CS 395T: Visual Recognition

Exploiting Context for Object Detection

5th October 2012

Aashish Sheshadri

Components Analyzed

1. Scene Classification using GIST Descriptors.
2. Contextual Priming.

Scene Classification

- Dataset : 15 Scene Categories - The Ponce Research Group [1].
  - Indoor and Outdoor Scenes.
- Descriptor : GIST Discriptor.
  - Matlab code by A. Oliva [2].

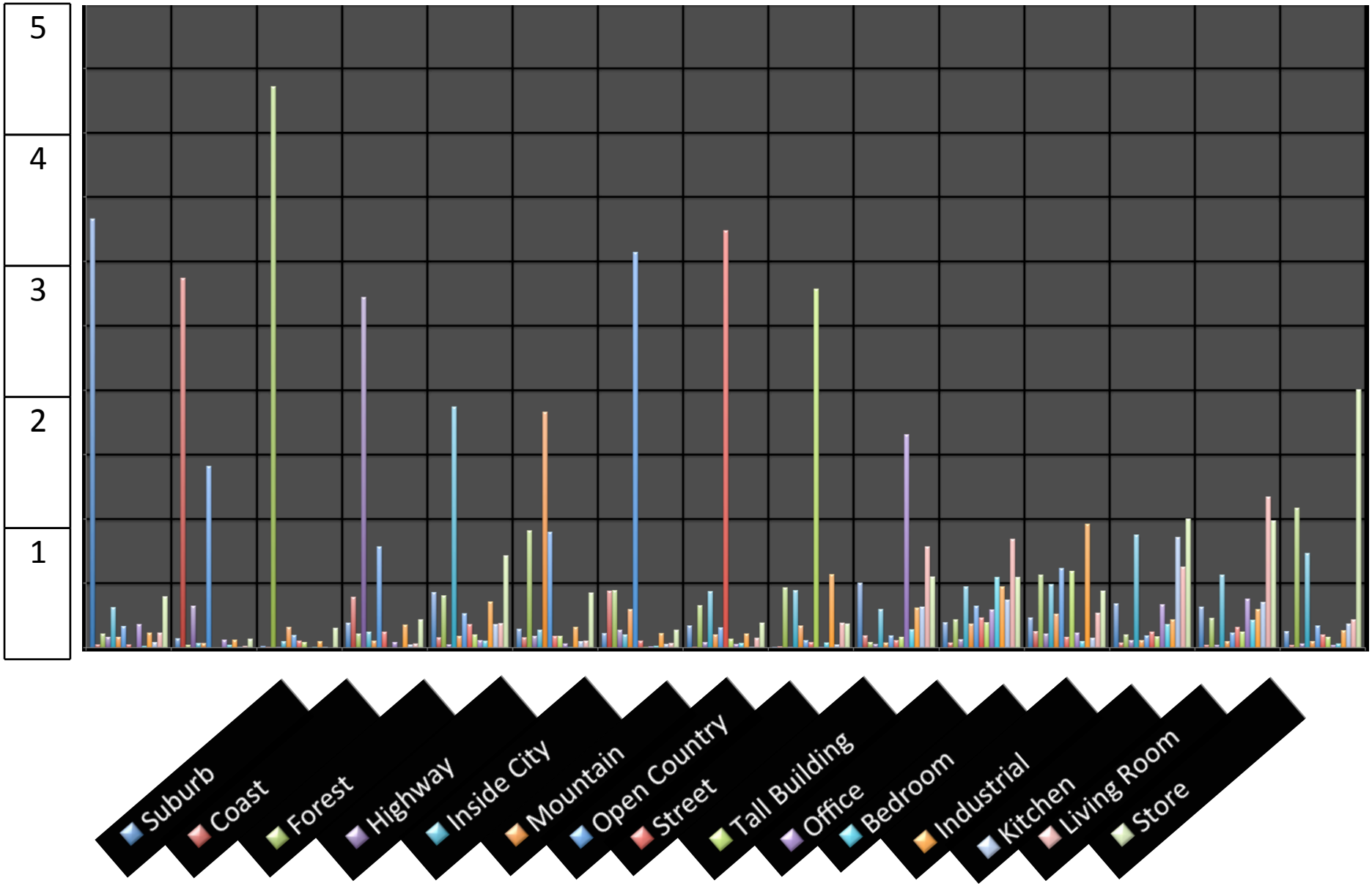
- [1] http://www-cvr.ai.uiuc.edu/ponce_grp/data/
- [2] http://people.csail.mit.edu/torralba/code/spatialenvelope/

Scene Classification

- Classifiers :
  - K-Nearest Neighbors (KNN)
    - Consensus among five neighbors.
    - Euclidean distance.
    - Netlab Toolbox for Matlab [1].
  - Support Vector Machine (SVM)
    - One vs All.
    - RBF Kernel.
    - LIBSVM package for Matlab [2].

