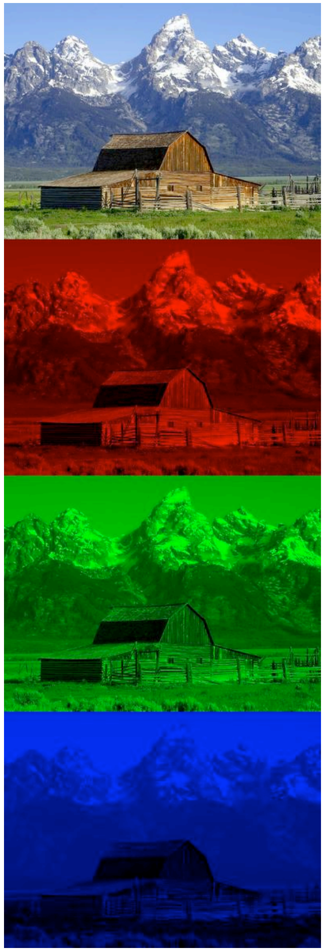
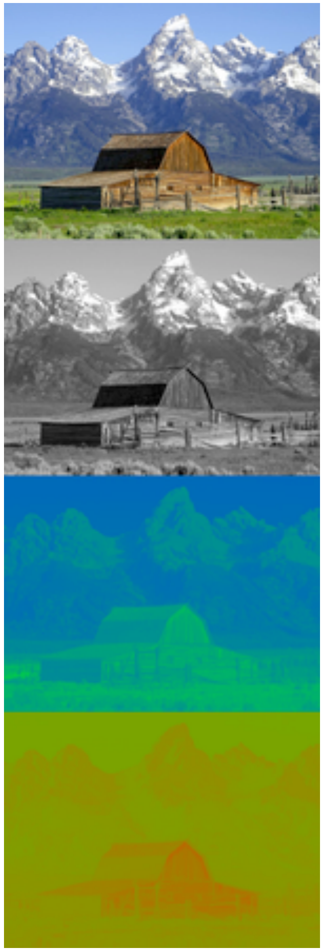
RGBD Tutorial

14210240041 Gu Pan

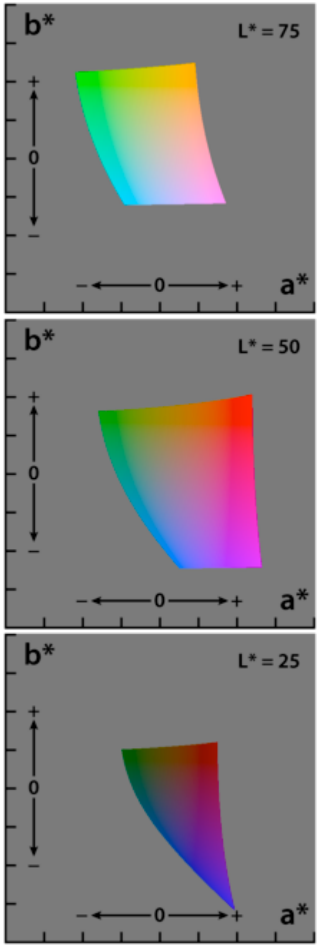
Image



RGB

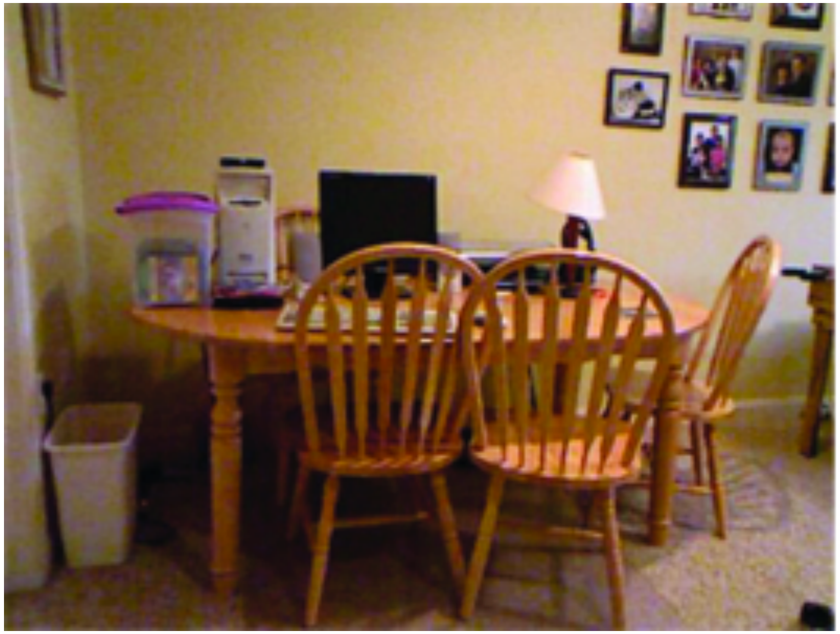


YUV

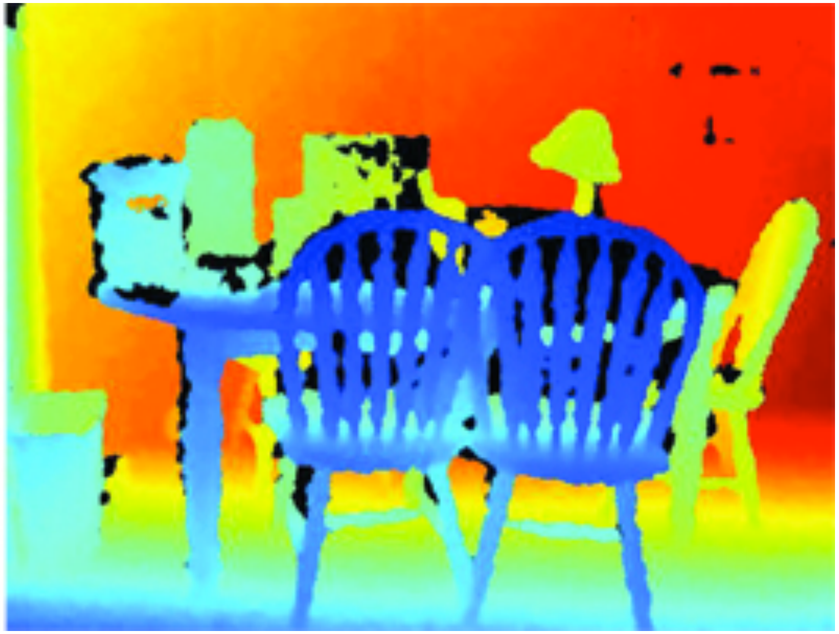


Lab

Depth Image



RGB image



Depth image

Each pixel in depth image shows the distance to camera

Device

- Kinect
- Kinect2 (we use)
- SoftKinetic
- Leapmotion

Kinect

- Depth camera developed by Microsoft in 2010 for XBOX360
- Mainly for entertainment (Motion Sensing Game)

