3D Convolutional Neural Networks for Efficient and Robust Hand Pose Estimation from Single Depth Images

Lihao Ge, Hui Liang, Junsong Yuan, and Daniel Thalmann
Institute for Media Innovation, Nanyang Technological University, Singapore

Introduction

- **Objective**: to estimate 3D hand pose in real-time from single depth images.
- **Motivation**: image-based features extracted by 2D CNNs are not directly suitable for 3D hand pose estimation due to the lack of 3D spatial information.
- **2D CNN**:
  - Depth image
  - 3D Joint Locations
  - 3D Points
  - Volumetric Representation

- **Multi-view CNNs** [1]: still cannot effectively exploit 3D spatial information in the depth image; the computational complexity will be increased when using more views.

- **Our Approach**: we propose a 3D CNN-based approach that can capture the 3D spatial structure of the input and accurately regress full 3D hand pose in a single pass.

Methodology

- **Volumetric Representation**
- **3D Data Augmentation**

Experiments

- **Self-comparisons**
- **Comparisons with State-of-the-art**
- **Qualitative Results**

Acknowledgement This research is supported by the BeingTogether Centre, a collaboration between Nanyang Technological University and University of North Carolina at Chapel Hill. This work is also supported in part by Singapore Ministry of Education Academic Research Fund.

References