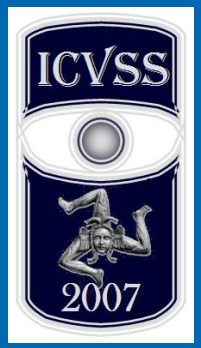


Recognizing Groceries *in situ* Using *in vitro* Training Data

Michele Merler

Advisor:
Prof. Serge Belongie

Outline

- Motivations
- Related Work
- Grozi -120
- Experiments
- Conclusions
- Future work

Motivations

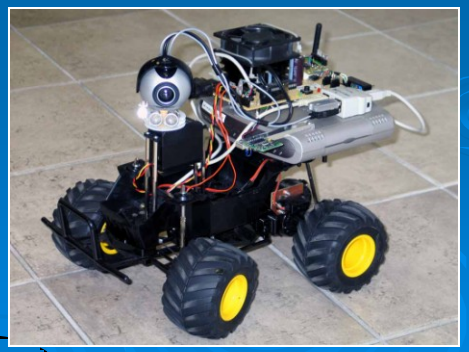
Applications

- Assistive vision technology for the blind

  e.g: Grozi project @ UCSD

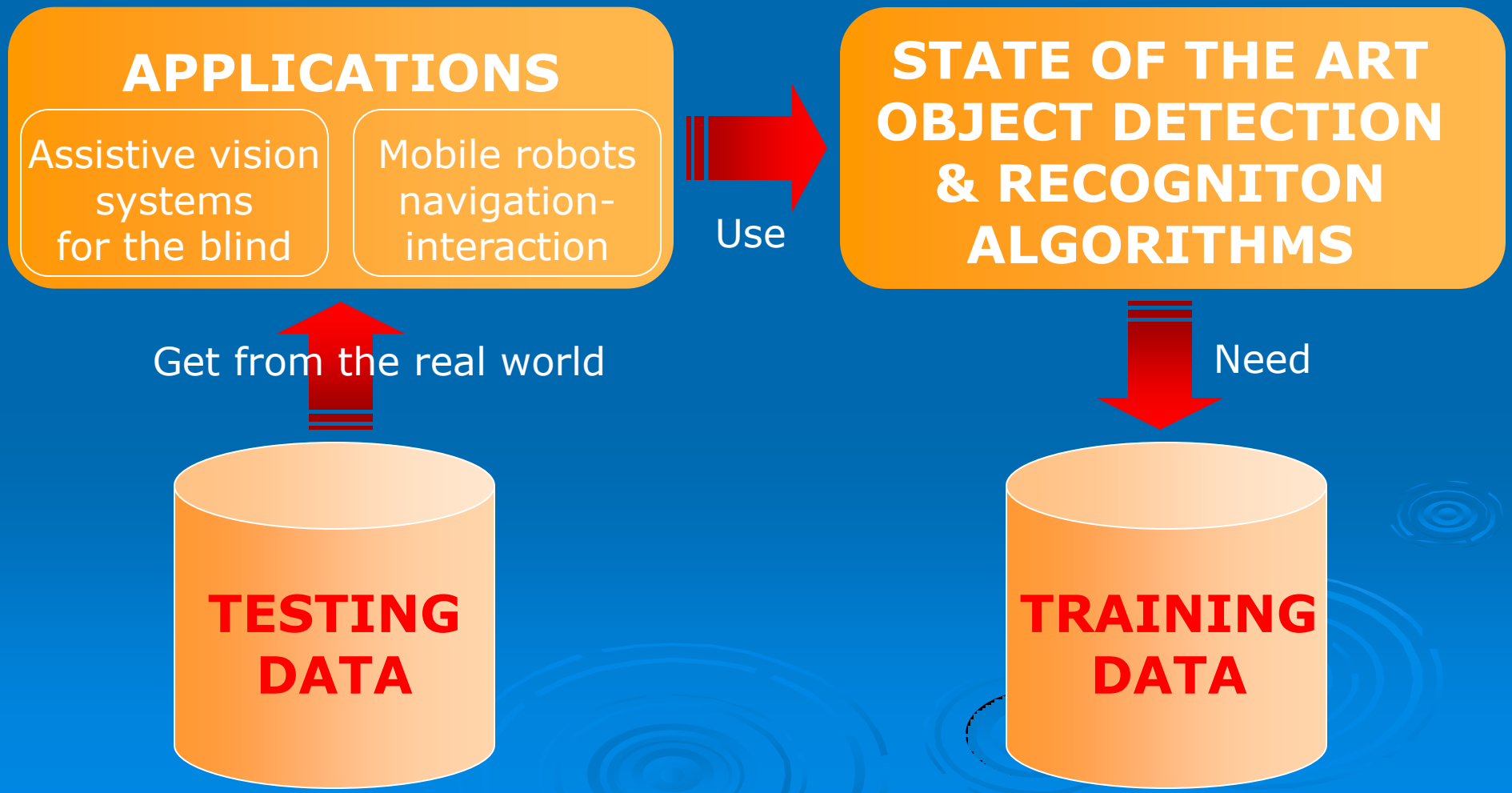
- Object recognition for mobile robots

  e.g: Semantic Robot Vision Challenge @ CMU.



Motivations

Problem Statement



Motivations

Problem Statement

APPLICATIONS

Assistive vision systems for the blind

Mobile robots navigation-interaction

Use

STATE OF THE ART OBJECT DETECTION & RECOGNITON ALGORITHMS

Get from the real world

TESTING DATA

Need

- acquired under different environmental conditions
- appearance drawn from different distribution than the test data

Motivations

Problem Statement

APPLICATIONS

Assistive vision systems for the blind

Mobile robots navigation-interaction

Use

STATE OF THE ART OBJECT DETECTION & RECOGNITON ALGORITHMS

Get from the real world

TESTING DATA

Need

Where to get them?

Motivations

Problem Statement

Use

STATE OF THE ART OBJECT DETECTION & RECOGNITON ALGORITHMS

Need

Where to get them?

World Wide Web

Related work

Object recognition databases

- PASCAL VOC
- Caltech 101(256)
- SOIL-47
- ALOI
- ETH-80
- LabelMe

Related work

Object recognition databases

Improvements needed*:

