Learning Spatial Context: Using Stuff to Find Things

Wei-Cheng Su

THE UNIVERSITY OF TEXAS AT AUSTIN
UT ECE
ELECTRICAL & COMPUTER ENGINEERING

Motivation

- Leverage contextual information to enhance detection
- Some context objects are non-rigid and are more naturally classified based on texture or color. e.g., sky, trees, road
- Find the relationships between the stuff of context and the object

Outline

- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

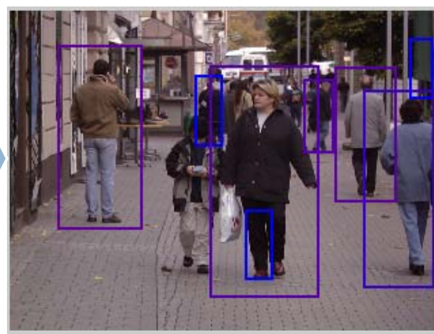
Training

Segmentation

Region features & centroids

Learning

Detection

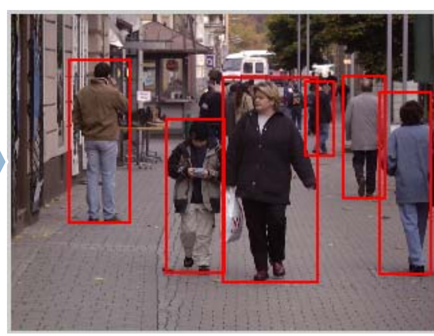


Candidate boxes & scores

Things-and-stuff relationships

Model parameters

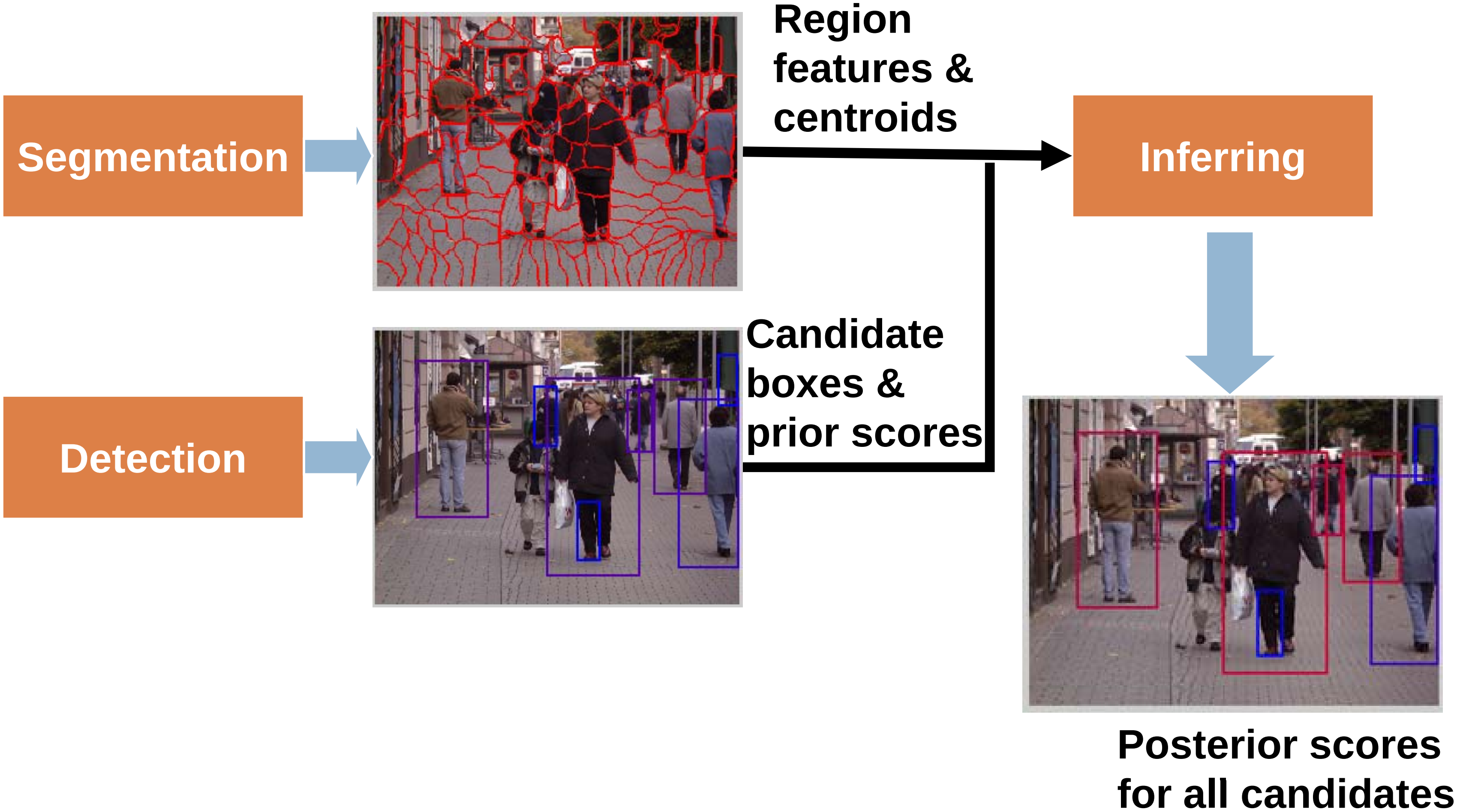
Annotation



Ground truths

*Red boxes indicate high scores
Blue boxes indicate low scores

Inferring


Segmentation
Region features & centroids
Inferring
Detection
Candidate boxes & prior scores
Posterior scores for all candidates

Outline

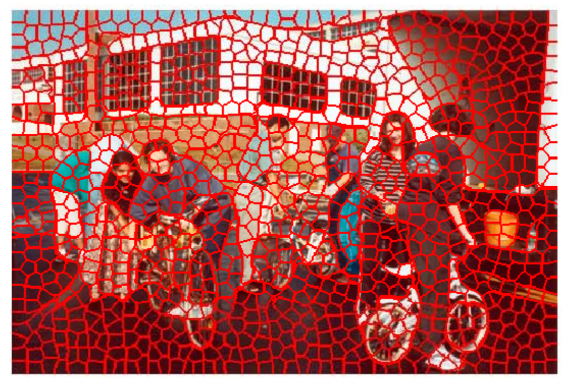
- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

