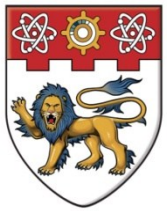


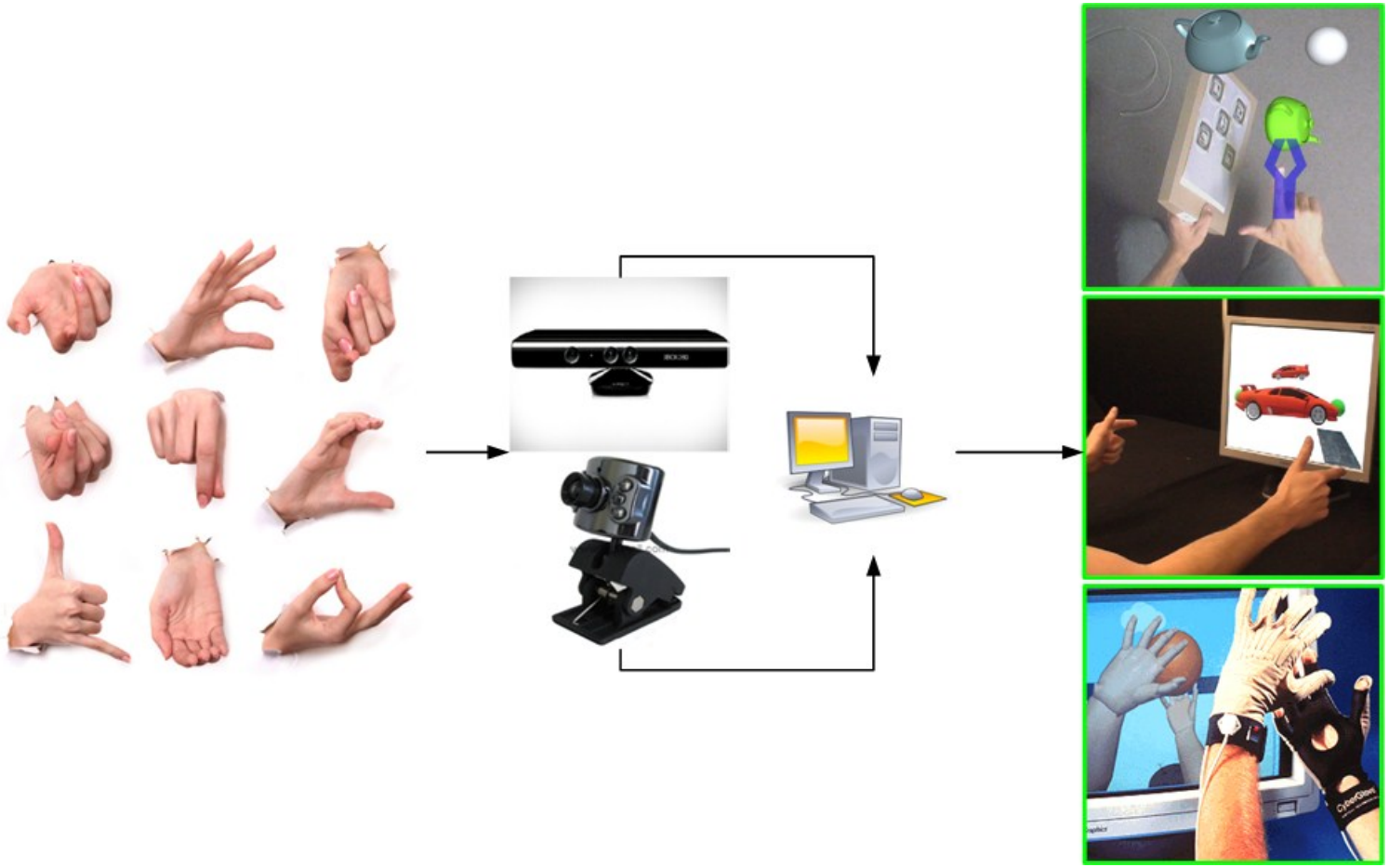
NANYANG TECHNOLOGICAL UNIVERSITY

Hand Parsing and Gesture Recognition with the Depth Camera

LIANG HUI
– School of Electrical & Electronic Engineering
– Institute for Media Innovation

