基于模型匹配的
室内物体重建与追踪

CONTANTS

研究背景

Introduction

Microsoft Kinect

Intel RealSense

RGB-D camera

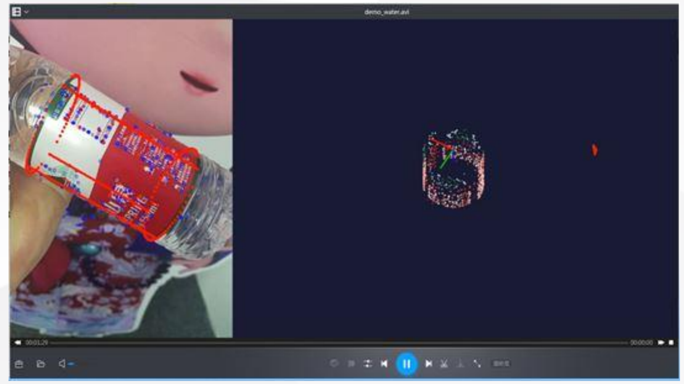
ASUS Xtion

Google Tango

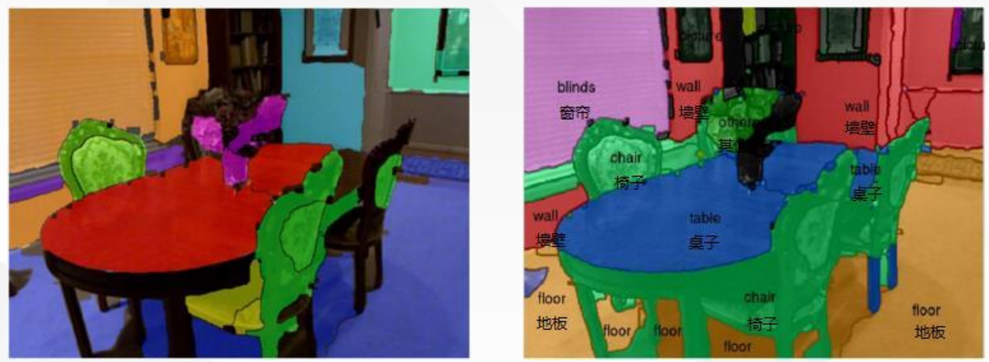
工作原理:

- 双目视觉
- 3D结构光 (Structured Light)
- 飞行时间 (Time of Flight, ToF)

1



2


blinds
窗帘
wall
墙壁
wall
墙壁
chair
椅子
table
桌子
wall
墙壁
table
桌子
floor
地板
floor
floor
chair
椅子
floor
floor
地板