Preprocessing

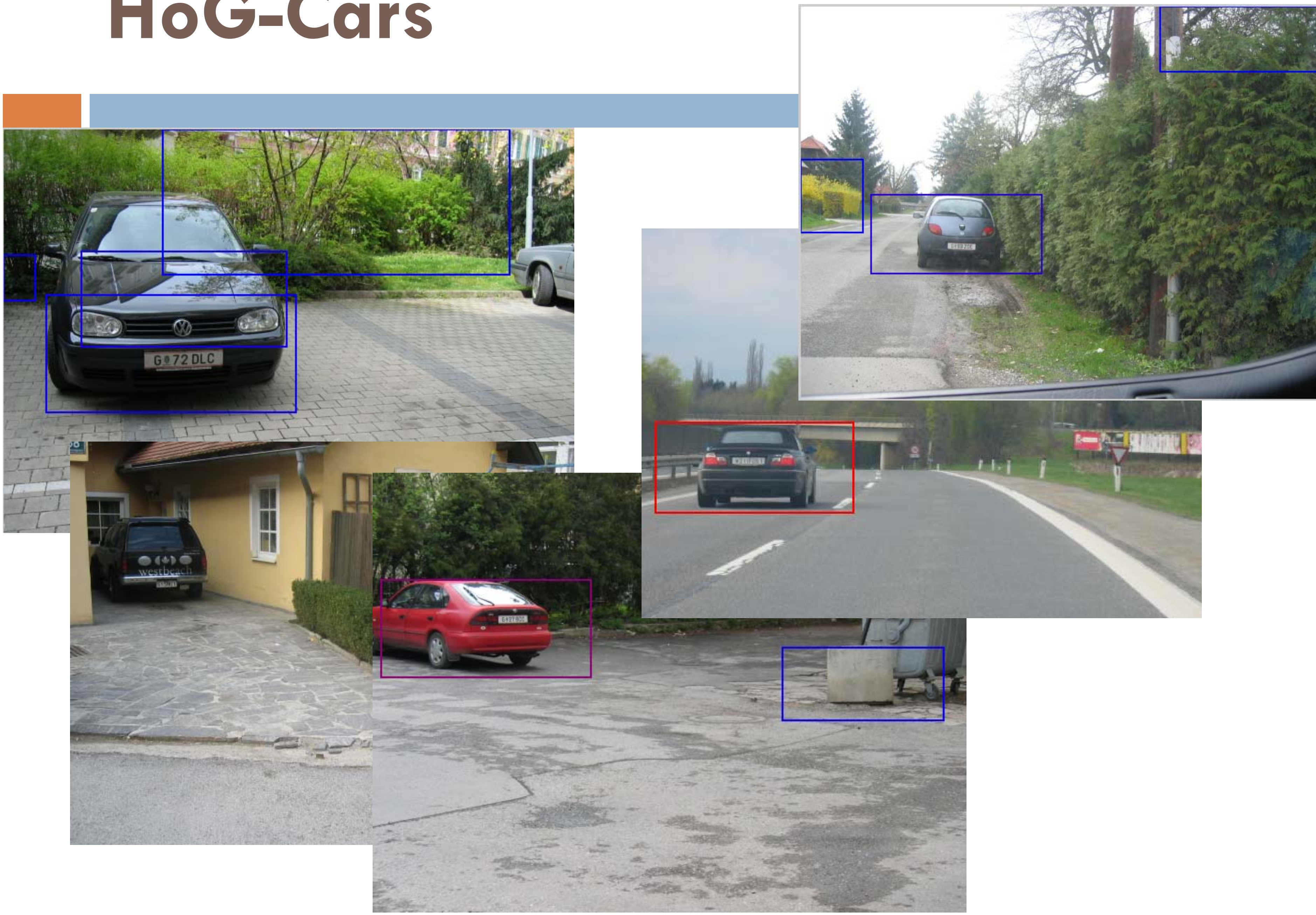
- Segmentation
  - Superpixel
  - Pentium-D 2.4 GHz, 4G RAM
  - Run out of memory with a 792x636 image
  - ~6.4 minutes for a 480x321 image
- Detection
  - HOG for detecting humans, cars, bicycles, and motorbikes
  - Patch-based boosted detector for detecting cars in satellite images

Segmentation

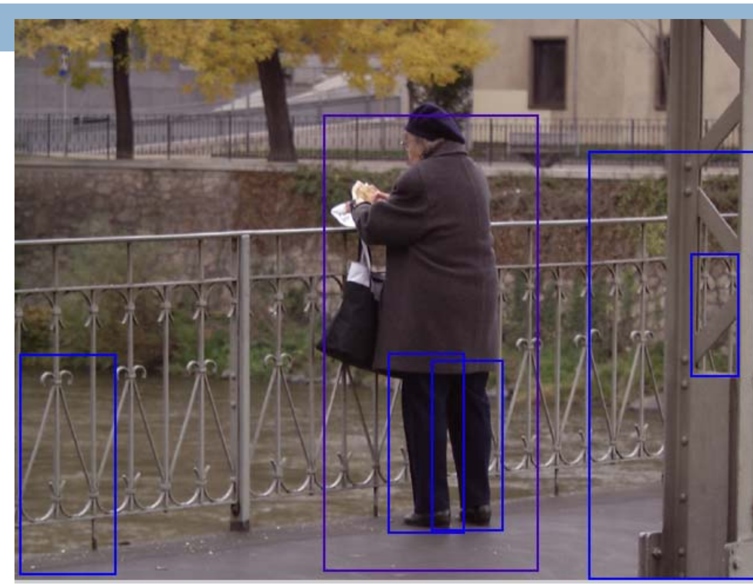
This level of segmentation result is used

