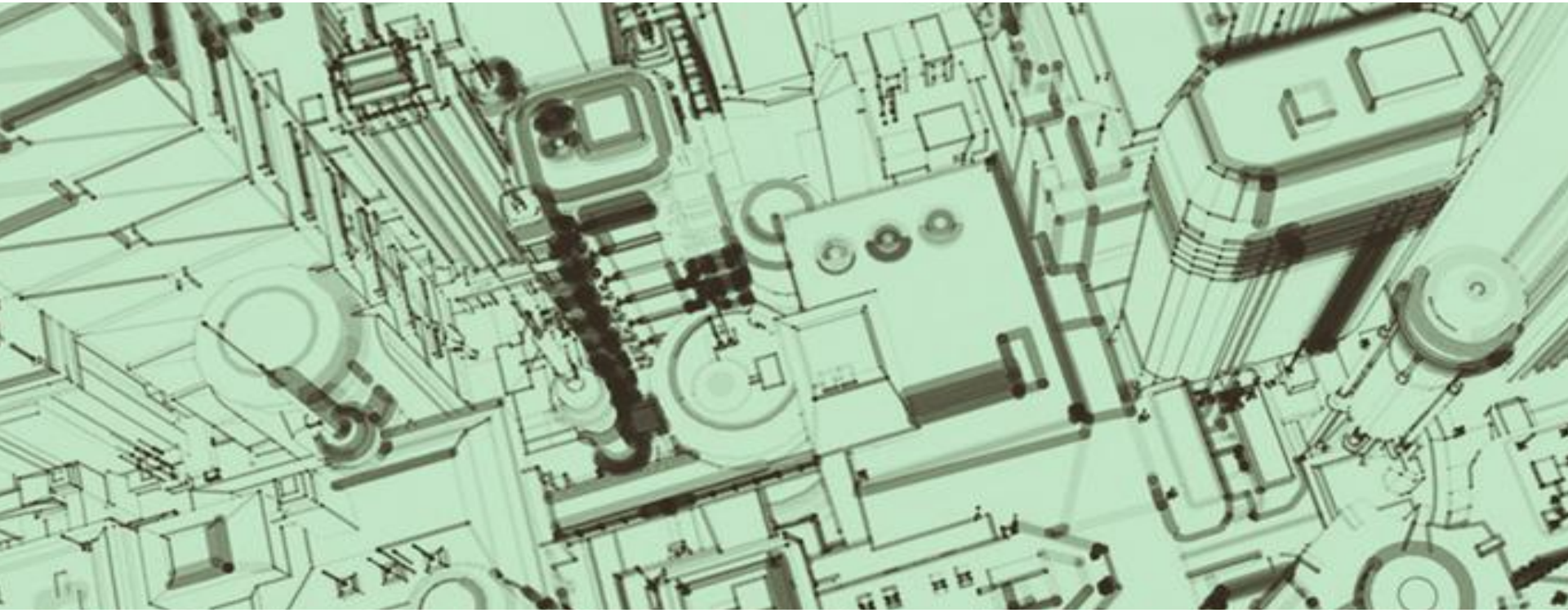
Semantic Localization of Indoor Places

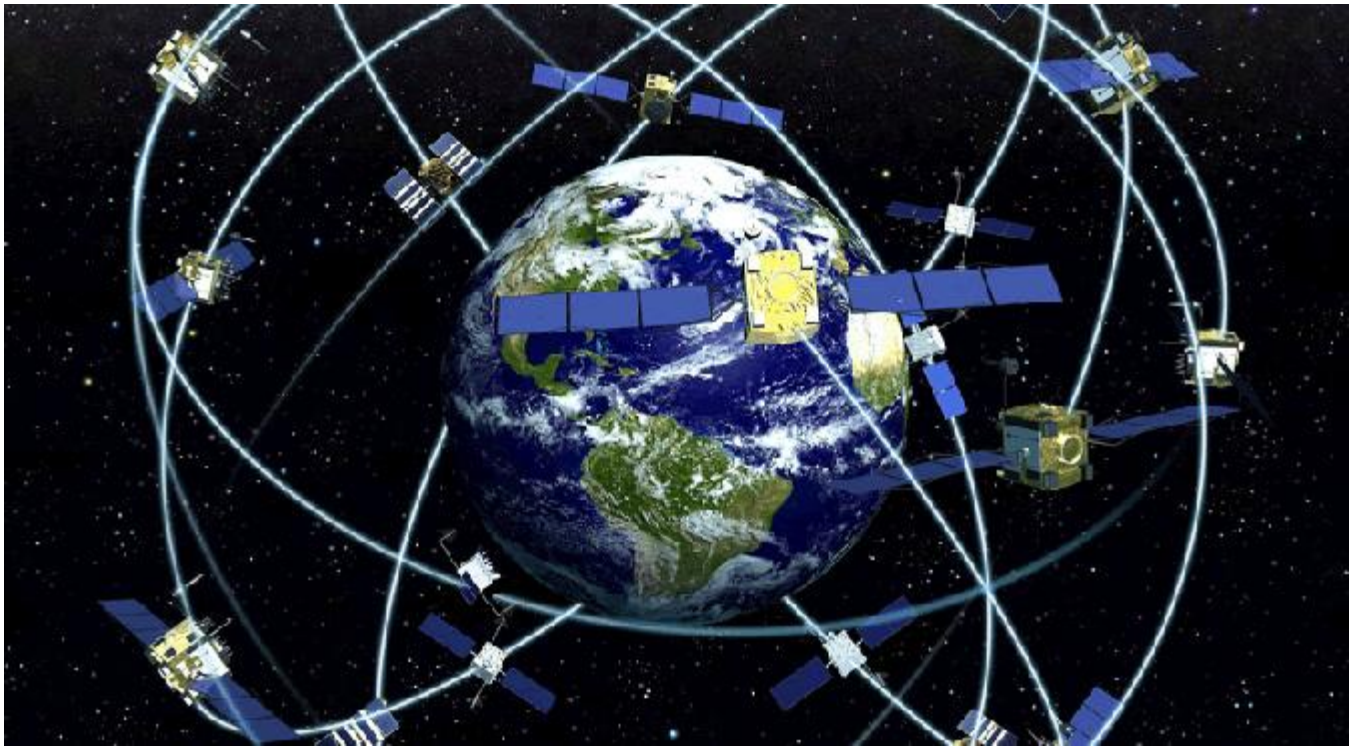
Lukas Kuster



institute for pervasive computing

Motivation

- GPS for localization



[7]

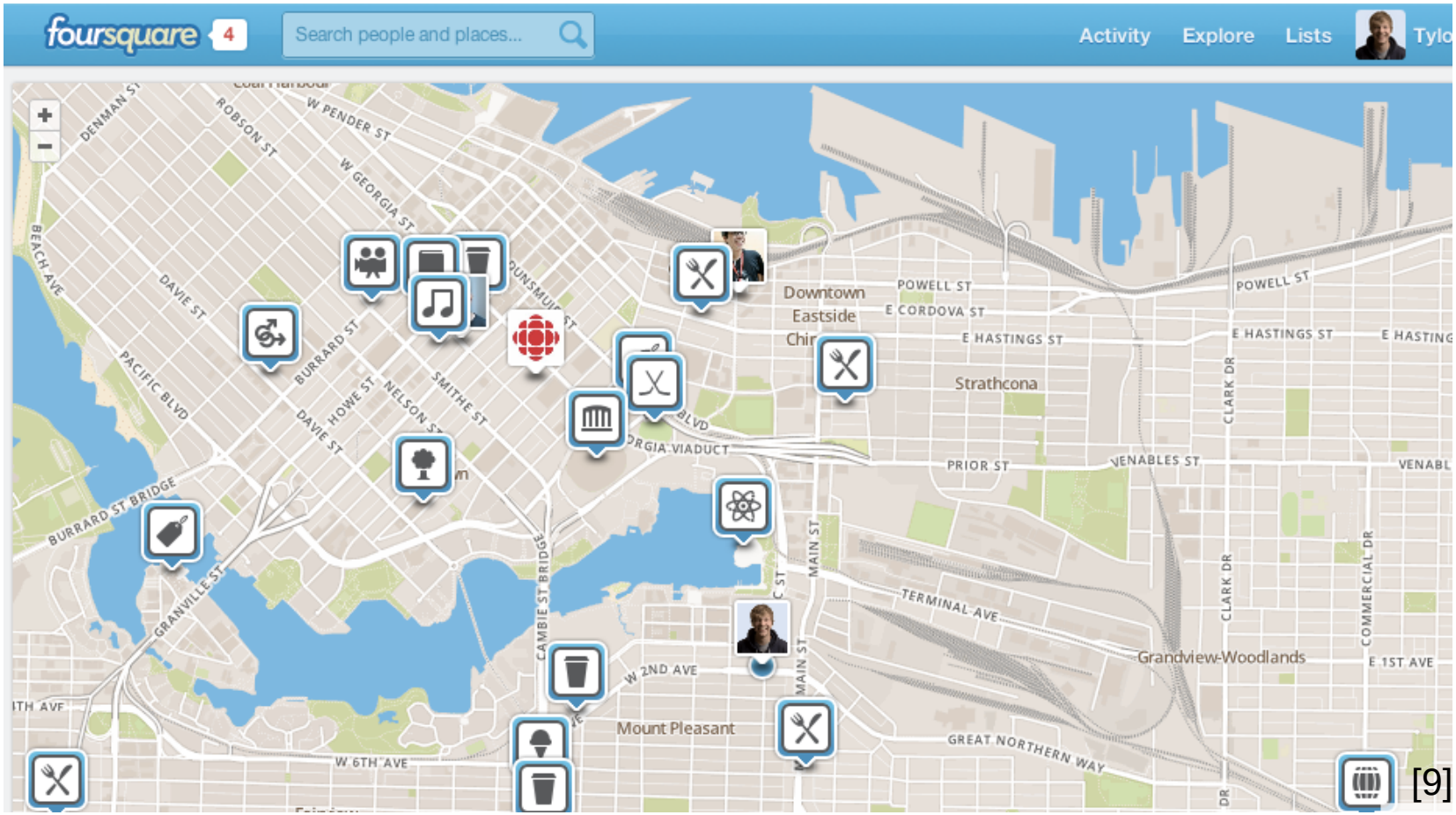
Motivation

- Indoor navigation

[8]