- Multiple object class instances within a single image
- Partial occlusion and truncation
- Size, viewpoint and orientation variations
- High degree of intra-class variability
- Exclude pre-segmented objects

* J. Ponce et al. *Dataset Issues in Object Recognition. Toward Category-Level Object Recognition*, Springer-Verlag Lecture Notes in Computer Science., 2006.

Related work

Object recognition databases

Improvements needed*:

- Multiple object class instances within a single image
- Partial occlusion and truncation
- Size, viewpoint and orientation variations
- High degree of intra-class variability
- Exclude pre-segmented objects

* J. Ponce et al. *Dataset Issues in Object Recognition. Toward Category-Level Object Recognition*, Springer-Verlag Lecture Notes in Computer Science., 2006.

Related work

Object recognition databases

Improvements needed*:

- Multiple object class instances within a single image

Training and testing data often come from the same distribution

- High degree of intra-class variability
- Exclude pre-segmented objects

* J. Ponce et al. *Dataset Issues in Object Recognition. Toward Category-Level Object Recognition*, Springer-Verlag Lecture Notes in Computer Science., 2006.

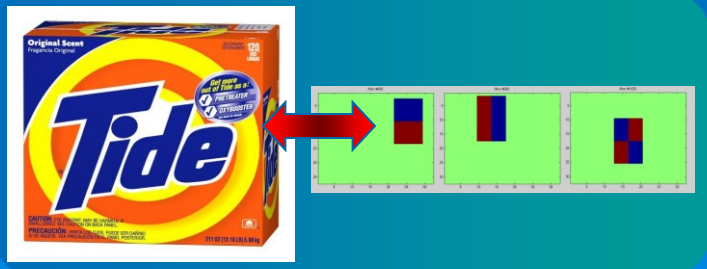
Related work

Object detection/recognition algorithms

- Color Histogram Matching (Lab) [Swain & Ballard '91] with Integral Histogram [Porikli `06]

- SIFT Descriptor [Lowe '99]

- Boosted Haar-like Features [Viola & Jones '01].