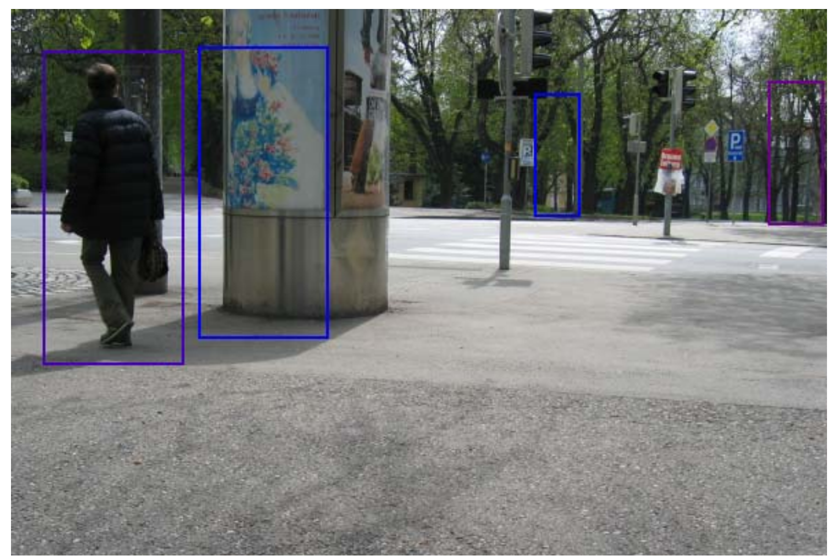


HoG-Cars

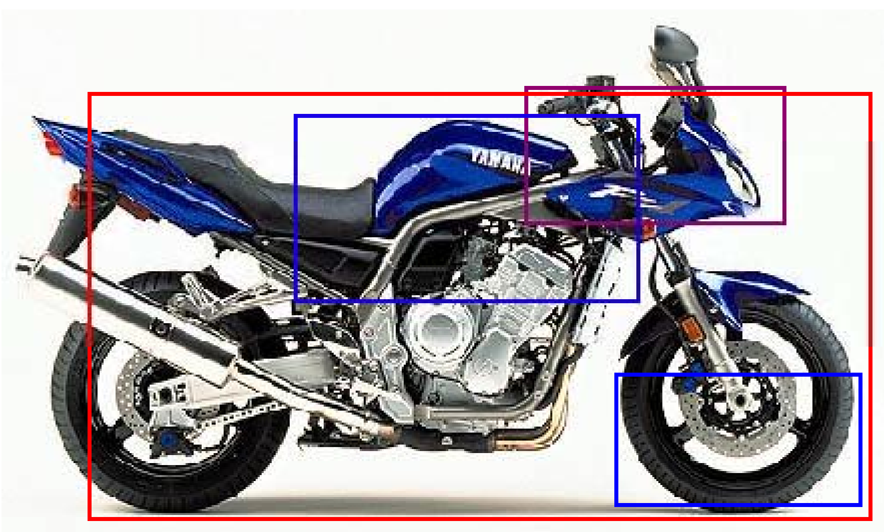


