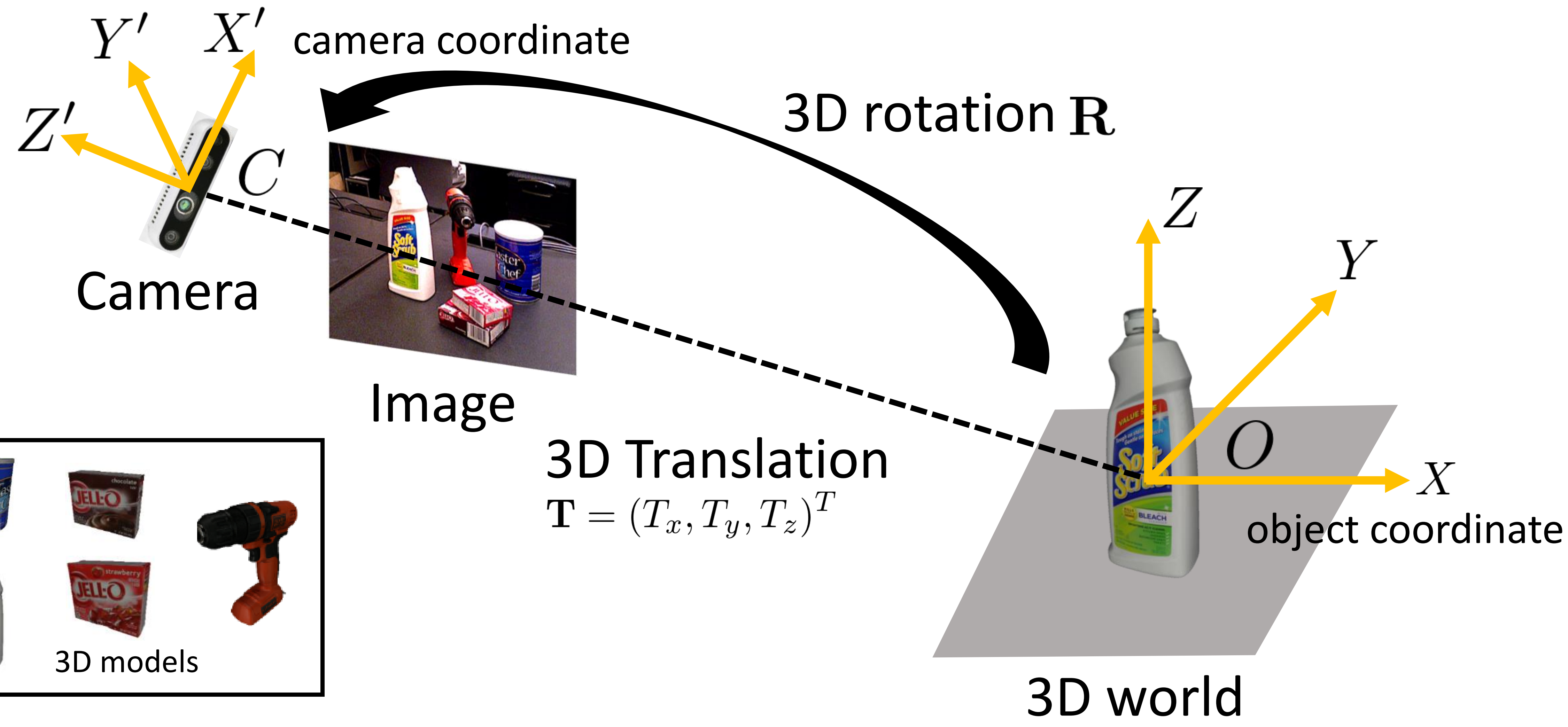
Pose Estimation of Objects, Humans and Hands

CS 6384 Computer Vision

Professor Yu Xiang

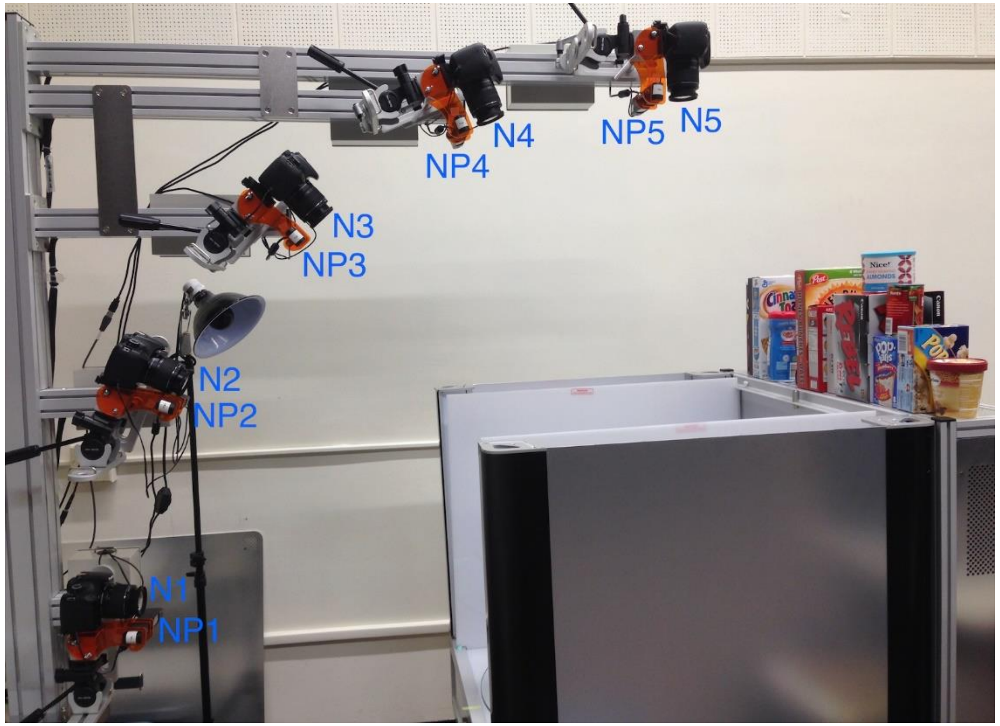
The University of Texas at Dallas

6D Object Pose Estimation



Building 3D Object Models

- 3D reconstruction from multiple images



Berkeley Instance Recognition Dataset. Singh et al., ICRA, 2014

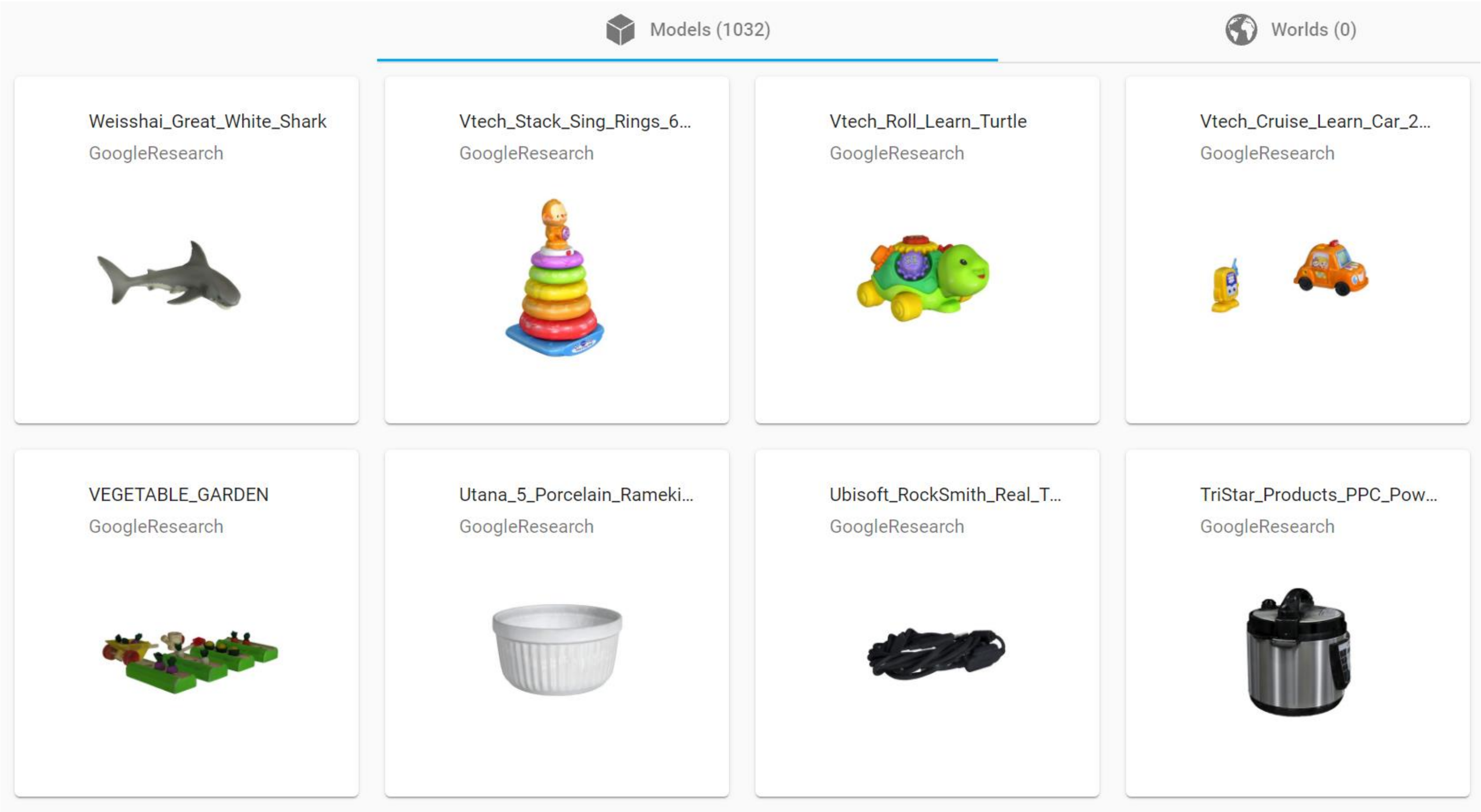
Building 3D Object Models

- A 3D reconstruction example

https://blog.kitware.com/3d-reconstruction-from-smartphone-videos/

Building 3D Object Models

- 3D Scanning



https://app.ignitionrobotics.org/GoogleResearch/fuel/collections/Google%20Scanned%20Objects

