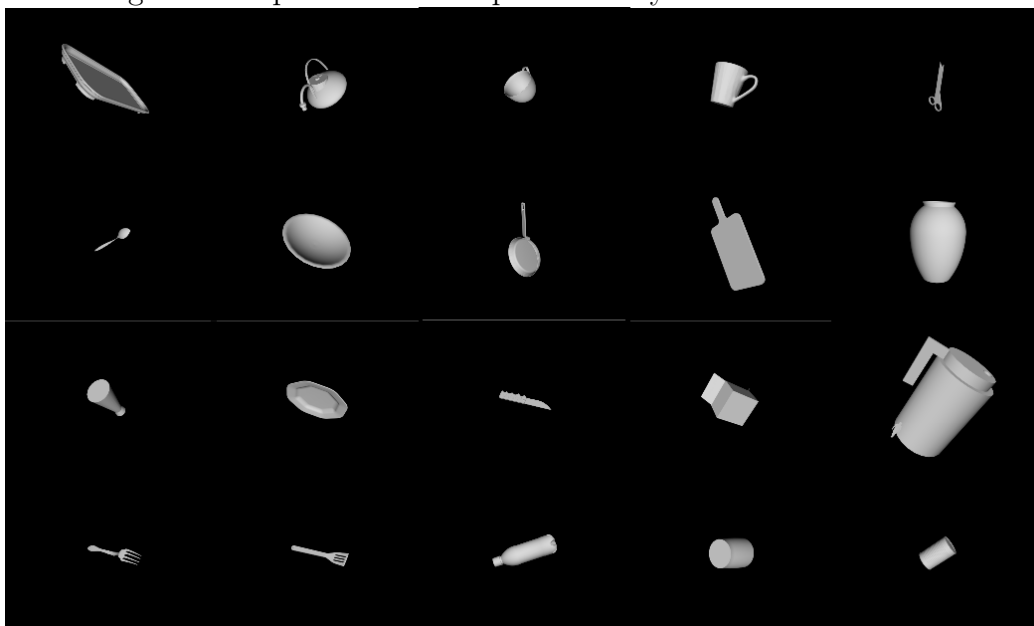


Dataset Description

February 29, 2016

In PaCMan Shape Dataset, we have collected 400 objects consisting of 20 categories, each having 20 models. The categories consist of household objects: Mug, cup, fork, jug, plate, bowl, frying pan, tray, teapot, tea cup, cutting board, box, knife, vase, can, fork, spatula, scissors, bottle and shaker. The objects have realistic sizes, and their longest dimension range from 80 mm to 200 mm, depending on the category. One representative image out of every class is depicted in Fig. 1.

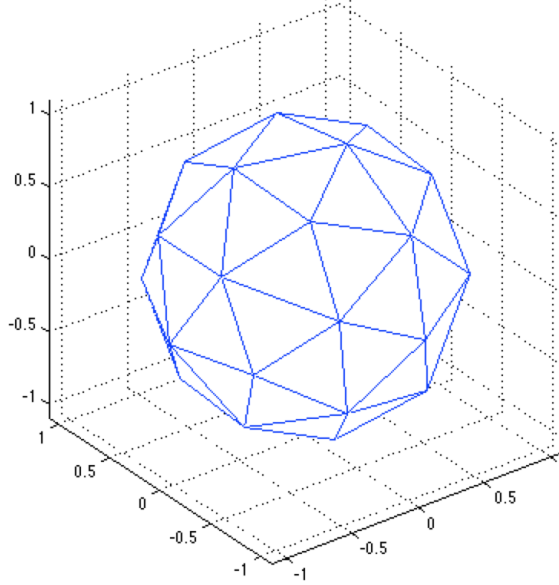
Figure 1: Representative shapes for every class in the dataset.



In order to render the objects from multiple viewpoints, we have used 32 uniformly sampled points on the view sphere. The azimuth and elevation of these samples are given below. Fig 2 shows the triangulation of the view sphere, where every vertex corresponds to a camera position, with the viewing vector directed to the centre of the sphere. We have rendered every object from each viewpoint using a simulated RGB camera and a depth camera, which has the characteristics of a Carmine 1.09 sensor.

- azimuth =(358, 62, 238, 272, 195, 167, 274, 50, 230, 242, 273, 92, 347, 314, 58, 134, 297, 336, 94, 308, 93, 254, 178, 156, 26, 184, 206, 4, 15, 128, 74, 117)
- elevation = (58, -38, -2, -26, -18, 12, -68, 39, -39, 38, 12, 26, -12, -38, 2, 38, 42, 24, 68, -1, -12, 75, -58, -24, -17, 50, 17, 50, 18, 1, -75, -42)

Figure 2: View sphere triangulation.



The image acquisition process proceeds as follows:

1. Object is placed such that its centre of gravity is at the origin.
2. Camera is located on the y axis in various distances from the origin, namely, [0, 600mm, 0] or [0, 900mm, 0] or [0, 1200mm, 0].

3. Two light sources have been placed in the scene. The more powerful light source is located above the camera, while the weaker one is below the camera, with respect to the camera's up vector. This creates enough intensity differences to reveal intensity edges on the object.
4. Camera is directed towards the origin. Camera has horizontal Field of View of 57.5 degrees.
5. Near range of the camera is 350mm, far range is 1400mm. That means that depth values in range from 350mm to 1400mm are encoded in 2 byte values. Larger values correspond to closer distances.

Rendered images of the objects are placed in a hierarchical structure, and named in a specific way. For example, for a rendered image that has the name "3_P2_IP3.png":

- 3 is the id of the object in the dataset (from 1 to 400)
- P2 - that means that values $a(2)$ and $b(2)$ were used to rotate the object (we rotate the object, not the camera).
- IP3 - id of the in-plane rotation. There are 8 ids in total, describing 8 rotation of the camera around y axis, with step size of 45 degrees. IP1 would correspond to zero orientation of the camera.

The category labels are stored in a 400x1 array that contains the label of every model in the dataset, distributed along with the images.