3

研究现状

Related Work

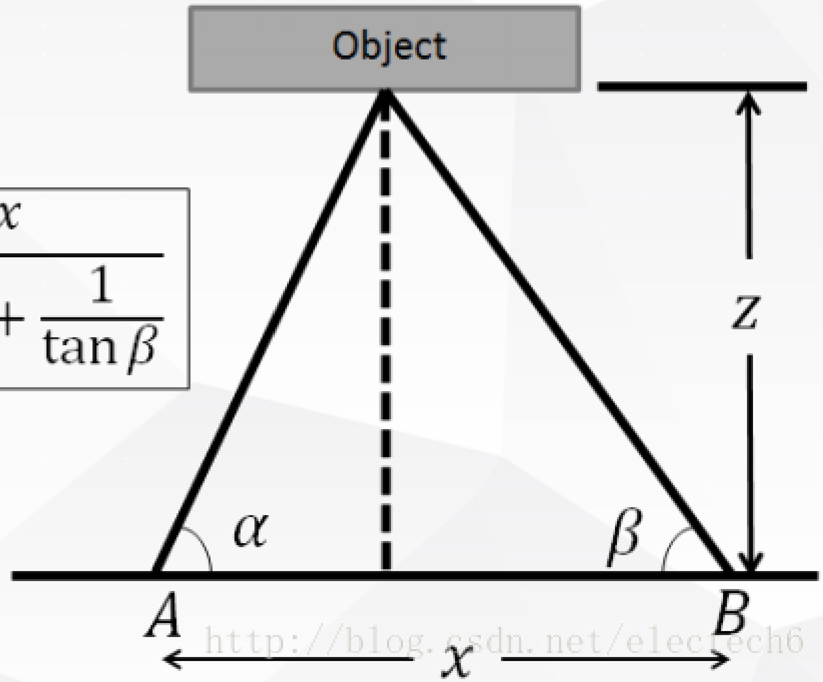
A RGB-D Camera

B SLAM

C 2D-3D Alignment

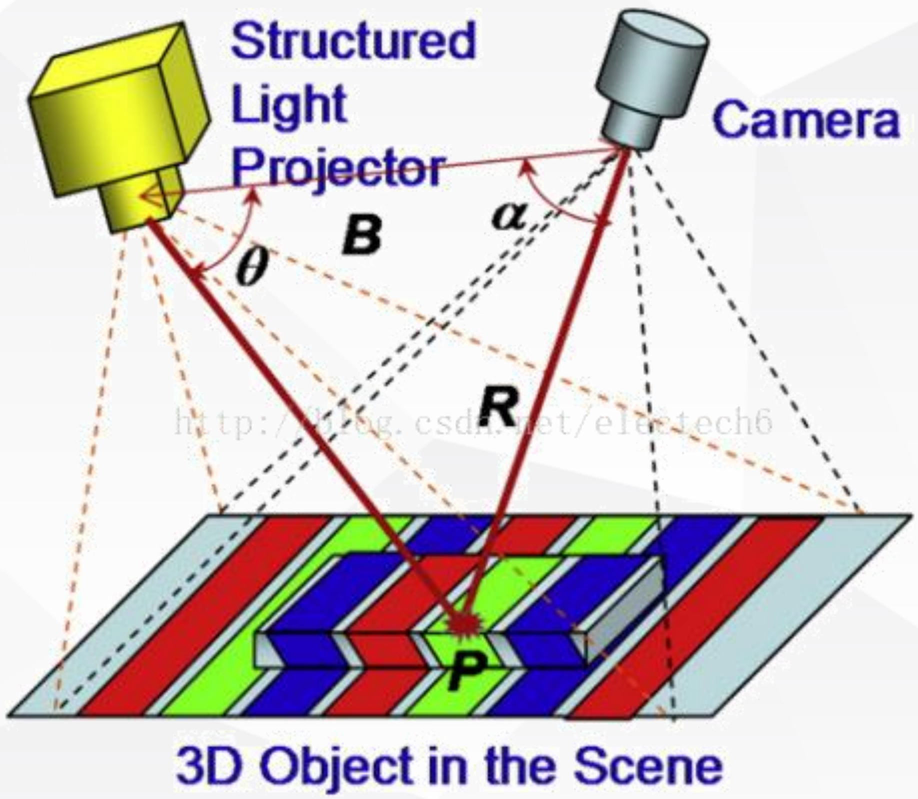
RGB-D Camera

$$z = \frac{x}{\frac{1}{\tan\alpha} + \frac{1}{\tan\beta}}$$



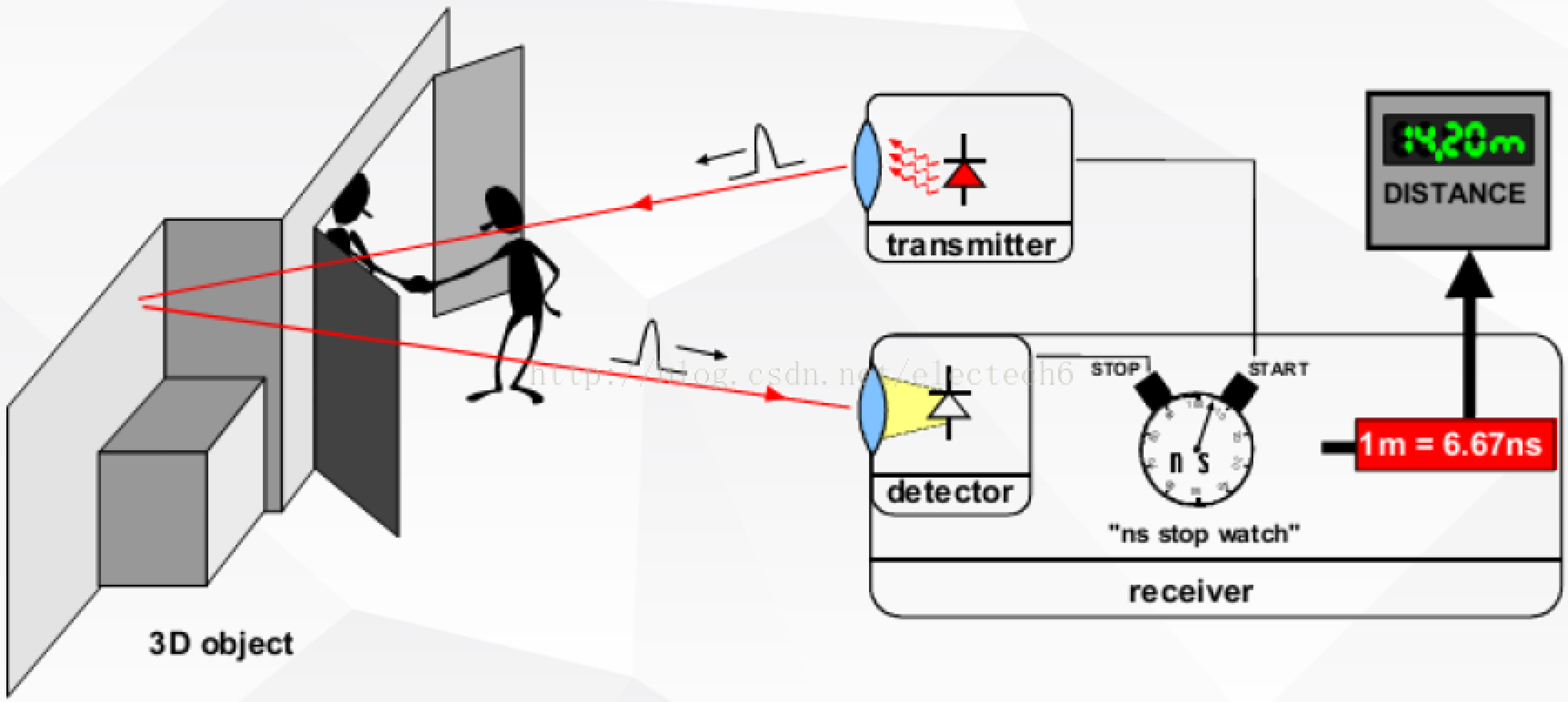
双目立体视觉法

- 原理与人眼类似
- 通过计算空间中同一个物体在两个相机成像的视差，根据三角关系计算物体与相机的距离



3D结构光法

- 投射特殊结构的图案（离散光斑、条纹光、编码结构光等）
- 使用另外一个相机观察在三维物理表面成像的畸变情况



飞行时间法 (Time of Flight, ToF)