Kinect2

- A new version of Kinect published in 2014
- Two different type for Windows and XBOX

Kinect for Windows

SoftKinectic

- Belgian company which develops gesture recognition hardware and software for real-time range imaging cameras

DS311
(2012)

Leapmotion (厉动)

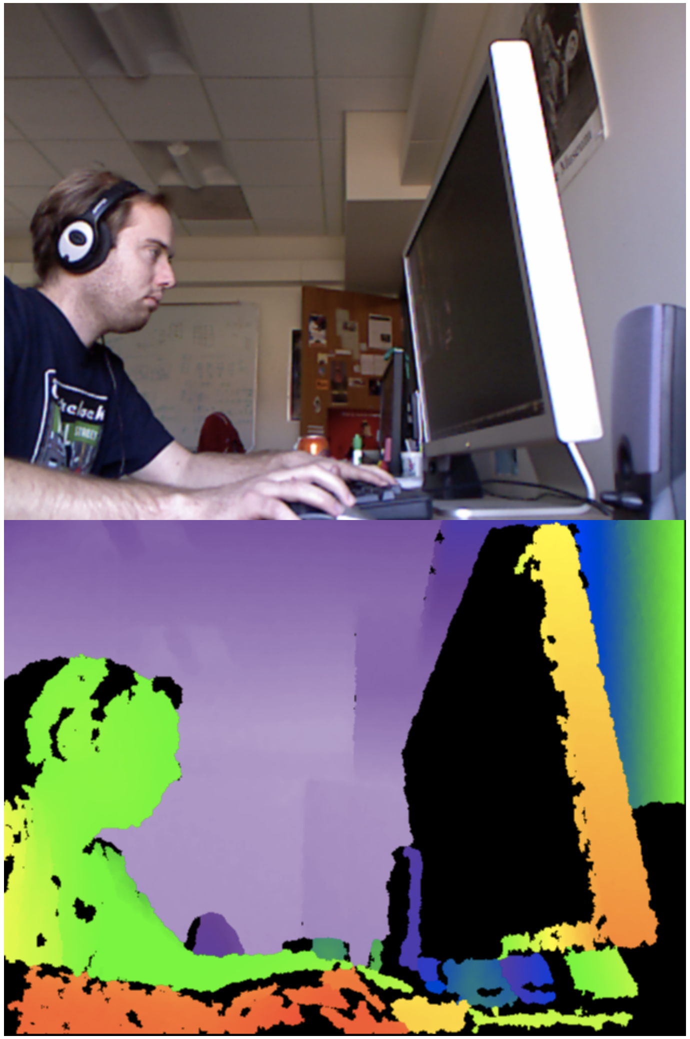
- A small USB peripheral device which is designed to be placed on a physical desktop

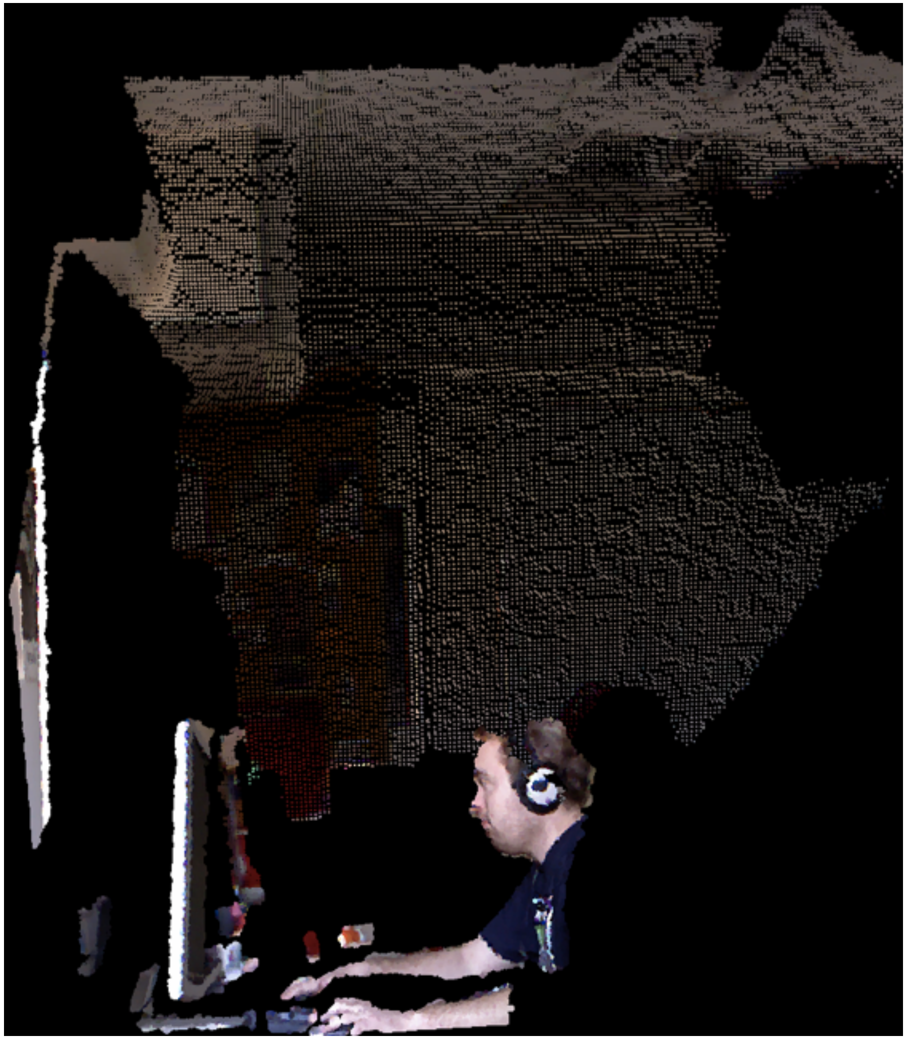
Depth Image 3D Reconstruction

- Depth Image shows the distance between object to camera

- 3D position of each pixel is the best
  - point cloud(点云)
  - triangular facet(面片)