Motivation

- Crowd sensing



[9]

Motivation

- Targeted Advertisement

[10]

Motivation

- Tourist guidance

[12]

Semantic Localization

- GPS
- WiFi
- Images
- Sound
- Mobility

Semantic Localization

- GPS
- WiFi
- **Images**
- Sound
- Mobility

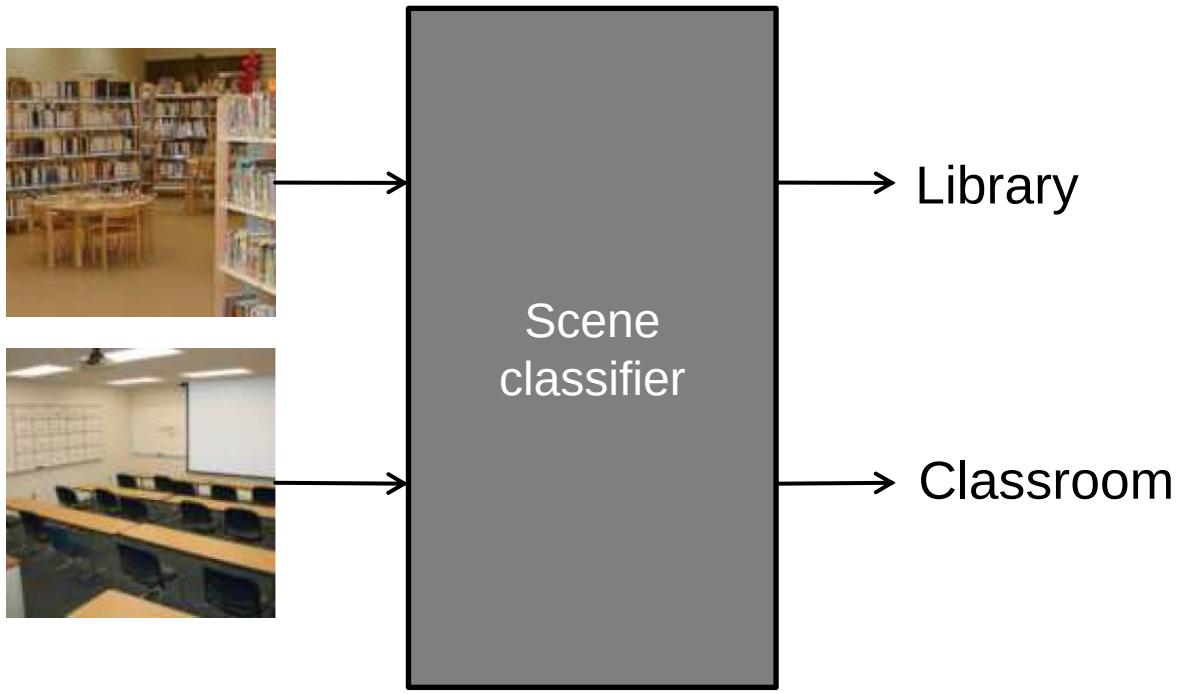
- Works for unseen places
- Outdoor and indoor
- Rich in information
- User's point of view
- No special hardware

Overview

- Motivation
- Image Indoor Scene Recognition
  - Recognizing Indoor Scenes – 2009
  - Unsupervised Discovery of Mid-Level Discriminative Patches – 2012
  - Blocks that Shout – 2013
- Semantic Localization in full Systems
- Conclusions

Scene classification in computer vision

- Goals:
  - Assign a scene category to an input image



Challenges in scene recognition

- Outdoor scenes
  - Global properties
  - Geometric
- Indoor scenes
  - Local properties
  - Semantic meaningful objects
  - Arrangement of Objects

Scene Classification

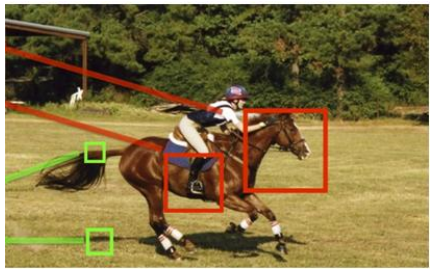
2009 2012 2013

Recognizing Indoor Scenes

Quattoni et al.

1

...



Unsupervised Discovery of Mid-Level Discriminative Patches

Singh et al.

2

Blocks that Shout: Distinctive Parts for Scene Classification

Juneja et al.

3

Recognizing Indoor Scenes - Quattoni et al. (2009)

- Two different Image feature descriptors
  - Global information – Gist descriptors
  - Local informations – Sift descrptors

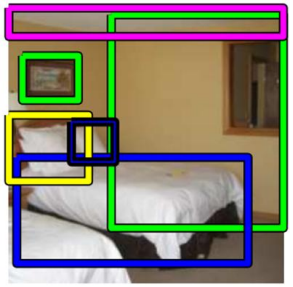
- MIT Scene 67 dataset

Recognizing Indoor Scenes - Quattoni et al. (2009)

Random Prototypes

Recognizing Indoor Scenes - Quattoni et al. (2009)

Random Prototypes

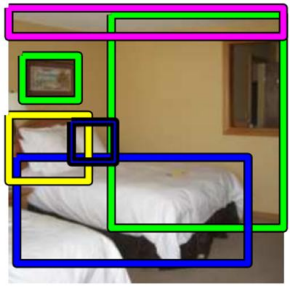


Segmentation

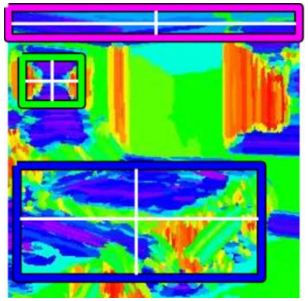
- Manual and automatic segmentation into ROI

Recognizing Indoor Scenes - Quattoni et al. (2009)

Random Prototypes



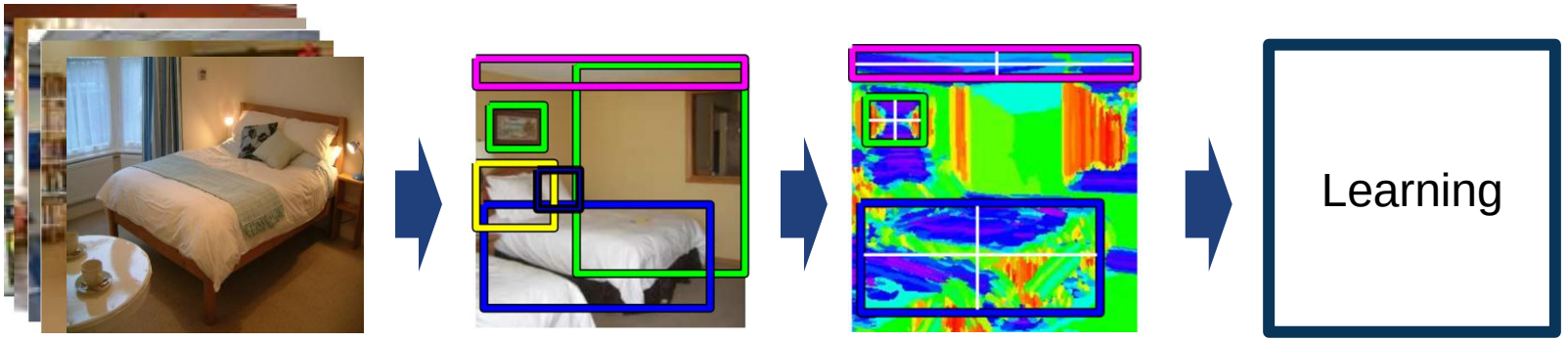
Segmentation



ROI descriptors

- Manual and automatic segmentation into ROI
- 2x2 Histogram of Visual Words

Recognizing Indoor Scenes - Quattoni et al. (2009)



Random Prototypes

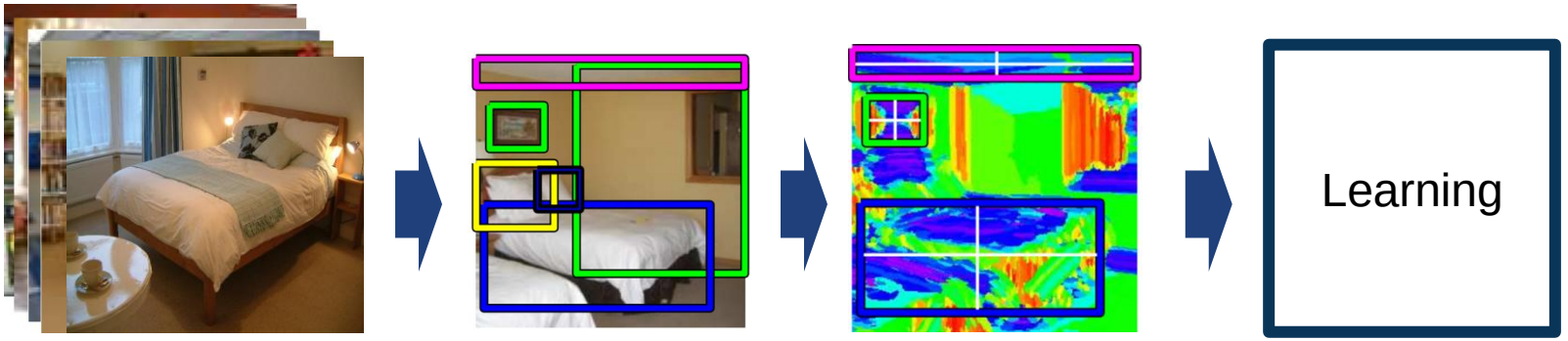
Segmentation

ROI descriptors

- Manual and automatic segmentation into ROI
- 2x2 Histogram of Visual Words
- Optimize parameters on test set

$$h(x) = \sum_{k=1}^{p} \beta_k \exp^{-\sum_{j=1}^{m_k} \lambda_{kj} f_{kj}(x) - \lambda_{kG} g_k(x)}$$

Recognizing Indoor Scenes - Quattoni et al. (2009)



Random Prototypes

Segmentation

ROI descriptors

- Manual and automatic segmentation into ROI
- 2x2 Histogram of Visual Words
- Optimize parameters on test set

$$h(x) = \sum_{k=1}^{p} \beta_k \exp\left(-\sum_{j=1}^{m_k} \lambda_{kj} f_{kj}(x) - \lambda_{kG} g_k(x) \right)$$

