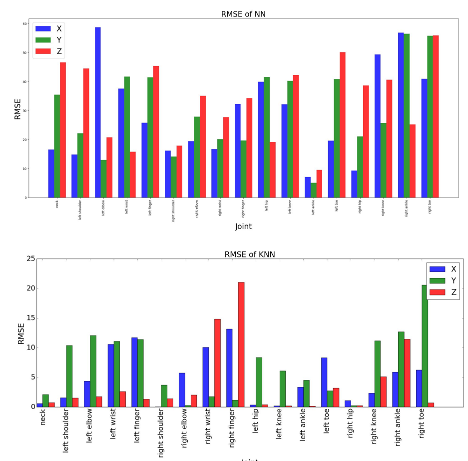

Fig. 4: 2D to 3D Error (mm) of NN & KNN

Findings

- Average euclidean image pixel distance error in joint estimations less than 100
- KNN accuracy = 5.23 mm
- NN accuracy = 29.6 mm

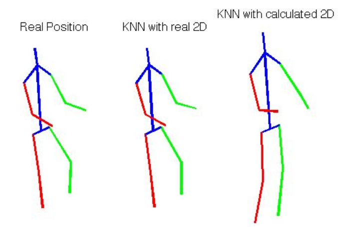

Fig. 3: 3D frame of pose estimation corresponding to Fig. 2

Conclusion

- Achieved accurate joint heat map predictions
- Compared methods for 2D to 3D prediction
- Marker-less pose estimation with cm level accuracy

1. Rhodin, Helge, et al. "EgoCap: egocentric marker-less motion capture with two fisheye cameras." *ACM Transactions on Graphics* 2016
2. Andriluka, Mykhaylo, et al. "2d human pose estimation: New benchmark and state of the art analysis." *IEEE Conf. on CV and Pattern Recognition*. 2014.