Grozi-120

- Multimedia database of 120 grocery products
- Objects vary in color, size, opacity, shape and rigidity. They are found in different lighting conditions and in presence of clutter and occlusion
- ***In vitro*** and ***in situ*** image representations (for training and testing data respectively)

Grozi-120

In vitro data

- Isolated images captured under ideal imaging conditions
  (e.g stock photography studio or lab)
- They can be found in the web (e.g Froogle, Amazon, etc)

Grozi-120

In vitro data

- 676 training images (average 6 images per object)
- Obtained from the web (Froogle, Shopwiki, Amazon Groceries, Yahoo images) using a list of 4000 UPC codes
- Clear foreground–background distinction via binary mask (if desired)

Grozi-120

In situ data

- Images from objects captured in natural environments (real world)
- They were shot inside a grocery store, using a MiniDV camcorder and includes every ***in vitro*** object
- 29 videos containing all products
- Product location in coordinates saved every 5 frames
- A total of 11194 ***in situ*** images (average 93 per product)

Size, viewpoint and orientation variations

Grozi-120
In situ data

Grozi-120
In situ data

Honey Nut Cheerios

04

016000665903

Colgate plus toothbrush

103

035000553003

GM HNY NUT CHEERIOS CEREAL CUP

95

016000141551

Morton Salt, Iodized

27

024600010030

Chap Stick Lip Balm

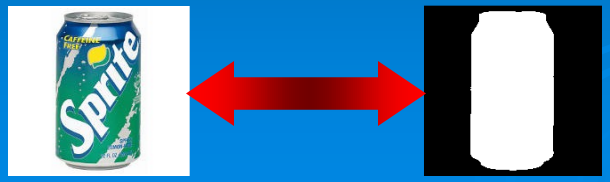
33

036600813313

Grozi-120
In situ data

Grozi-120

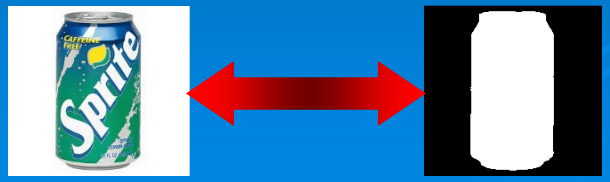
In vitro data
Training



In situ data
Testing

Grozi-120

In vitro data
Training

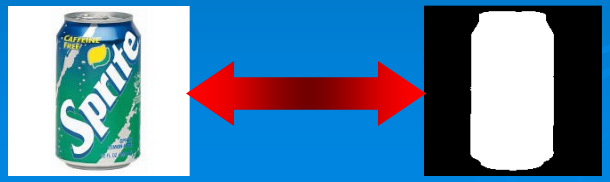


In situ data
Testing

Big difference in quality!

Grozi-120

In vitro data
Training



In situ data
Testing

Big difference in quality!

- Illumination
- Deformations
- Clutter
- Occlusion
- Truncation

Experiments

- **Color histogram matching** (CHM): Histogram template (16 bins in ab from Lab) + integral histogram + L1 distance

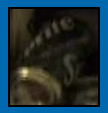
- **SIFT: bag of features approach**
  One bag per object + L2 distance

- **Boosted Haar-like features** (BHaar): Data + synthetic data + Haar-like features + cascades (OpenCv)