Prototype weight

Local features

Global feature

MIT Scene 67 dataset

- 15620 labeled images
- 67 indoor scenes categories

Test Setup – Quattoni et al. (2009)

- 67 * 80 images for training
- 67 * 20 images for testing
- Performance metric: Standard average multiclass prediction accuracy

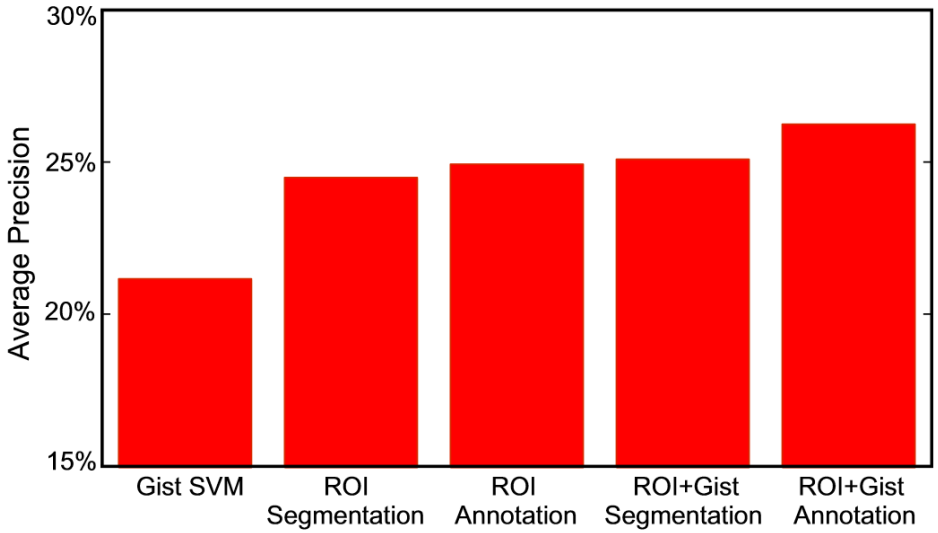
	Category 1 (Predicted)	Category 2 (Predicted)	Category 3 (Predicted)	Category 4 (Predicted)	Category 5 (Predicted)
Category 1 (Actual)	90.12%	0.00%	9.88%	0.00%	0.00%
Category 2 (Actual)	0.00%	100.00%	0.00%	0.00%	0.00%
Category 3 (Actual)	0.00%	0.00%	92.66%	0.00%	7.34%
Category 4 (Actual)	37.20%	0.00%	10.34%	52.46%	0.00%
Category 5 (Actual)	0.00%	0.00%	12.69%	0.00%	87.31%

Results – Quattoni et al. (2009)

Evaluation – Quattoni et al. (2009)

- Segmentation Methods:
  - Segmentation: automatic
  - Annotation: manual

- Features:
  - Only ROI
  - ROI + Gist



Conclusion – Quattoni et al. (2009)

- ✓ Indoor Scene classification
- ✓ Local and global features

- – Low accuracy (26%)
- – Manual annotation

Scene Classification

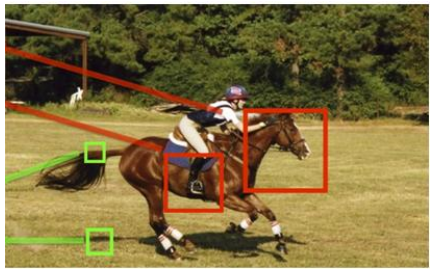
2009 2012 2013

Recognizing Indoor Scenes

Quattoni et al.

1

...



Unsupervised Discovery of Mid-Level Discriminative Patches

Singh et al.

2

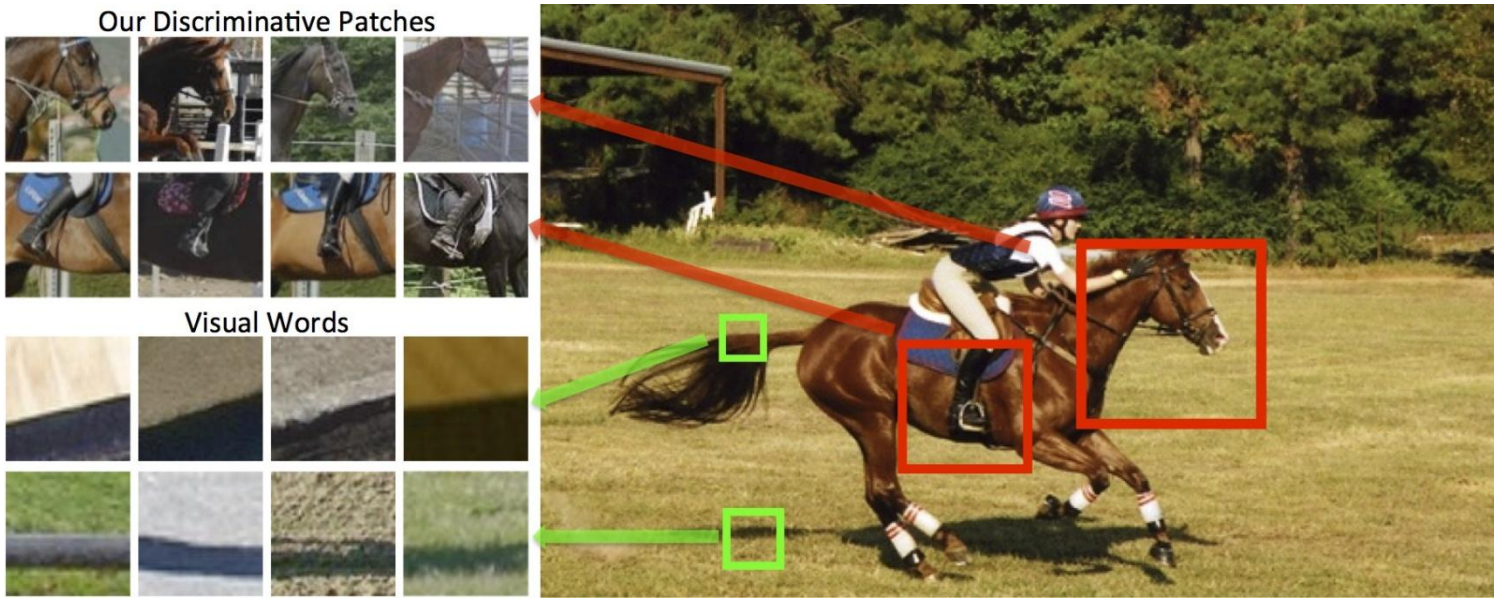
Blocks that Shout: Distinctive Parts for Scene Classification

Juneja et al.

3

Unsupervised Discovery of Mid-Level Discriminative Patches – Singh et al. (2012)

- Mid-Level patches
  - Representative: frequent occurence in world
  - Discriminative: diffrent enough from rest of the world



Singh et al. (2012)

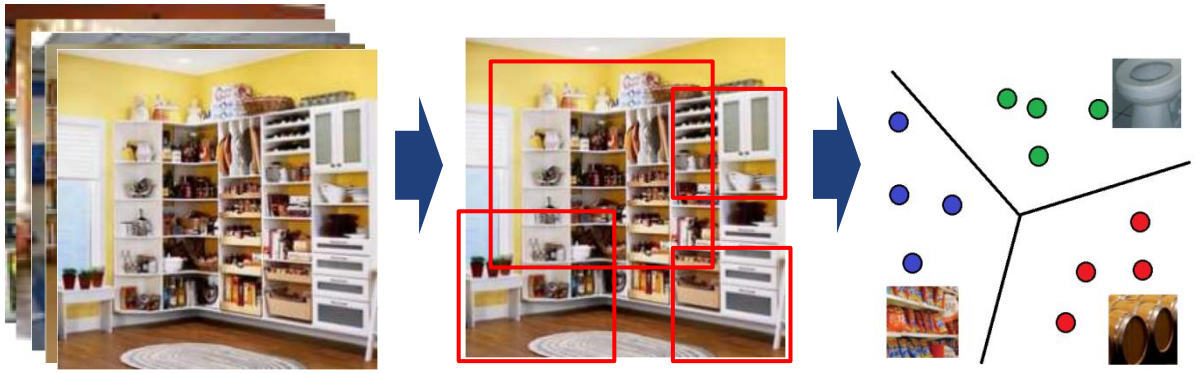
Random discovery set

Singh et al. (2012)

Random discovery set

Random patches

Singh et al. (2012)



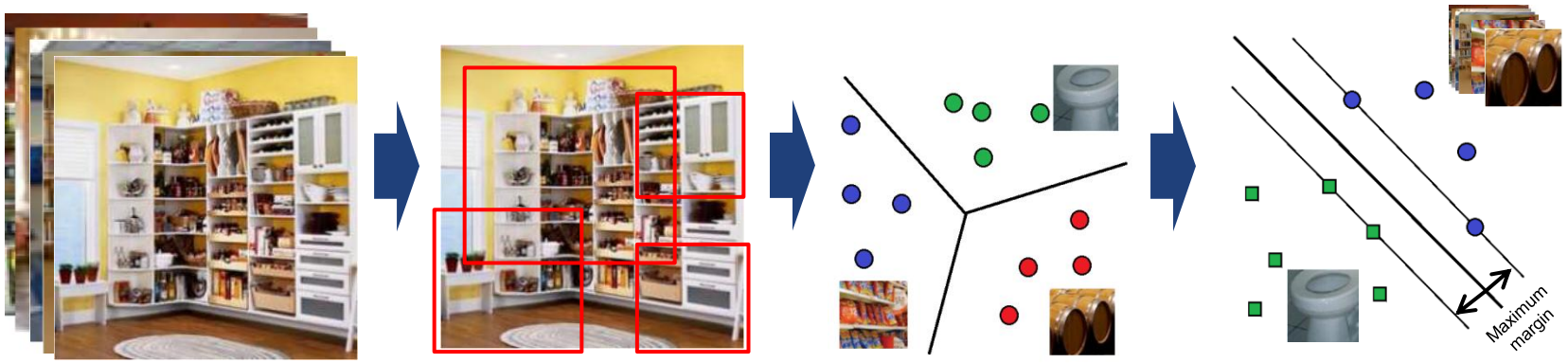
Random discovery set

Random patches

Kmeans clustering

- Cluster patches in HOG space

Singh et al. (2012)



