

CS365 Project Proposal

Group 6

Ankit Agarwal, Ripu Singla

Vision Based Grasping

- Our project is to find grasp point for mugs using vision only and to use Nao for physical grasping.
- We were able to predict grasp points in synthetic as well as real images with high precision.
- Little work of planning on Nao is left otherwise most of the planning part is also complete.

Dataset Used

- Used 300 synthetic images from the following dataset
<http://ai.stanford.edu/~asaxena/learninggrasp/data/mug.tar.gz>
- All the images are of size 640 * 480 PX
- Every image in dataset contains grasp label
- No depth information has been used
- Other images (about 600) in dataset are used for testing purpose.

Training

- First we prepare the feature vector for each training image.
- Feature Vector (after converting to YCbCr color space)
 - 6 oriented edge filters applied to intensity channel
 - 9 laws mask applied to intensity channel
 - 2 features from 2 color channels
 - Calculate above 17 feature for 3 different scales (1,2,5) and also for 24 neighbours in 5*5 window
- Then we create a generalized linear model to fit the feature vectors obtained above.

Testing

- First we resize the image to $640 * 480$.
- Then divide the image into grids of $10 * 10$.
- Create the feature vector for the test image as for training images.
- Then predict values from generalized linear model created beforehand.

Planning

- Our planning algorithm is fairly simple
- We detect the grasp point on the images taken from two different positions.
- Then using triangulation, we detect the 3-d coordinate of grasp point.
- Then we go up to a threshold distance towards grasp point.
- At this position, we again detect the grasp point.
- If we are in the “standard position” , then we grasp the mug.

Results

- We tested on about 600 images in the dataset
- Accuracy is about 60 %.
- For the real images, accuracy is quite high, probably due to correct orientation of mug.

Test Image



Test Image



Random Image from Google



CSE Lounge



Problems Faced

- The paper assumed availability of stereo camera which was not available to us so we were not able to use any depth information.
- The quality of image produced by Nao is very low however the results for high quality images (Digicam 10MP) is good.
- Manipulation of Nao is not accurate.