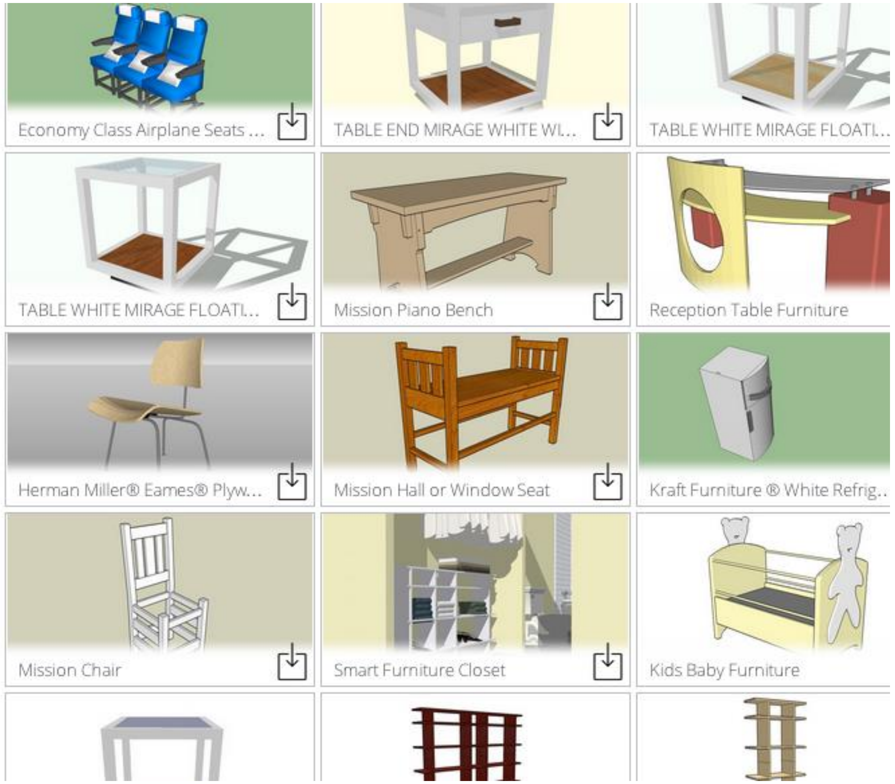
Building 3D Object Models

- 3D Scanning

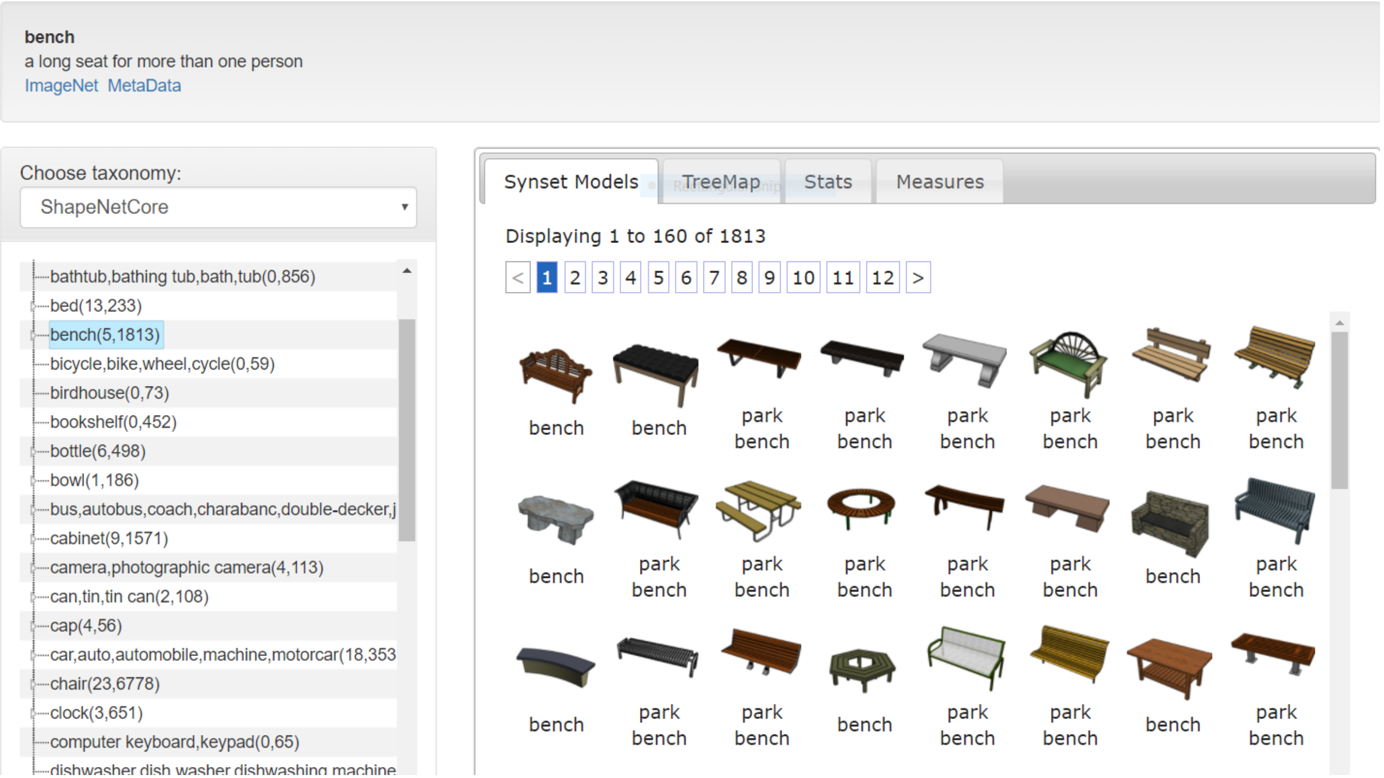
https://3dscanexpert.com/shining-3d-einscan-pro-3d-scanner-review/

Building 3D Object Models

- 3D CAD models



Trimble 3D Warehouse
https://3dwarehouse.sketchup.com

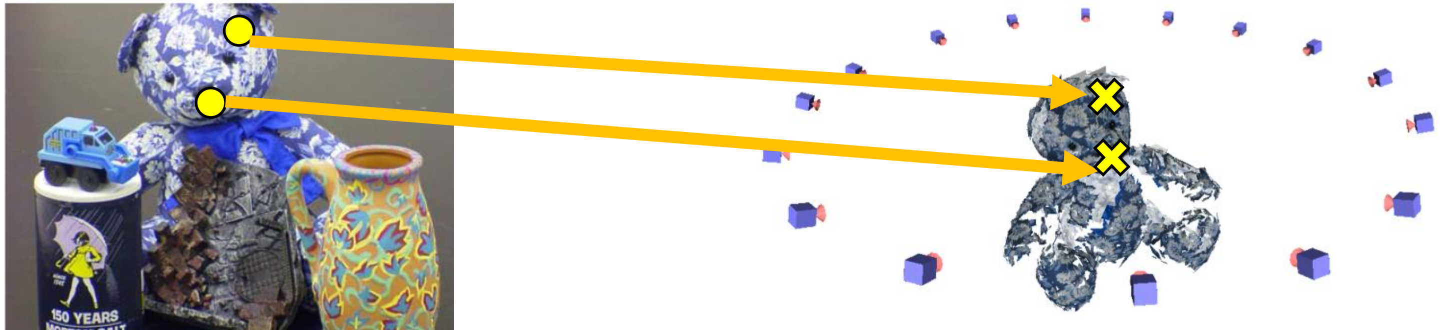


ShapeNet
https://www.shapenet.org/

6D Object Pose Estimation

- Feature matching-based methods

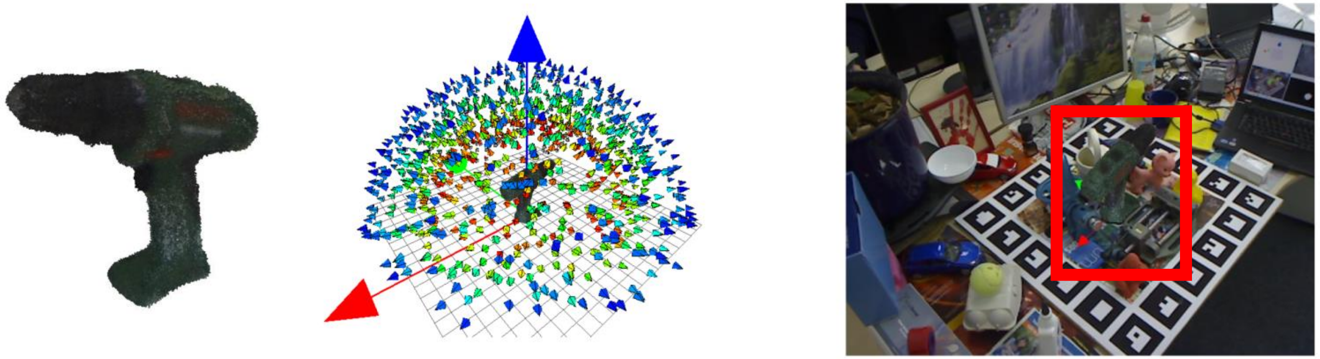
Rothganger et al., IJCV, 2006



2D image

3D model

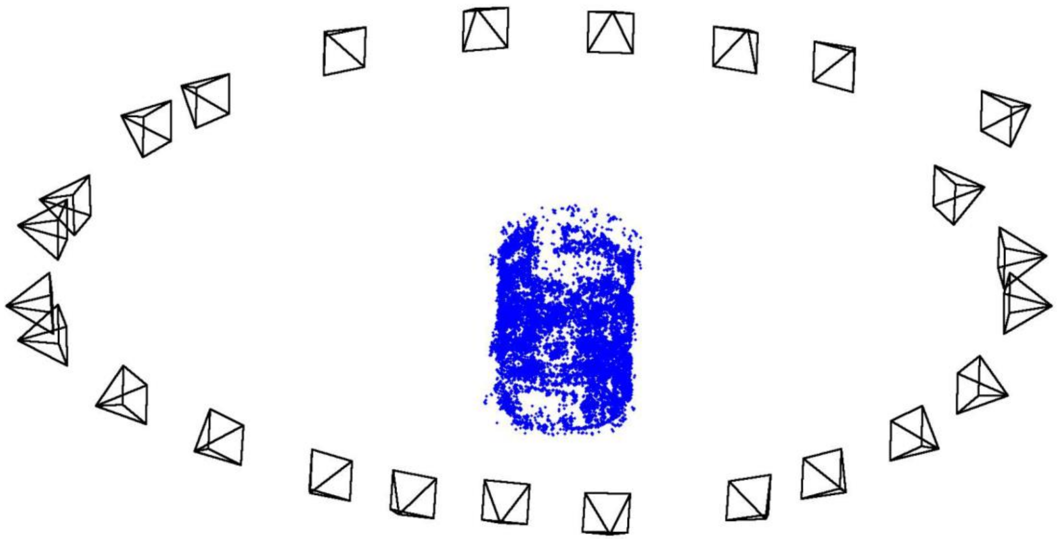
- Template matching-based methods



Hinterstoisser et al., ACCV, 2012

A Case Study for Feature Matching

- 3D Models of Objects using Structure from Motion
  - 3D points with SIFT descriptors (each 3D point can have a list of descriptors or use the mean of the descriptors)



Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR’10.