Random discovery set

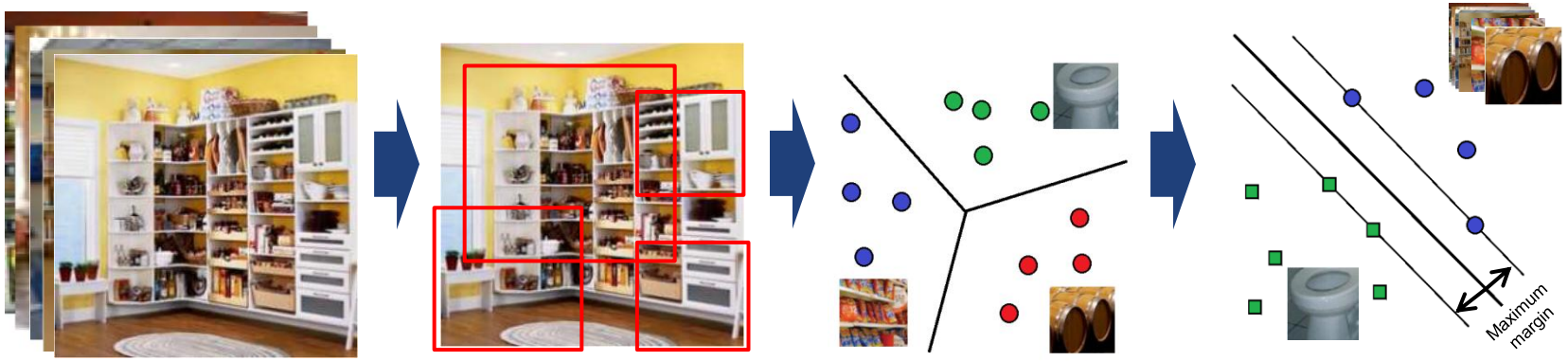
Random patches

Kmeans clustering

SVM train

- Cluster patches in HOG space
- Train detector for each cluster

Singh et al. (2012)



- Cluster patches in HOG space
- Train detector for each cluster
- Use detector on validation set
  - Get top 5 matches for new cluster
  - Kill clusters that have less than 2 matches

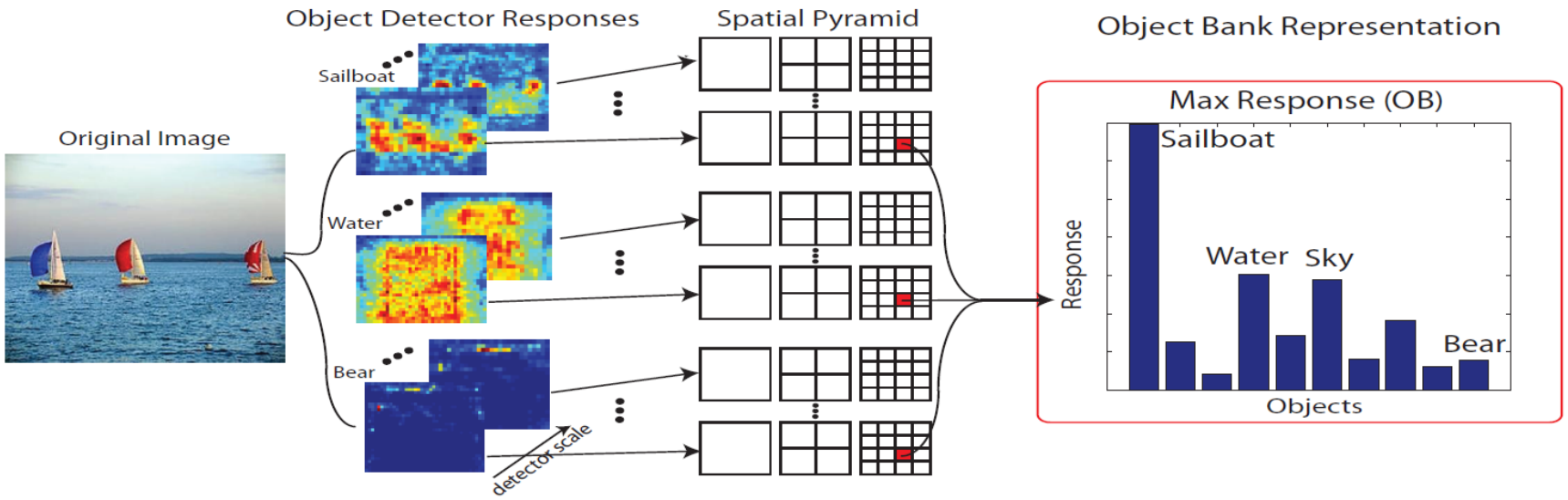
Ranking Detectors – Singh et al. (2012)

- Purity
  - Same visual concept
  - Sum of top r detection scores

- Discriminativeness
  - Detected rarely in natural world
  - $$\frac{\#\text{detections in training set}}{\#\text{detections in (training set} \cup \text{natural world)}}$$

Image descriptor – Singh et al. (2012)

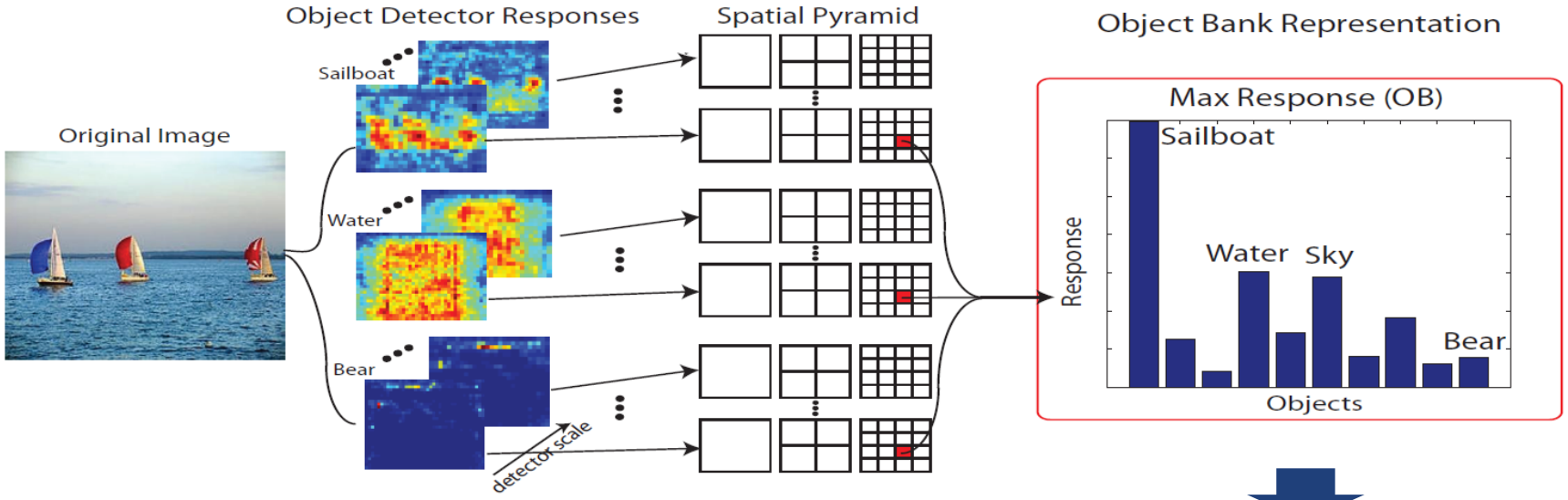
Object Bank Image representation – Li, L-J et al. (2010)



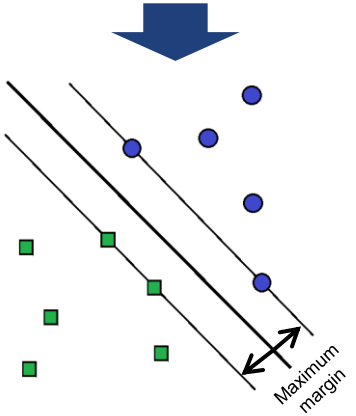
- Detect Patches on diffrent scales and diffrent spatial pyramid levels
- Train classifier with SVM

Image descriptor – Singh et al. (2012)

Object Bank Image representation – Li, L-J et al. (2010)



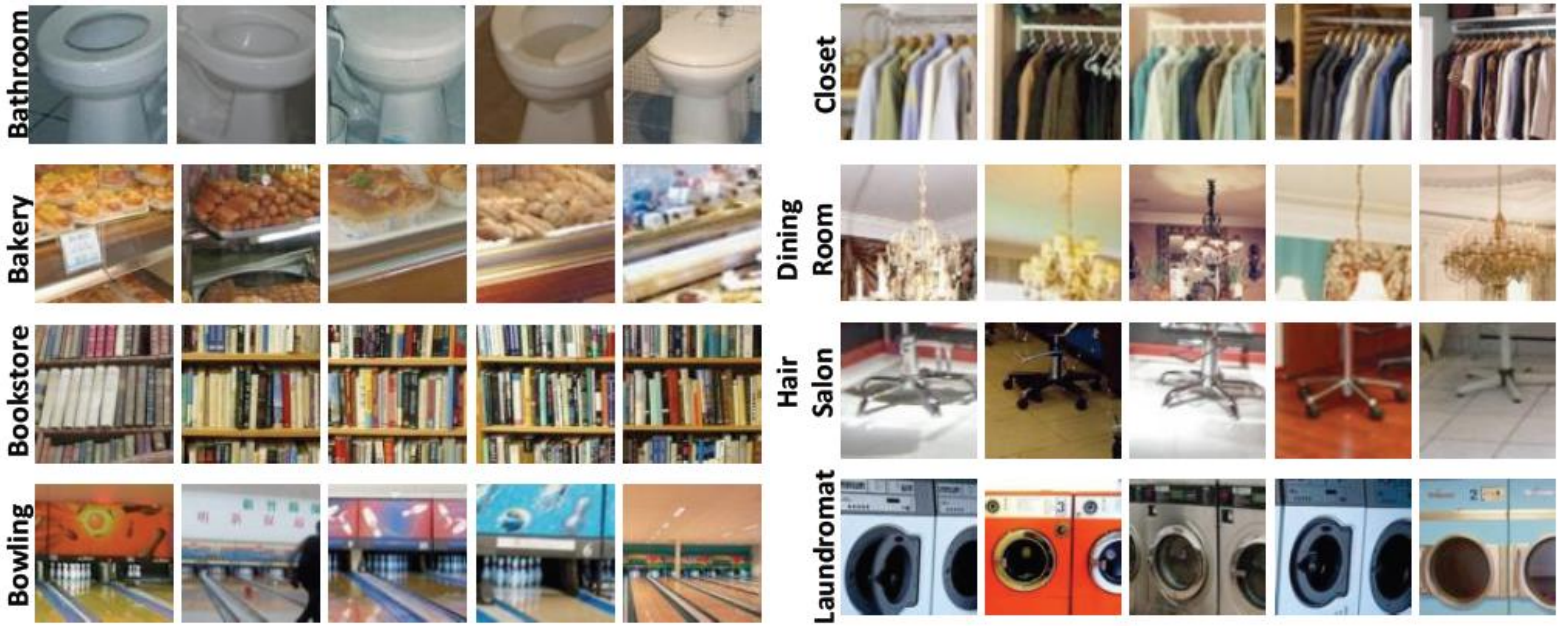
- Detect Patches on diffrent scales and diffrent spatial pyramid levels
- Train classifier with SVM



