

Real Scale Hungry Networks: Real Scale 3D Reconstruction of a Dish and a Plate using Implicit Function and a Single RGB-D Image

Shu Naritomi¹, Keiji Yanai²

¹NTT Communications Corporation

²The University of Electro-Communications, Tokyo, Japan



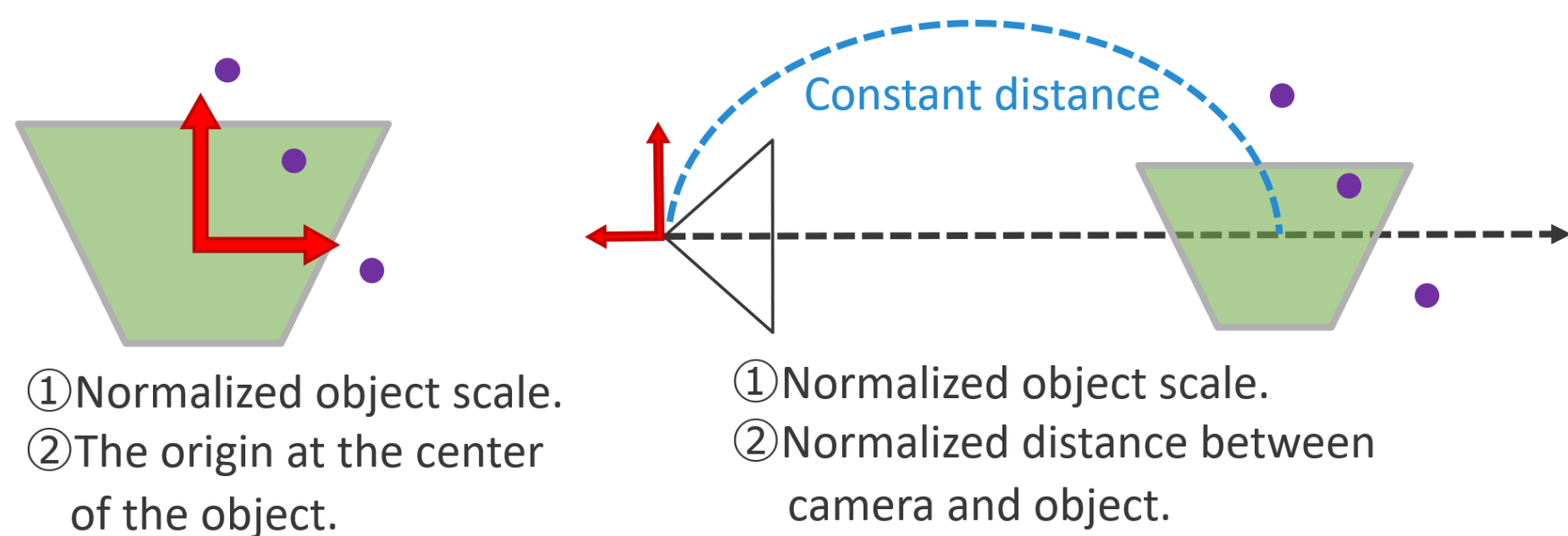
Objective

Reconstruct the 3D shapes of the dish and plate in real scale from a single RGB-D image.

The volume of the reconstruction results can be used directly for estimating the caloric content of food.

Previous method Issues

Hungry Networks [1], and other 3D reconstruction methods using implicit function representation, the output 3D shape is normalized.



Proposed Method

Depth image + perspective projection model
→ Actual size can be calculated

To maintain actual size...