A Case Study for Feature Matching

- Ratio test

Query Image

3D Model

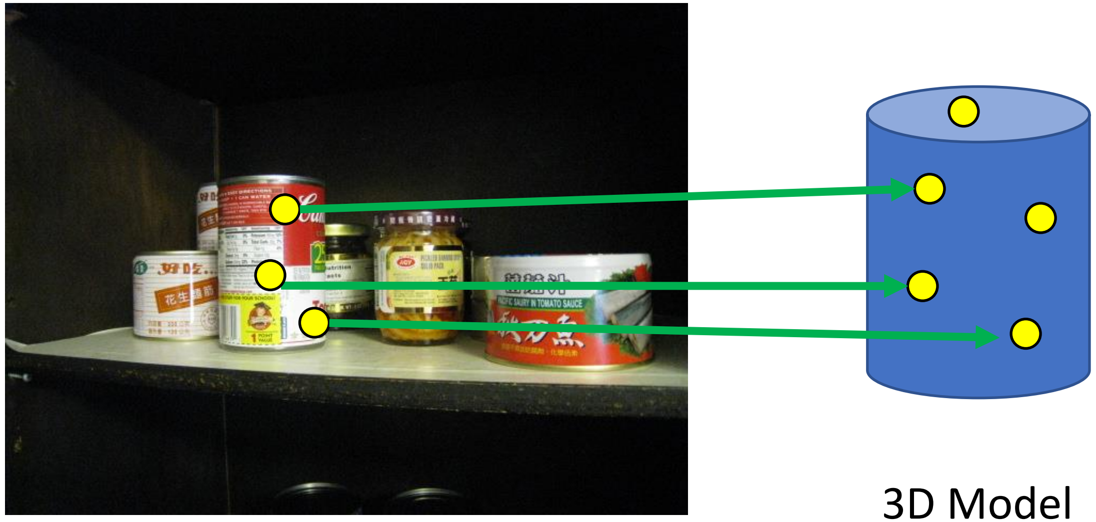
Distance to closest 3D point

$$\text{ratio} = \frac{d_1}{d_2} < 0.8$$

Distance to second closest 3D point

A Case Study for Feature Matching

- 3D-2D correspondences from feature matching $(\mathbf{X}_i, \mathbf{x}_i)_{i=1}^{N}$



Query Image

3D Model

Option 1: minimizing reprojection error

- Levenberg-Marquardt

$$g(\mathbf{R}, \mathbf{T}) = \sum_{i=1}^{N} \| P(\mathbf{X}_i, \mathbf{R}, \mathbf{T}) - \mathbf{x}_i \|^2$$

Option 2: solve the PnP problem

- EPnP (lecture 10)

Random Sample Consensus (RANSAC)

- An iterative method for parameter estimation from a set of observed data that contains **outliers**

RANSAC Algorithm {

1. Selects N data items as random
2. Estimates parameter $\vec{x}$
3. Finds how many data items (of M) fit the model with parameter vector $\vec{x}$ within a user given tolerance. Call this K.
4. If K is big enough, accept fit and exit with success.
5. Repeat step 1 until 4 (as Ltimes)
6. Algorithm will be exit with fail

}

Sample N 3D-2D correspondences $(\mathbf{X}_i, \mathbf{x}_i)_{i=1}^{N}$

Estimate $(\mathbf{R}, \mathbf{T})$

Find how many $(\mathbf{X}_i, \mathbf{x}_i)$ obeys $(\mathbf{R}, \mathbf{T})$

A Case Study for Feature Matching

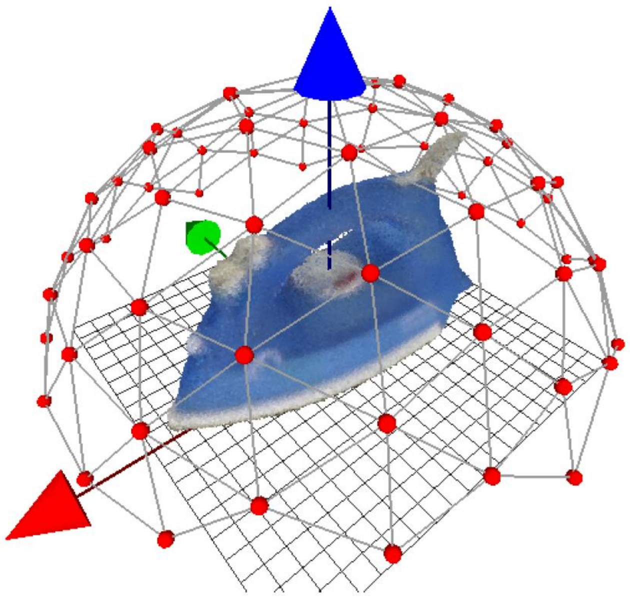
- Pose estimation examples

3D models

Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR'10.

A Case Study for Template Matching

- Render 3D models of objects to obtain template images

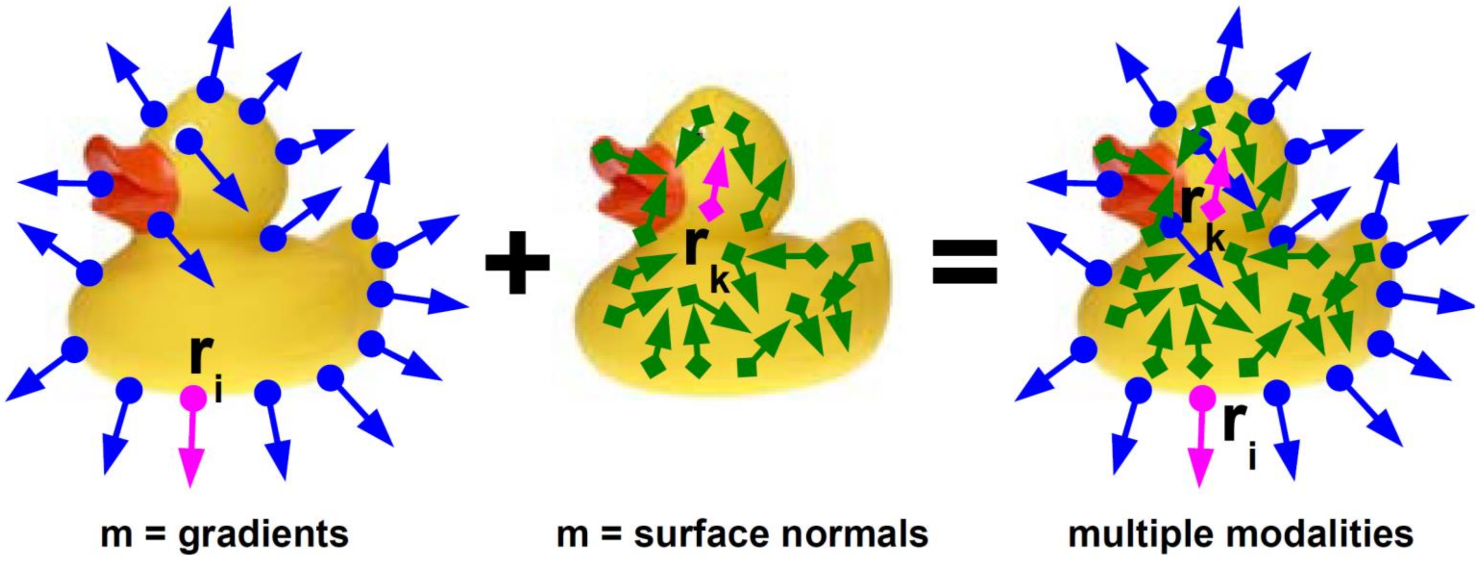


Viewpoint sampling

Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. Hinterstoisser et al., ACCV'12.

A Case Study for Template Matching

- Compute color and depth features for each template image



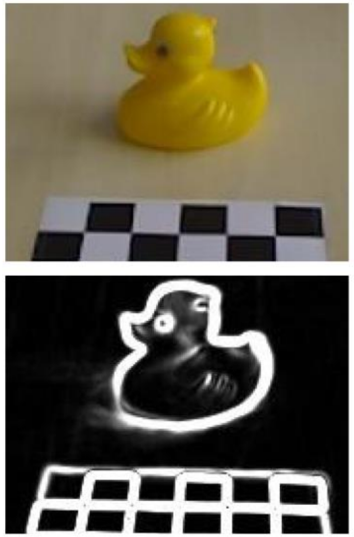
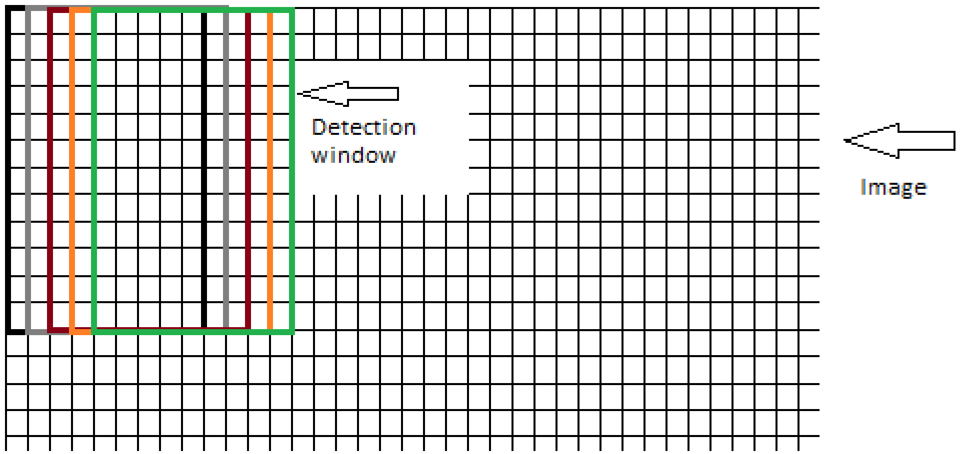


Image gradients

Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. Hinterstoisser et al., ACCV'12.