- 连续发射经过调制的特定频率的光脉冲（一般为不可见光）到被观测物体上
- 接收从物体反射回去的光脉冲
- 探测光脉冲的飞行（往返）时间来计算被测物体离相机的距离

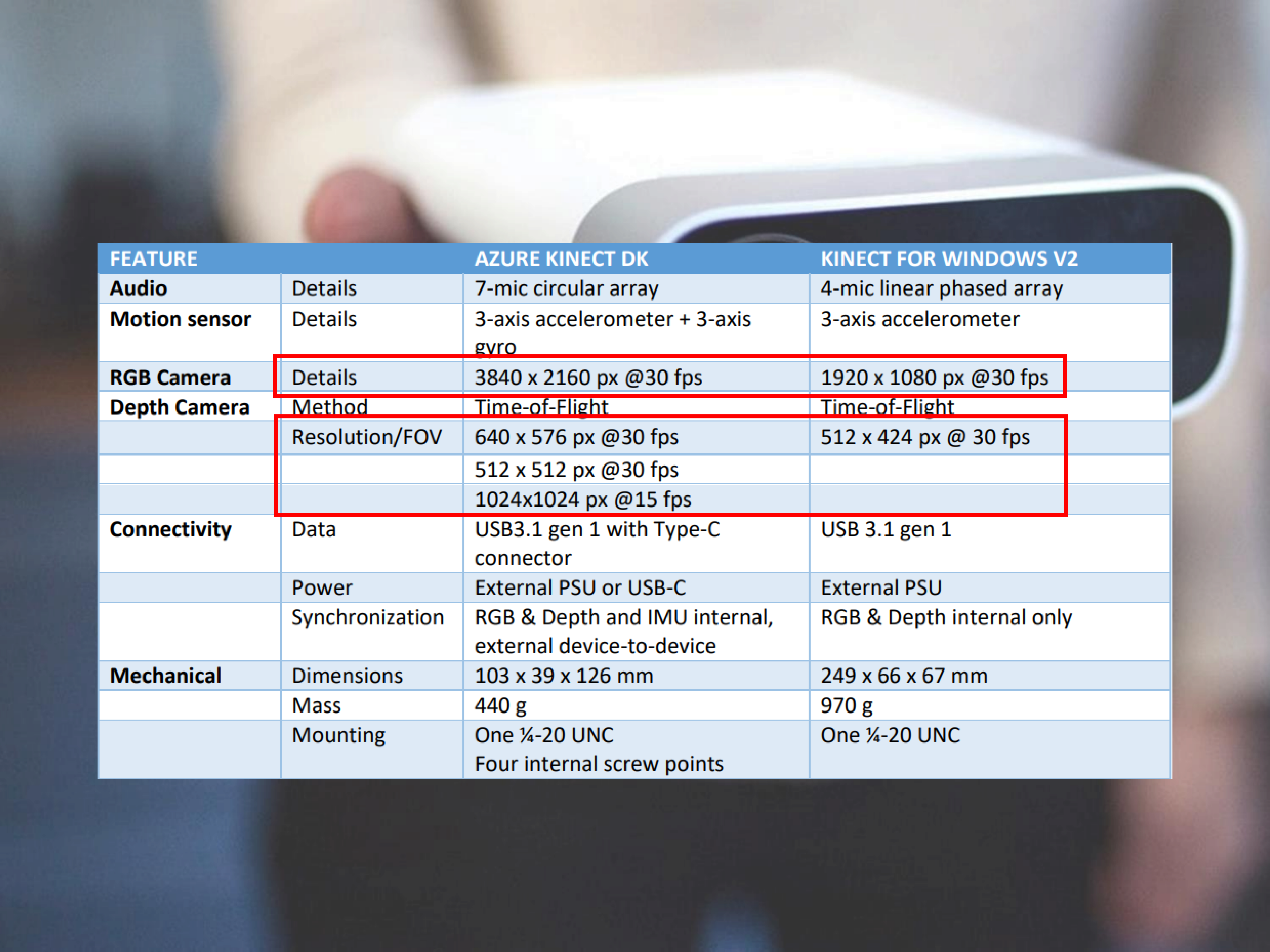
方案	双目	3D结构光	ToF
基本原理	视差算法	散斑结构光	飞行时间
光源	无（被动式）	15000个散斑	均匀面光源
工作距离	≤2m	0.2m-1.2m	**0.4m-5m**
低光表现	差	**良好**、取决于光源	**良好**（红外激光）
强光表现	**良好**	差	中等
深度精度	差 误差5%-10%	**高** **误差0.1%-0.5%**	中 误差0.5%-1%
平面分辨率	中	**高**	低
代表应用	Leap Motion 背景虚化	iPhone X Kinect v1	ASUS Xtion 2 Kinect v2