→ Cannot use normalization

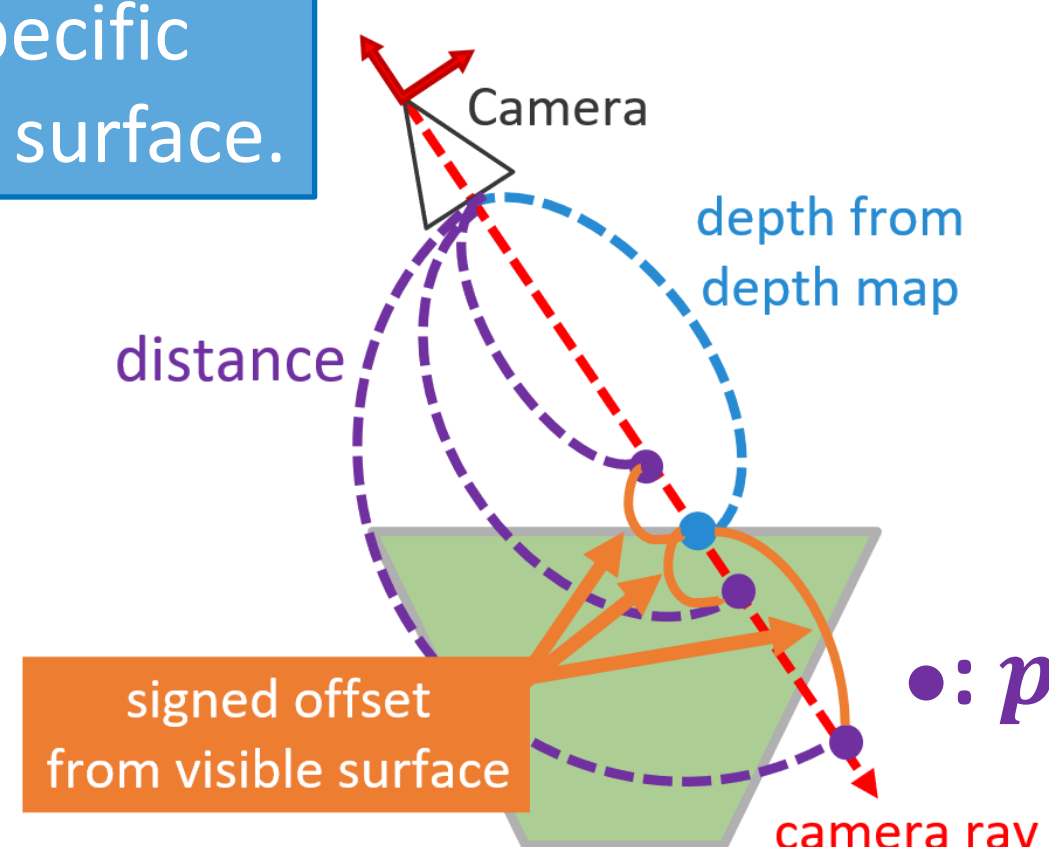
→ No existing methods can estimate actual size using implicit function representation

In this Research

We needed to train a learning model using implicit function representation without normalization.

→ **Focus on the signed offset that indicates how far away the point p from the surface of the object visible to the camera that you want to infer occupancy.**