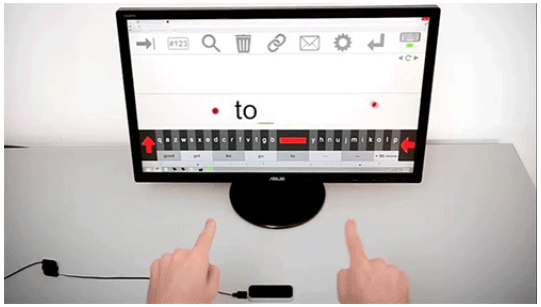
Supervisors:
Prof. Junsong Yuan
– School of Electrical and Electronic Engineering
Prof. Daniel Thalmann
– Institute for Media Innovation

Introduction



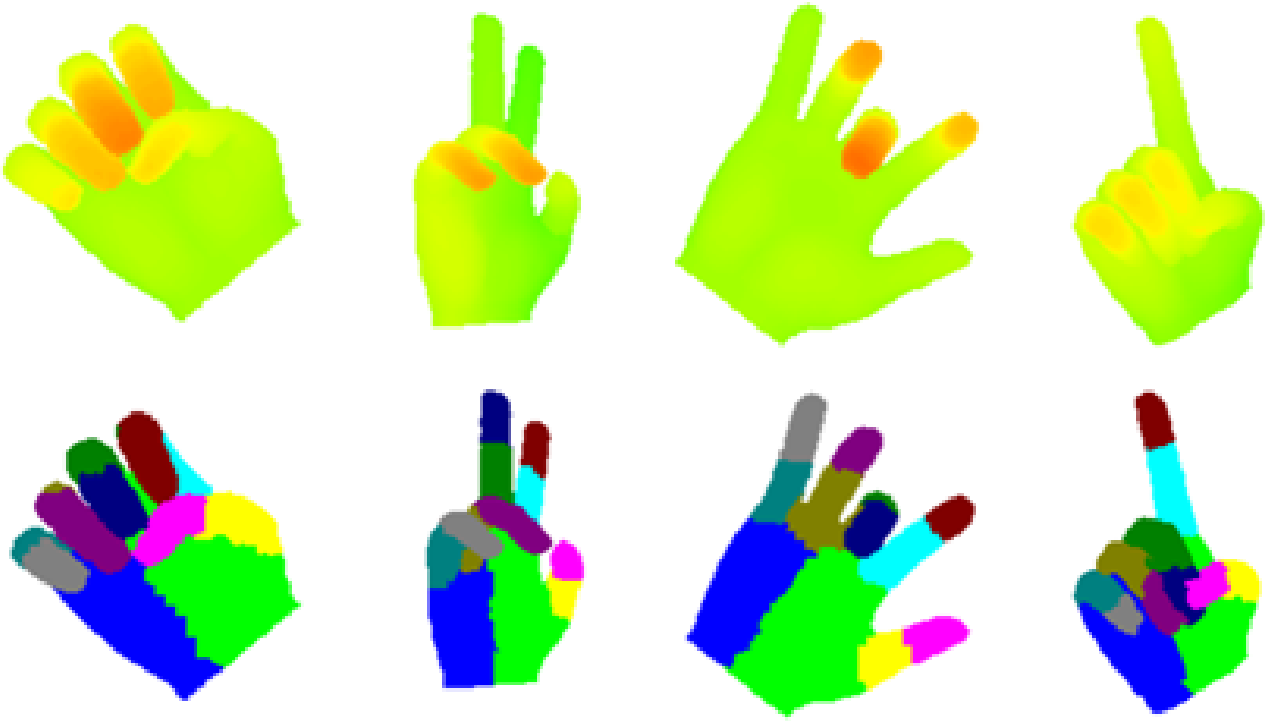
Applications

- Remote gesture control
  - Don’t have to look for remote
  - No battery to run down
- Sign language recognition
  - Enable the disabled to communicate with the computers
- Virtual keyboard
  - Compact in size
  - Convenient for mobile devices