Point Cloud of Depth Image





Triangular Facet of Depth Image

Depth Image Applications

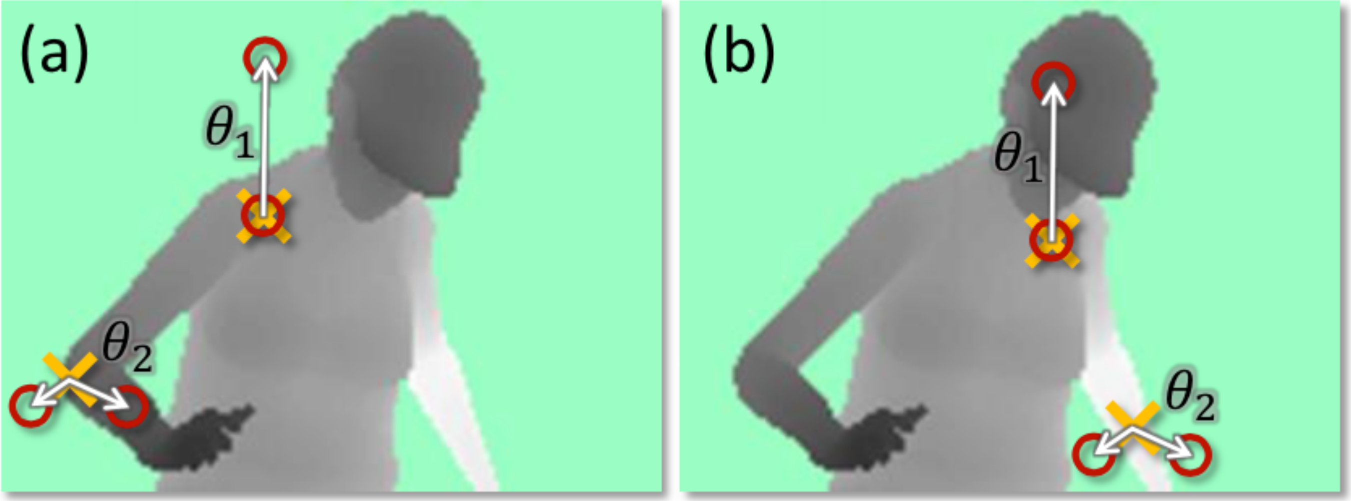
- Depth feature
- Human pose recognition
- Semantic segmentation
- Salient region detection
- Hand tracking

Depth Feature

- Depth comparison features:

$$f_\phi(I,x) = d_I\left(x + \frac{u}{d_I(x)}\right) - d_I\left(x + \frac{v}{d_I(x)}\right)$$

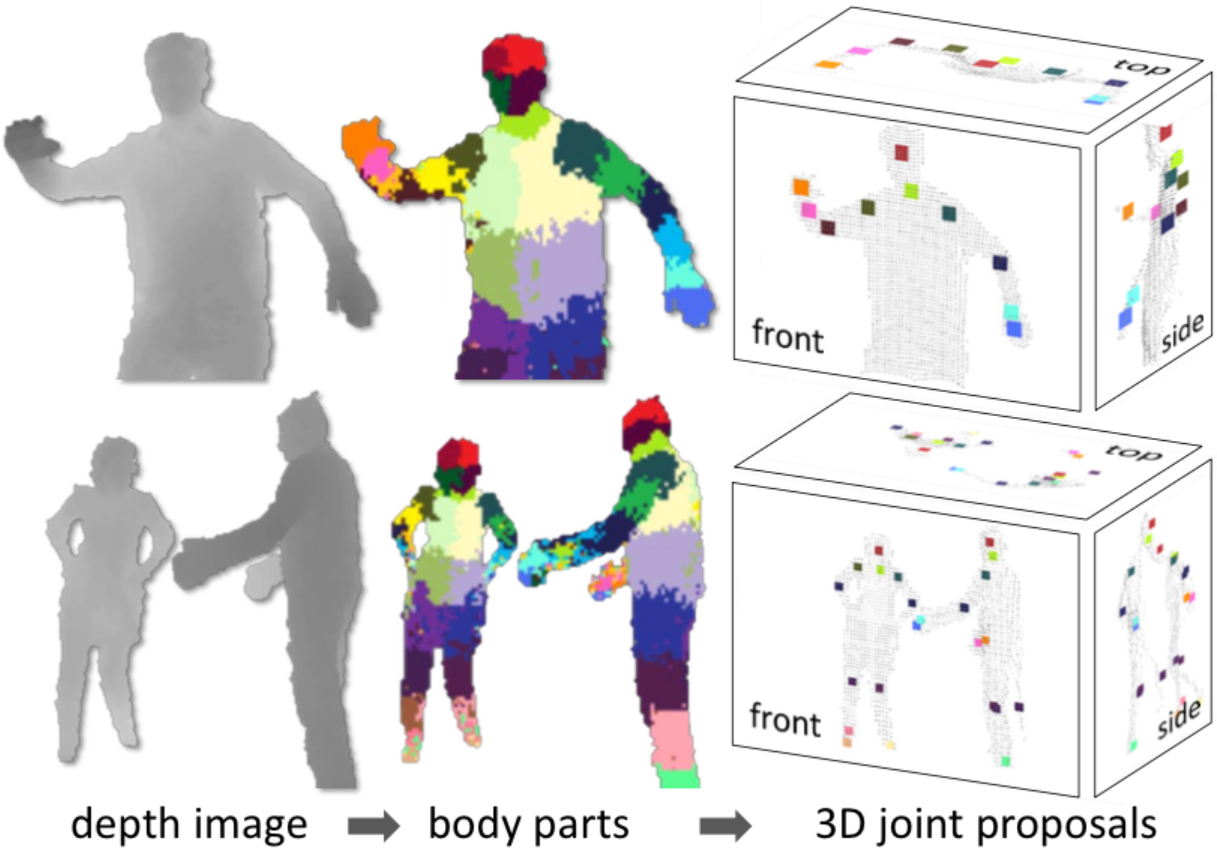
  - $d_I(x)$ is the depth at pixel x in image I
  - $\varphi = (u,v)$ describe offsets u and v