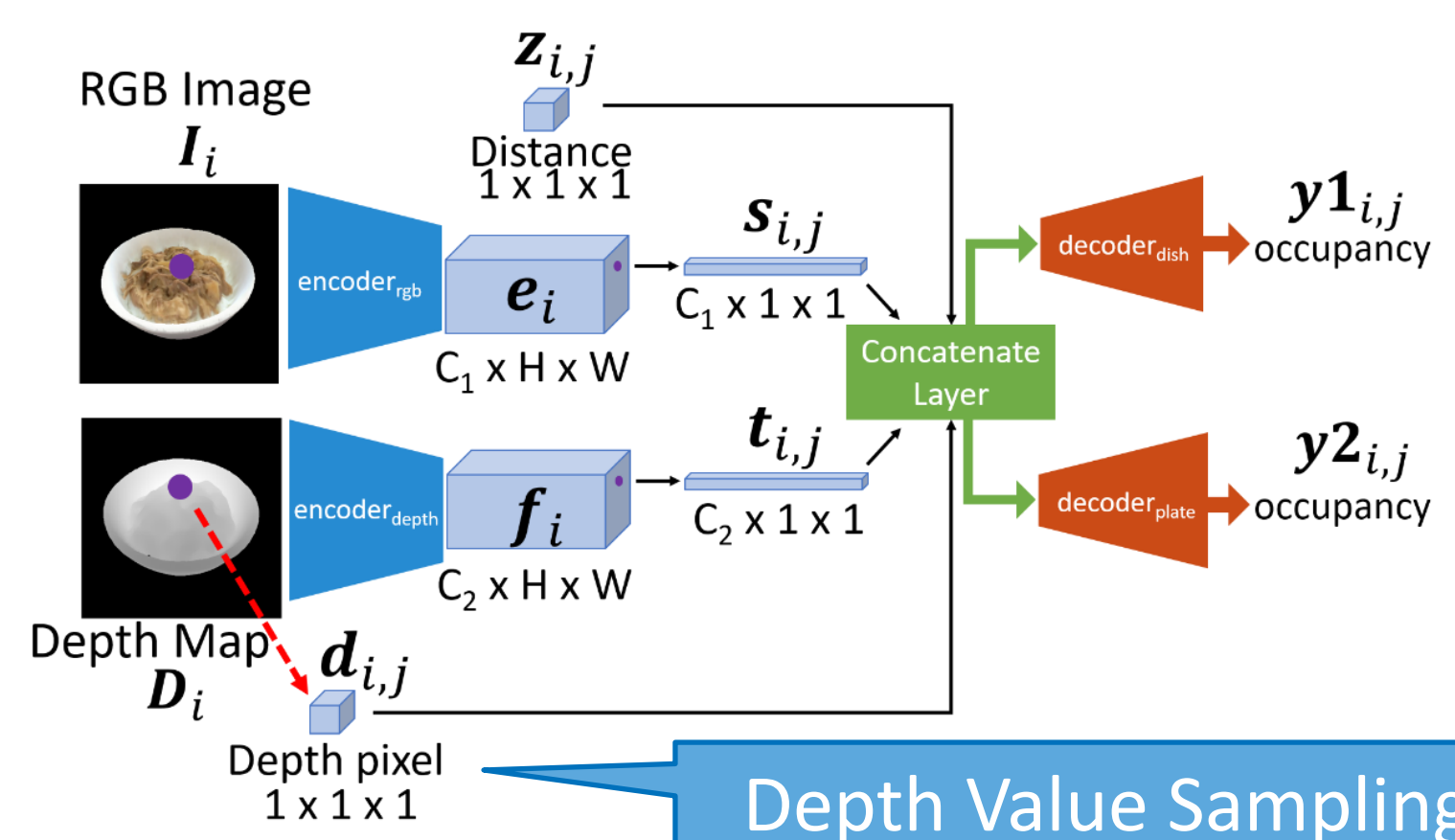
Offsets act like normalization because it fits within a specific range based on the object's surface.



