Multi-View Silhouette and Depth Decomposition for High Resolution 3D Object Representation

Edward Smith, McGill University; Scott Fujimoto, McGill University; David Meger, McGill University

ABSTRACT

A new approach for efficiently generating high resolution 3D objects.

- Practically impossible to generate high resolution 3D structure directly in voxel space using 3D convolutions.
- Instead employ a novel multi-view decomposition framework to predict high resolution orthographic depth maps (ODMS).
- The first effective 3D super-resolution method.
- State of the art results for 3D reconstruction.
- Highest current generation resolution.

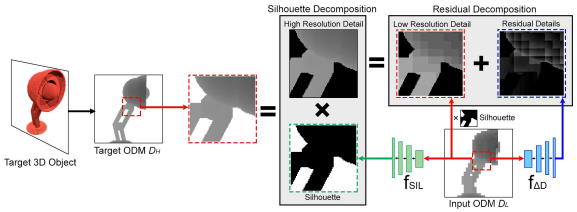
METHODS

ODMs are extracted from low resolution objects and fed through two specialized networks to predict corresponding high resolution ODMs.

- Network $f_{\Delta D}$ predicts a constrained depth map from the low resolution ODM input D_L.
  - Only residual depth is predicted, and only within a small window r to greatly constrain the prediction and take advantage of known properties in the objects.

$$C_H = r\sigma\big(f_{\Delta D}(D_L)\big) + g(D_L)$$

- Network f_{SIL} predicts an occupancy map or silhouette to capture the full structure of the object.



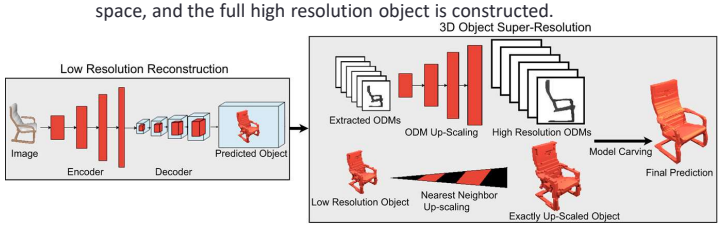
- Network outputs are then combined to produce a complete high resolution prediction $\widehat{D}_H$.

$$\widehat{D}_H = C_H \circ f_{SIL}(D_L)$$

- By working entirely in 2D space for the network predictions $O(n^2)$ complexity is achieved.
- The outputs of the system can be trivially converted into either voxel or mesh models for the full object prediction.

The 3D super resolution method is appended to the end of a 3D convolutional auto-encoder to predict high resolution objects from images.

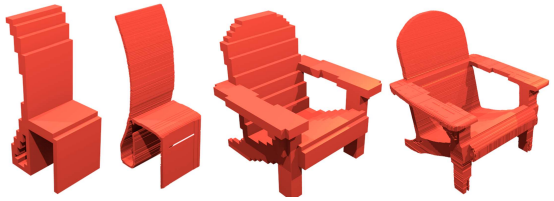
- The auto-encoder predicts the general object shape in low resolution voxel space.
- The super resolution method then predicts the fine details in 2D space, and the full high resolution object is constructed.



RESULTS: 3D SUPER-RESOLUTION

- Example super-resolution results when increasing from 32^3 resolution to 256^3 resolution.

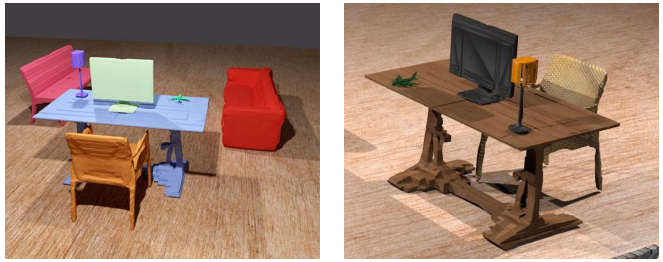
- Example super resolution results when increasing from 32^3 resolution to 512^3 resolution. This is first method capable of generating objects at this resolution.



RESULTS: IMAGE RECONSTRUCTION

- Reconstruction results from single RGB images at 256^3 resolution.

- Scene created from objects reconstructed from images at 256^3 resolution.



- 3D object reconstruction F1 scores when sampling points from object surfaces. Comparisons made to voxel, point-cloud, and mesh reconstruction algorithms.

Category	3D-R2N2	PSG	N3MR	Pixel2Mesh	MVD (Ours)
Plane	41.46	68.20	62.10	71.12	**87.34**
Bench	34.09	49.29	35.84	57.57	**69.92**
Cabinet	49.88	39.93	21.04	60.39	**65.87**
Car	37.80	50.70	36.66	**67.86**	**67.69**
Chair	40.22	41.60	30.25	54.38	**62.57**
Monitor	34.38	40.53	28.77	51.39	**57.48**
Lamp	32.35	41.40	27.97	**48.15**	**48.37**
Speaker	45.30	32.61	19.46	48.84	**53.88**
Firearm	28.34	69.96	52.22	73.20	**78.12**
Couch	40.01	36.59	25.04	51.90	**53.66**
Table	43.79	53.44	28.40	66.30	**68.06**
Cellphone	42.31	55.95	27.96	70.24	**86.00**
Watercraft	37.10	51.28	43.71	55.12	**64.07**
Mean	39.01	48.58	33.80	59.72	**66.39**

- 3D object reconstruction IoU scores at 256^3 resolution. Comparisons made to octree voxel methods.

Category	AE	HSP	MVD (Ours)
Car	55.2	70.1	**72.7**
Chair	36.4	37.8	**40.1**
Plane	28.9	56.1	**56.4**

Category	AE	OGN	MVD (Ours)
Car	68.1	78.2	**80.7**
Chair	37.6	-	**43.3**
Plane	34.6	-	**58.9**