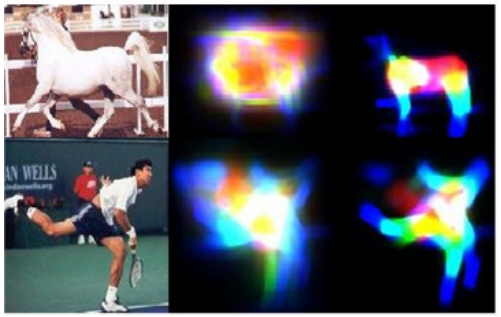
Problem Formulation

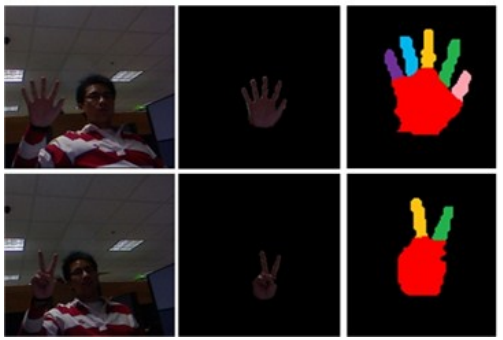
- How to classify each pixel in the input image into a pre-defined hand part set?



Related Work

- RGB image based approaches
  - Use of color markers [1]
  - Shape and contour analysis [2]
  - Inference with pictorial structures [3, 4]
- Disadvantages
  - Inconvenient for users
  - Low performance due to ambiguity in RGB images





[1] R. Y. Wang et al., Real-time hand-tracking with a color glove, in ACM ToG, 2009.
[2] Z. Ren, et al., Minimum near-convex decomposition for robust shape representation, in ICCV 2011.
[3] D. Ramanan, Learning to parse images of articulated bodies, in NIPS 2006.
[4] M. Eichner, et al., 2D Articulated human pose estimation and retrieval in (almost) unconstrained still images, in IJCV 2012.

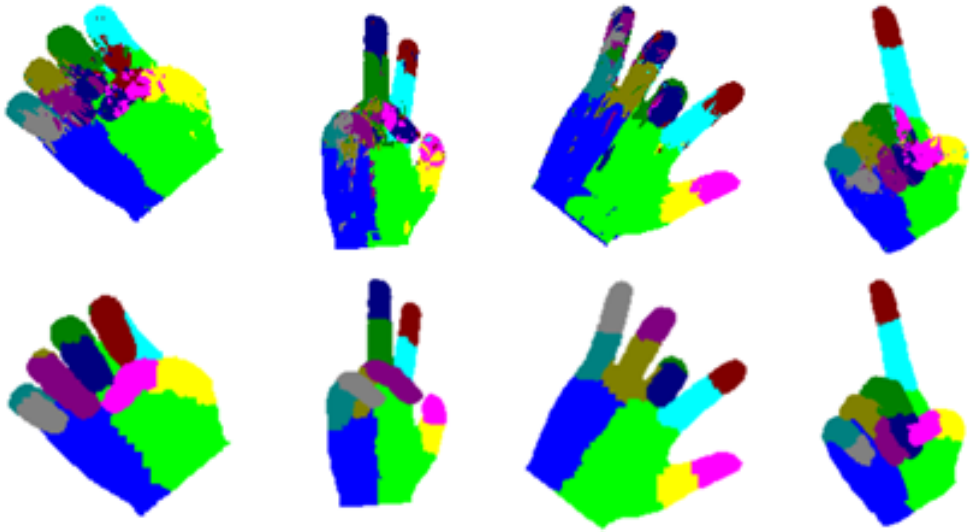
Related Work

- Depth image based approaches
  - Per-pixel classification with the depth context [5, 6]
  - Fuse temporal and spatial constraints for per-pixel classification [7]
- Problem
  - Per-pixel classification is still quite noisy