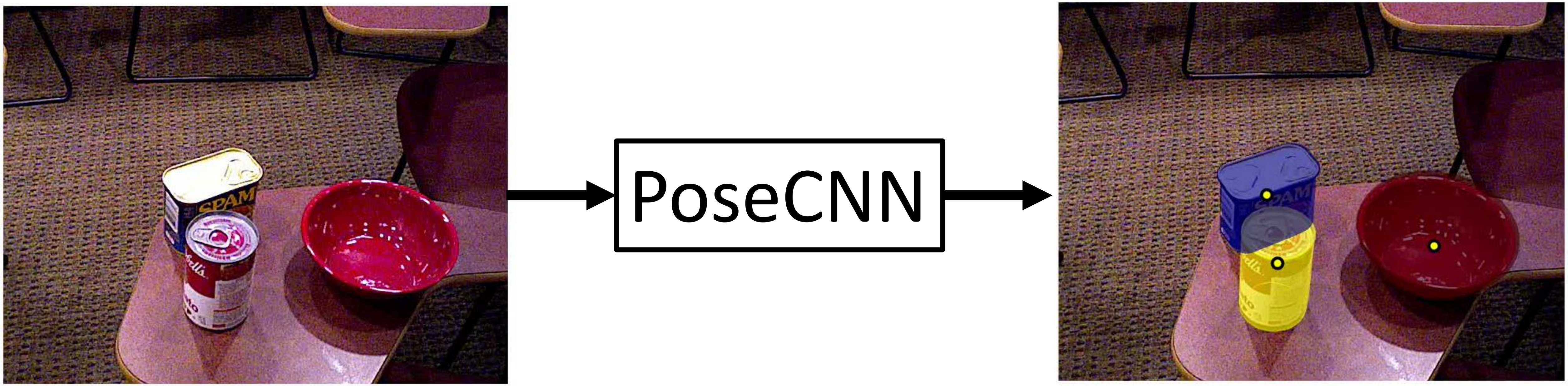
A Case Study for Template Matching

- Apply the templates to an input image for detection and pose estimation (sliding window)
  - Each template is associated with a 6D pose



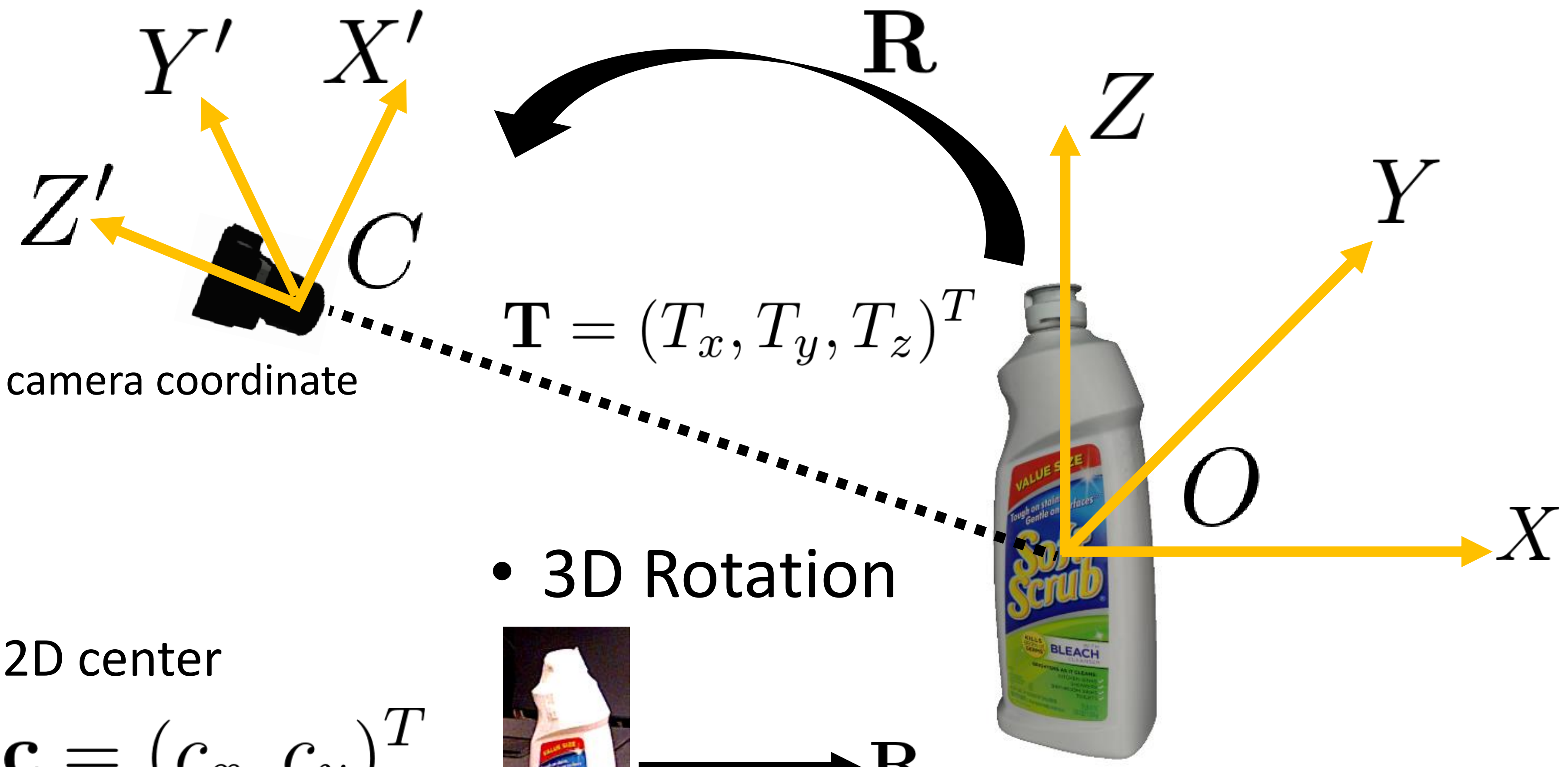
Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. Hinterstoisser et al., ACCV'12.