Network

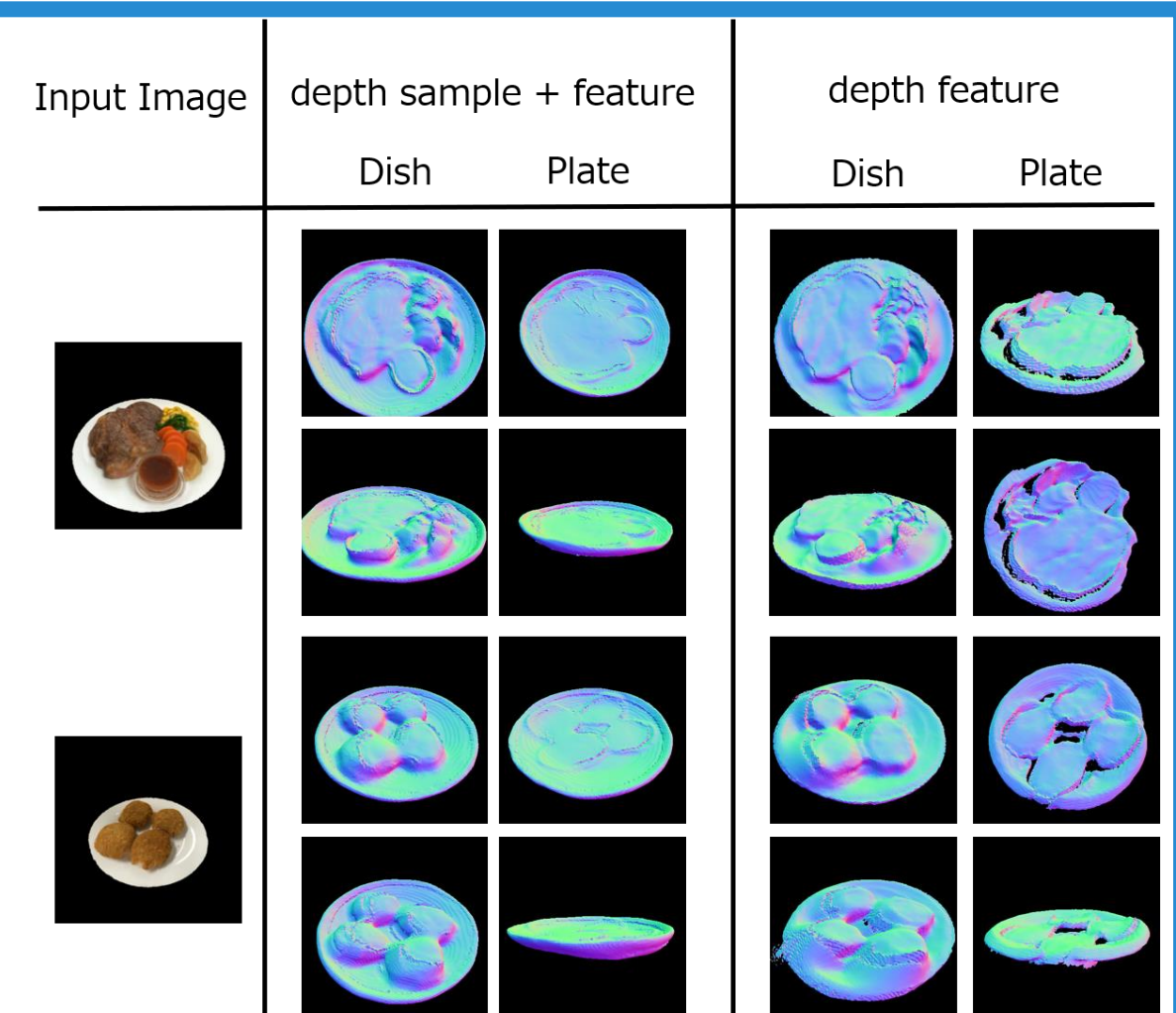
Infers occupancy of point p from the following inputs

- ① Distance z from camera to point p
- ② Depth value d of the coordinate (u, v) at which point p is projected onto the image.
- ③ s sampled from (u, v) on RGB CNN features.
- ④ t sampled from (u, v) on Depth CNN features.



Evaluation

The idea of this research realizes real scale 3D reconstruction using implicit function.



Depth	Valid	IoU (dish)↑	IoU (Plate)↑	Food Volume Error (cm ³)↓	Relative Food Volume Error↓	
					Dish	Plate
C + S	24/24	0.567	0.407	51.24	15.0 %	
C	24/24	0.534	0.337	90.291	25.9 %	
S	23/24	0.365	0.124	104.122	32.0 %	
None	1/24	invalid	invalid	invalid	invalid	

C: Depth CNN Feature S: Depth Value Sampling

Reconstruction Result.