Table 3.1 Comparison of the main 3D camera commercially available

Device	Technology	Range (m)	Resolution	Frame rate (fps)	Field of view
PMD CamCube 2.0™	Time-of-Flight	0–13	200 × 200	80	40° × 40°
PMD CamBoard™	Time-of-Flight	0.1–4.0	224 × 171	45	62° × 45°
MESA SR 4000™	Time-of-Flight	0.8–8.0	176 × 144	30	69° × 56°
MESA SR 4500™	Time-of-Flight	0.8–9.0	176 × 144	30	69° × 55°
ASUS Xtion™	Structured-light	0.8–4.0	640 × 480	30	57° × 43°
Occipital™	Structured-light	0.8–4.0	640 × 480	30	57° × 43°
Sense 3D scanner™	Structured-light	0.8–4.0	640 × 480	30	57° × 43°
Kinect V1™	Structured-light	0.8–4.0	640 × 480	30	57° × 43°
Kinect V2™	Time-of-Flight	0.5–4.5	512 × 424	30	70° × 60°
Creative Senz 3D™	Time-of-Flight	0.15–1.0	320 × 240	60	74° × 58°
SoftKinetic DS325™	Time-of-Flight	0.15–1.0	320 × 240	60	74° × 58°
Google Tango™ Phone	Time-of-Flight	–	–	–	–
Google Tango™ Tablet	Structured-light	0.5–4.0	160 × 120	10	–
Orbbec Astra S™	Structured-light	0.4–2.0	640 × 480	30	60° × 49.5°
Intel SR300™	Structured-light	0.2–1.5	640 × 480	90	71.5° × 55°
Intel R200™	Active stereoscopy	0.5–6.0	640 × 480	90	59° × 46°
Intel Euclid™	Active stereoscopy	0.5–6.0	640 × 480	90	59° × 46°
Intel D415™	Active stereoscopy	0.16–10	1280 × 720	90	63.4° × 40.4°
Intel D435™	Active stereoscopy	0.2–4.5	1280 × 720	90	85.2° × 58°
StereoLabs ZED™	Passive stereoscopy	0.5–20	4416 × 1242	100	110°(*diag.*)

Giancola, Silvio, Matteo Valenti, and Remo Sala. A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies. Springer, 2018.

名称	Azure Kinect
尺寸	126.00 x 103.00 x 39.00 mm
重量	440 g
深度摄像头	100 万像素 ToF
RGB 摄像头	1200 万像素，卷帘快门 CMOS 传感器



FEATURE		AZURE KINECT DK	KINECT FOR WINDOWS V2
Audio	Details	7-mic circular array	4-mic linear phased array
Motion sensor	Details	3-axis accelerometer + 3-axis gyro	3-axis accelerometer
RGB Camera	Details	3840 x 2160 px @30 fps	1920 x 1080 px @30 fps
Depth Camera	Method	Time-of-Flight	Time-of-Flight
	Resolution/FOV	640 x 576 px @30 fps	512 x 424 px @ 30 fps
		512 x 512 px @30 fps	
		1024x1024 px @15 fps	
Connectivity	Data	USB3.1 gen 1 with Type-C connector	USB 3.1 gen 1
	Power	External PSU or USB-C	External PSU
	Synchronization	RGB & Depth and IMU internal, external device-to-device	RGB & Depth internal only
Mechanical	Dimensions	103 x 39 x 126 mm	249 x 66 x 67 mm
	Mass	440 g	970 g
	Mounting	One ¼-20 UNC Four internal screw points	One ¼-20 UNC

SLAM

SLAM

- **S**imultaneous **L**ocalization **a**nd **M**apping
- 同时定位与地图构建
- 它是指搭载特定**传感器**的主体，在**没有环境先验信息**的情况下，于**运动过程中**建立**环境**的模型，同时估计自己的**运动**

视觉SLAM

- 传感器主要为**相机**
  - 单目相机 (Monocular)
  - 双目相机 (Stereo)
  - **深度相机 (RGB-D)**
    - 室内场景

