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

Michele Merler University of Trento (michele.merler@studenti.unitn.it)
Advisor : Prof. Serge Belongie CSE Department, UC San Diego

Using pictures of objects captured under ideal imaging conditions (*in vitro*) to recognize objects in natural environments (*in situ*) is an emerging area of interest in computer vision and pattern recognition. We propose a new multimedia database of 120 grocery products, GroZi-120.

For every product, two different recordings are available: *in vitro* images extracted from the web, and *in situ* images extracted from camcorder video collected inside a grocery store. We present the results of three commonly used object recognition/detection algorithms on the dataset.

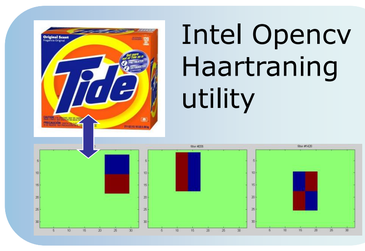
General Problem in Computer Vision

Our Work

16 bins, ab components from Lab colorspace

David Lowe's approach

Intel Opencv Haartraning utility



Color Histogram Matching

SIFT

Boosted Haar like Features

APPLICATIONS

Assistive vision systems for the blind

Mobile robots navigation-interaction

STATE OF THE ART OBJECT DETECTION - RECOGNITION ALGORITHMS

Get from the real world

need

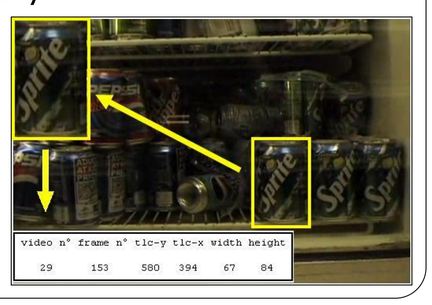
Where to get them?

Database "in situ"

Videos Colleceted in Store

- 29 Divx 5.2.1 files, 30 fps, 2kbps
- Cluttered background, different products per frame
- Multiple instances of same object per frame, partially occluded
- Rotated,different illumination, angle of view,affine and projective distortion
- Product location saved every 5 frames

	n. samples
total	11194
avg	9.33
max	814
min	14



TESTING DATA

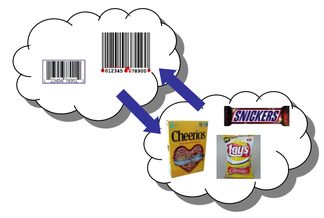
Do detection and localization work as well as with other databases and problems?

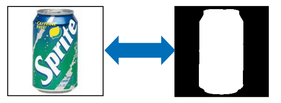
TRAINING DATA

Database "in vitro"

Online Images

- Web : Froogle, Shopwiki, Amazon Groceries, Yahoo images
- General + specialized (UPC code) queries
- Include a variety of sizes, poses and illuminations (coming from different online vendors and stock photo suppliers)
- Easy to analyze
- Clear foreground–background distinction (binary mask)





	n. samples
total	676
avg	5.63
max	14
min	2

NO

Future Work Needed

- Use more precise and elaborate detection/recognition algorithms
- Dynamically increase the dataset
- Use context information about physical object proximity to improve localization

Results - Localization

14 frames per product with highest number of keypoints as True Positives

100 frames with none of the dataset products as True Negatives

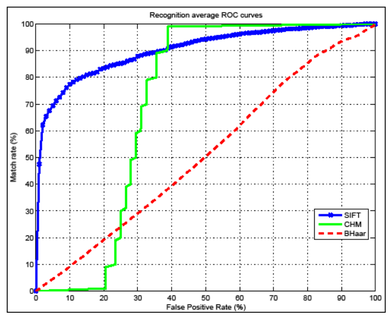
CHM	%Rec	%Pre	%TP	%FP
Mean	15	17	18	65
Std Dev	28	16	35	32
Best	71	82	100	4
Worst	0.7	0.2	0	100
SIFT	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	72	18	22	62
Std Dev	20	17	26	28
Best	14	83	93	25
Worst	26	0.9	0	64
BHaar	**%Rec**	**%Pre**	**%TP**	**%FP**
Mean	15	17	18	65
Std Dev	13	13	19	24
Best	35	74	50	38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

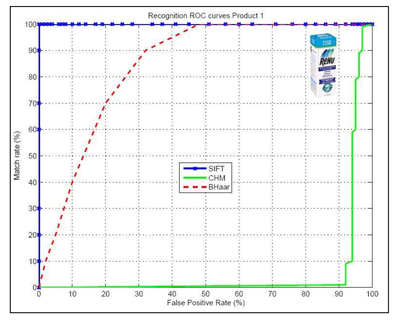
Rec = Overall Recall, Pre = Overall Precision

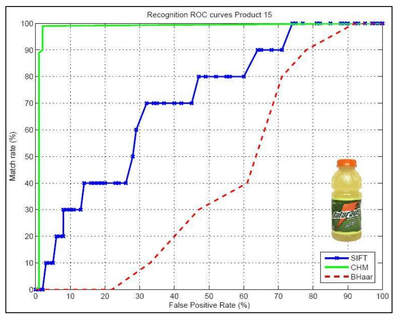
Results - Recognition

10 *in situ* images per product with highest n. of keypoints as Positives.

110 samples of the remaining products as Negatives







Acknowledgments

Special thanks to Carolina Galleguillos and all the people involved in the GroZi project at UCSD. Michele Merler was supported by the California Institute for Telecommunication and Information Technology (CalIt2) 2006 Summer Undergraduate Scholarship program.