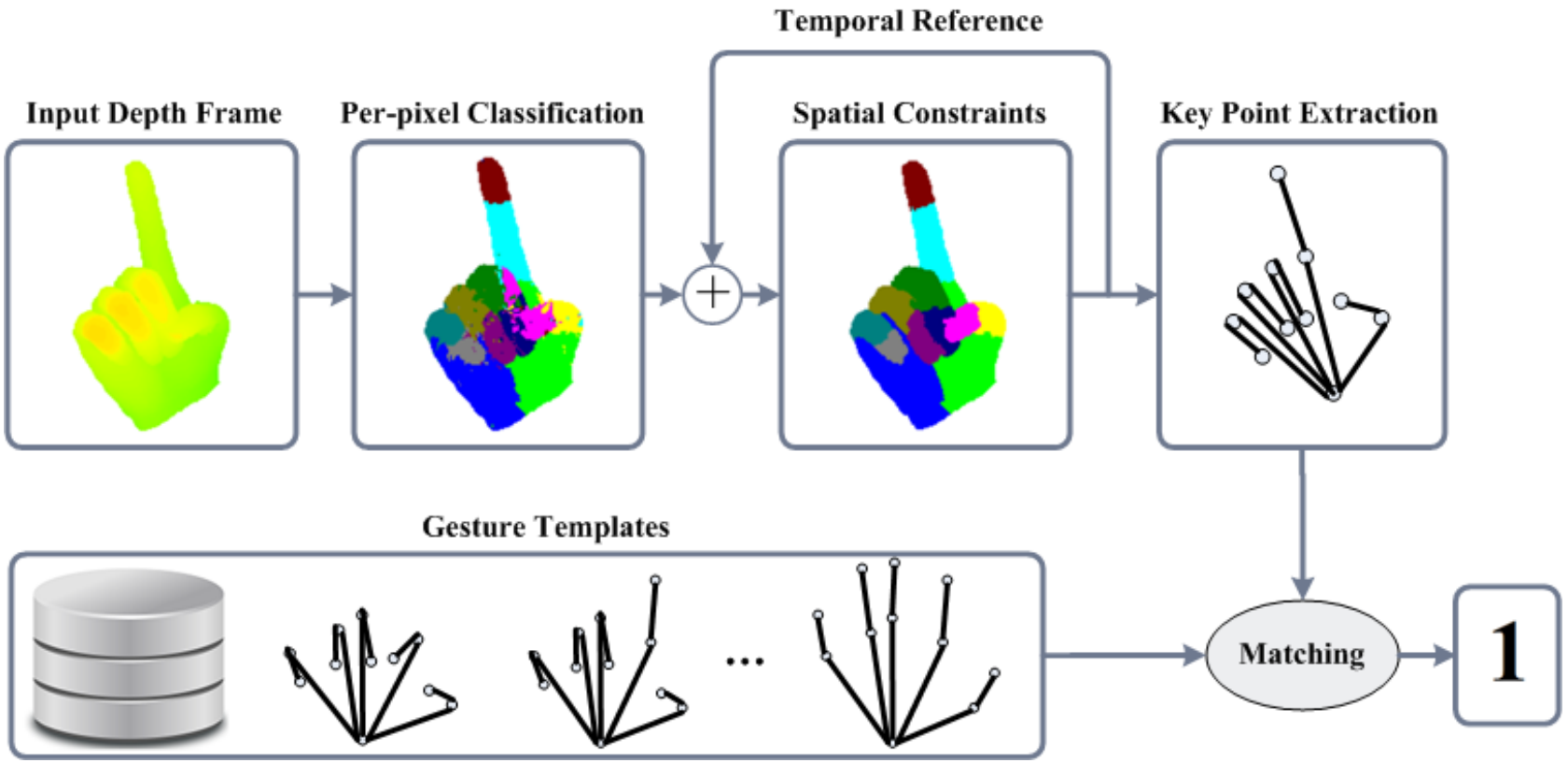
[5] Y. Yao, et al., Real-time hand pose estimation from RGB-D sensor, in Proc. of ICME, 2012.
[6] C. Keskin, et al., Real-time hand pose estimation using depth sensors, in ICCV 2011.
[7] H. Liang, et al., Model-based Hand Pose Estimation via Spatial-temporal Hand Parsing and 3D Fingertip Localization, in VCJ 2013.