- [1] http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/
- [2] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Neighbor Presence



Nearest Neighbors

Suburb

Coast

Forest

Highway

Inside City

Nearest Neighbors

Mountain

Open Country

Street

Tall Building

Office

Nearest Neighbors

Bedroom

Industrial

Kitchen

Living Room

Store

Confusion Matrix (SVM)

	Suburb	Coast	Forest	Highway	Inside City	Mountain	Open Country	Street	Tall Building	Office	Bedroom	Industrial	Kitchen	Living Room	Store
Suburb	84	1	8	0	0	0	0	0	3	4	0	0	0	0	0
Coast	0	75	9	2	0	2	4	3	0	4	0	0	0	0	1
Forest	0	0	83	0	0	4	10	2	0	0	0	0	0	0	1
Highway	0	5	4	66	12	1	0	2	3	6	0	0	0	0	1
Inside City	3	0	10	1	50	5	8	18	3	0	0	0	1	0	1
Mountain	0	0	5	0	2	73	12	5	1	2	0	0	0	0	0
Open Country	0	5	10	1	0	3	72	4	0	4	0	0	0	0	1
Street	0	6	2	0	11	1	0	78	0	2	0	0	0	0	0
Tall Building	0	5	1	1	0	3	0	2	76	0	0	11	0	0	1
Office	2	0	6	0	0	5	0	3	0	59	0	0	1	20	4
Bedroom	0	2	10	1	1	7	0	5	0	15	25	0	10	23	1
Industrial	0	5	15	3	3	5	8	10	16	11	0	16	0	0	8
Kitchen	0	1	18	4	2	8	0	2	0	20	0	0	25	10	10
Living Room	0	2	14	6	0	6	0	4	3	8	2	0	16	23	0
Store	0	10	10	3	1	5	1	8	2	22	1	0	0	0	37

Average Classification Rate : 56.13%

Inferring Object Presence and Location

- Identifying scene category enables object inference.
- Using scene information to infer object location.
- Statistical inference of object location using GIST of the scene to enable contextual priming [1].

- [1] Contextual Priming for Object Detection by Antonio Torralba.

Mixture Density Networks (MDN)

- Combination of mixture model and a neural network.
- Learning conditional distributions by training the network.
- Input GIST vector and train network to learn desired probability distribution.
- MDN implementation used from Netlab Toolbox [1].

- [1] http://www1.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/

Segmented and Annotated Dataset [1]

[1] http://labelme.csail.mit.edu/Release3.0/

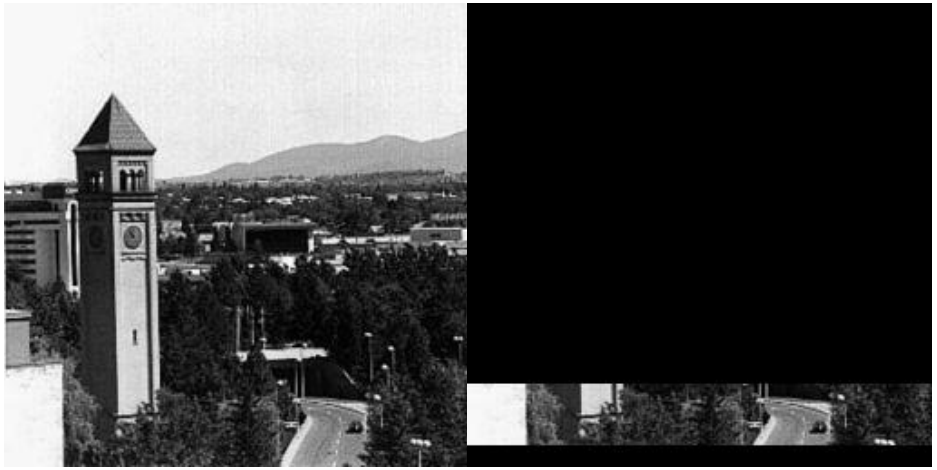
Inferring Location of Cars

- Scene categories with cars - Mountains, Street, Open Country - We've Travelled Everywhere!

Learning Distributions

- 566 Training Examples.
- Distributions Learnt:
  - P(Y|g).
  - P(s|g).
- Set P(X|g) to be uniform across the image.

