Fuzzy Semantics Understanding for natural HRI

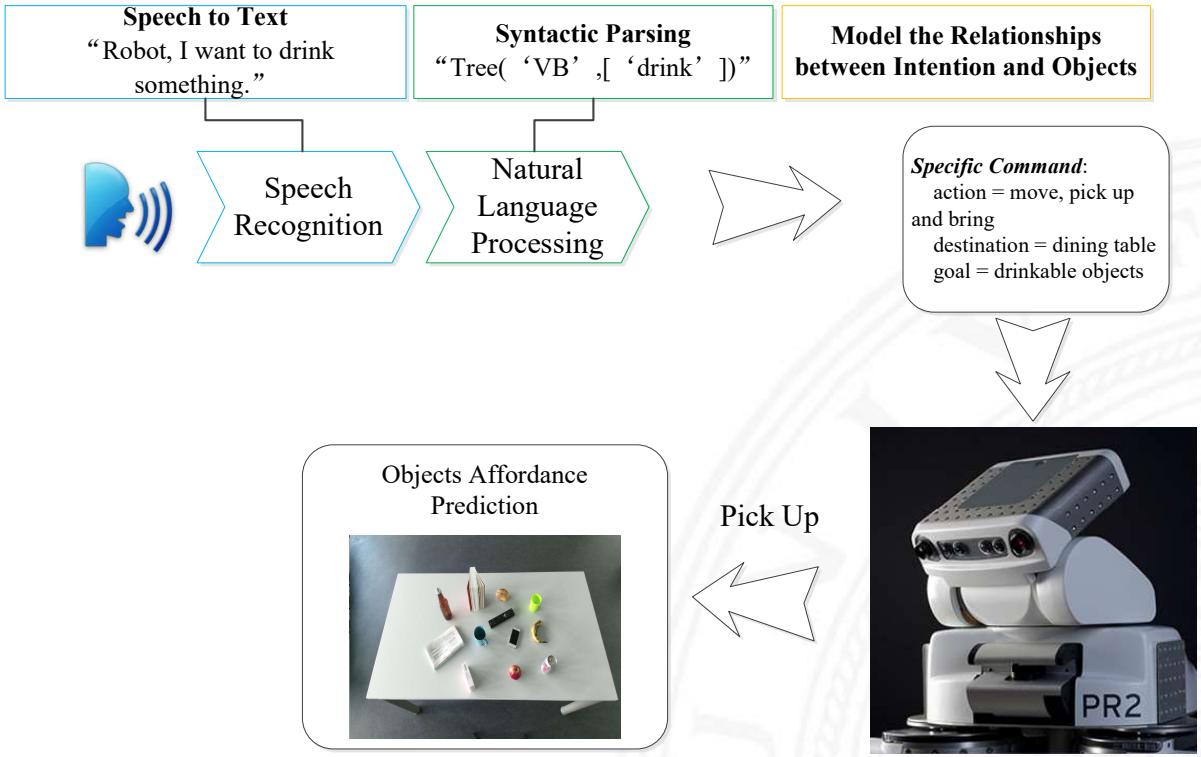
Jinpeng Mi

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme

27. June 2017

Gliederung

Basic Architecture

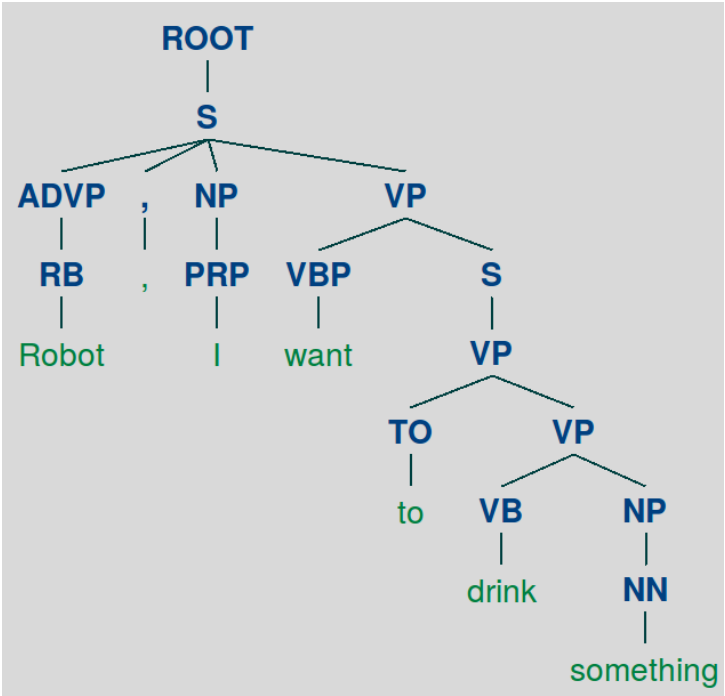


Online Speech recognizer

- `toolkit`: Kaldi
- `corpus`: WSJ0 and WSJ1
- `Acousic Model (AM)`: HMM + DNN
- `Language Model (LM)`: n-gram, n =3
- `Lexicon`: pronunciation dictionary