回环
检测

回环检测

判断机器人是否曾经到达过先前的位置。如果检测到回环，则把信息提供给后端进行处理。

传感器
数据

前端
视觉里程计

后端
非线性优化

建立
地图

传感器信息读取

在视觉SLAM中主要为相机图像信息的读取和预处理。

视觉里程计

估算相邻图像间相机的运动，以及局部地图的样子。视觉里程计（**V**isual **O**dometry, VO)又称为前端 (Front End)。

后端优化

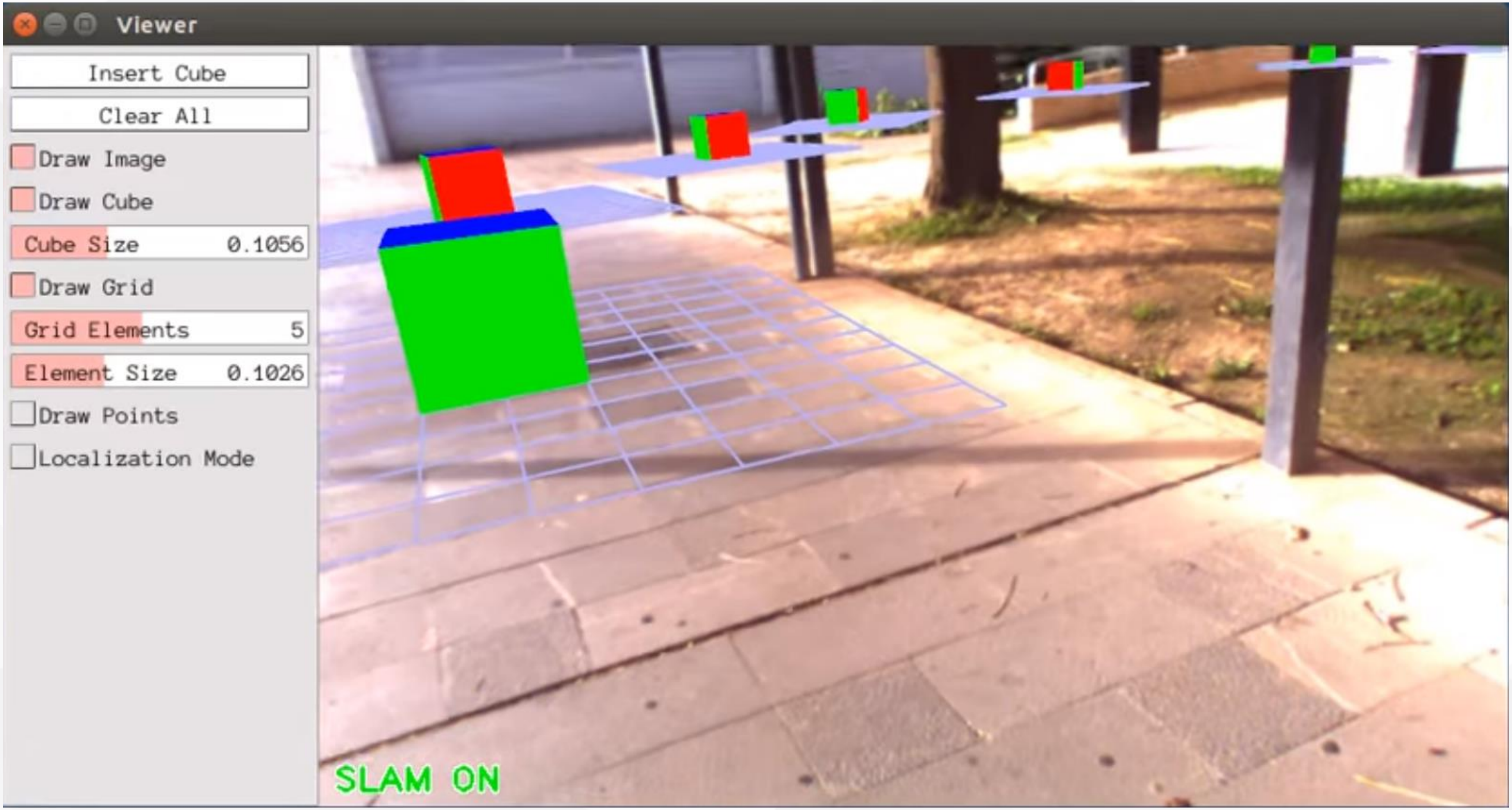
接受不同时刻VO测量的相机位姿，以及回环检测的信息，对它们进行优化，得到全局一致的轨迹和地图。由于接在VO之后，又称为后端 (Back End)。

建图

根据估计的轨迹，建立与任务要求对应的地图。

ORB-SLAM2

- ORB (Oriented FAST and Rotated BRIEF) 是一种快速特征点提取和描述的算法，具有旋转和尺度不变性，并且能够迅速地提取特征并进行匹配
- 基于单目、双目以及RGB-D的完整开源方案
- 支持地图重用、回环检测和重新定位
- 能够在标准的CPU上进行实时工作
- 包含了一个轻量级的定位模型，能够利用视觉里程计来追踪未建图的区域并且匹配特征点
- 由三个并行的线程组成
  - 跟踪：通过每一帧图像定位相机，选择是否加入关键帧
  - 局部建图：处理新的关键帧，完成重建
  - 回环检测：对新加入的关键帧进行回环检测


Viewer
Insert Cube
Clear All
Draw Image
Draw Cube
Cube Size 0.1056
Draw Grid
Grid Elements 5
Element Size 0.1026
Draw Points
Localization Mode
SLAM ON

2D-3D Alignment

Fast Alignment of 3D Geometrical Models and 2D Color Images using 2D Distance Maps