Single Instance

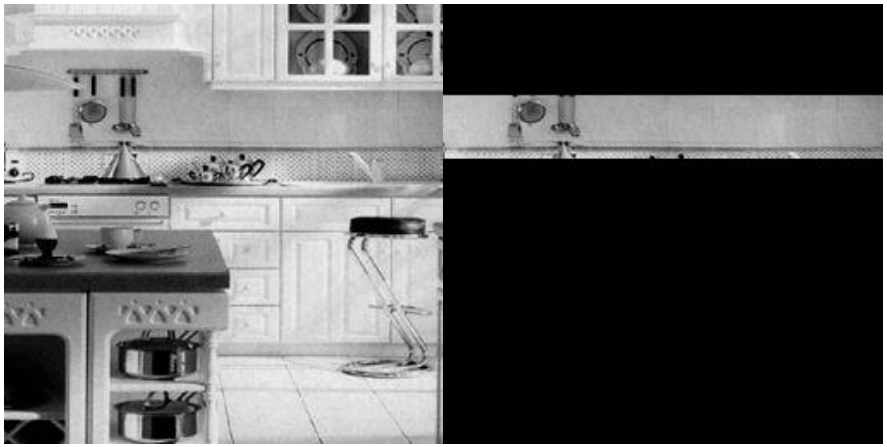


Multiple Instances

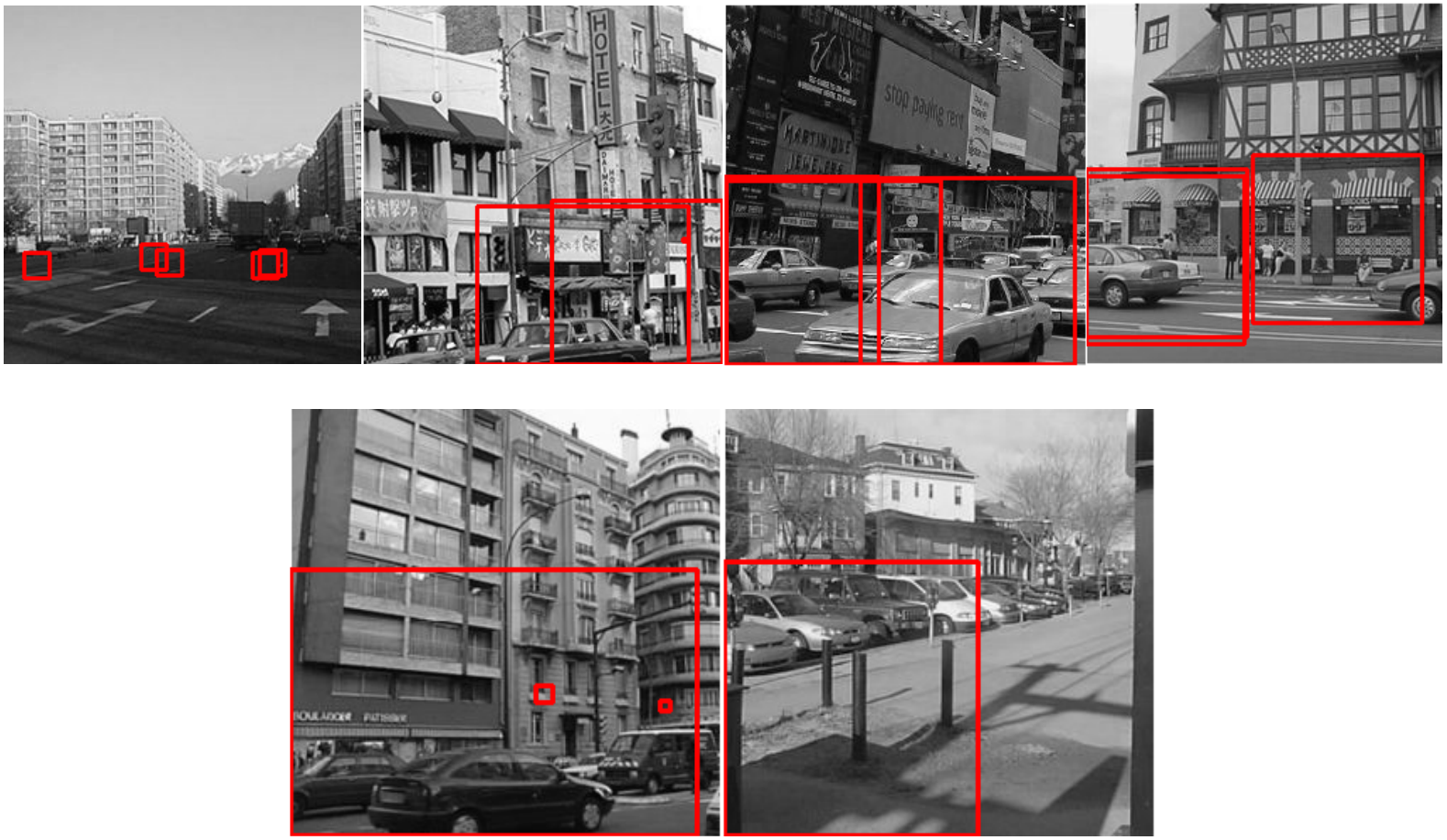
- Multiple modes ?