Syntactic Parsing

Syntax tree of "Robot, I want to drink something."



- ADVP: adverb phrase
- NP: noun phrase
- VP: verb phrase
- RB: adverb
- PRP: pronoun, personal
- VBP: verb, present tense, not 3rd person singular
- NN: noun, common, singular or mass

Objects Affordance Prediction

- Existing Method
- Data Collection and Features Extraction
- Attribute Classification
- Affordance Prediction

Existing Method

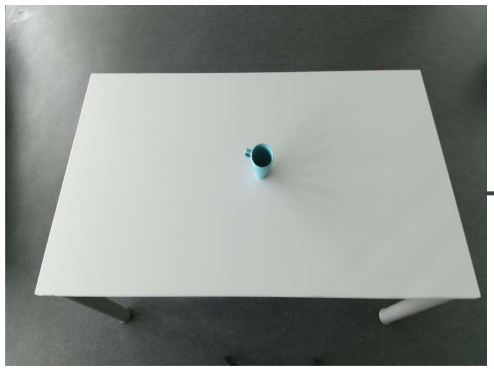
- physical and visual attributes based affordance prediction[1]
- statistical learning based affordance prediction [2]
- through human-object interaction to learning the affordance [3]
- knowledge based representation [4]

[1] Hermans T, Rehg J M, Bobick A. Affordance prediction via learned object attributes. ICRA 2011.
[2] Moldovan B, Moreno P, van Otterlo M, et al. Learning relational affordance models for robots in multi-object manipulation tasks. ICRA 2012.
[3] Koppula H S, Gupta R, Saxena A. Learning human activities and object affordances from rgb-d videos. The International Journal of Robotics Research, 2013.
[4] Zhu Y, Fathi A, Fei-Fei L. Reasoning about object affordances in a knowledge base representation. European conference on computer vision, 2014.

Data Collection

- Dataset: 12 objects, every object takes 10 photos

Features Extraction



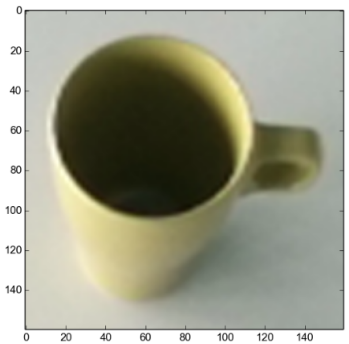
Original
(960*720)

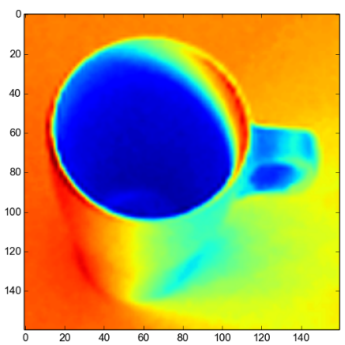
ROI and Zoom-in
(160*160)

- Canny Edge
- HOG Features
- LAB Hist
- LBP Hist

images show way define
cv2.show()