The Proposed Parsing Scheme

- Input with a single depth camera
- No color markers are required
- Enforce spatial & temporal constraints on per-pixel results
- Linear computation complexity with the number of pixels



The Framework



Quantitative results on synthetic data

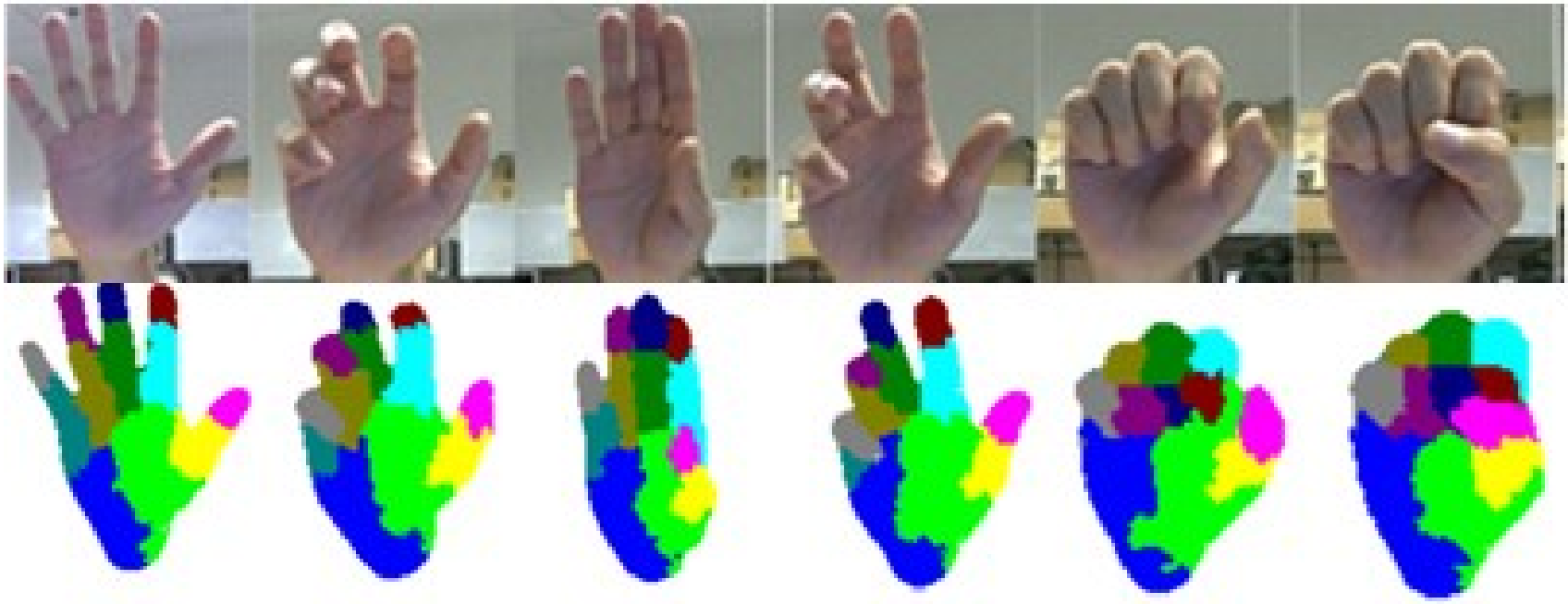