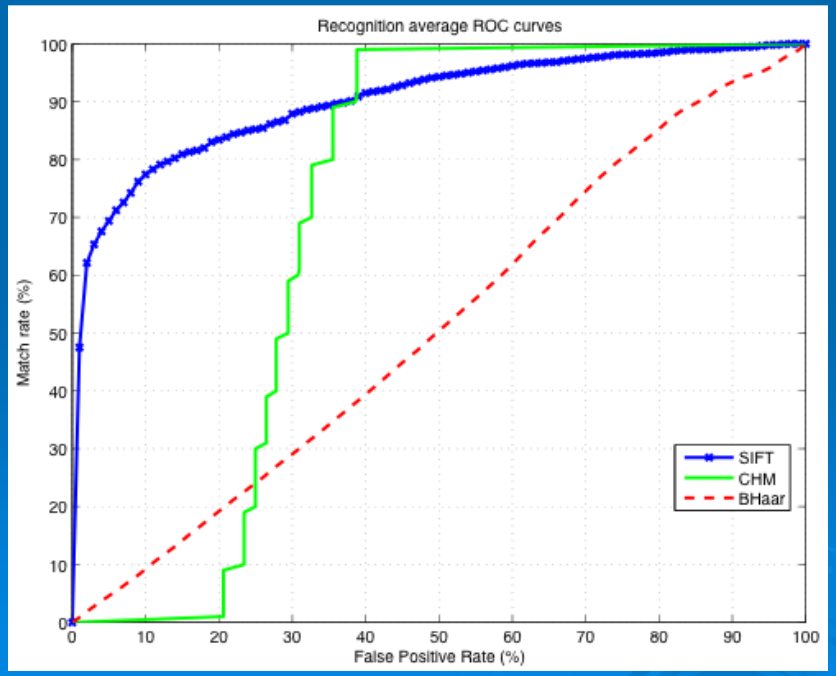
Experiments

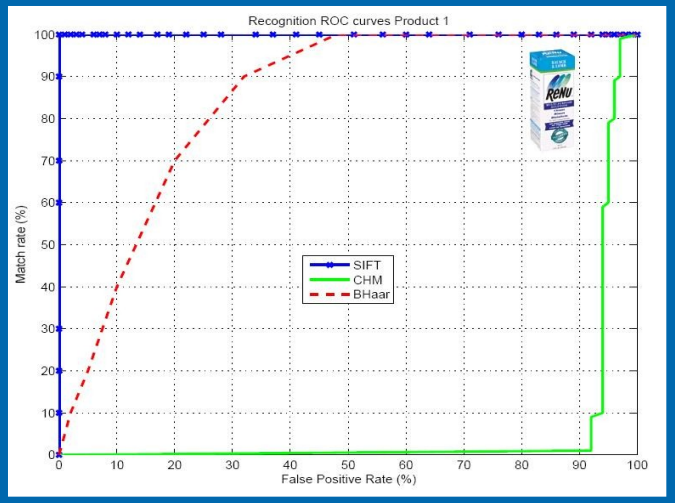
Recognition:

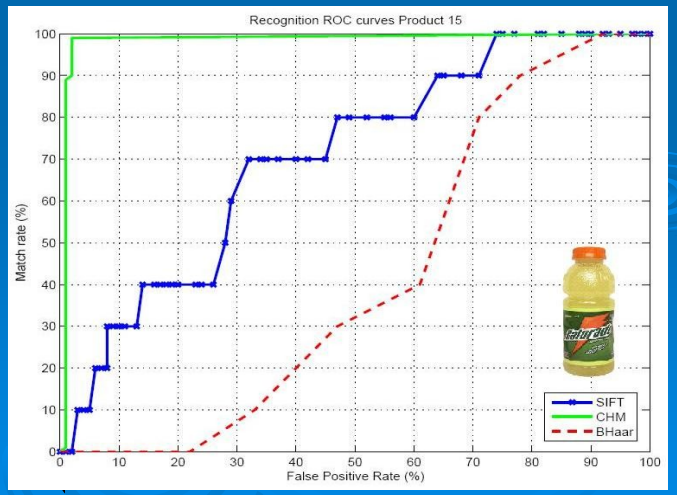
10 *in vitro* samples per object for training (200 for BHaar)



10 *in situ* images per object for testing







Experiments

Localization:

- 10 *in vitro* samples per object for training
  (200 for BHaar)
- 14 *in situ* frames per object for true positives
- 100 frames with no product on the database as true negatives