PoseCNN



Y. Xiang, T. Schmidt, V. Narayanan and D. Fox. PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. In RSS’18.

PoseCNN: Decouple 3D Translation and 3D Rotation



- 3D Translation

2D center

$\mathbf{c} = (c_x, c_y)^T$

Distance T_z

2D Center Localization

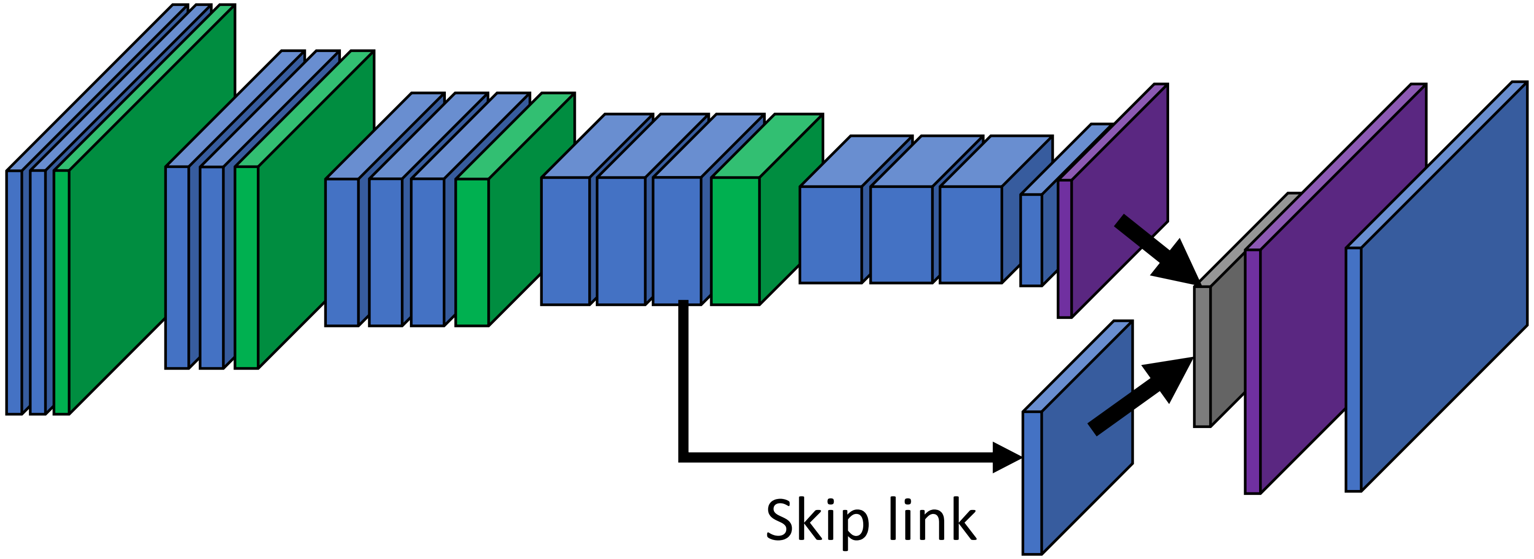
- 3D Rotation

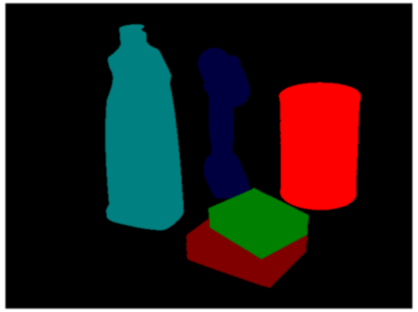
$\rightarrow \mathbf{R}$

3D Rotation Regression

PoseCNN: Semantic Labeling

Input image

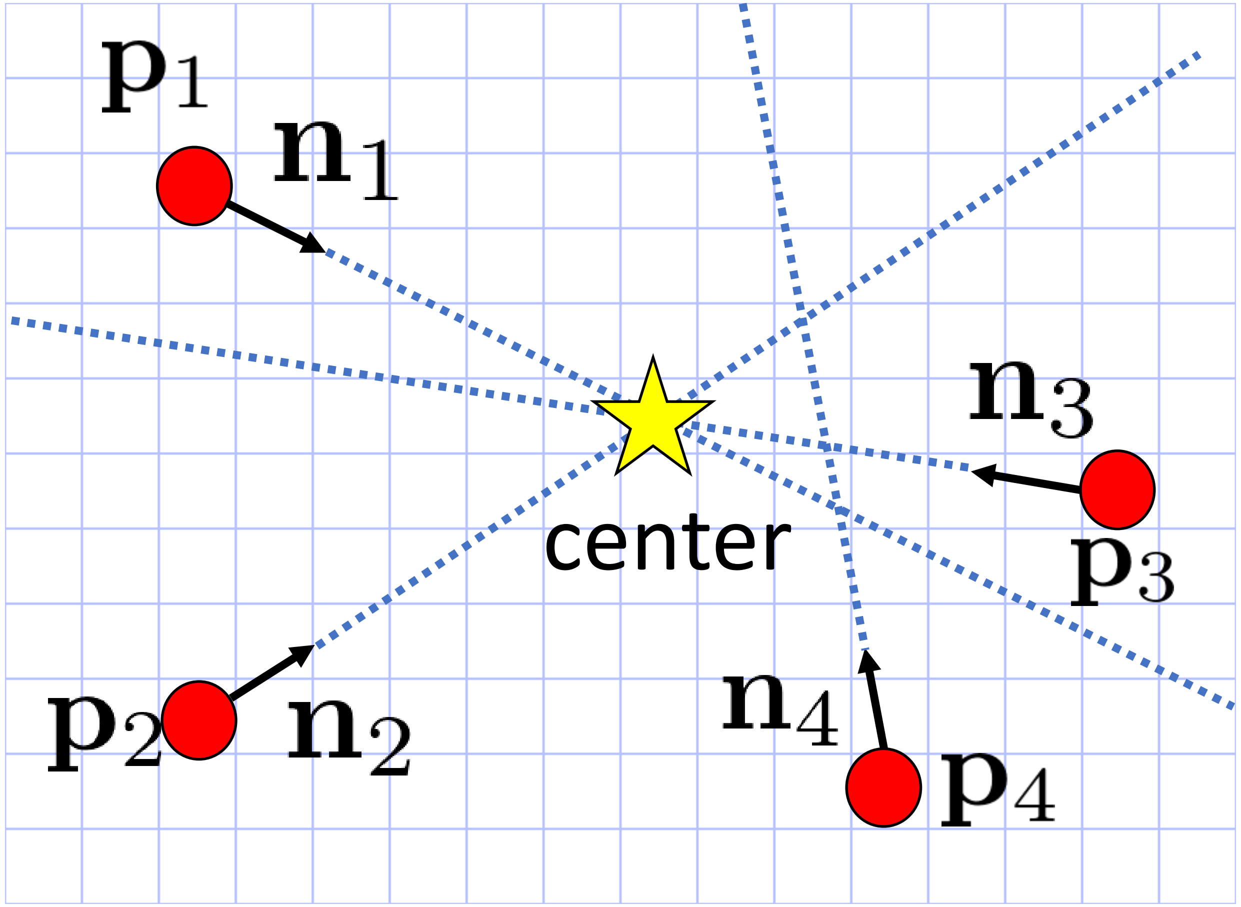




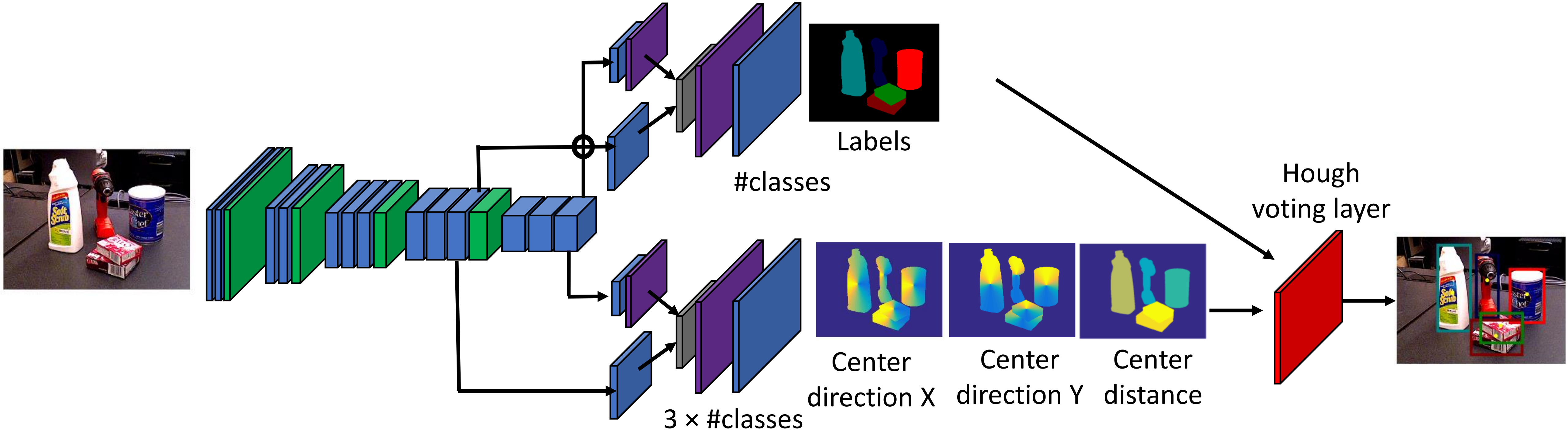
Labels

Fully convolutional network

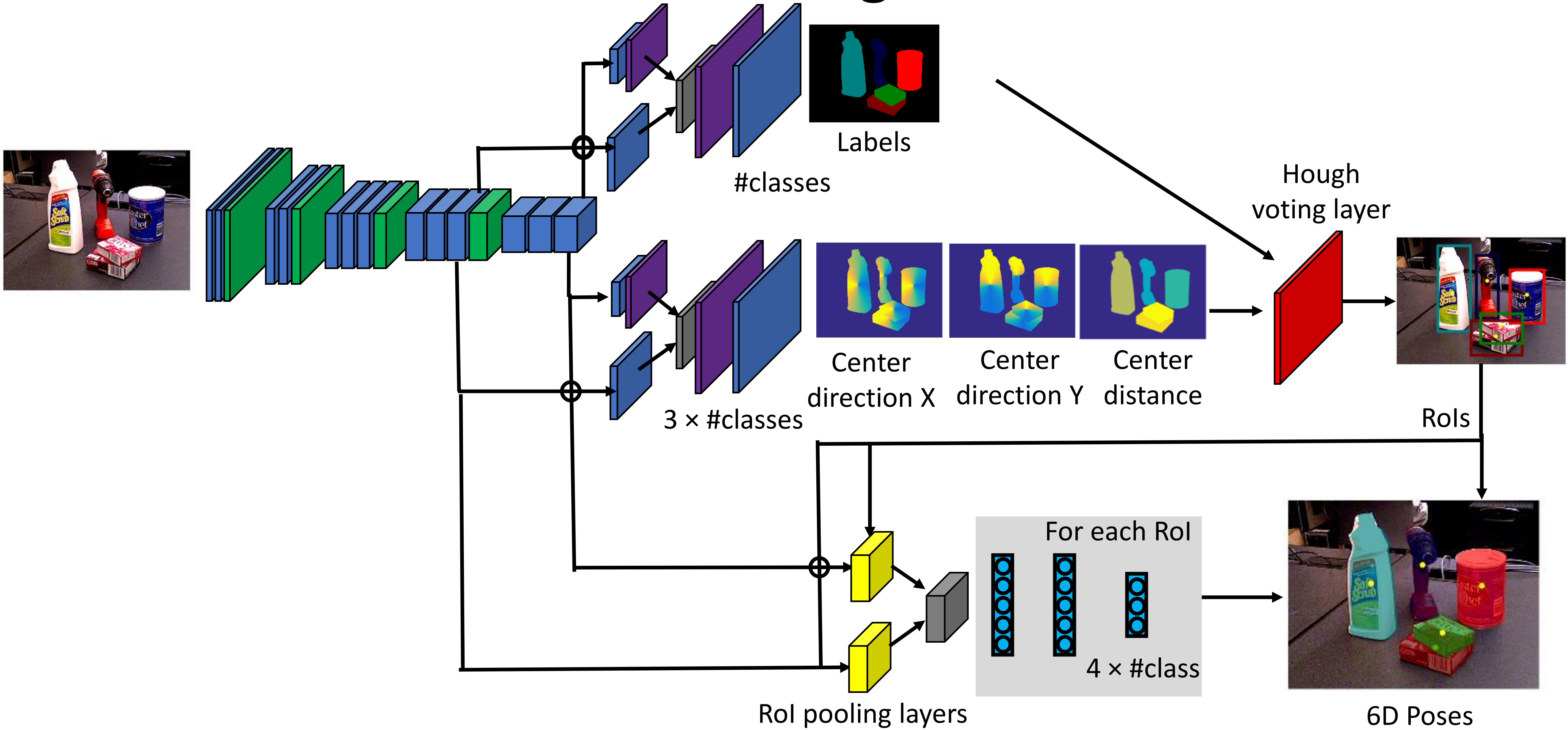
PoseCNN: 2D Center Voting for Handling Occlusions