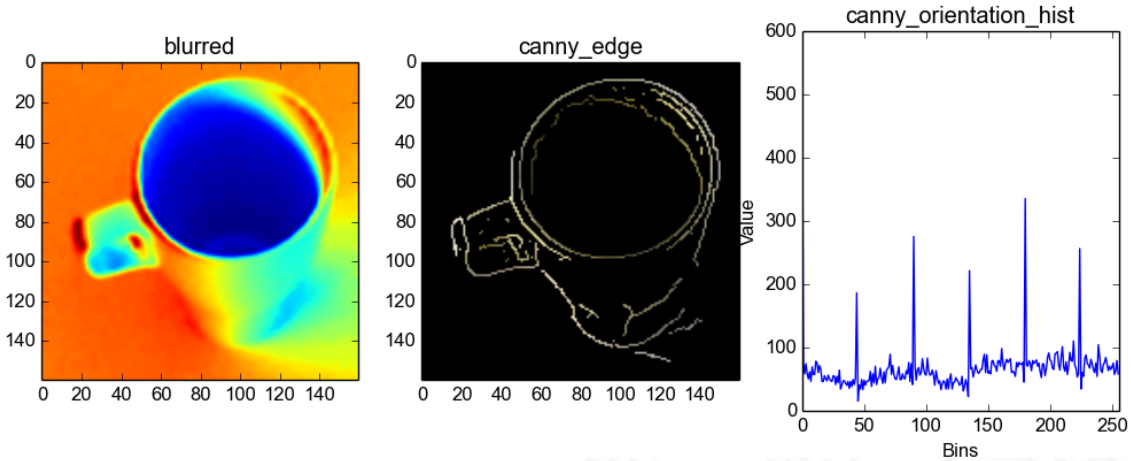
matplotlib.show()





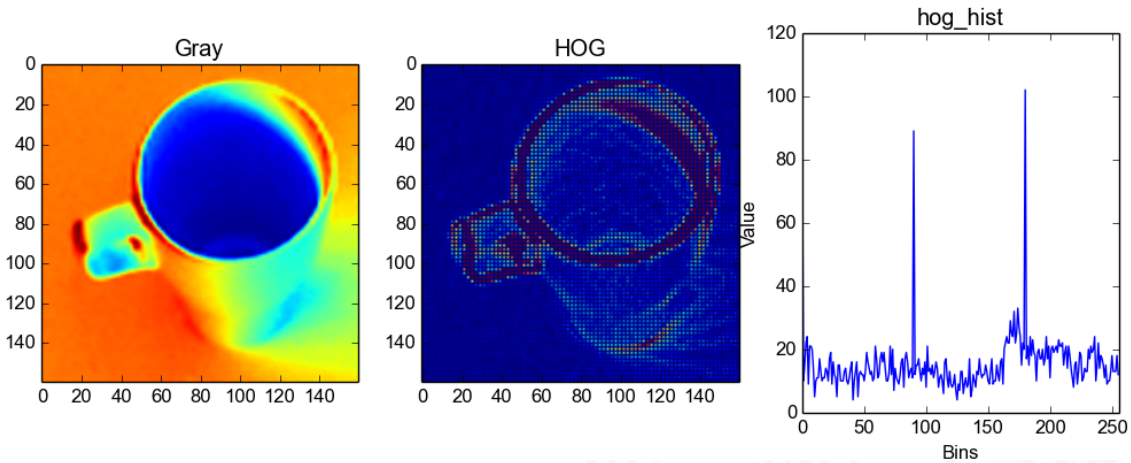
Features Extraction

Canny Edge: the direction of gradient shows how the edge is oriented.