SVM

Top Ranked patches – Singh et al. (2012)

- MIT 67 Benchmark



Evaluation – Singh et al. (2012)

Accuracy:

Spatial Pyramid HOG	29,8
Spatial Pyramid SIFT (SP)	34,4
ROI-GIST (Quattoni et al.)	26,5
Object Bank	37,6
Patches	38,1

Evaluation – Singh et al. (2012)

Accuracy:

Spatial Pyramid HOG	29,8
Spatial Pyramid SIFT (SP)	34,4
ROI-GIST (Quattoni et al.)	26,5
Object Bank	37,6
Patches	38,1

Combination approaches:

GIST+SP+DPM	43,1
Patches+GIST+SP+DPM	49,4

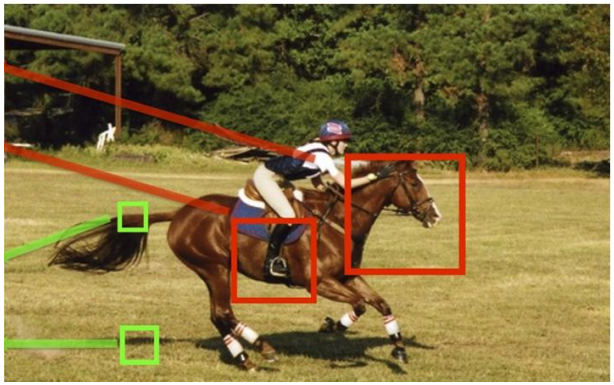
Conclusion

Quattoni et al. (2009)

- ✓ Indoor Scene classification
- ✓ Local and global features

- – Low accuracy (26%)
- – Manual annotation

Singh et al. (2012)



- ✓ Low supervision
- ✓ Better accuracy

- – Low accuracy (49%)
- – Inefficient

Scene Classification

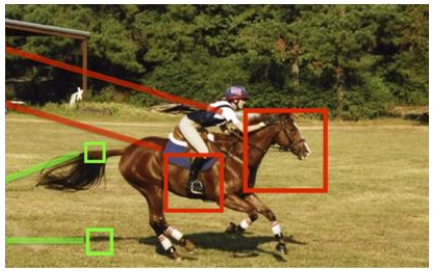
2009 2012 2013

Recognizing Indoor Scenes

Quattoni et al.

1

...



Unsupervised Discovery of Mid-Level Discriminative Patches

Singh et al.

2

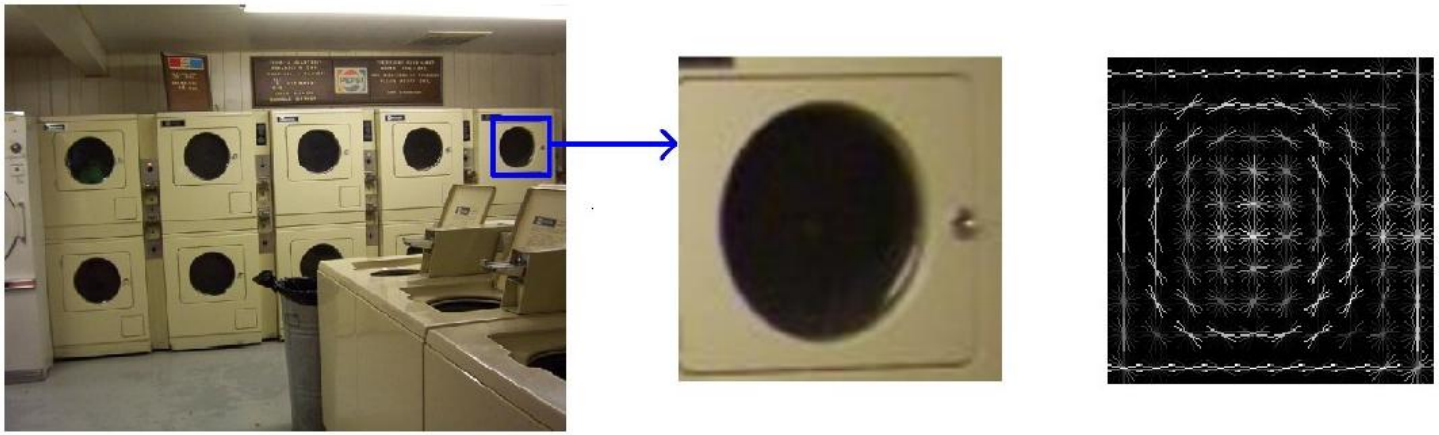
Blocks that Shout: Distinctive Parts for Scene Classification

Juneja et al.

3

Blocks that Shout: Distinctive Parts for Scene Classification – Juneja et al. (2013)

- More efficient
- Distinctive patches



Original Image

Block

HOG Representation

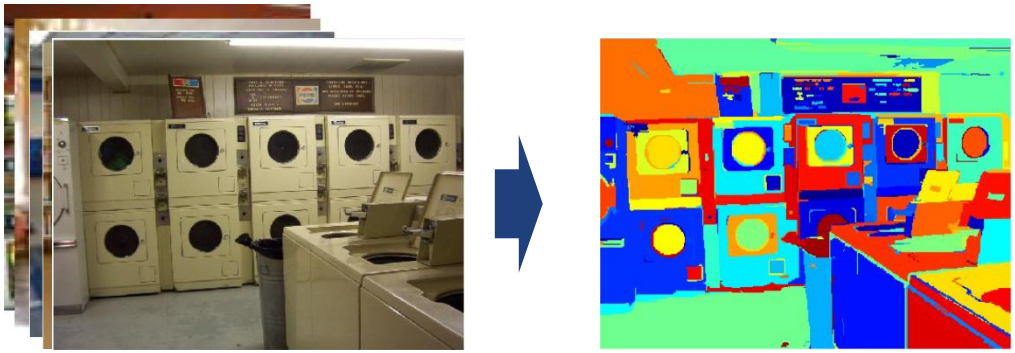
Blocks that Shout – Juneja et al. (2013)

Seeding

Initial training set

Blocks that Shout – Juneja et al. (2013)

Seeding



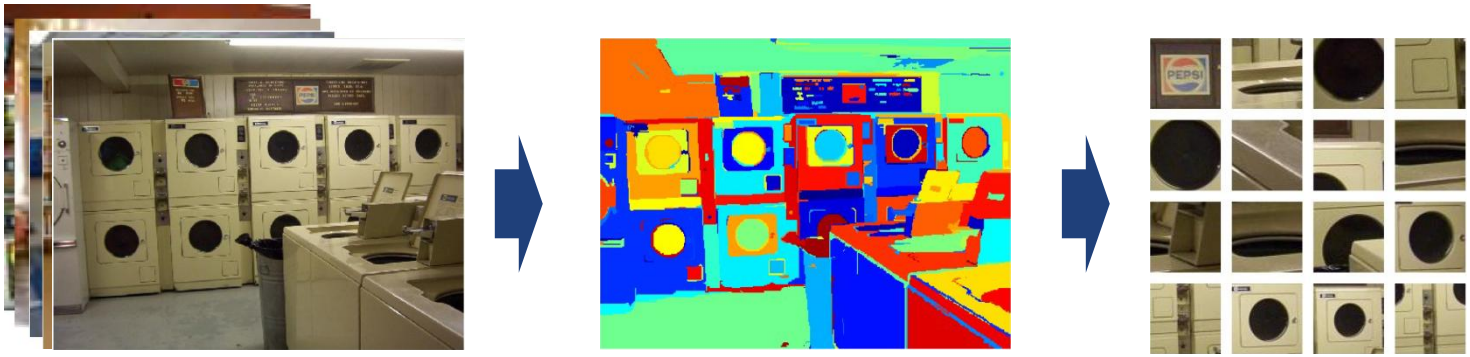
Initial training set

Superpixels

- Automatic segmentation into superpixels

Blocks that Shout – Juneja et al. (2013)

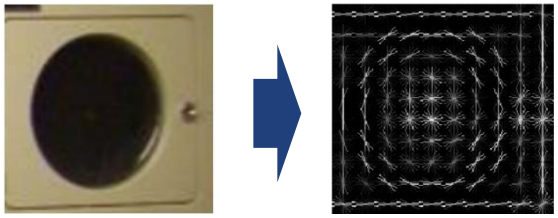
Seeding



- Automatic segmentation into superpixels
- Seedblocks:
  - Intermediate sized superpixels
  - Image variation

Blocks that Shout – Juneja et al. (2013)

Seeding → Expansion

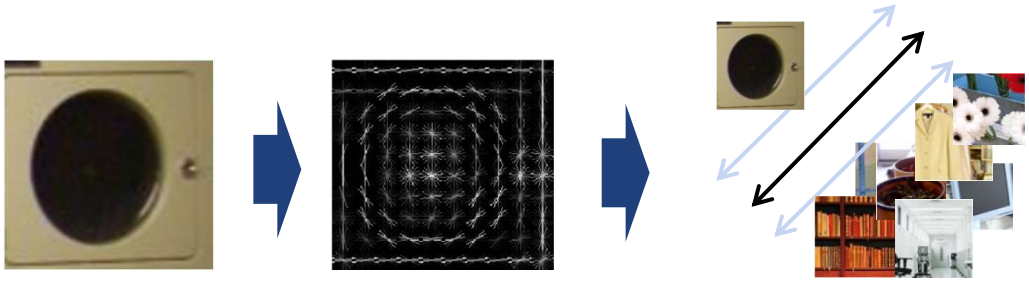


Seed Block HOG descriptor

- 8x8 HOG cells of 8x8 pixels

Blocks that Shout – Juneja et al. (2013)

