

399 Appendices

400 A Implementation Details

401 A.1 Real Robot Experiment

402 **Perception:** Our perception pipeline uses two OAK-D Pro cameras to estimate the 6D (x, y, z, roll, pitch, yaw) pose of objects. The side camera, oriented towards the shelf, determines the poses of the objects on the shelf. In contrast, the top camera determines the pose of the object situated on the table. We maintain a database encompassing comprehensive details of all objects in the experiment, including their size and texture. We approximate all objects as cuboids, defined by three parameters: length, width, and height. Side and top view images of these objects are captured to form a collection of ground truth images. These images aid the identification of objects via the SIFT (Scale Invariant Feature Transform) feature matching algorithm. This algorithm is executed for each object in the scene to match keypoints, which are subsequently used to calculate the homography matrix. We then use this matrix to calculate the 2D coordinates and 1D orientation for each object in the image space. Given the setup of our experiment, it suffices to estimate 3 degrees of freedom (DOF) for all objects: (y, z, roll) for objects on the shelf and (x, y, yaw) for the object on the table. The other 3 DOF is deterministic based on the known shelf location and object size. Incorporating the 3D pose of each object in the image space with our prior knowledge of the experiment setup, we can obtain 415 the 6D pose for all the objects in the scene for robot manipulation.

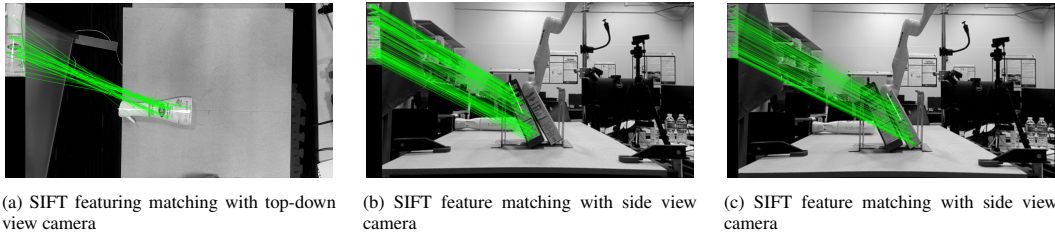




Figure 6: SIFT results with setup-(f). The ground truth images collected before the experiment are shown in the upper-left corner of each subfigure. The green lines show the keypoints correspondence.




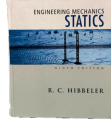





417 A.2 More information on the objects

Table 4: Object Dimensions

Object	Length (m)	Width (m)	Height (m)	Image
Sugar box	0.089	0.040	0.177	
Lint roller	0.110	0.050	0.230	


Continued on next page

Table 4 – Continued from previous page

Object	Length (m)	Width (m)	Height (m)	Image
Clorox bottle	0.065	0.255	0.055	
Cheez-it box	0.152	0.047	0.190	
Charger box	0.168	0.051	0.240	
Book	0.203	0.028	0.245	
Amazon box	0.210	0.288	0.045	
Keto box	0.193	0.259	0.046	
Kellogg's box	0.190	0.045	0.284	
Chex box	0.195	0.050	0.285	
Cheerios box	0.195	0.052	0.285	

Continued on next page

Table 4 – *Continued from previous page*

Object	Length (m)	Width (m)	Height (m)	Image
Bran flakes box	0.198	0.297	0.057	
Power drill box	0.250	0.073	0.279	