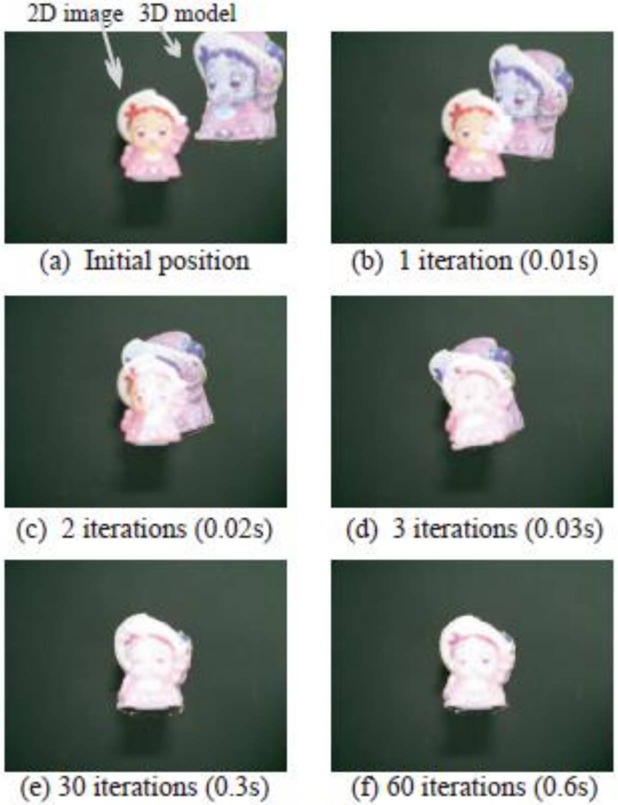
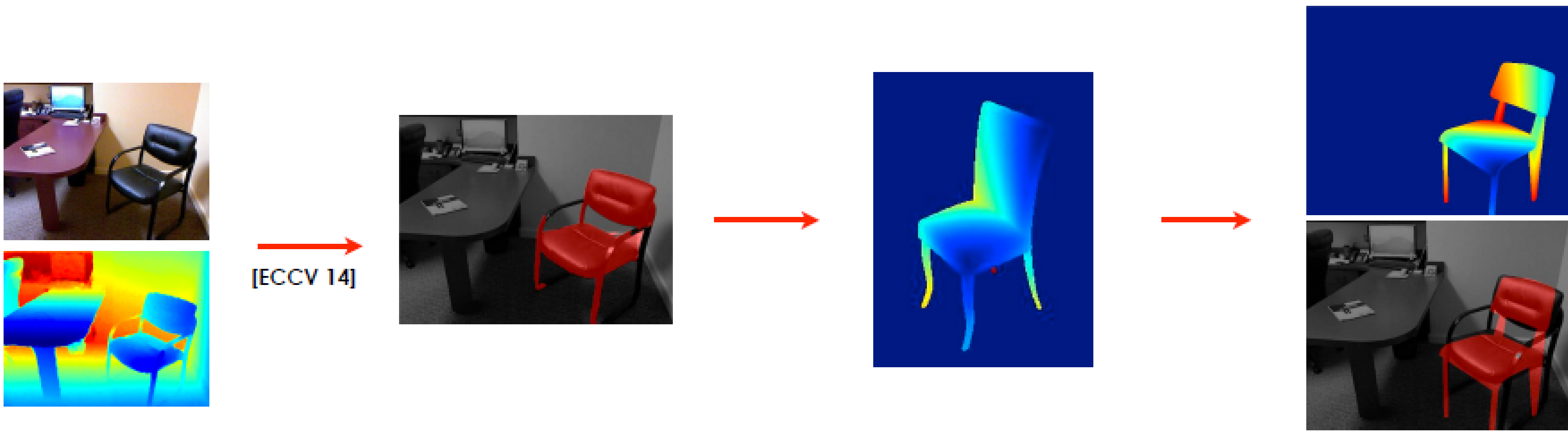


Figure 9. 2D-3D registration of simulation images.

- [Yumi Iwashita et. 2005]
- 基于2D图像与3D模型的轮廓线进行匹配
- 使用主动轮廓线模型 (Active Contour Model) 提取2D图像轮廓线
- 需要手动标记初始轮廓线
- 只能对齐与图像对应的模型
- 只适用于不规则模型
- 无法应用于实时场景

Aligning 3D Models to RGB-D Images of Cluttered Scenes

Input **Overview** Output



Input **Instance Segmentation** **Estimate Coarse Pose** **Align to data**

3D reasoning by initial 2D processing and then 'lifting' to 3D

Learning from synthetic data and generalizing to real data

Starting with weak annotation (instance segmentation) able to produce a much richer output

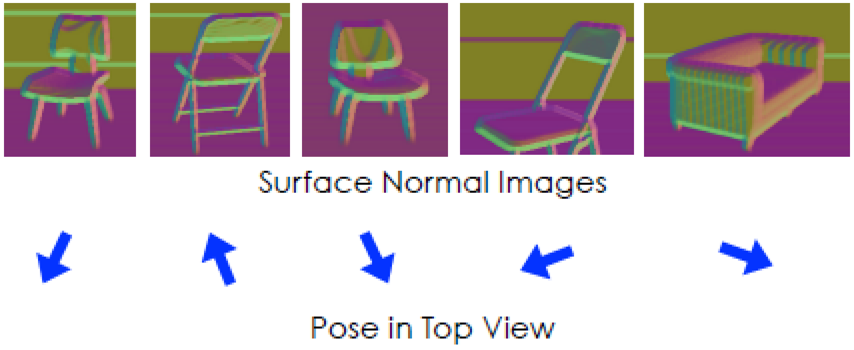
3 layer CNN on **normal images** trained on **synthetic** data

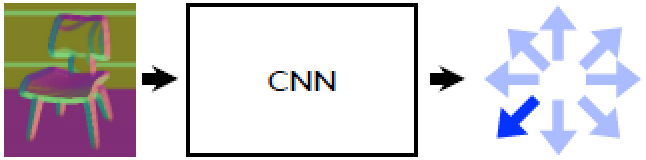
Search over **scale, placement** and **sub-type** to minimize re-projection error

Aligning 3D Models to RGB-D Images of Cluttered Scenes

Coarse Pose Estimation

- Train on **synthetic data** (pose aligned CAD models [Wu et al.] rendered in scales and positions they occur in scenes)
- **Input representation**
  - HHA (depth, height above ground, angle with gravity) images don't have azimuth information
  - Normal Images
- Desirable to be **robust to occlusion**
- Depth images are 'simpler', so we use a **shallow network**





Use a shallow 3 layer fully convolutional network (average pooling to predict)