Human pose recognition

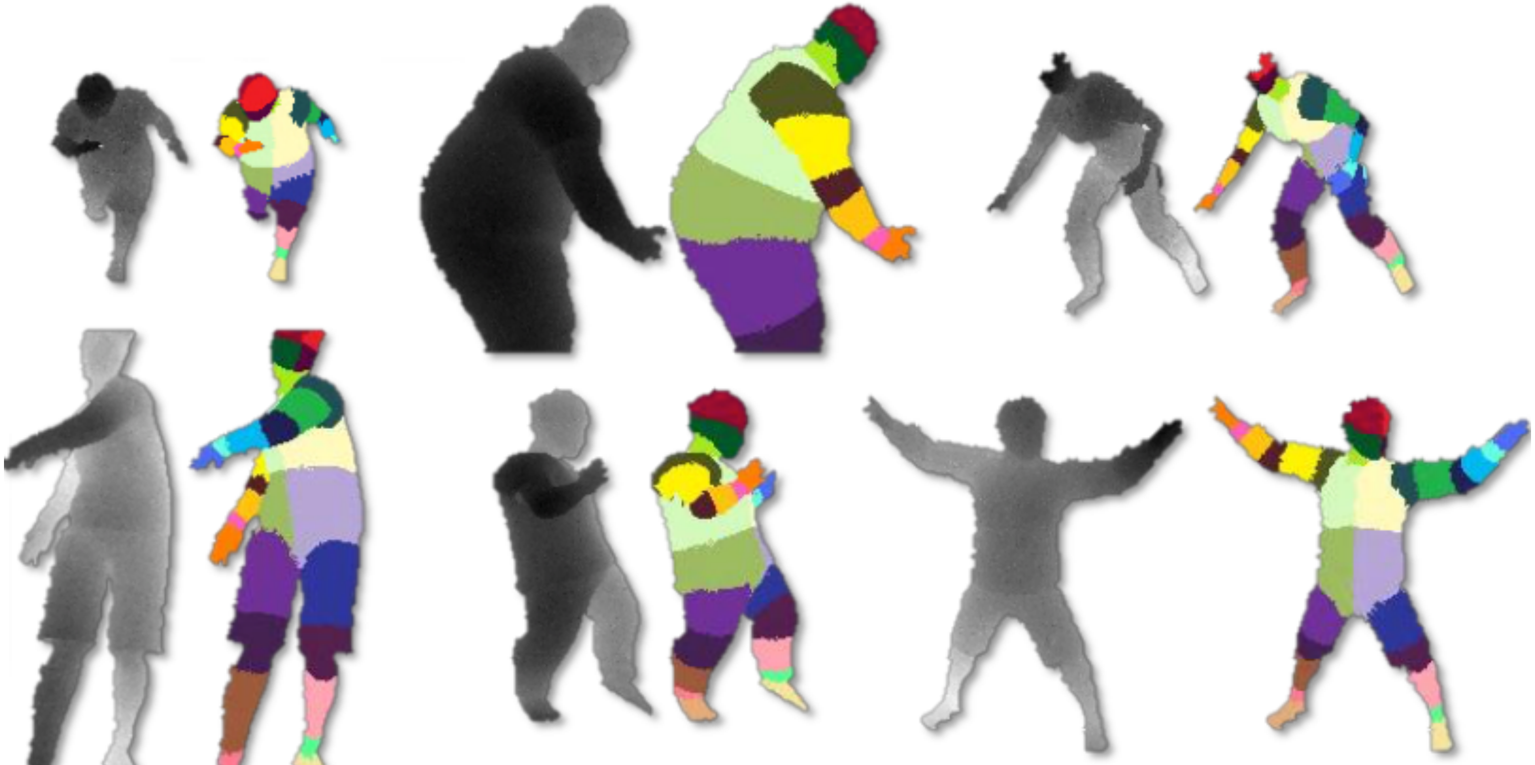
Real-time Human Pose Recognition in Parts from Single Depth Images, CVPR2011

- Recognition body parts in depth image



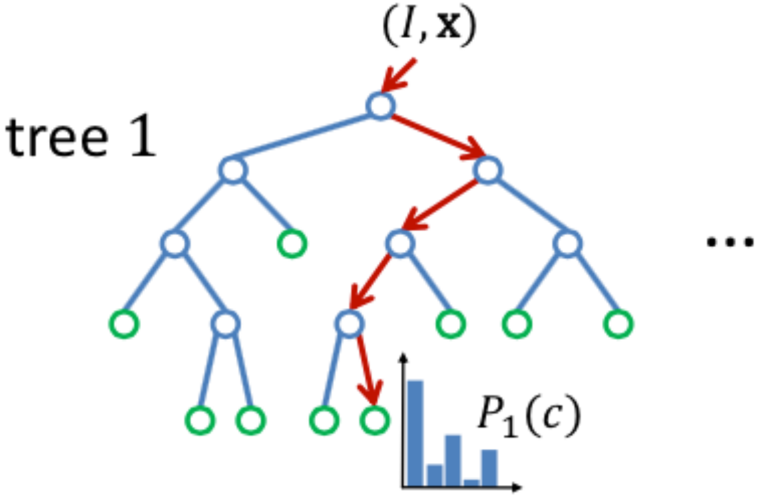
Pose Recognition – Body part labeling

- 31 body parts: LU/RU/LW/RW head, neck, L/R shoulder, LU/RU/LW/RW arm, L/R elbow, L/R wrist, L/R hand, LU/RU/LW/RW torso, LU/RU/LW/RW leg, L/R knee, L/R ankle, L/R foot (Left, Right, Upper, loWer)



Pose Recognition – Random Forest

- Each split node consists of a ***depth feature*** and threshold to classify pixel in image
- Each leaf node learned distribution $P_t(c|I,x)$ means the probability of pixel x belongs to body parts c