Difficult Scenes

Where are the Cars?



Predicting Scale



Failed Scenes

What's Important

- Car side view.
- Present but occluded.
- Frontal view.
- Just right.

Finding People ?

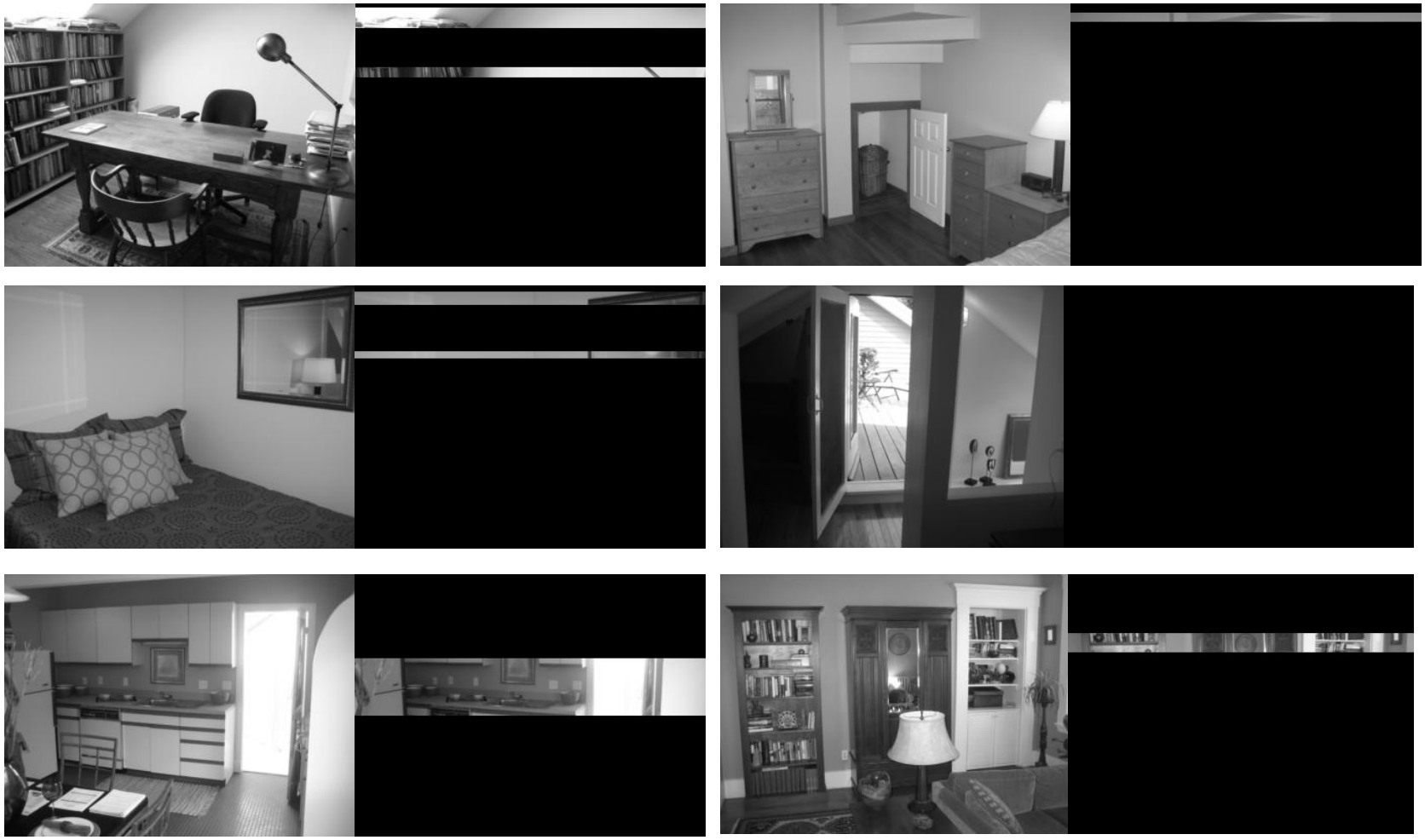
Pedestrians

Faces




Failed Instances

Something Challenging.. Lamps?



Lamps Better Results?

Closing Points

- When does it work ?
- Why does it work ?
- How can we improve inference?