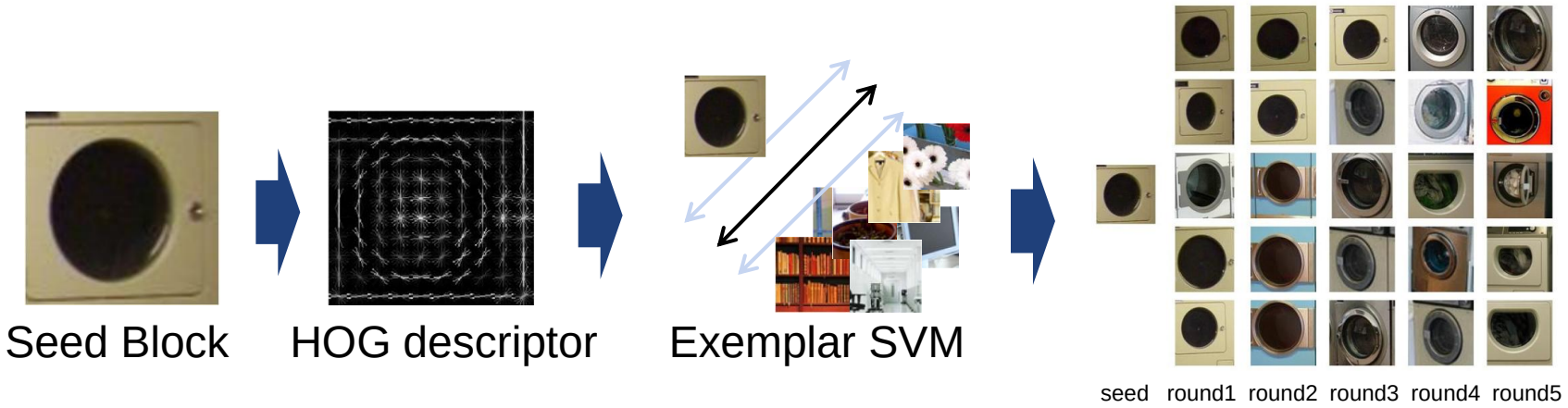
Seeding → Expansion



- 8x8 HOG cells of 8x8 pixels
- Detect similiar blocks

Blocks that Shout – Juneja et al. (2013)

Seeding → Expansion



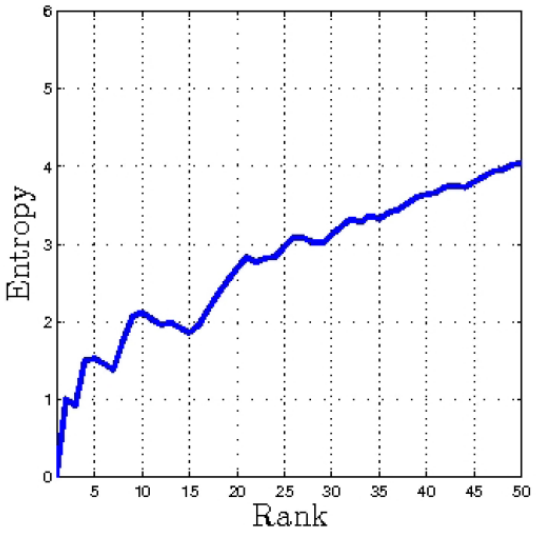
- 8x8 HOG cells of 8x8 pixels
- Detect similiar blocks
- 5 iterations for final part detector

Blocks that Shout – Juneja et al. (2013)

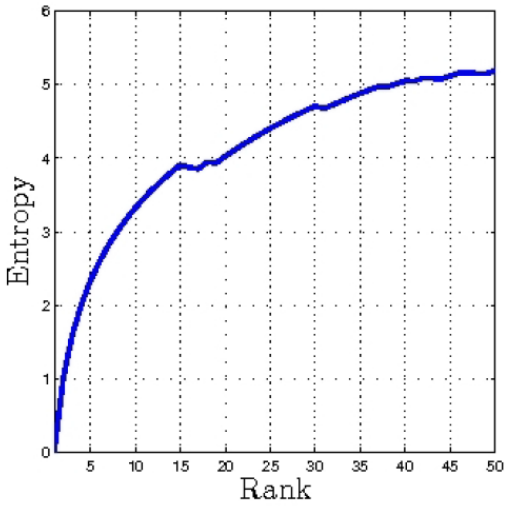
Seeding → Expansion → Selection

- Select most distincitve part detectors

- Entropy: $$H(Y,r) = -\sum_{y=1}^{N} p(y,r) \log_2 p(y,r)$$



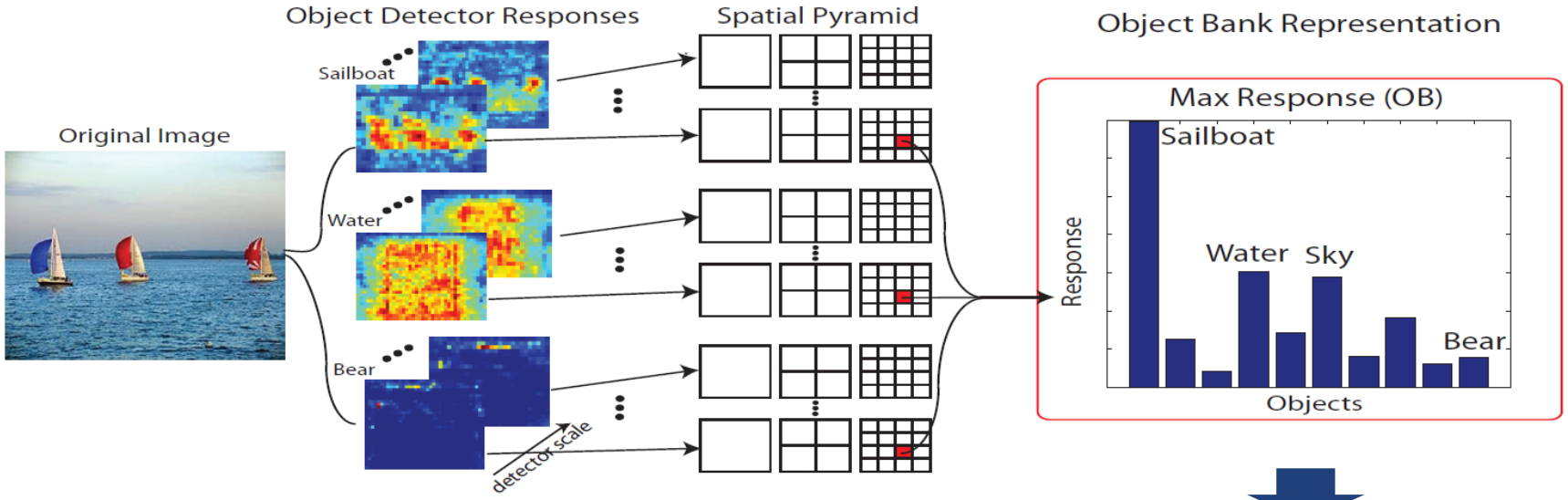
(a) Discriminative detector



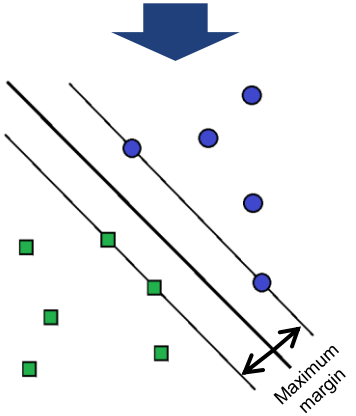
(b) Non-Discriminative detector

Image descriptor – Blocks that Shout (2013)

Object Bank Image representation – Li, L-J et al. (2010)



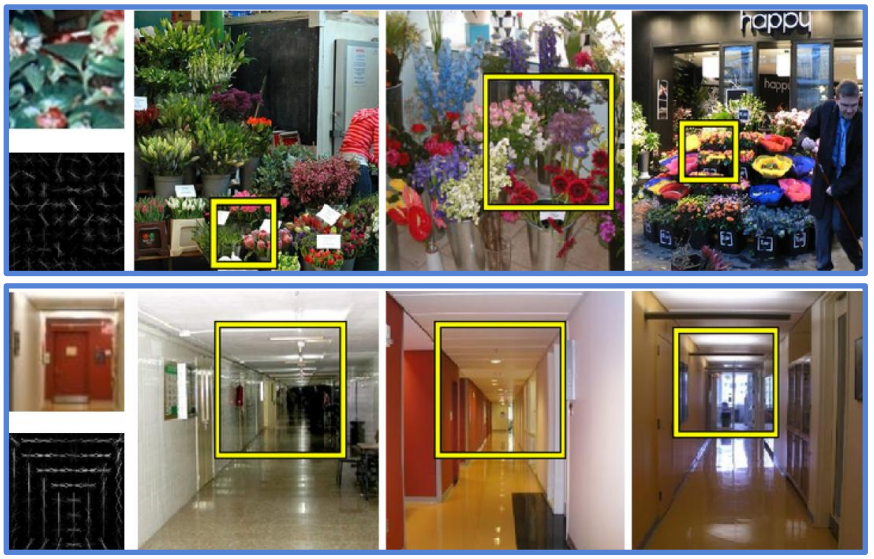
- Detect Patches on diffrent scales and diffrent spatial pyramid levels
- Train classifier with SVM



SVM

Blocks that Shout – Juneja et al. (2013)

Results



Blocks that Shout – Juneja et al. (2013)

Evaluation

Accuracy:

ROI-GIST (Quattoni et al.)	26,5
Object Bank	37,6
Patches (Singh et al.)	38,1
BoP	46,1

Blocks that Shout – Juneja et al. (2013)

Evaluation

Accuracy:

ROI-GIST (Quattoni et al.)	26,5
Object Bank	37,6
Patches (Singh et al.)	38,1
BoP	46,1

Combination approaches:

Patches+GIST+SP+DPM (Singh et al.)	49,4
IFV + BoP	63,1

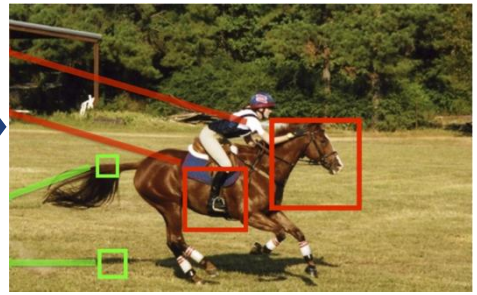
Conclusion

Quattoni et al. (2009)

- ✓ Indoor Scene classification
- ✓ Local and global features

- – Low accuracy (26%)
- – Manual annotation

Singh et al. (2012)



- ✓ Low supervision
- ✓ Better accuracy

- – Low accuracy (49%)
- – Inefficient

Juneja et al. (2013)

- ✓ Low supervision
- ✓ More efficient
- ✓ Distinctive Parts
- ✓ Even better accuracy

- – Low accuracy (63%)

Overview

- Motivation
- Image Indoor Scene Recognition
  - Recognizing Indoor Scenes – 2009
  - Unsupervised Discovery of Mid-Level Discriminative Patches – 2012
  - Blocks that Shout – 2013
- Semantic Localization in full Systems
- Conclusions

Systems Overview

CrowdSense@Place

- 2012
- Crowd sensing
- Link visits with place categories
- Share output with location sensitive applications

Systems Overview

CrowdSense@Place

- 2012
- Crowd sensing
- Link visits with place categories
- Share output with location sensitive applications

Place Naming System

- 2013
- Crowd sensing

Output:

- Functional name (eg. Food place)
- Business name (eg. Starbucks)
- Personal name (eg. My home)

Systems Overview

CrowdSense@Place

- 2012
- Crowd sensing
- Link visits with place categories
- Share output with location sensitive applications

Place Naming System

- 2013
- Crowd sensing

Output:

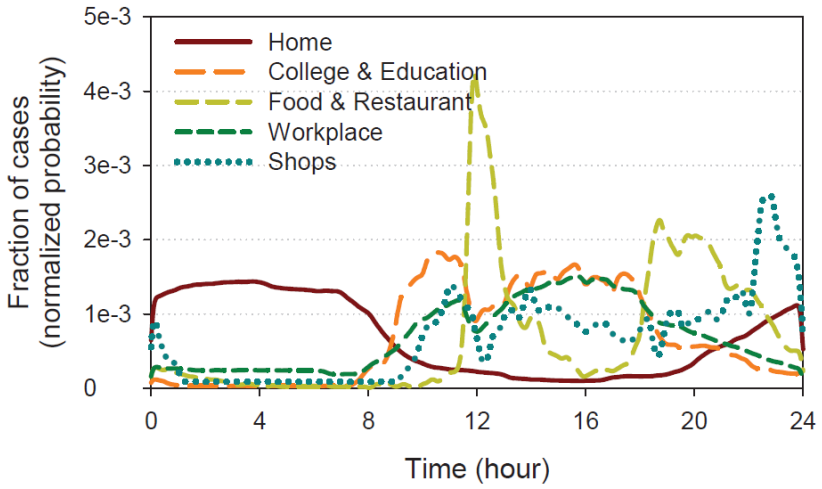
- Functional name (eg. Food place)
- Business name (eg. Starbucks)
- Personal name (eg. My home)

CheckInside

- 2014
- Location-based Social Network
- Improved venues list in Check-ins

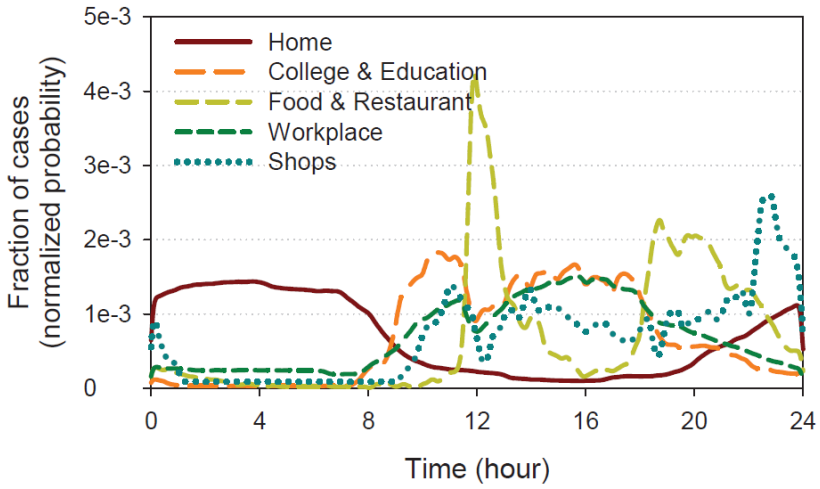
Sensor Data

- Mobility:
  - GPS
  - WiFi
  - Trajectory

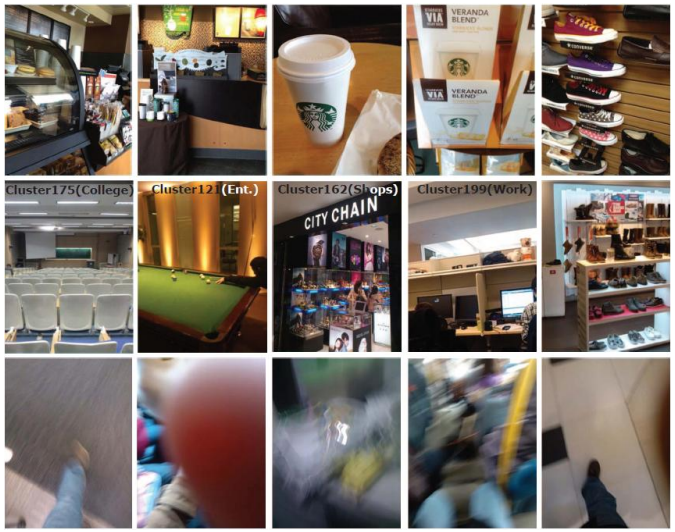


Sensor Data

- Mobility:
  - GPS
  - WiFi
  - Trajectory

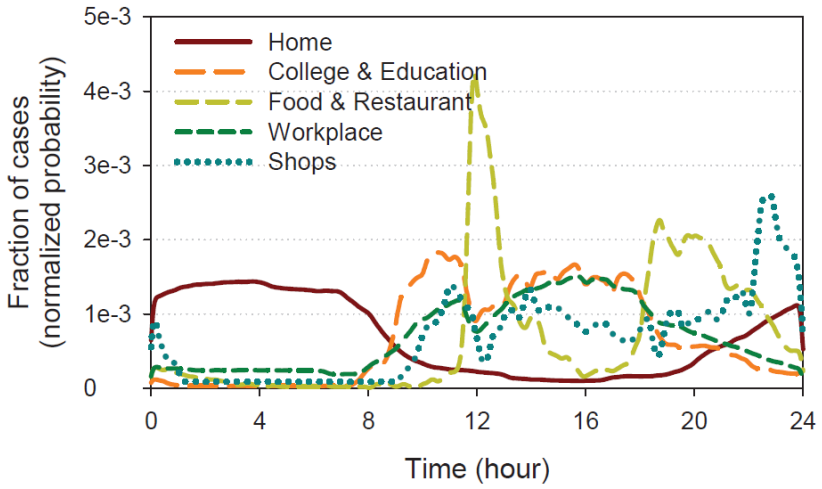


- Visual Classifiers:
  - Text Recognition
  - **Indoor Scene Classification**
  - Object Recognition

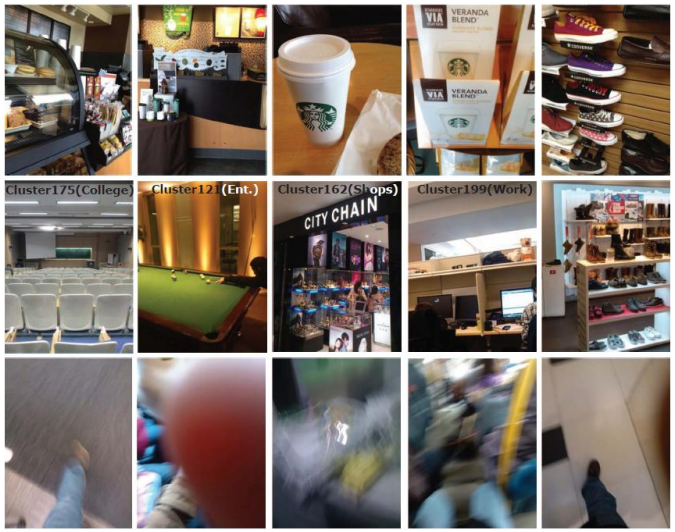


Sensor Data

- Mobility:
  - GPS
  - WiFi
  - Trajectory



- Visual Classifiers:
  - Text Recognition
  - **Indoor Scene Classification**
  - Object Recognition

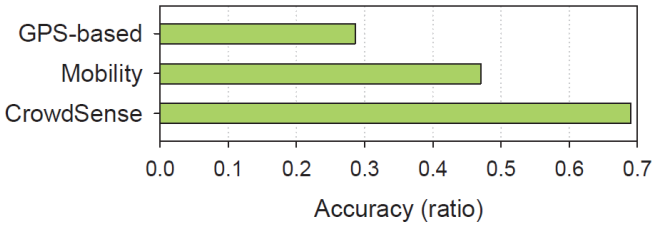


- Sound Classifiers:
  - Speech Recognition
  - Sound Classification

Evaluation

CrowdSense@Place

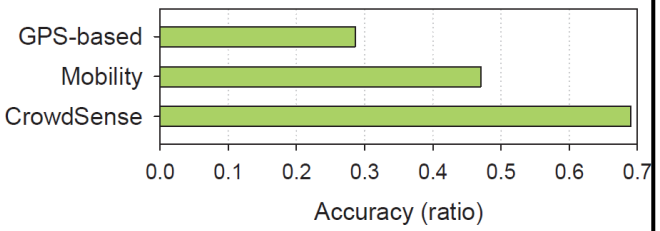
- 1241 places
- 6 categories
- Accuracy:
  ~ 40% - 95%
- Overall : ~ 69%