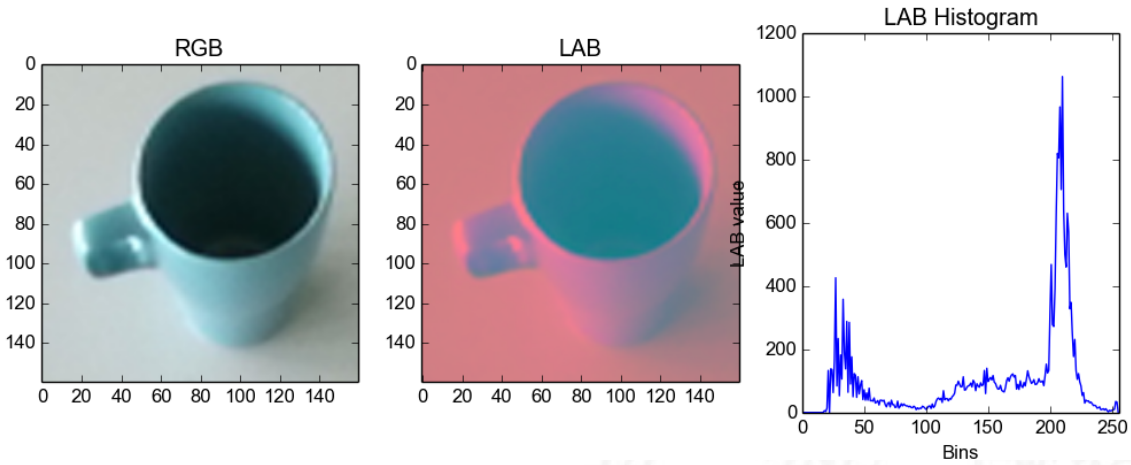
Features Extraction

HOG Features: the magnitude of gradients is large around edges and corners



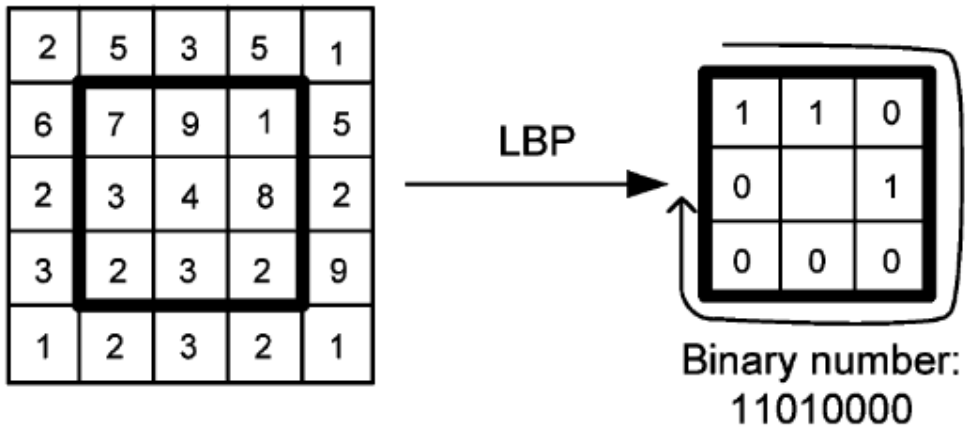
Features Extraction

LAB: another representation of color feature

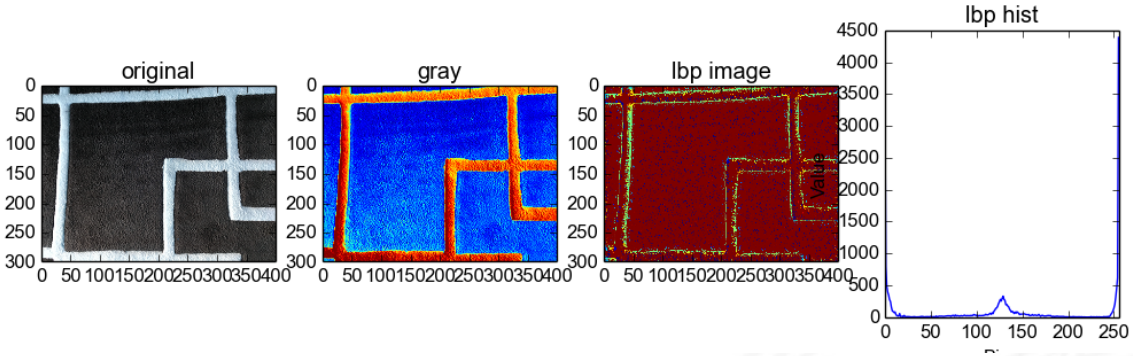


Features Extraction

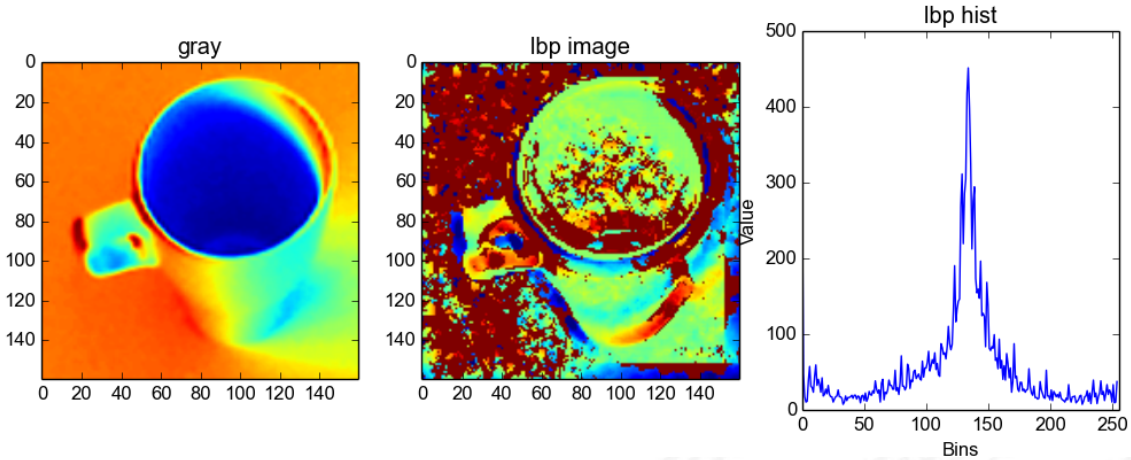
LBP(Local Binary Patterns) is used to extract texture feature.