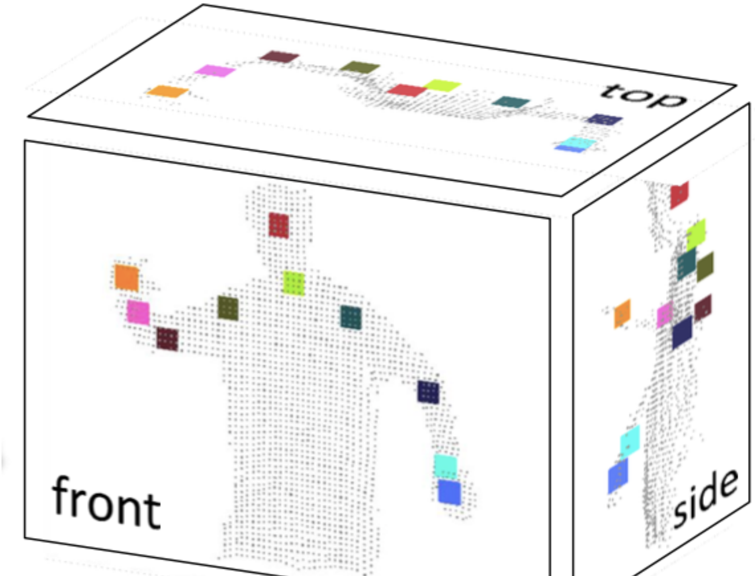
$$P(c|I,x) = \frac{1}{T}\sum_{t=1}^{T} P_t(c|I,x)$$

Pose Recognition – Joint Position

- *Mean-shift* to find center for each body part
- Density function:

$$f_c(\hat{x}) \propto \sum_{i=1}^{N} w_{ic} \exp\left(-\left\|\frac{\hat{x}-x_i}{b_c}\right\|^2\right)$$