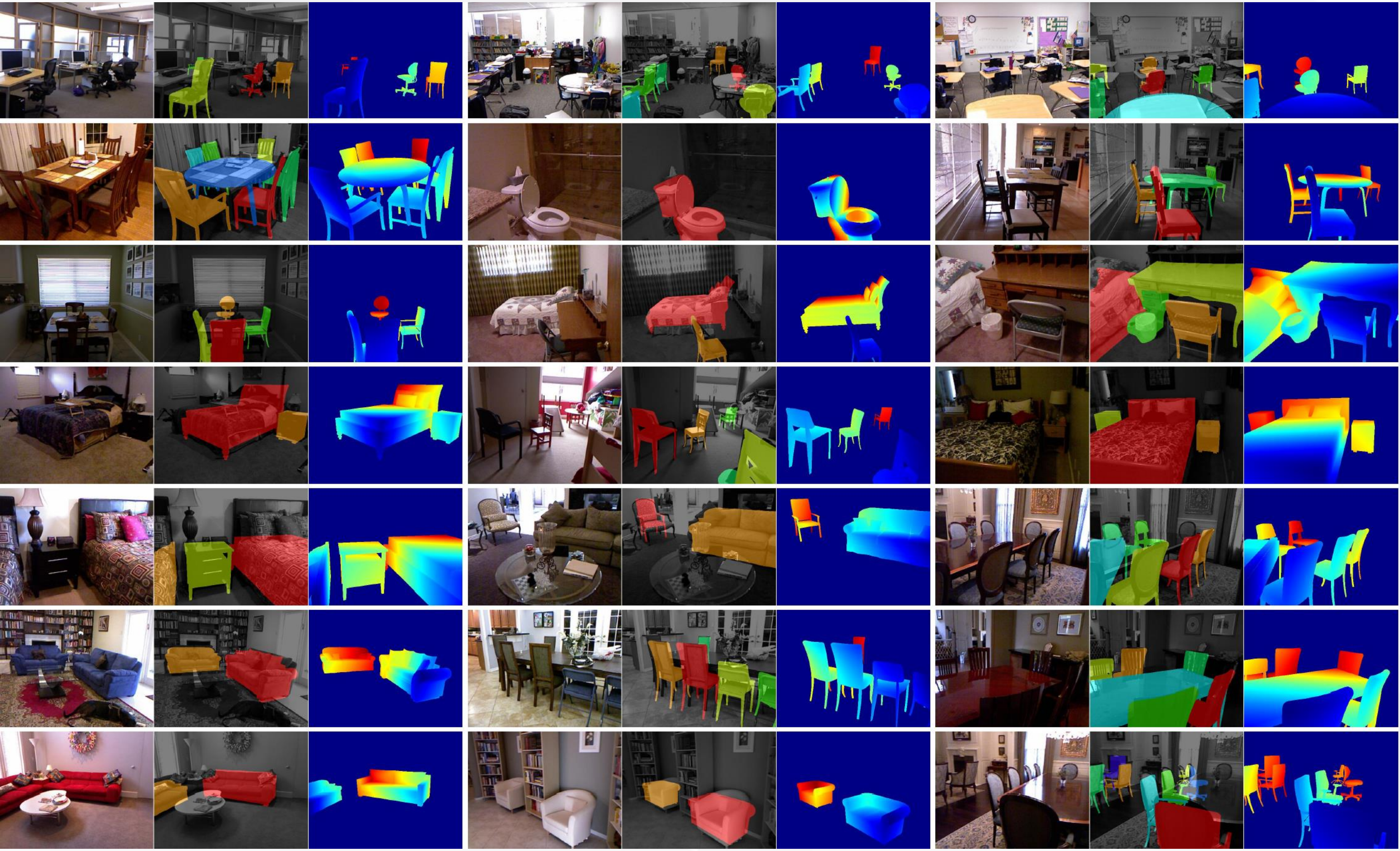
Aligning 3D Models to RGB-D Images of Cluttered Scenes

Fine Pose Estimation

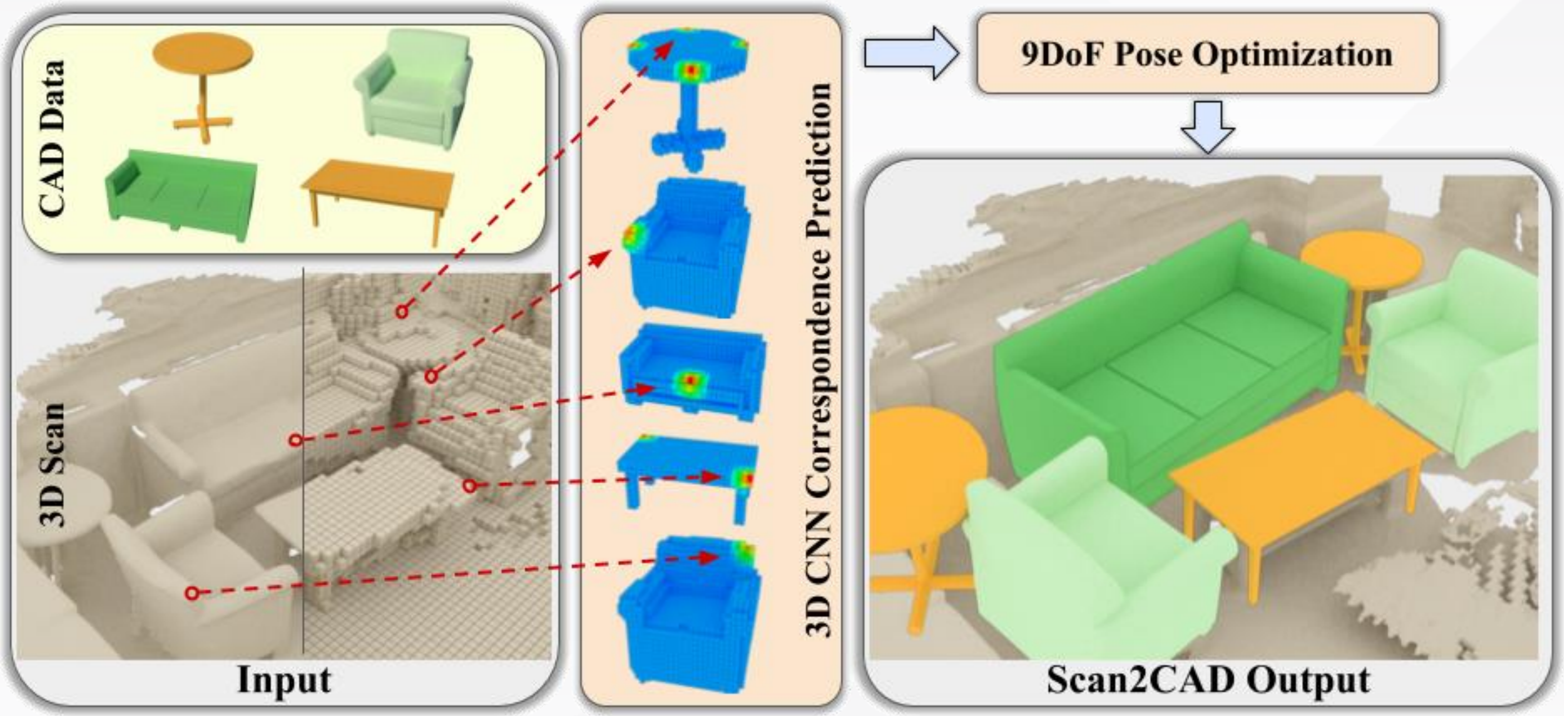
- Start with a model ***M***, at scale ***s***, an initial pose estimate ***R***
  - Iterative Closest Point (ICP) to optimize for ***R***, ***t*** (that aligns best to data)
    - Render model, use visible points, run ICP between these points, and points in the segmentation mask, re-estimate ***R***, ***t***, repeat
- Pick best model ***M****, scale ***s**** and pose ***R****, ***t**** based on fit to the data