LBP examples



mug's LBP Hist



Objects Attributes

- attributes

  - shape: 6 faces, cylindrical, rounded, curved
  - color: red, green, blue, yellow, brown, black, other
  - texture: organic, paper, plastic, glass/ceramic, metal

- example: objects in different shape class

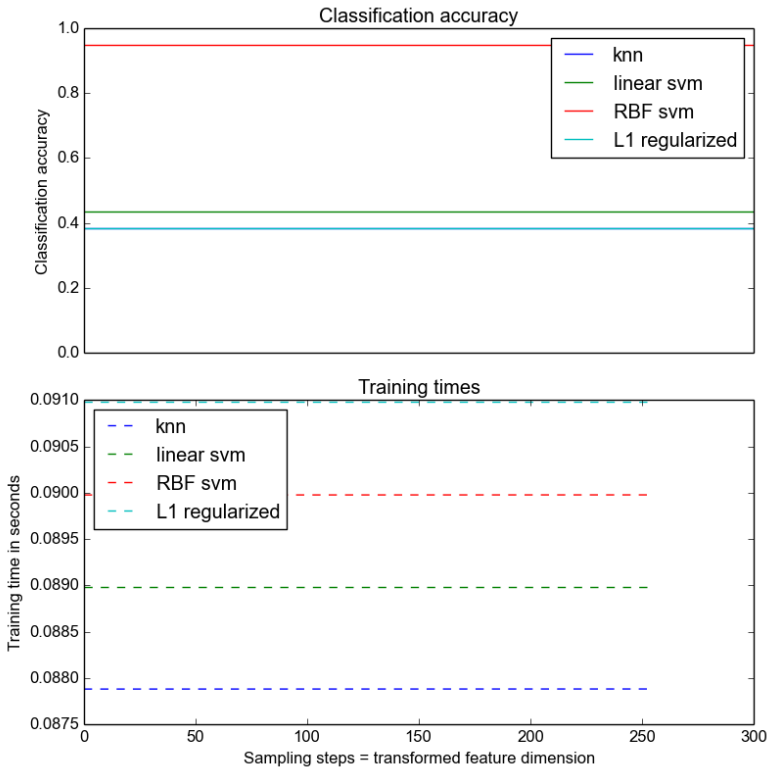
  - 6 faces: book, iphone, medicine, newspaper, remote
  - cylindrical: bottle, cup, cola, mug
  - rounded: apple, bread
  - curved: banana

Attributes Classification

- classifiers: kNN (k = 4), linear SVM, RBF SVM, L1 regularized logistic regression classifier
- feature matrix: features are quantized into 256 bins and constituent 120 * 256 matrix
- exiting problem:
  - dataset is too small, classification accuracy is low
  - some class only have one object
  - classifier with better performance