CHM	%Rec	%Pre	%TP	%FP
Mean	15	17	18	65
Std Dev	28	16	35	32
Best (20)	71	82	100	4
Worst (32)	0.7	0.2	0	100
SIFT	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	72	18	22	62
Std Dev	20	17	26	28
Best (34)	14	83	93	25
Worst (9)	26	0.9	0	64
BHaar	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	15	17	18	65
Std Dev	13	13	19	24
Best (92)	35	74	50	38
Worst (5)	0.5	0.2	0	92

Rec = Overall Recall, Pre = Overall Precision

Experiments

Localization:

- 10 *in vitro* samples per object for training
  (200 for BHaar)
- 14 *in situ* frames per object for true positives
- 100 frames with no product on the database as true negatives

CHM	%Rec	%Pre	%TP	%FP
Mean	15	17	18	65
Std Dev	28	16	35	32
Best (20)	71	82	100	4
Worst (32)	0.7	0.2	0	100
SIFT	%Rec	%Pre	%TP	%FP
Std Dev	20	17	26	28
Best (34)	14	83	93	25
Worst (9)	26	0.9	0	64
BHaar	%Rec	%Pre	%TP	%FP
Mean	15	17	18	65
Std Dev	13	13	19	24
Best (92)	35	74	50	38
Worst (5)	0.5	0.2	0	92

Rec = Overall Recall, Pre = Overall Precision

Experiments

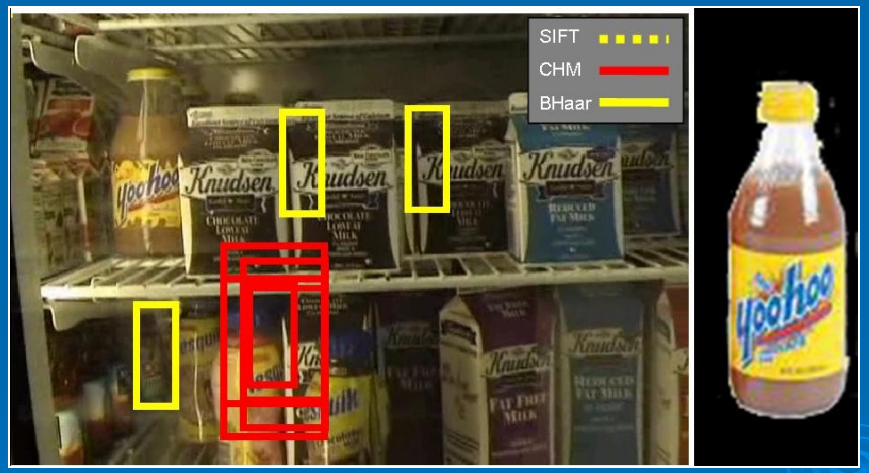
Localization:

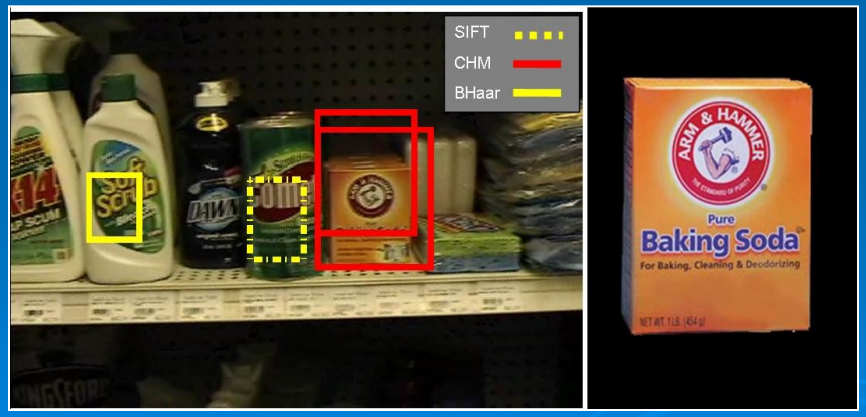
- 10 *in vitro* samples per object for training
  (200 for BHaar)
- 14 *in situ* frames per object for true positives
- 100 frames with no product on the database as true negatives