- Single-frame based hand parsing

Method	Accuracy
Per-pixel DR [6]	72.0%
The Proposed scheme	89.2%

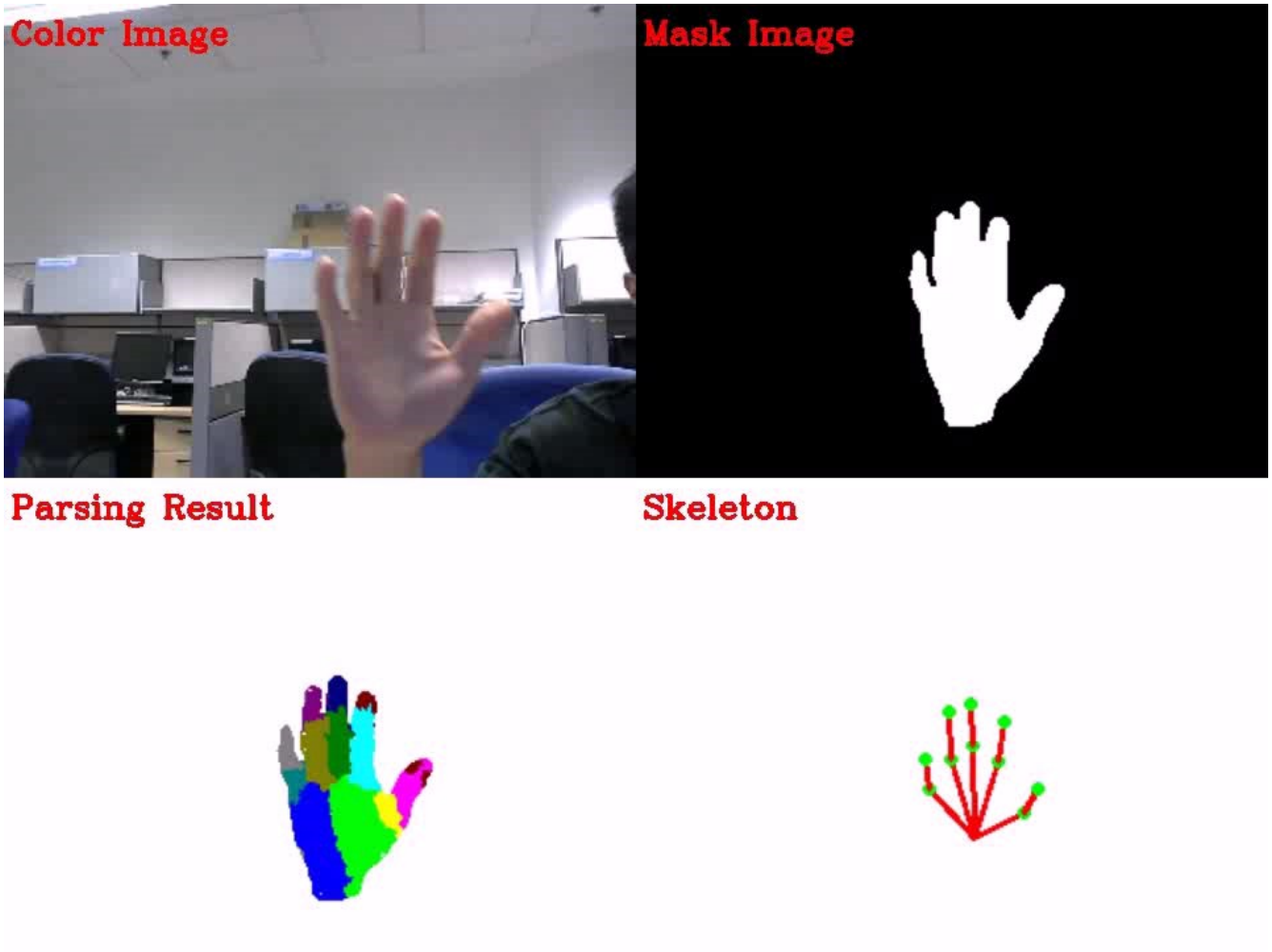
- Hand parsing in continuous sequences

	Per-pixel DR [6]	Single-frame parsing	Spatial-temporal parsing
Seq. 1	77.2%	87.1%	89.6%
Seq. 2	81.7%	90.8%	91.5%