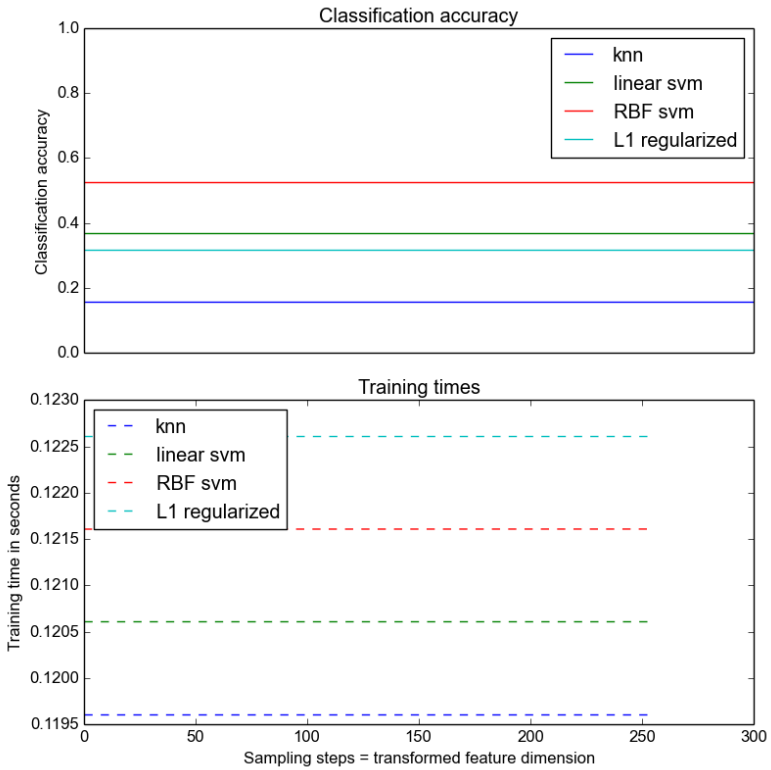
Classification Results

shape classify



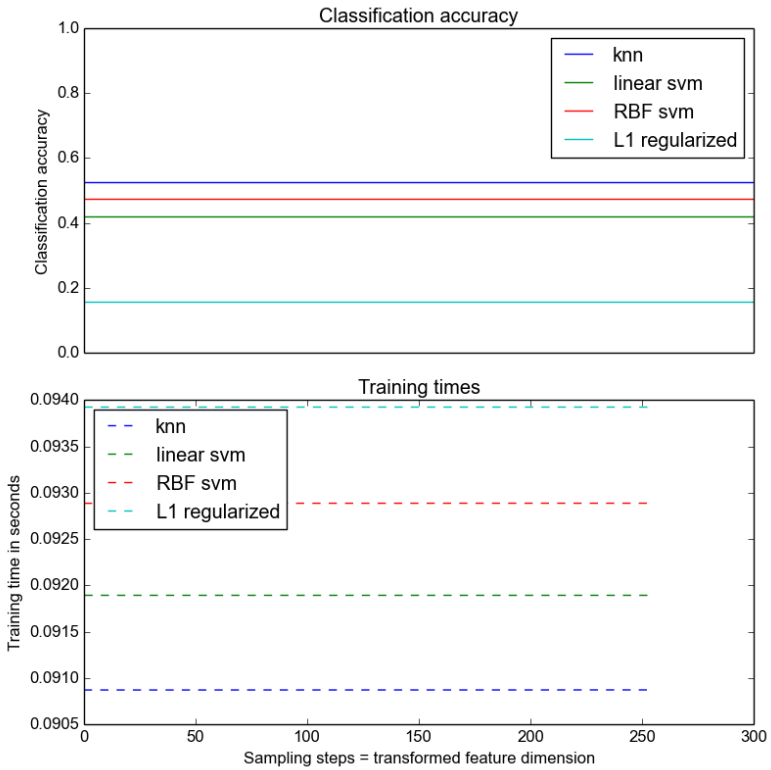
Classification Results

color classify



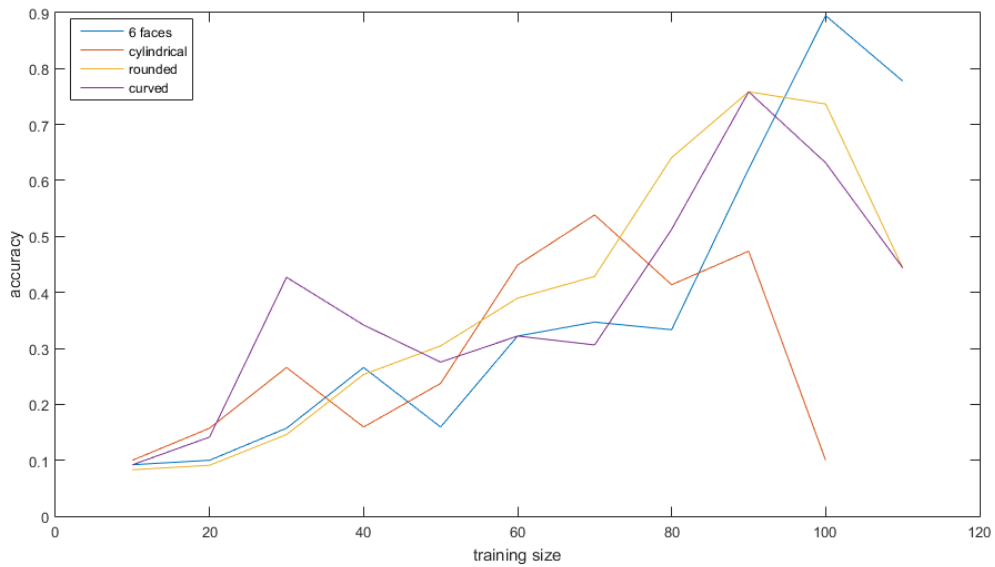
Classification Results

texture classify



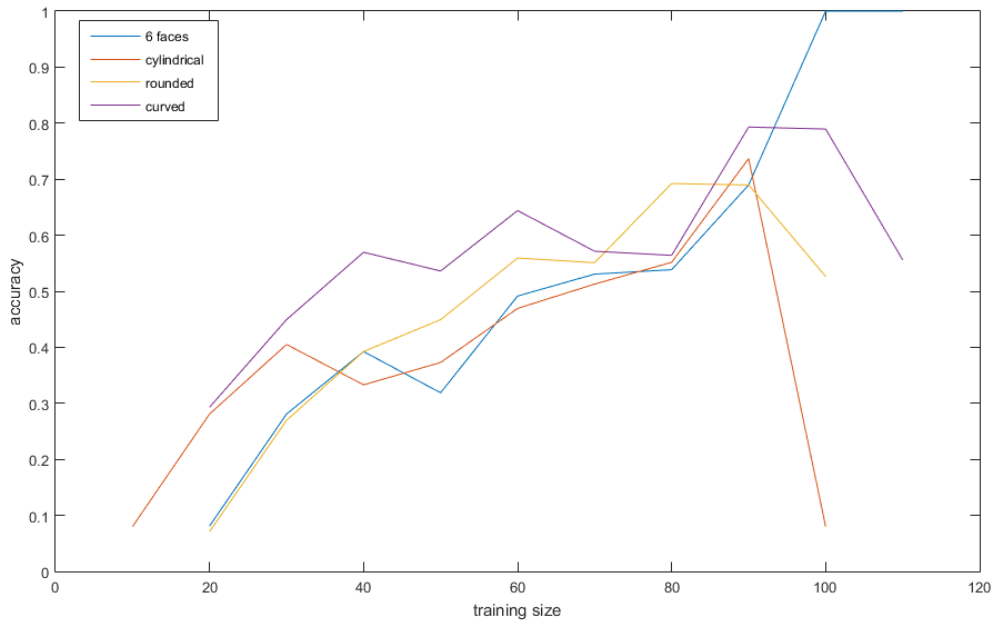
Classification Results