PoseCNN: 3D Translation Estimation



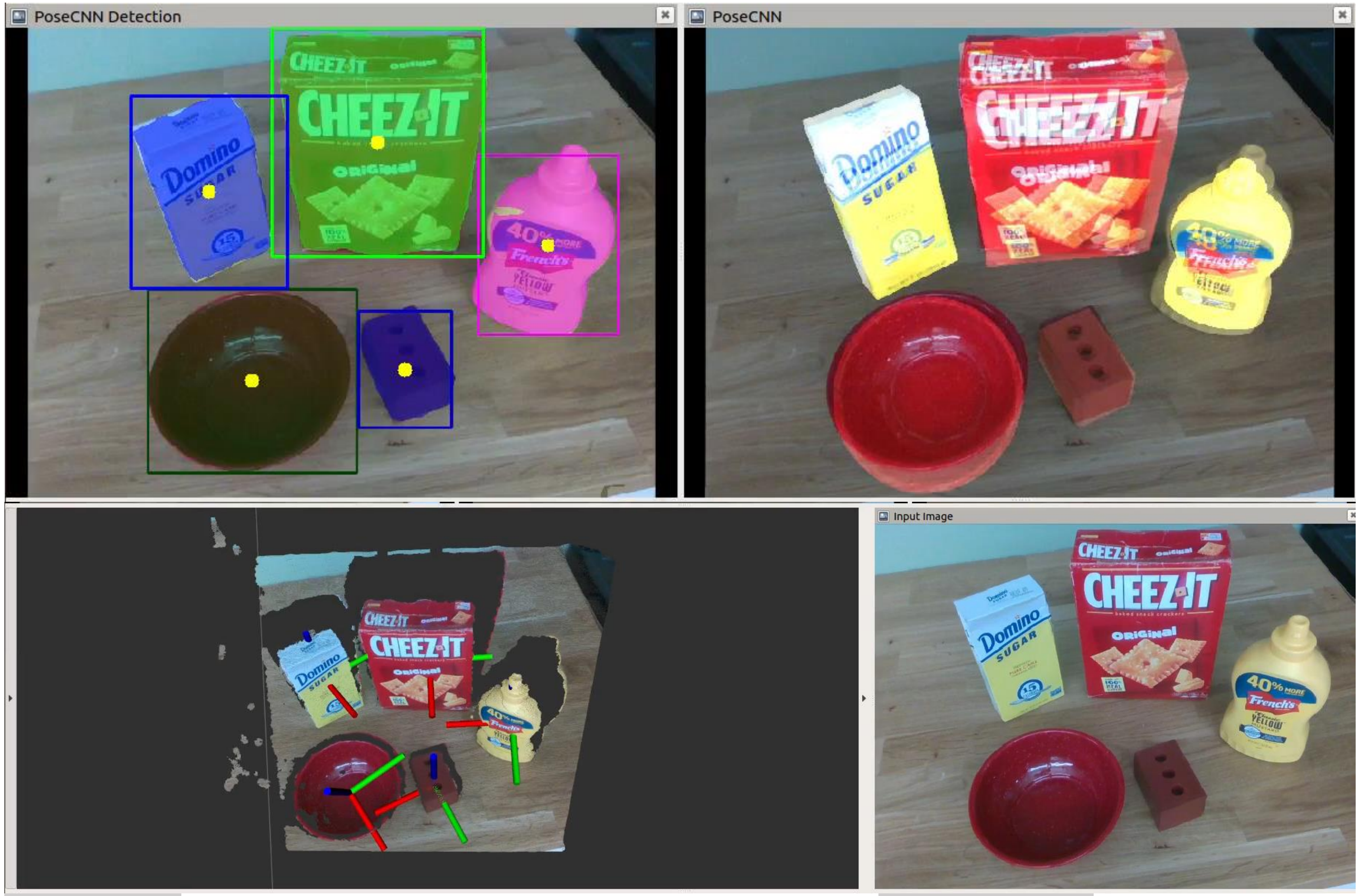
PoseCNN: 3D Rotation Regression



PoseCNN

Segmentation and Detection

Poses

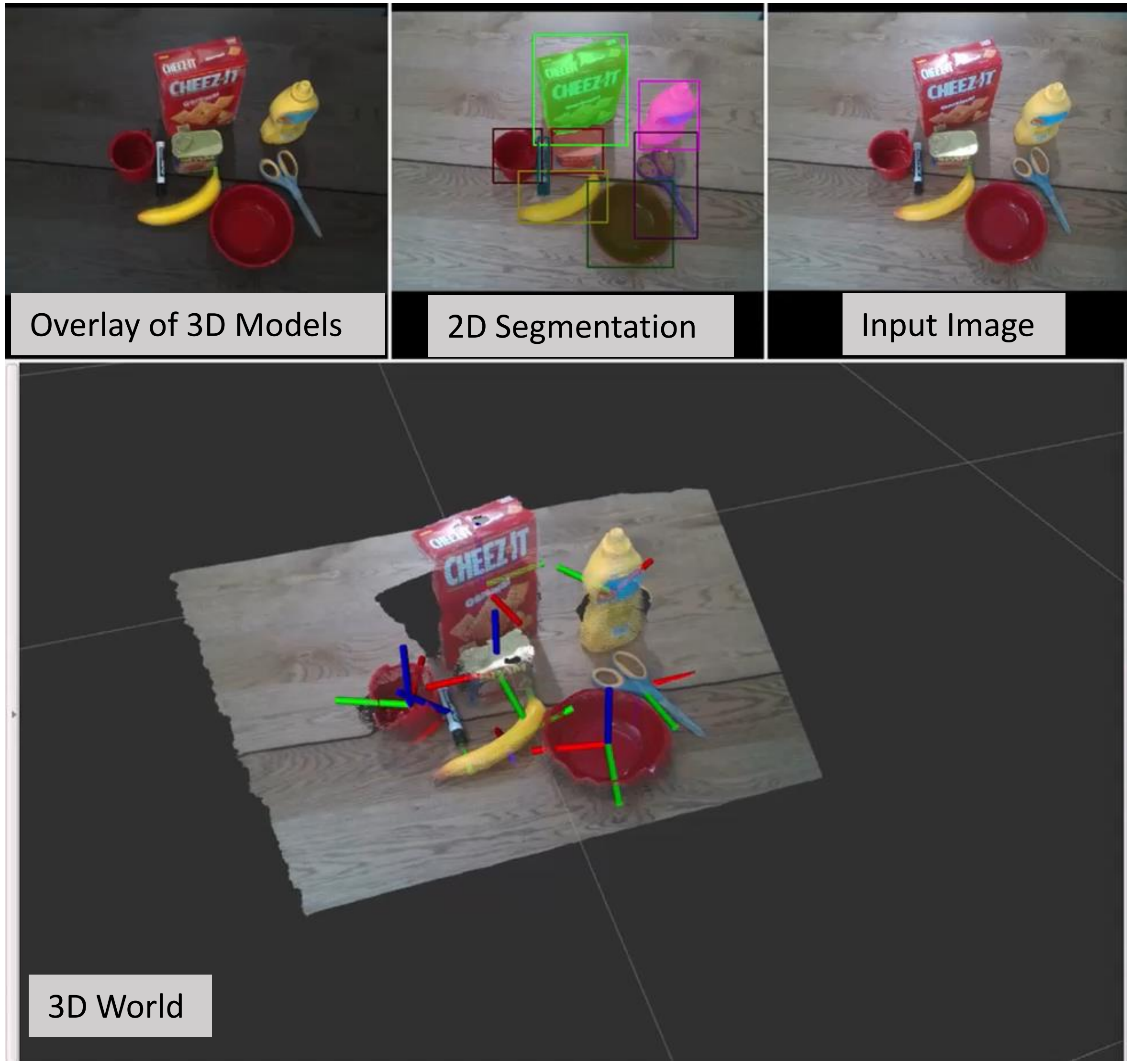


3D World

Input image

6D Object Pose Tracking

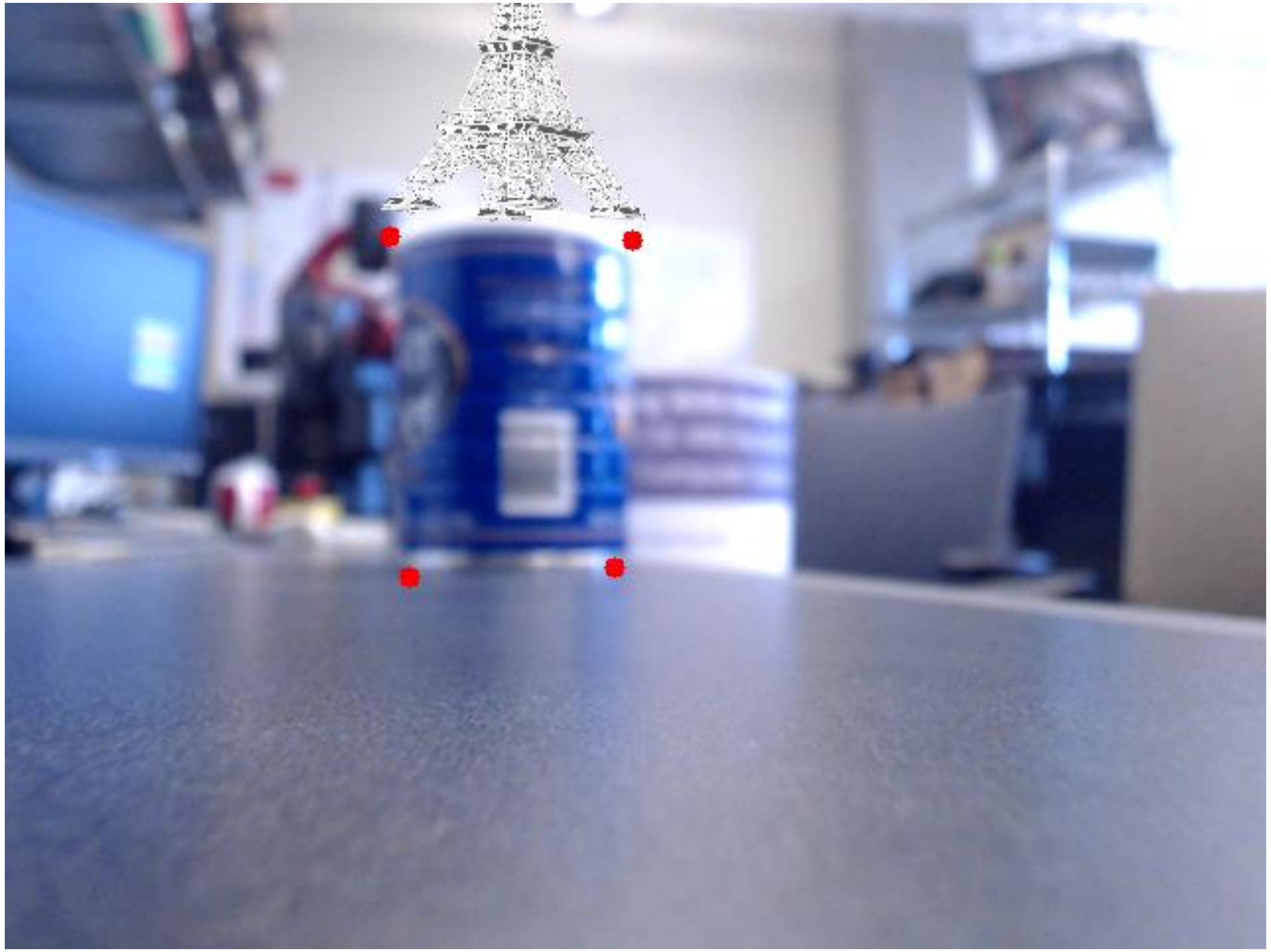
3D models



PoseRBPF: Deng et al., RSS'19

AR Demo with 6D Pose Estimation

DeepIM, Li et al., IJCV'19



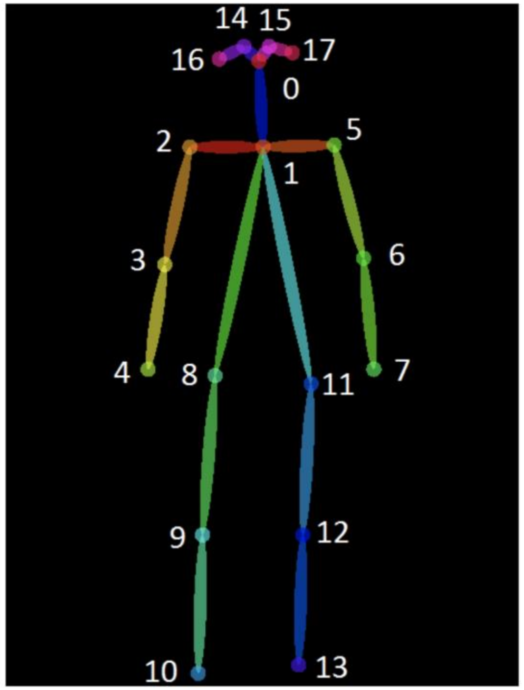
Credit: Lirui Wang

Human Pose Estimation

- Localizing human joints in images or videos

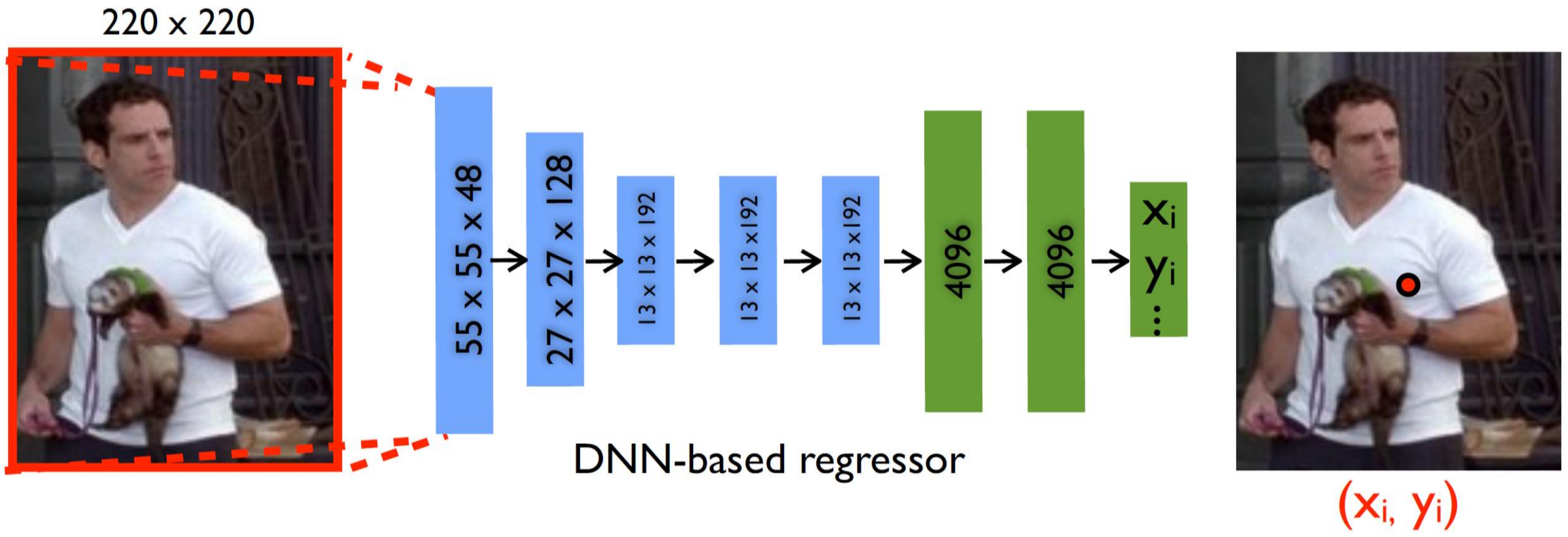
- 2D human pose estimation
  - Detect human joints in images (x, y)

- 3D human pose estimation
  - Detect human joints in 3D (x, y, z)



Human Pose Estimation

- Body joint detection/regression

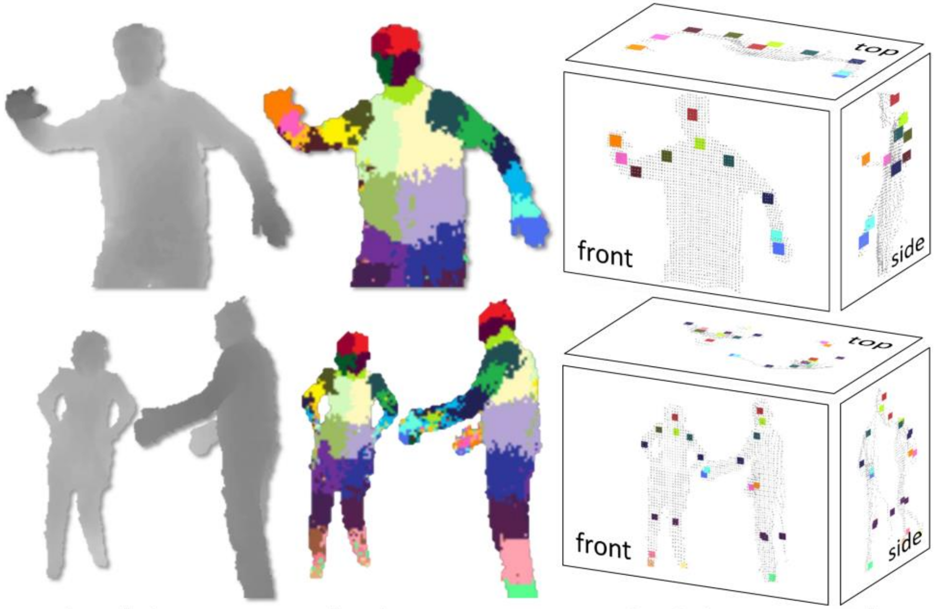


K joint locations