CHM	%Rec	%Pre	%TP	%FP
Mean	15	17	18	65
Best (20)	71	82	100	4
Worst (32)	0.7	0.2	0	100
SIFT	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	72	18	22	62
Best (34)	14	83	93	25
Worst (9)	26	0.9	0	64
BHaar	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	15	17	18	65
Best (92)	35	74	50	38
Worst (5)	0.5	0.2	0	92

Rec = Overall Recall, Pre = Overall Precision

Experiments





Conclusions

We presented:

- A new multimedia database for studying object recognition in presence of ***in vitro***/***in situ*** dichotomy
- Baseline performances for three widely used algorithms

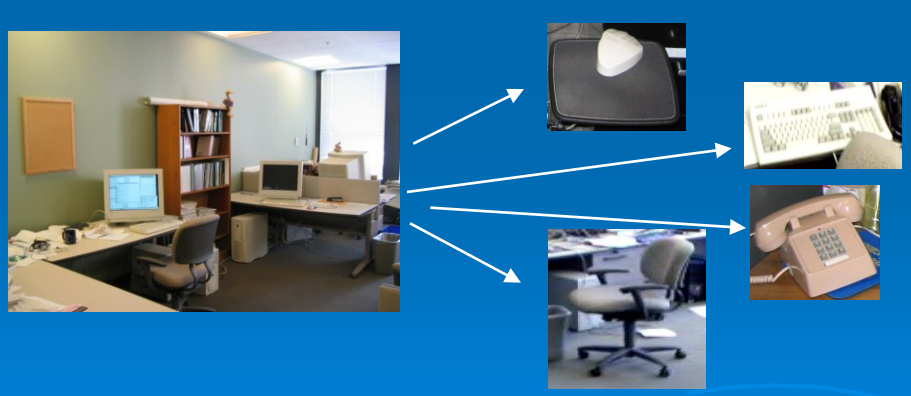
The results suggest the need of more precise and elaborate recognition algorithms.

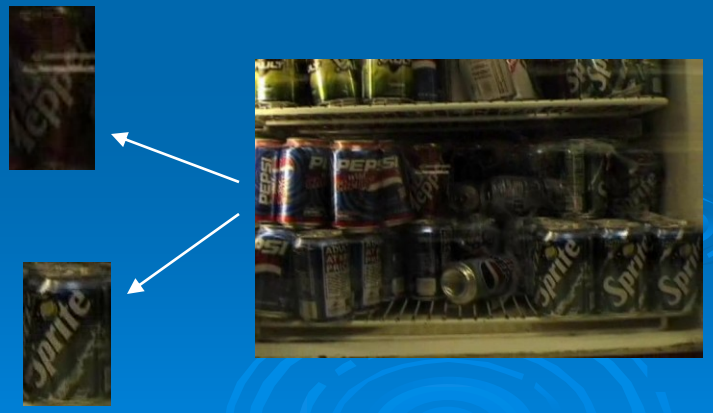
Future Work

- Develop new algorithms that fuse these different approaches in order to improve results

Future Work

- Develop new algorithms that fuse these different approaches in order to improve results
- Make use of context information based on physical object proximity to improve localization of objects in natural scenes

