Reliable Attribute-Based Object Recognition Using High Predictive Value Classifiers

Wentao Luan[1], Yezhou Yang[2], Cornelia Fermüller[2], John S. Baras[1]

[1] Institute for Systems Research, [2] Computer Vision Lab, University of Maryland, College Park

Introduction

Goal: Reliable 3D object recognition using an ensemble of attribute-based classifiers for active agents.

Motivation:

- Insufficient representative training samples make it difficult to learn the optimal positive and negative detection rate.
- The viewing conditions can have a strong influence on classification performance.

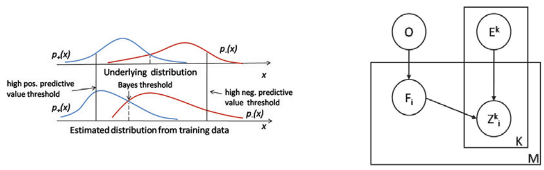

Figure 1: Left: Illustration of common conditional probability density functions of the positive and negative class. Top: ground truth distribution of the two classes; bottom: a possible distribution represented by the training data. Blue line: positive class; red line: negative class. Dashed line: (estimated) Bayes threshold; solid line: high PPV and NPV thresholds. Right: The relationship of objects (O), attributes (F_i), environmental variables (E_k) and observations (Z_i^k) in our model.

Methods and Intuition:

- Classifiers use two thresholds with one aiming for a positive predictive value (PPV), giving high precision for positive classes, and the other aiming for a negative predictive value (NPV), giving high precision for negative classes.
  - High PPV and NPV thresholds should be easier to obtain than the classical Bayes threshold (minimizing the classification error), when the number of training samples is too small to represent well the underlying distribution.
- Incorporating environmental factors (distance) into decision making. Defining a reliable working region for each basic attribute classifier, indicating a fair separation of the distributions of positive and negative classes.
  - Hence our approach can actively select "safe" samples and discard "unsafe" ones in unreliable regions.
- Allows using simple attributes.
  - Simple attributes are usually more robust to viewing conditions, though less discriminative.

Framework

Learning phase (offline): We learn distance-dependent attribute classifiers and determine a reliable range of distance intervals for each attribute classifier.

Testing (online): The system decides the distance interval for the RGBD images from active agents; combines classifier measurements from multiple images via maximum a posteriori probability (MAP) estimation.

As illustrated in Figure 1 (right),

$$P(O=o_j|\mathbb{Z}^K,\mathbb{E}^K)=\lambda P(O=o_j)\prod_{k=1}^{K}\prod_{i=1}^{M}\frac{P(F_i=f_{ij}|Z_i^k,E^k)}{\sum_{\{t|f_{it}=f_{ij}\}}P(O=o_t)}.$$

The recognition $\mathbb{A}$ is derived using MAP estimation as:

$$\mathbb{A}\triangleq\underset{o_j}{\operatorname{argmax}}\ P(O=o_j|\mathbb{Z}^K,\mathbb{E}^K).$$

Only takes high predictive value observations in reliable working region:

$$P(F_i=1|Z_i^k,E^k)=\begin{cases}p_i^+ & \text{if } e_k\in\mathbb{R}_i\ \&\ z_i^k=1,\\ 1-p_i^- & \text{if } e_k\in\mathbb{R}_i\ \&\ z_i^k=0,\\ \sum_{t|f_{it}=f_{ij}}P(O=o_t) & \text{o.w.}\end{cases}$$

$\mathbb{R}_i$ is the set of reliable working regions for the i-th classifier.

Theoretical Results

Theorem 1: The system is guaranteed to provide correct recognition if

- The recognized attributes can differentiate an object from others;
- The component classifiers' predictive values are larger than specified values (details see paper).

Theorem 2: The MAP estimation will converge to the correct result if

- The attribute classifiers' PPV and NPV are high enough in their reliable working region, where a lower bound of detection rate exists;
- The inputs are sampled randomly such that each attribute classifier gets the same chance to work in its reliable region.

Experimental Validation

Set of 9 objects and 10 attributes in the range of 0.6-1.6m

Object ID	plane surface	cylinder	gable top carton shape	box shape	wide mouth bottle shape	cup shape	bottle shape	red color	blue color	yellow color
1	✓	-	✓	-	-	-	-	-	✓	-
2	✓	-	✓	-	-	-	-	✓	-	-
3	✓	-	✓	-	-	-	-	-	-	✓
4	✓	-	-	✓	-	-	-	✓	-	-
5	-	✓	-	-	✓	-	-	-	-	-
6	-	✓	-	-	-	✓	-	-	✓	-
7	-	✓	-	-	-	-	✓	-	-	✓
8	-	✓	-	-	-	-	✓	✓	-	-
9	-	✓	-	-	-	-	✓	-	✓	-

Figure 2: Object IDs and their list of attributes.The dataset is available from `http://ece.umd.edu/~wluan/ECCV2016.html`

Attributes:

- Fine shape: 1) retrieve templates point cloud based on input's location w.r.t the camera 2) match with the templates using VFH features
- Coarse shape: plane and cylinder fitting using RANSAC
- Color: histogram matching in hue and saturation space

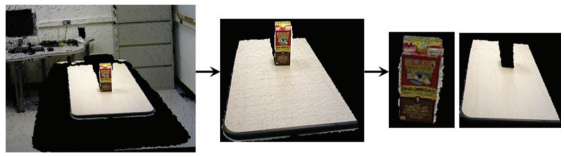

Figure 3: Illustration of the preprocessing pipeline. Left: input point cloud; Middle: point cloud after passthrough filtering; Right: segmented candidate object and removed table surface.

Experiment One: demonstrates the necessity of incorporating environmental factors (the recognition distance in our case) for object recognition

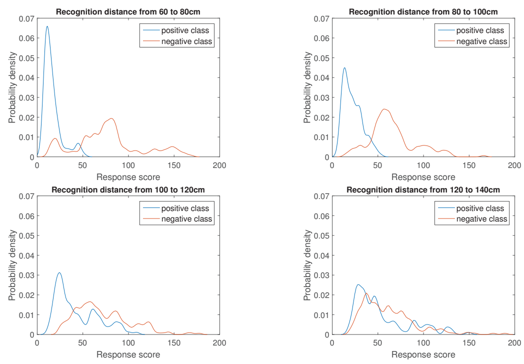

Figure 4: Estimated distribution of bottle shape classifier's response score under 4 recognition distance intervals. The classifiers' response score distributions are indeed distance variant.

Experiment Two: evaluates the performance of the high predictive value threshold classifier in comparison to the single threshold one

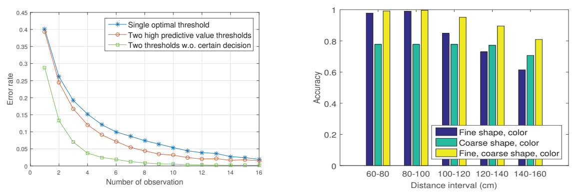

Figure 5: Left: error rate using classification with a single threshold (blue) and two high predictive value thresholds (red). The green line depicts the error component due to the cases where the two thresholds method picks randomly. Right: three systems' recognition accuracy for different working distance intervals.

Our method provides an alternative for cases when the training data in real world scenarios does not represent well the underlying distribution.

Experiment Three: demonstrates the benefits of using less discriminative attributes for extending the system's working range.

- The classification accuracy decreases with larger distances.
- At 120 cm to 160 cm, the system using fine shape attributes (blue) performs worse than the system using less selective coarse attributes (green). It validates that the coarse shape based classifier has a larger working region.
- The system using all attributes (yellow) achieves the best performance at each working region.