DeepPose: Human Pose Estimation via Deep Neural Networks. Toshev and Szegedy, CVPR'14

Human Pose Estimation

- Kinect: 3D human pose estimation from depth images



Real-Time Human Pose Recognition in Parts from Single Depth Images. Shotton et al, CVPR'11

- Randomized decision forests for part labeling
- Mean shift to find the modes of each part
- Push back modes to obtain joint positions

Human Pose Estimation

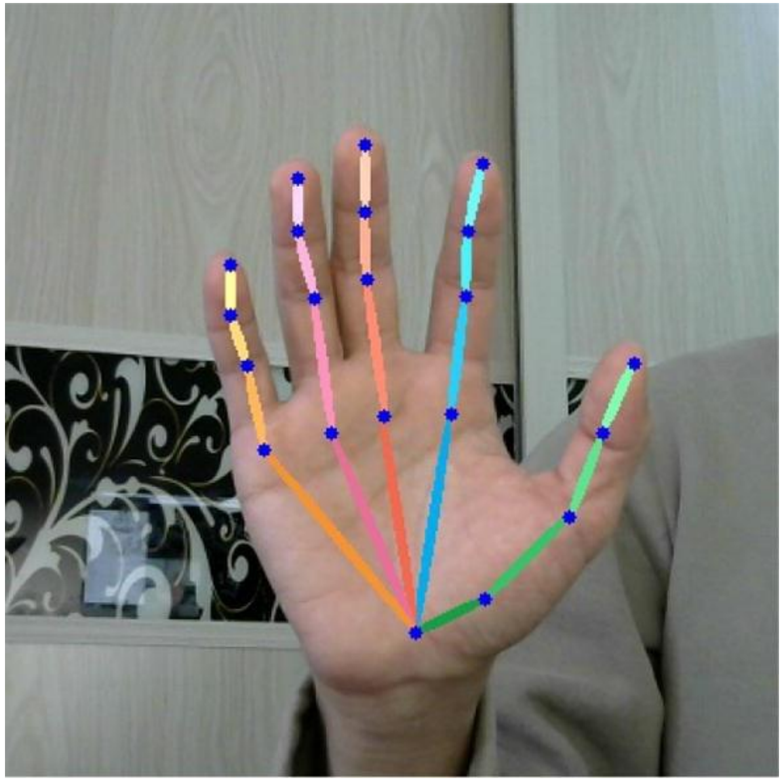
Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Cao et al, CVPR'17.

Human Pose Estimation

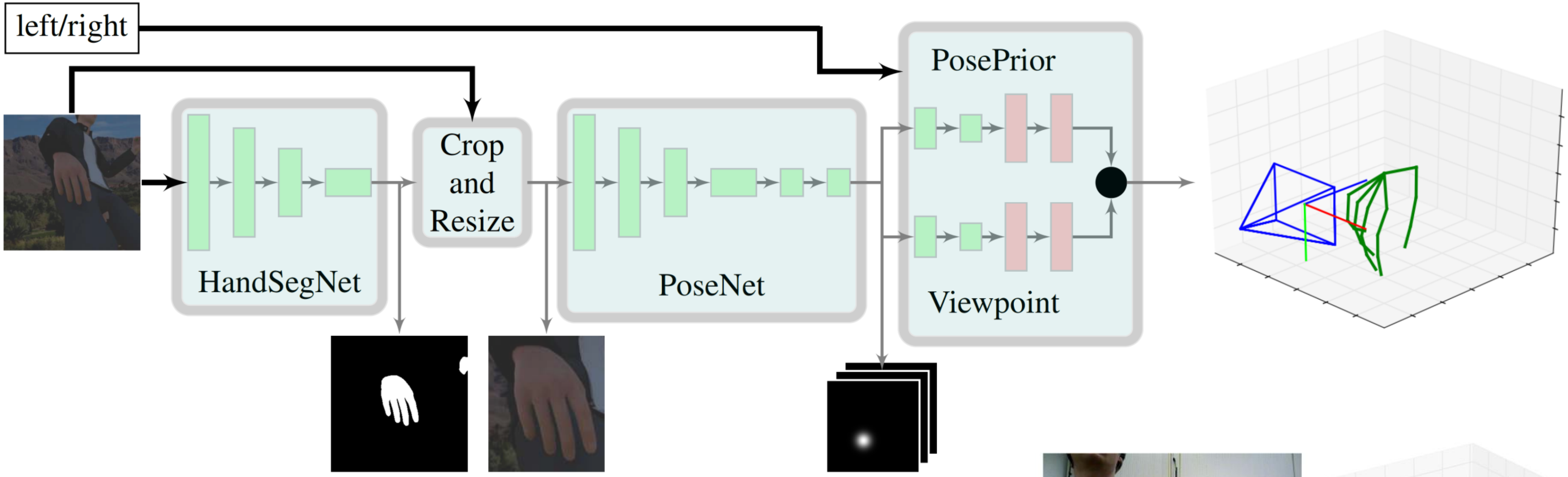
OpenPose: https://github.com/CMU-Perceptual-Computing-Lab/openpose

Hand Pose Estimation

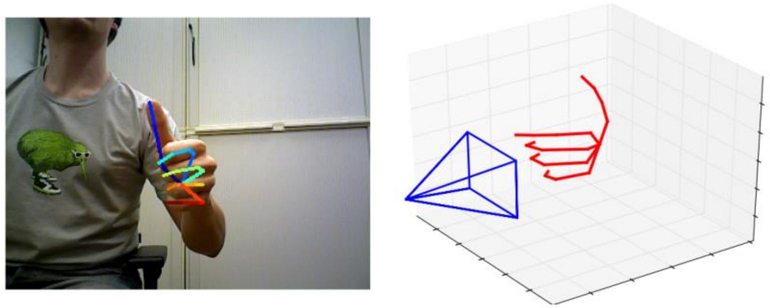
- Localizing hand joints in images or videos
- 2D hand pose estimation
  - Detect hand joints in images (x, y)
- 3D hand pose estimation
  - Detect hand joints in 3D (x, y, z)