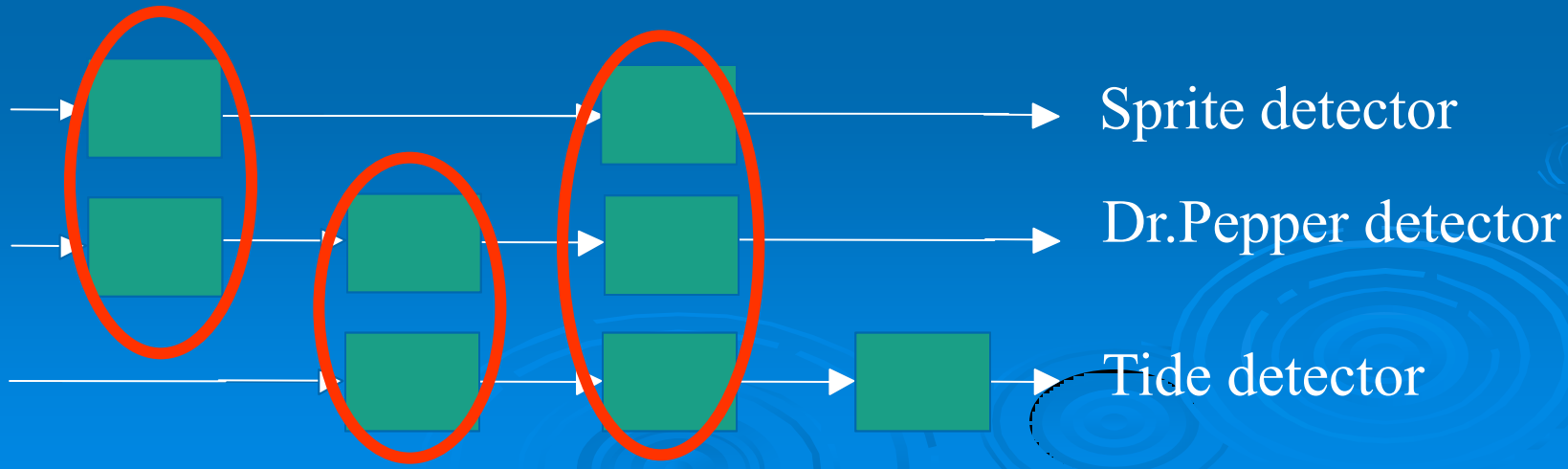




Future Work

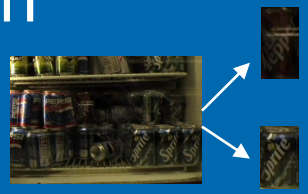
- Develop new algorithms that fuse these different approaches in order to improve results
- Make use of context information based on physical object proximity to improve localization of objects in natural scenes

- Use Shared features

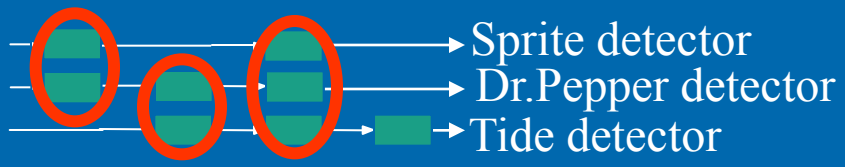


Future Work

- Develop new algorithms that fuse these different approaches in order to improve results
- Make use of context information based on physical object proximity to improve localization of objects in natural scenes



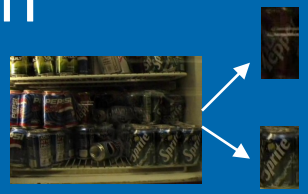
- Use Shared features



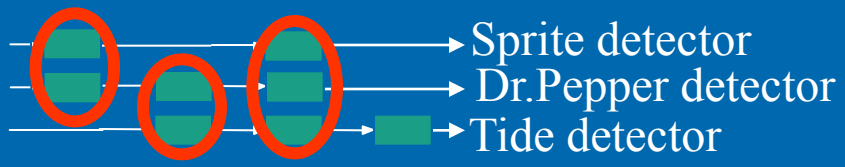
- Use temporal correlation (tracking)

Future Work

- Develop new algorithms that fuse these different approaches in order to improve results
- Make use of context information based on physical object proximity to improve localization of objects in natural scenes



- Use Shared features



- Use temporal correlation (tracking)
- Include more objects and grow the number of samples per product

Acknowledgments

Part of this work has been funded by the UCSD division of the California Institute for Telecommunications and Information Technology (Calit2) through the 2006 Summer Undergraduate Scholarship Program

Michele Merler, Carolina Galleguillos and Serge Belongie, Recognizing groceries *in situ* using *in vitro* training data, SLAM 2007, Minneapolis

Questions?

http://grozi.calit2.net