shape classify result with kNN

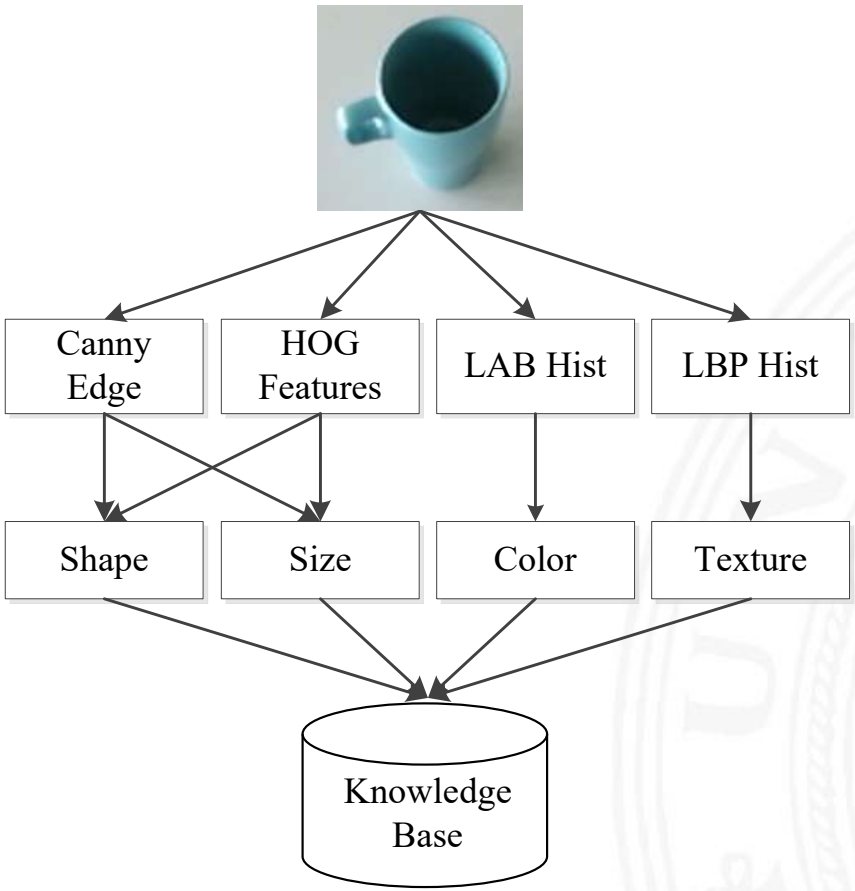


Classification Results

shape classify result with RBF SVM



Affordance Prediction Model



Features Extraction

Attributes Classification

Knowledge Base Query and Inference

Knowledge Base learning

phase 1: collect evidence for KB construction.