HoG-People

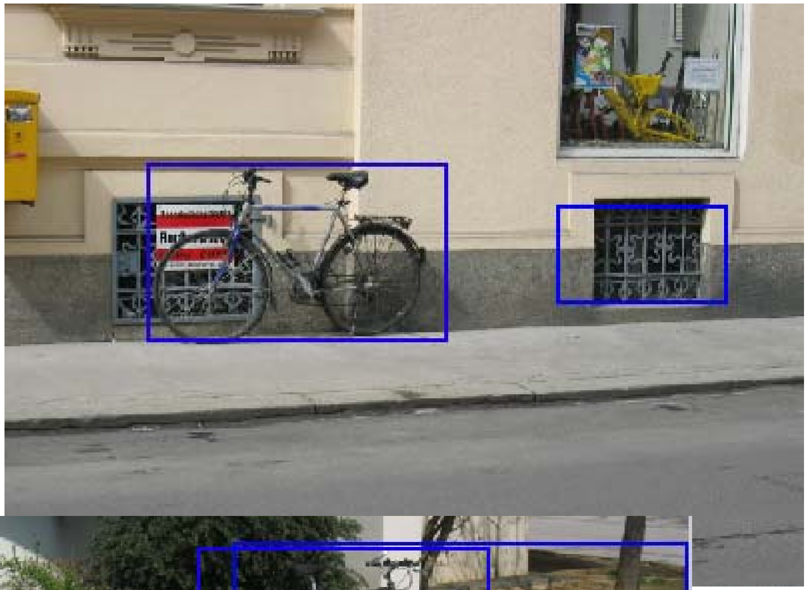


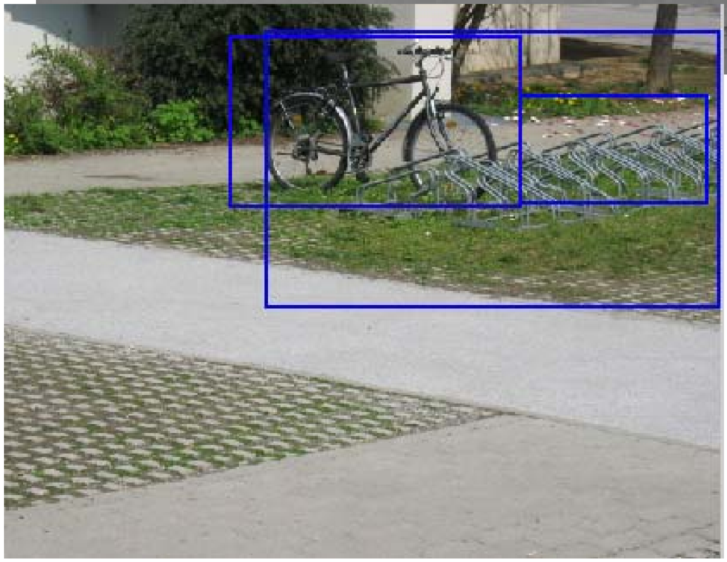