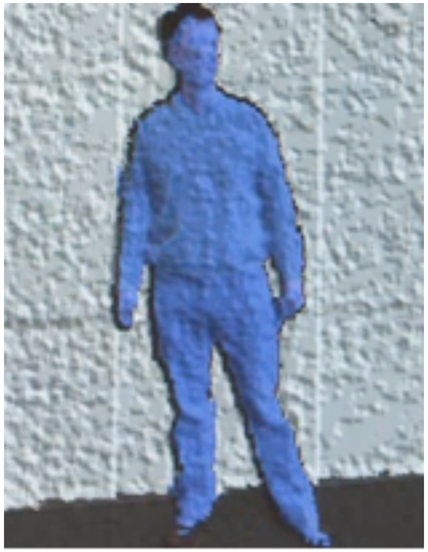
- 3D Reconstruction for each center



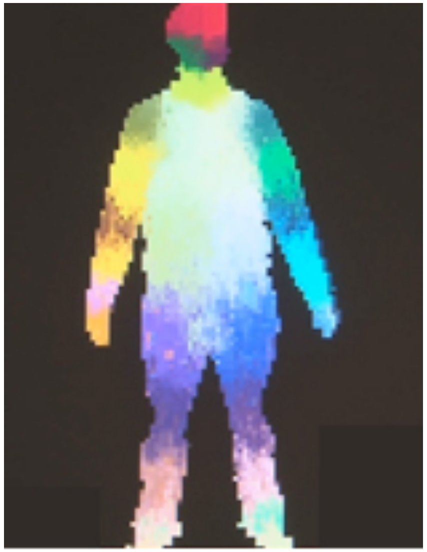
Pose Recognition - Result

http://research.microsoft.com/en-us/projects/vrkinect/

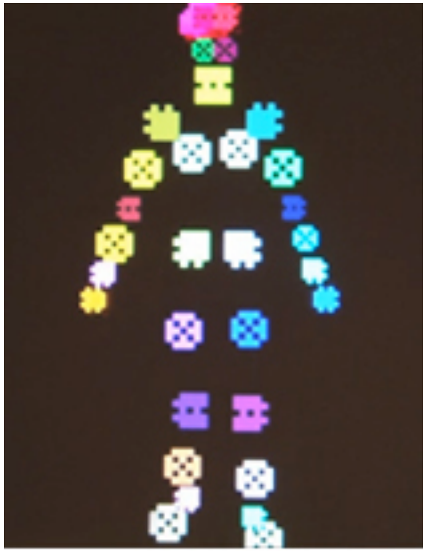
RGB image



Depth image



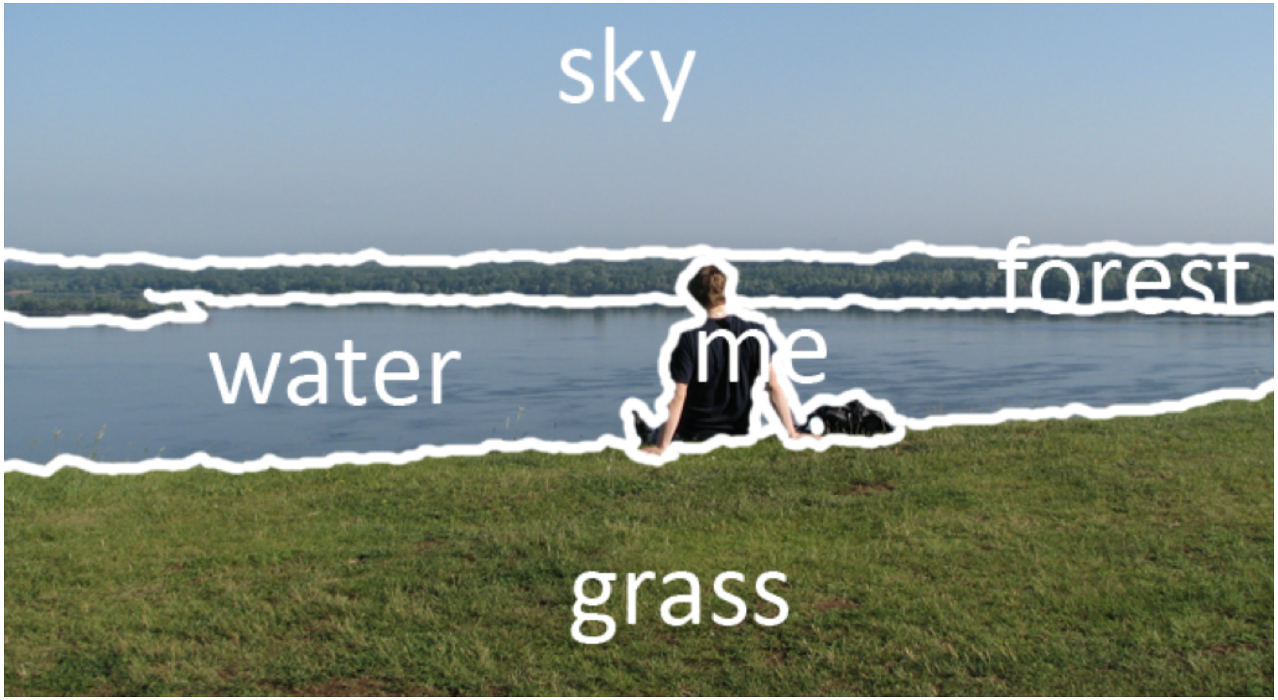
Body part inferred



Body part position

Semantic Segmentation

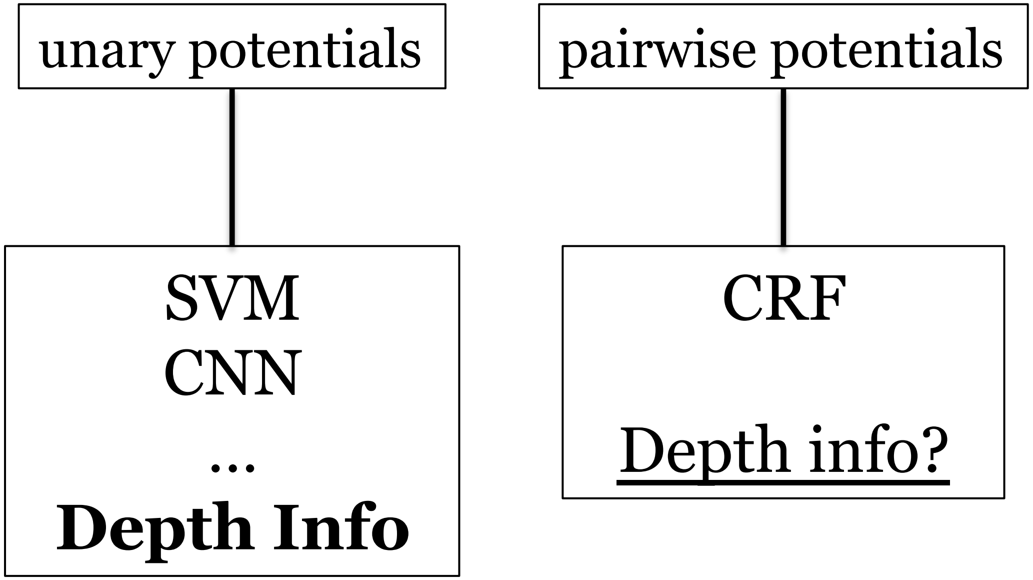
- Divide image into regions which correspond to the objects of the scene



Semantic Segmentation - Formulation

- The basic formulation is

$$E(c) = \sum_{i \in I} P(c_i | p_i) + \lambda \sum_{(i,j) \in \epsilon} P\left(c_i, c_j \middle| p_i, p_j\right)$$



unary potentials

SVM
CNN
…
Depth Info

pairwise potentials

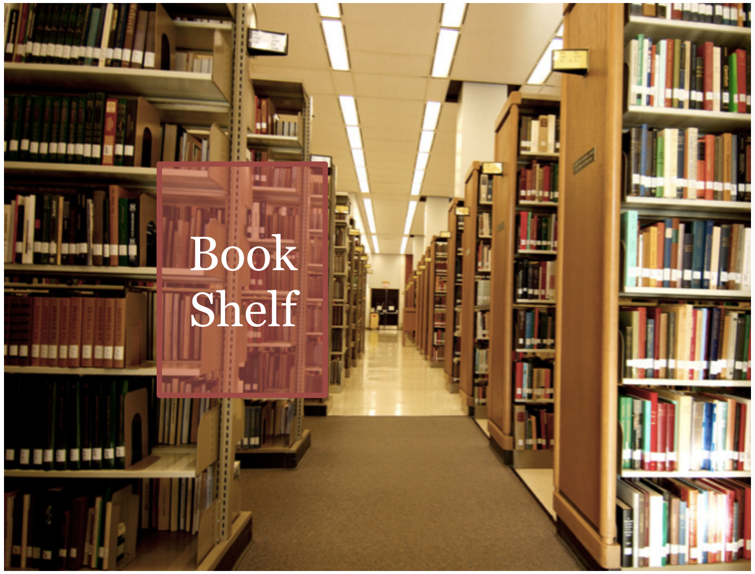
CRF

Depth info?

Semantic Segmentation - Idea

$$E(c) = \sum_{i \in I} P(c_i|p_i) + \lambda_1 \sum_{(i,j)\in\epsilon} P\left(c_i, c_j \middle| p_i, p_j\right) + \lambda_2 \boxed{\sum_{i} P\left(c_i, c_j \middle| p_i, p_i, d(p_i), d(p_j)\right)}$$