Hand Pose Estimation



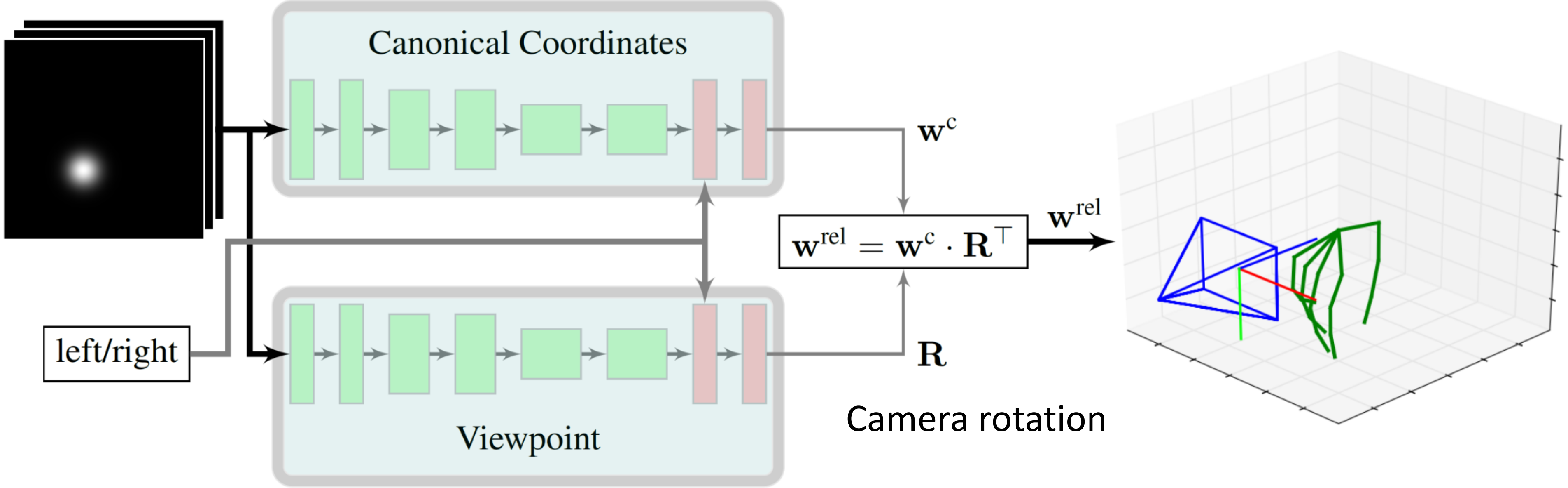
Hand keypoint score maps



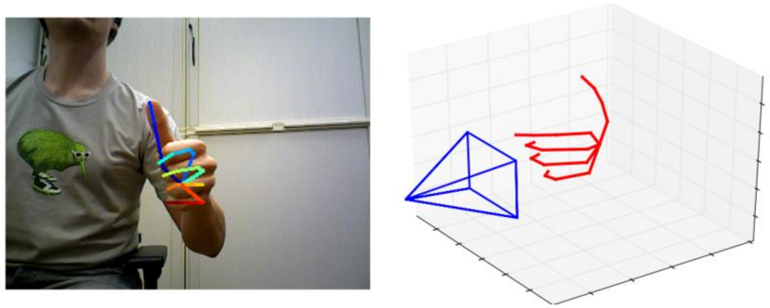
Learning to Estimate 3D Hand Pose from Single RGB Images. Zimmermann and Brox. ICCV'17.

Hand Pose Estimation

3D hand joints $\mathbf{w}_i = (x_i, y_i, z_i)$



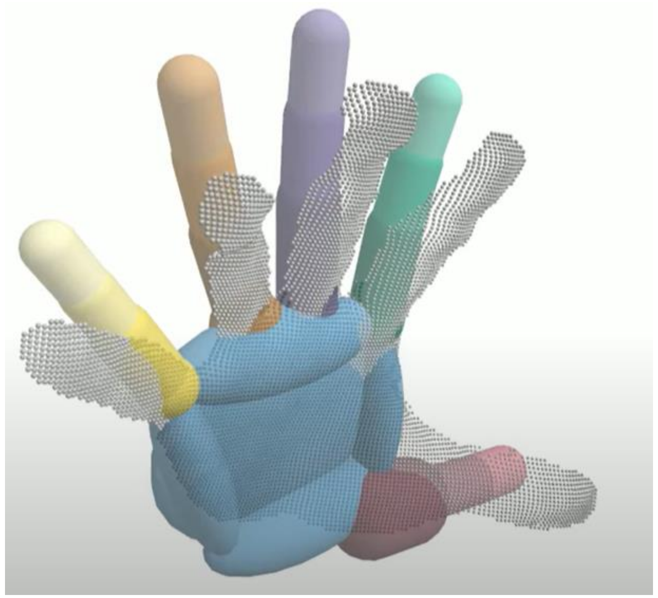
the PosePrior network



Learning to Estimate 3D Hand Pose from Single RGB Images. Zimmermann and Brox. ICCV’17.

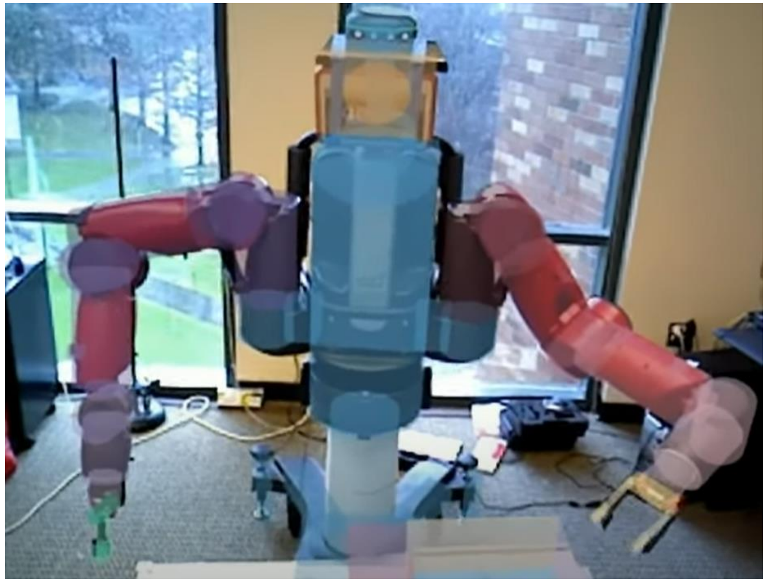
Model-based Articulated Object Tracking

- Given a 3D model of an articulated object, match the 3D model to the input image (RGB or depth)



Human hand

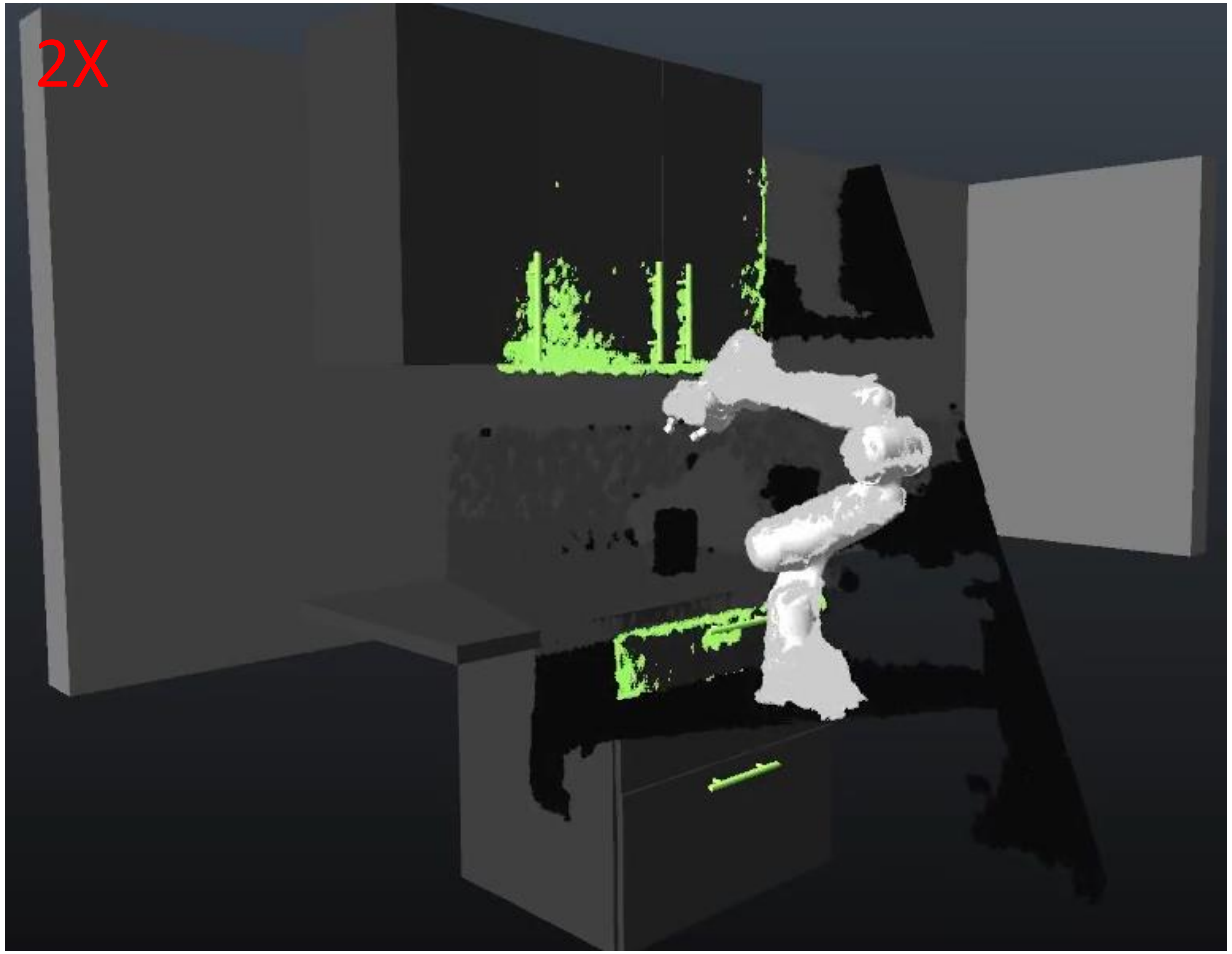
Human body



Robot

DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS'14.

Model-based Articulated Object Tracking



DART: Dense Articulated Real-Time Tracking Schmidt, Newcombe and Fox, RSS'14.

Summary

- Object pose estimation
  - Estimate 3D rotation and 3D translation of objects with respect to the camera
  - Feature-matching based methods and template-matching based methods

- Human pose estimation
  - Localizing human body joints
  - 2D or 3D

- Hand pose estimation
  - Localizing hand joints
  - 2D or 3D

Further Reading

- Making specific features less discriminative to improve point-based 3D object recognition. Hsiao, Collet and Hebert. CVPR’10. https://www.cs.cmu.edu/~ehsiao/ehsiao_cvpr10.pdf
- Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. Hinterstoisser et al., ACCV’12. http://www.stefan-hinterstoisser.com/papers/hinterstoisser2012accv.pdf
- PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. Xiang et al., RSS’18. https://arxiv.org/abs/1711.00199
- DeepPose: Human Pose Estimation via Deep Neural Networks. Toshev and Szegedy, CVPR’14 https://arxiv.org/abs/1312.4659
- Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Cao et al, CVPR’17. https://arxiv.org/abs/1611.08050
- Learning to Estimate 3D Hand Pose from Single RGB Images. Zimmermann and Brox. ICCV’17. https://arxiv.org/abs/1705.01389
- DART: Dense Articulated Real-Time Tracking. Schmidt, Newcombe and Fox, RSS’14. http://www.roboticsproceedings.org/rss10/p30.pdf