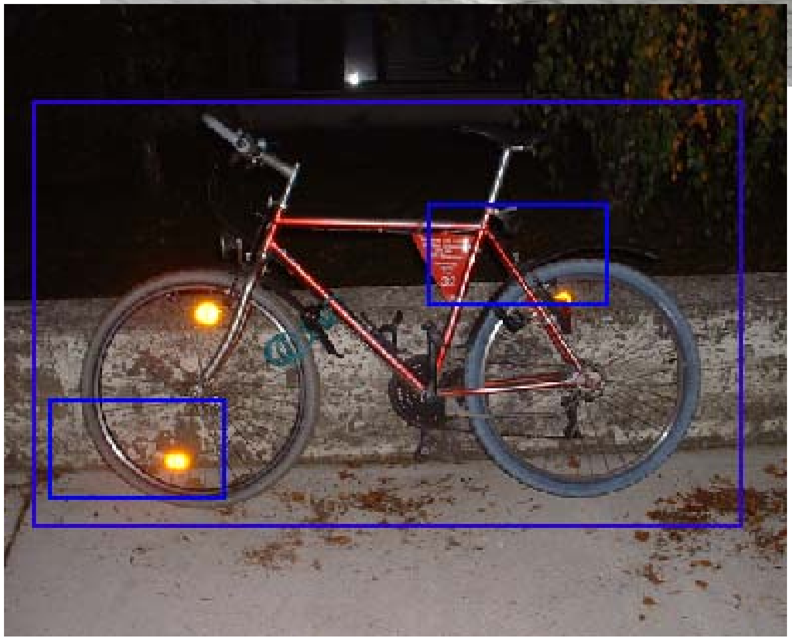
HoG-Motorbikes



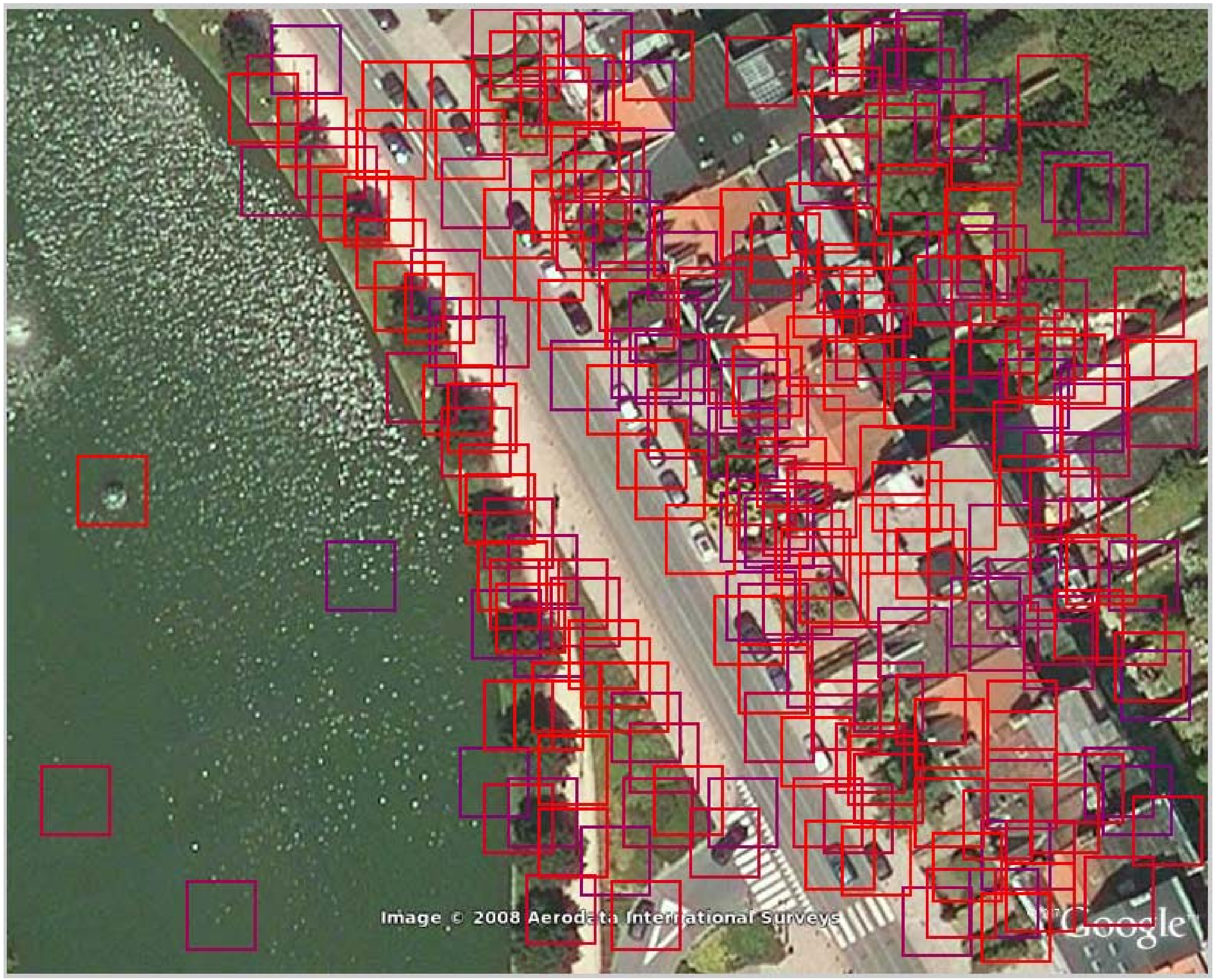