pairwise depth potentials

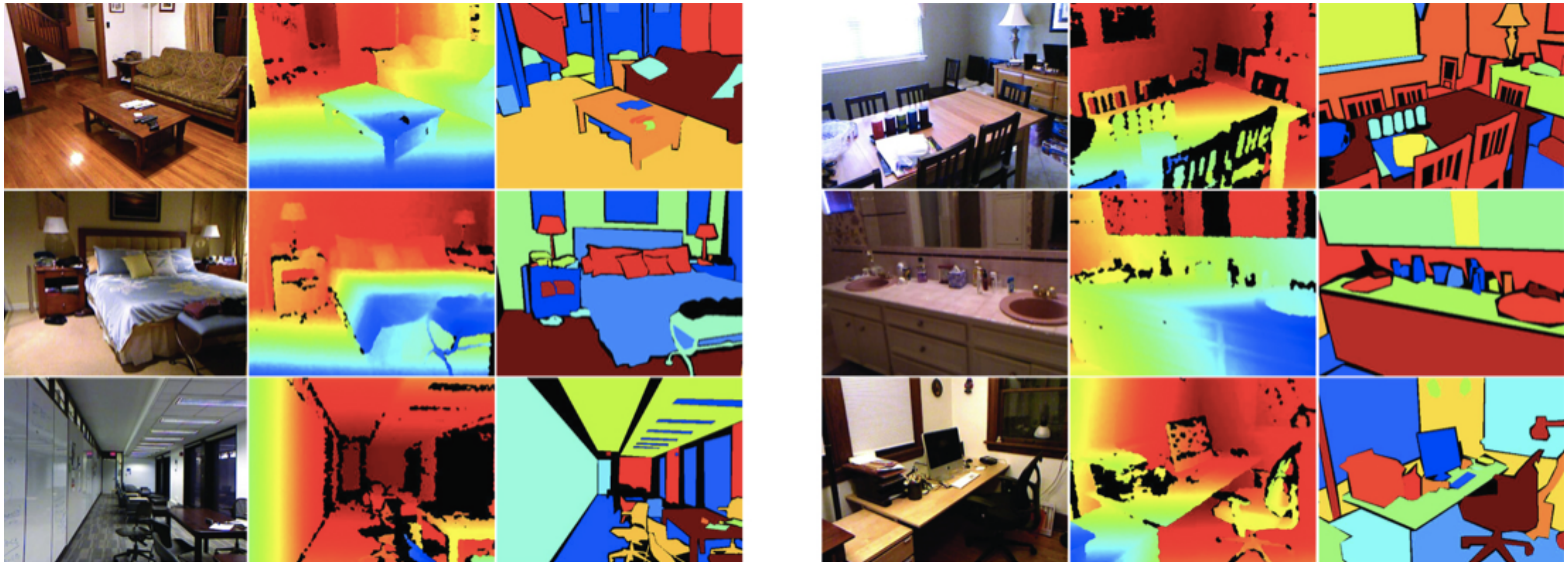


same label but
depth inconsecutive region

depth consecutive but
different label region

Semantic Segmentation - Dataset

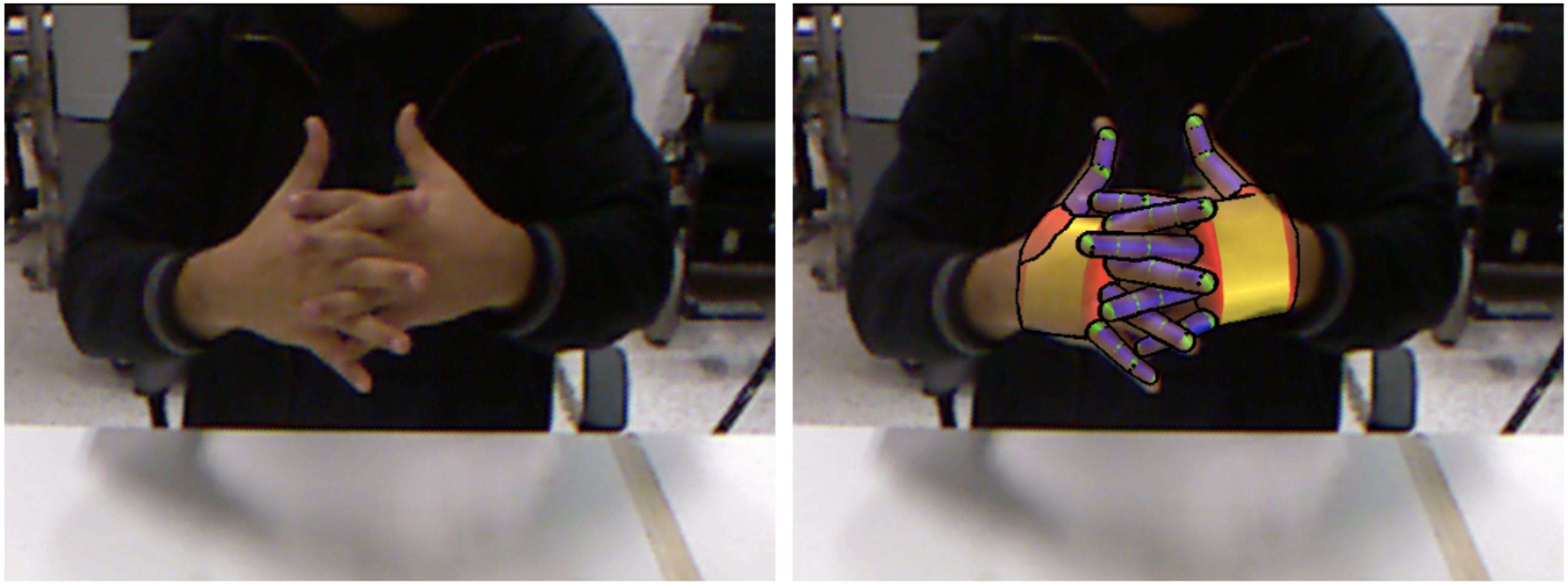
- NYU Depth Set V2
- http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html



Hand Tracking

Tracking the Articulated Motion of Two Strongly Interacting Hands, CVPR2012

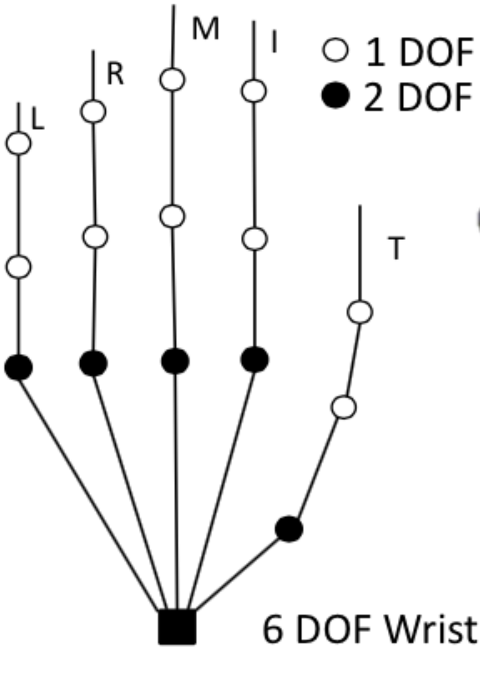
- Real-time tracking hands in video
- Not only estimate the position of hands but also construct hands model in 3D space