Qualitative results on real-world video

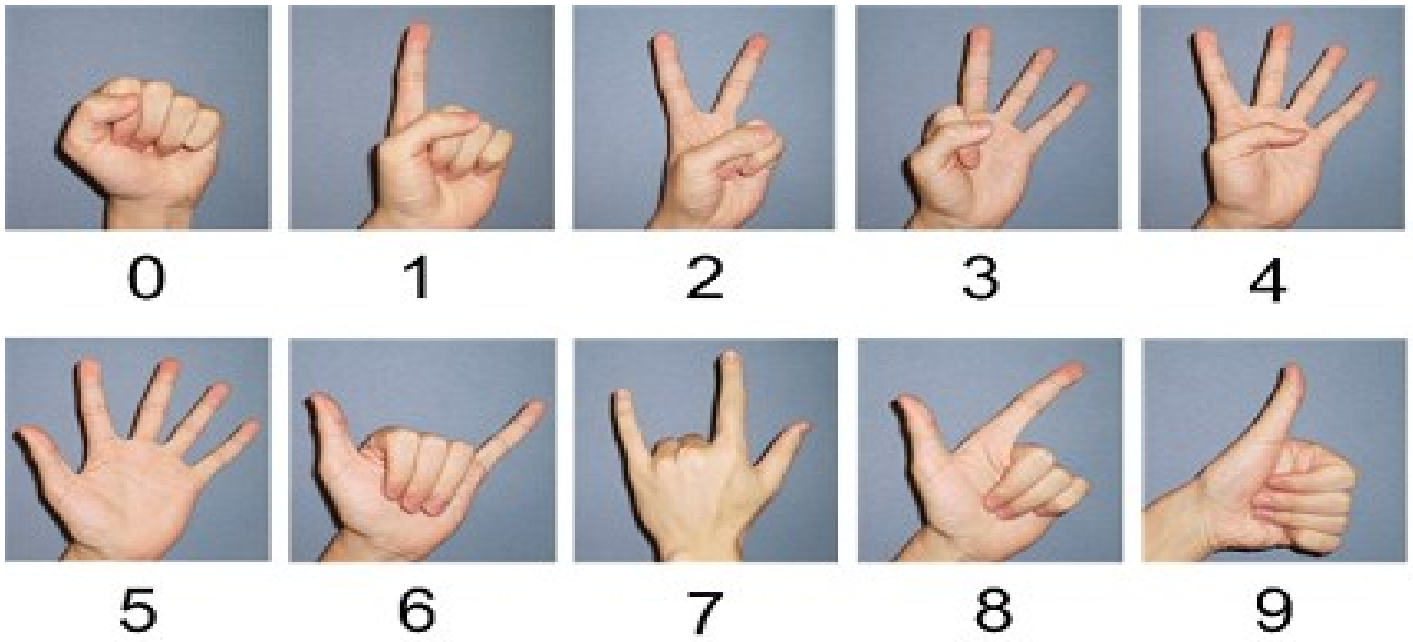


Qualitative results on real-world video

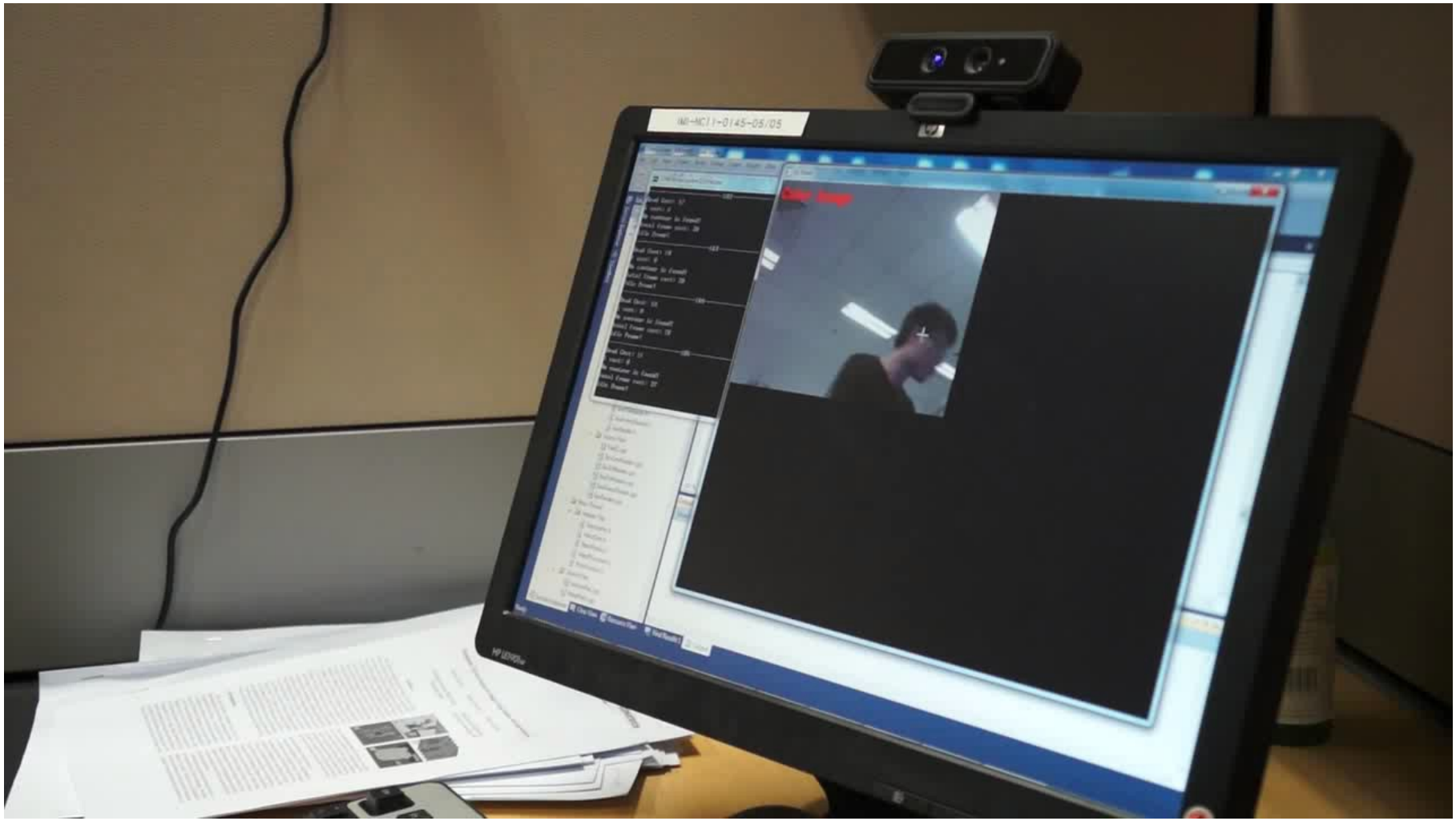


Demo

- Digit Gesture Recognition
  - Recognize ten digit gestures in real-time with high accuracy
  - The user can input with either one or two hands



Demos



Thank You!

NANYANG
TECHNOLOGICAL
UNIVERSITY