Evaluation

CrowdSense@Place

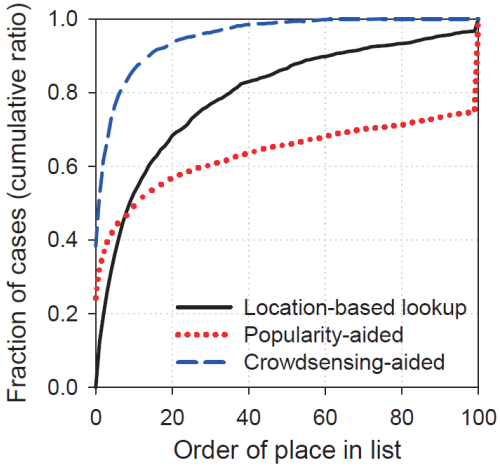
- 1241 places
- 6 categories
- Accuracy:
  ~ 40% - 95%
- Overall : ~ 69%



Place Naming System

- 3800 places
- 9 categories
- Functional name:
  ~ 20% - 90%

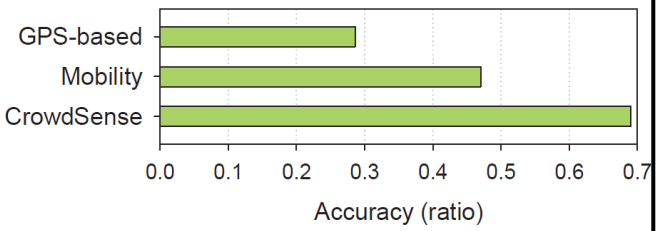
Business Name:



Evaluation

CrowdSense@Place

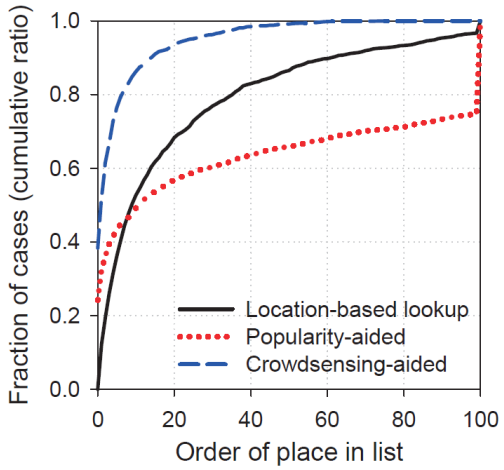
- 1241 places
- 6 categories
- Accuracy:
  ~ 40% - 95%
- Overall : ~ 69%



Place Naming System

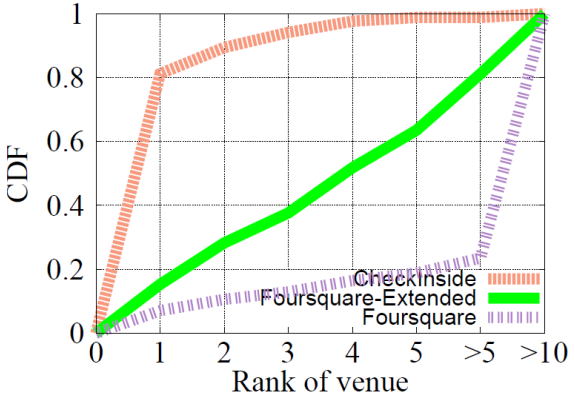
- 3800 places
- 9 categories
- Functional name:
  ~ 20% - 90%

Business Name:



CheckInside

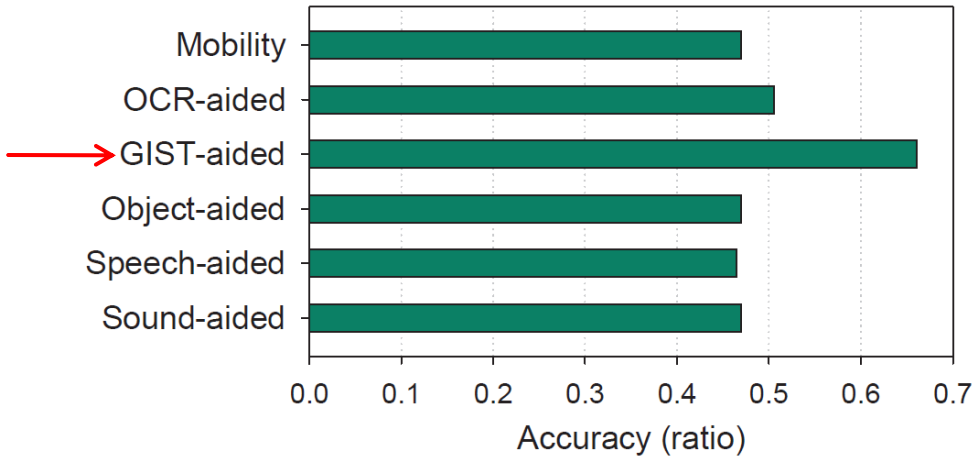
- 711 stores
- 99% in top 5



Visual Scene Recognition Evaluation

- Good for functional naming

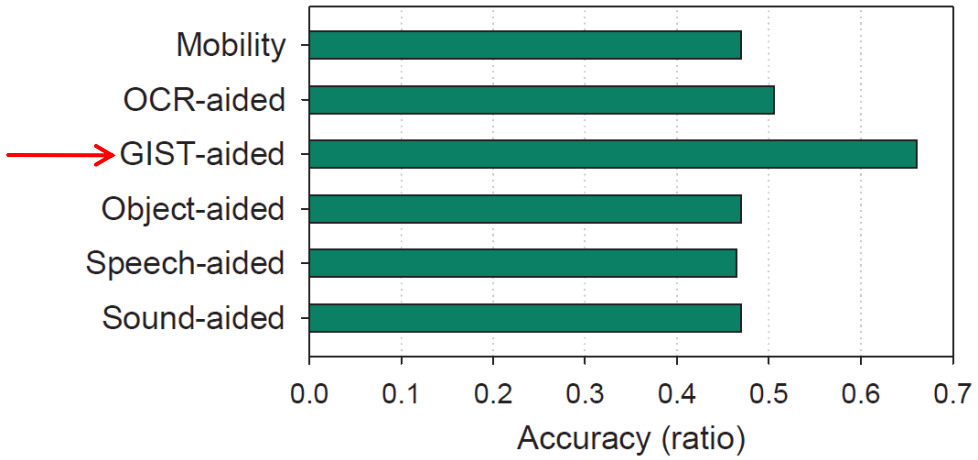
CrowdSense@Place accuracy:



Visual Scene Recognition Evaluation

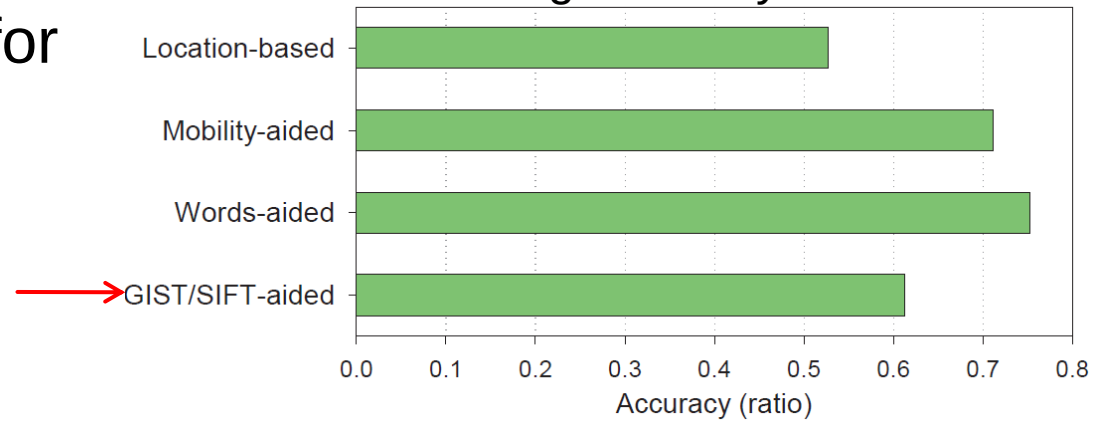
- Good for functional naming

CrowdSense@Place accuracy:



- Intermediate performance gain for business naming

Business Naming accuracy:



Conclusion

- Crowd sensing improves semantic localization
- Relatively low accuracy
- User interaction still needed

- Visual scene recognition:
  - Fast progress
  - State of the art could improve the systems

References

(1) Quattoni, A.; Torralba, A., "Recognizing indoor scenes," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.*

(2) Singh, S.; Gupta, A; Efros, A. A., “Unsupervised discovery of mid-level discriminative patches,” *European conference on Computer Vision* (ECCV)*, 2012.*

(3) Juneja, M.; Vedaldi, A.; Jawahar, C.V.; Zisserman, A., "Blocks That Shout: Distinctive Parts for Scene Classification," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.*

(4) Chon, Y.; Lane, N. D.; Li, F.; Cha, H.; Zhao, F., “Automatically characterizing places with opportunistic crowdsensing using smartphones,” *ACM Conference on Ubiquitous Computing* (UbiComp), 2012.

(5) Chon, Y.; Kim, Y.; Cha, H., “Autonomous place naming system using opportunistic crowdsensing and knowledge from crowdsourcing,” *International conference on Information processing in sensor networks* (IPSN), 2013.

(6) Elhamshary, M; Youssef, M., “CheckInside: a fine-grained indoor location-based social network,” *ACM International Joint Conference on Pervasive and Ubiquitous Computing* (UbiComp), 2014.

References

(7) http://www.extremetech.com/extreme/126843-think-gps-is-cool-ips-will-blow-your-mind/2

(8) http://free-wifi-service.com/

(9) http://denimandsteel.com/talks/polyglot/

(10) http://www.elatewiki.org/images/Special.jpeg

(11) Li, L.J., Su, H., Xing, E., Fei-fei, L., “Object bank: A high-level image representation for scene classication and semantic feature sparsication,” *Conference on Neural Information Processing Systems* (NIPS), 2010.

(12) https://www.flip4new.de/blog/nokia-lumia-920-review-was-kann-das-windows-phone/