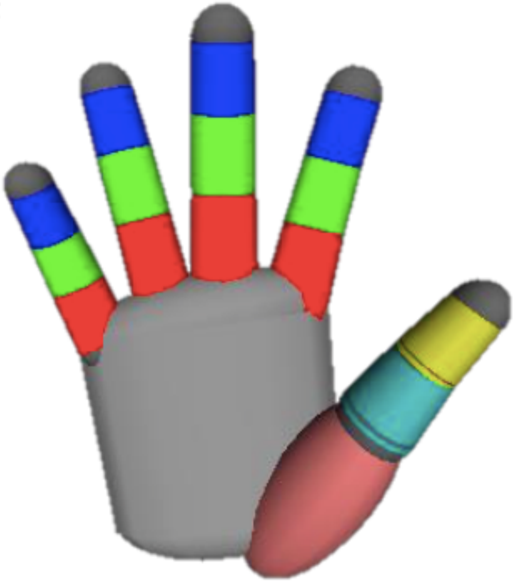
Hand Tracking – Hand Model

Construction and Animation of Anatomically Based Human Hand Models, SIGGRAPH

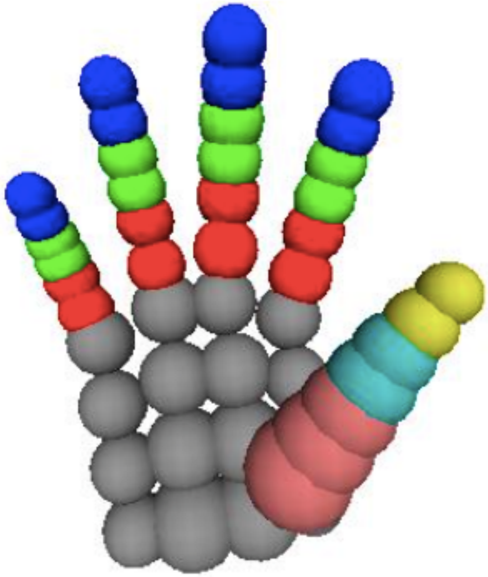
- There are 26 DoF(degree of freedom)
- 26 dimension feature show one hand in basic model



Basic model



Shape model



Sphere model
simplification of Shape model

Hand Tracking - Objective

- Our objective function

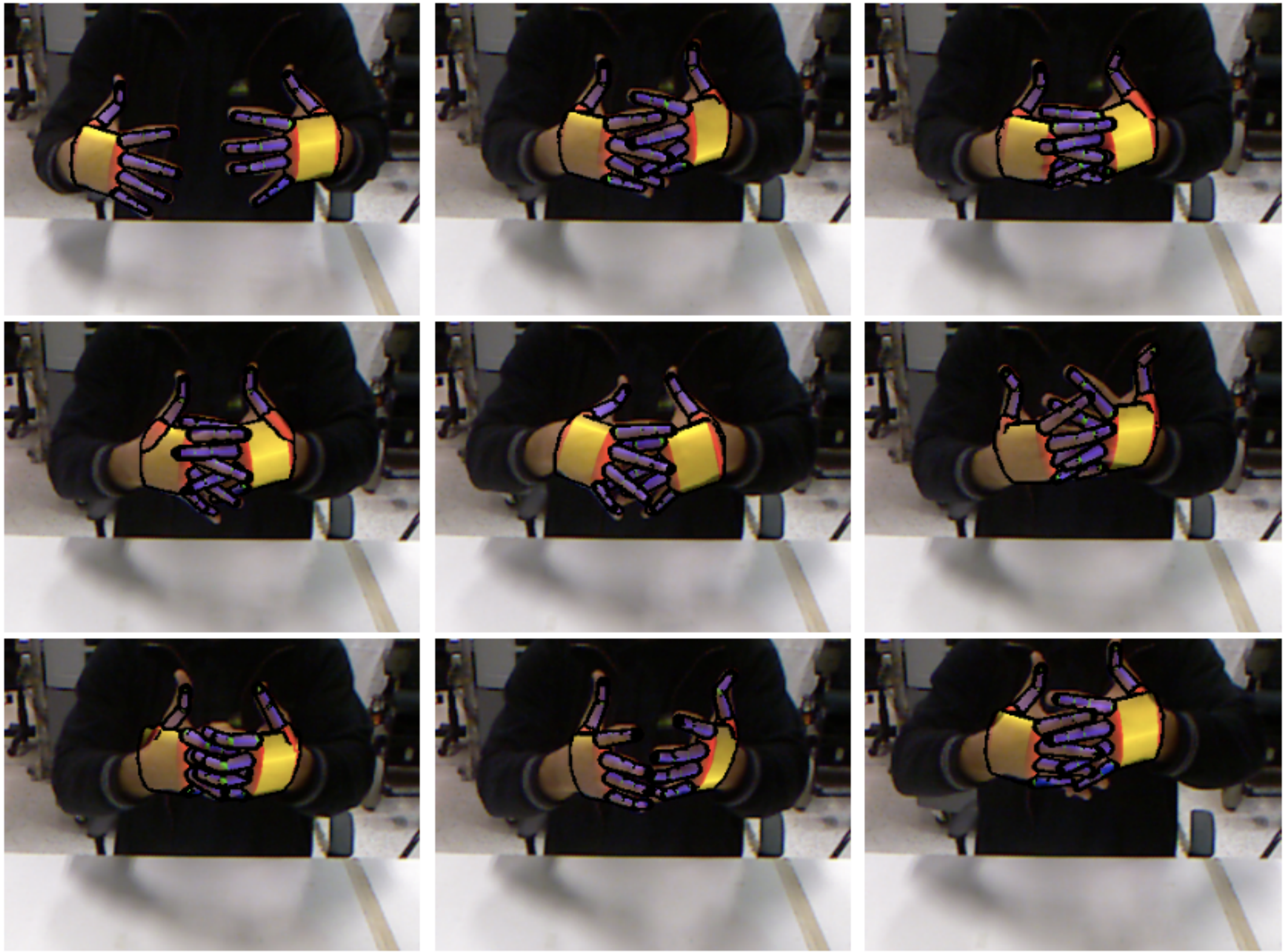
$$\underset{x}{\operatorname{argmin}} \boldsymbol{E}(x, o, h) = \left|\left|\boldsymbol{M}(x) - \boldsymbol{P}(o)\right|\right| + \lambda \boldsymbol{L}(x, h)$$

  - x is 26DoF hand feature
  - o is input RGBD image
  - h is tracking history
  - ***M**(.)* and ***P**(.)* is the function translate variable into same feature space
  - ***L**(.)* is self-constraint

Hand Tracking - PSO

- Particle Swarm Optimization is a randomized algorithms to find the approximate optimal parameter of objective function

Hand Tracking – Result



Hand Tracking – Some Problem

- Real-time
  - ICP-PSO
- Hand model for different hand
  - Robust Tracking
- Optimization Method
- Learning Method

- And so on

Q&A

THANKS