Works reasonably well even though

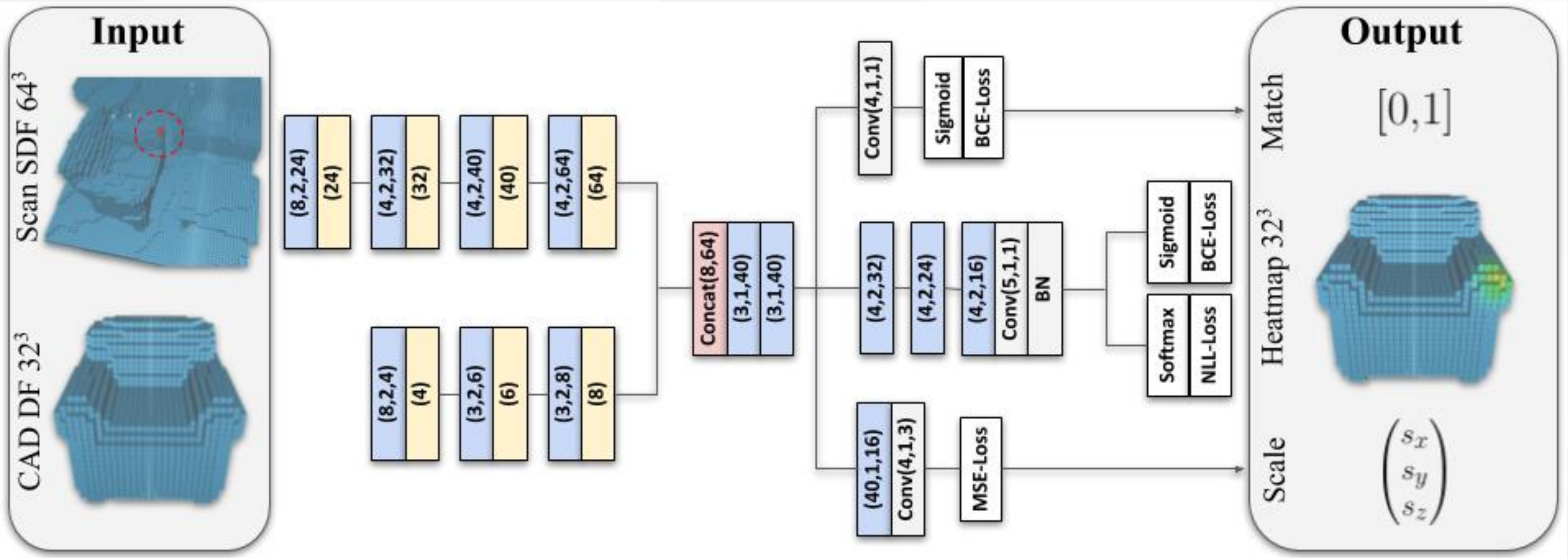
- Inaccurate models
- Imperfect segmentation masks



Scan2CAD: Learning CAD Model Alignment in RGB-D Scans



Scan2CAD: Learning CAD Model Alignment in RGB-D Scans



技术路线

1

- ↘ RGB-D相机
- ↘ SLAM（稠密重建）
- ↘ 点云分割
- ↘ 基于点云的模型匹配与追踪

2

- ↘ RGB-D相机
- ↘ SLAM（仅重建特征点用于定位）
- ↘ 基于RGB-D图像的特征提取
- ↘ 模型匹配与追踪

A 基于RGB-D相机的室内场景重建

B 在线特征提取与匹配

C 实时定位与追踪

THANKS