- visual attributes: shape, color, texture
- physical attributes: size
- categorical attributes
- affordance labels: edible, drinkable, readable, watch-able, call-able, medicine

Knowledge Base learning

phase 2: learning KB using Markov Logic Network (MLN).

Markov Logic Network (MLN)

- composed of pairs <F, w>
- widely used language in statistical relational learning

Markov Logic Network (MLN)

$$P(X = x) = \frac{1}{Z} exp(\sum_{i=1}^{n} w_i f_i(x_{\{i\}}))$$

Z: partition function,normalization constant
w_i: weight of formulae i
F: the set of first-order formulae in MLN
n: the number of formulae in F
$x_{\{i\}}$: the state of ground atoms appearing in the formulae F_i
$f_i(x_{\{i\}})$ =1, if $F_i(x_{\{i\}})$ = true; otherwise is 0.

Knowledge Base learning

KB schema and general rules

Schema

hasAffordance(object, affordance)
isA(object, category)
hasVisualAttribute(object, attribute)
hasSize(object, size)

General Rules

Attribute-Attribute relations

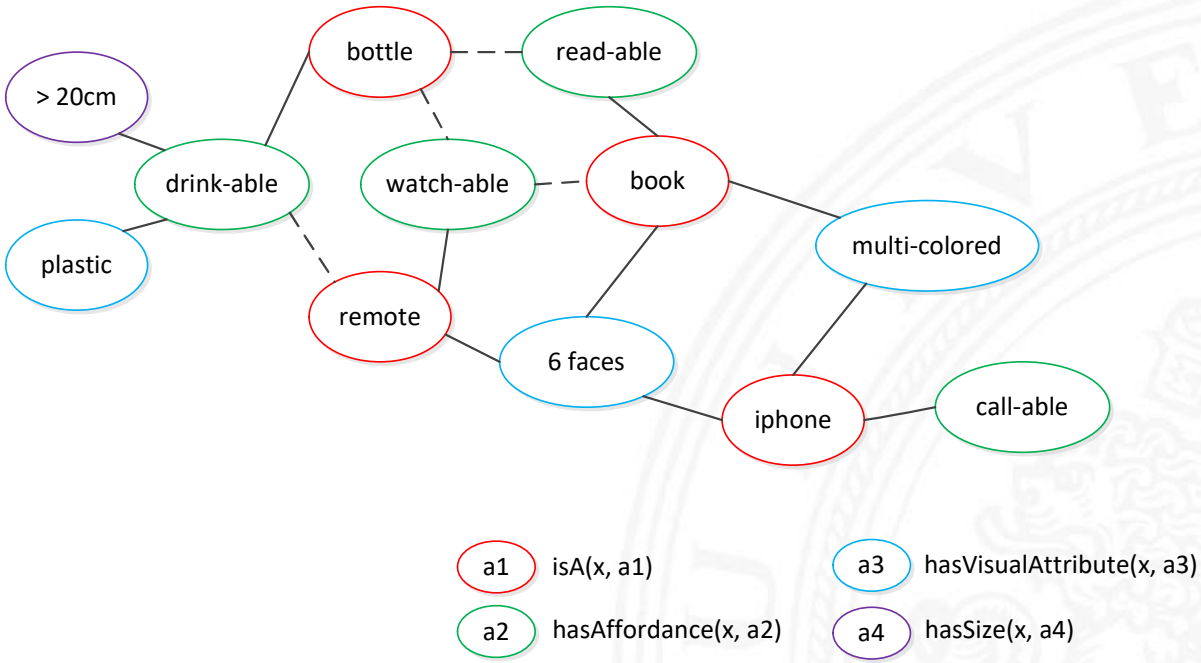
Attribute-Affordance relations

Example

hasVisualAttribute(x, cylindrical)
hasAffordance(x, drink-able) →

Knowledge Base representation

graphical illustration of the constructed KB



TODO

- afffordance prediction
- intention - objects relationship model

Thank you for your attention!