

Demo Video

End-to-end Dynamic Matching Network for Multi-view Multi-person 3d Pose Estimation

Anonymous ECCV submission

Paper ID 6258

This document is to demonstrate the environmental settings and model steps used in the demo video for paper 6258. The settings are listed as below:

- The recording environment is an indoor laboratory.
- Four cameras with 226 fps and 1440×1080 resolution are placed at different corners of the ceiling and connected with two Apple-mini servers.
- An NVIDIA DevBox remote server with four GeForce GTX 1080 GPUs is deployed to handle the images captured by these four cameras.
- A local machine which has an i7-9600 CPU, 16G RAM and two Geforce GTX 1080 GPUs is used to run the process control code and model proposed in the above paper and present the pose estimation results.

There are six steps of the system:

- The local machine warms up the camera system and starts the program.
- The cameras capture the images at the same time and send them to the DevBox for synchronization.
- The synchronized images of the same frame are sent to YOLOv3 and CPN for 2d human pose detection.
- Both the images and the results from the 2d estimator are sent to the local machine for matching, which groups the 2d poses and heatmaps by people's identities.
- On the local machine, the grouped heatmaps are sent to a learnable matrix-weighted triangulation module to estimate the 3d poses along with the previously detected 2d poses.
- The local machine shows the estimated 3d poses on its screen.

The video presents a testing demo of the proposed model under real settings. The average frame rate of this system in the video is around 5 fps.