HoG-Bicycles





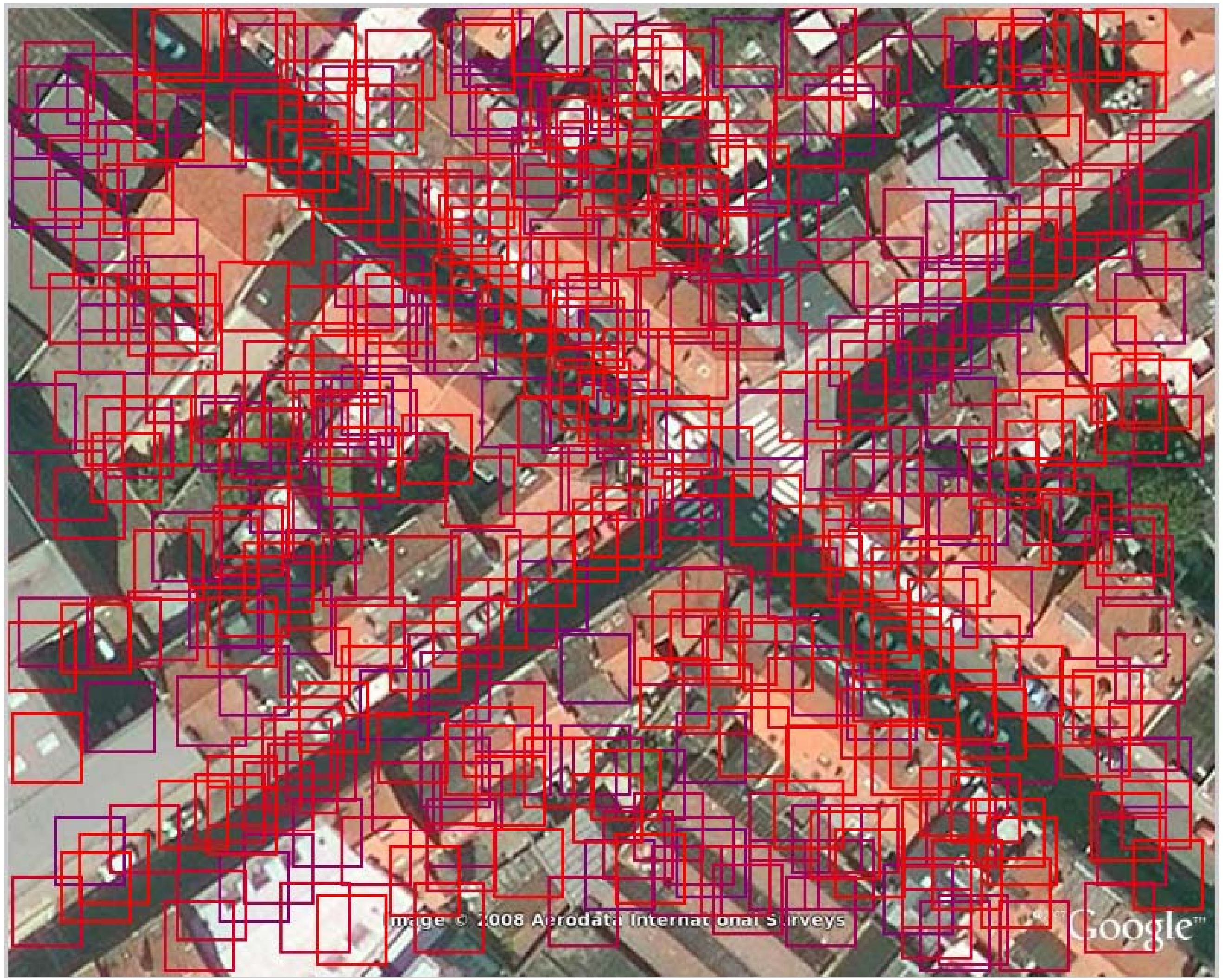


Satellite



Satellite

- Th=0

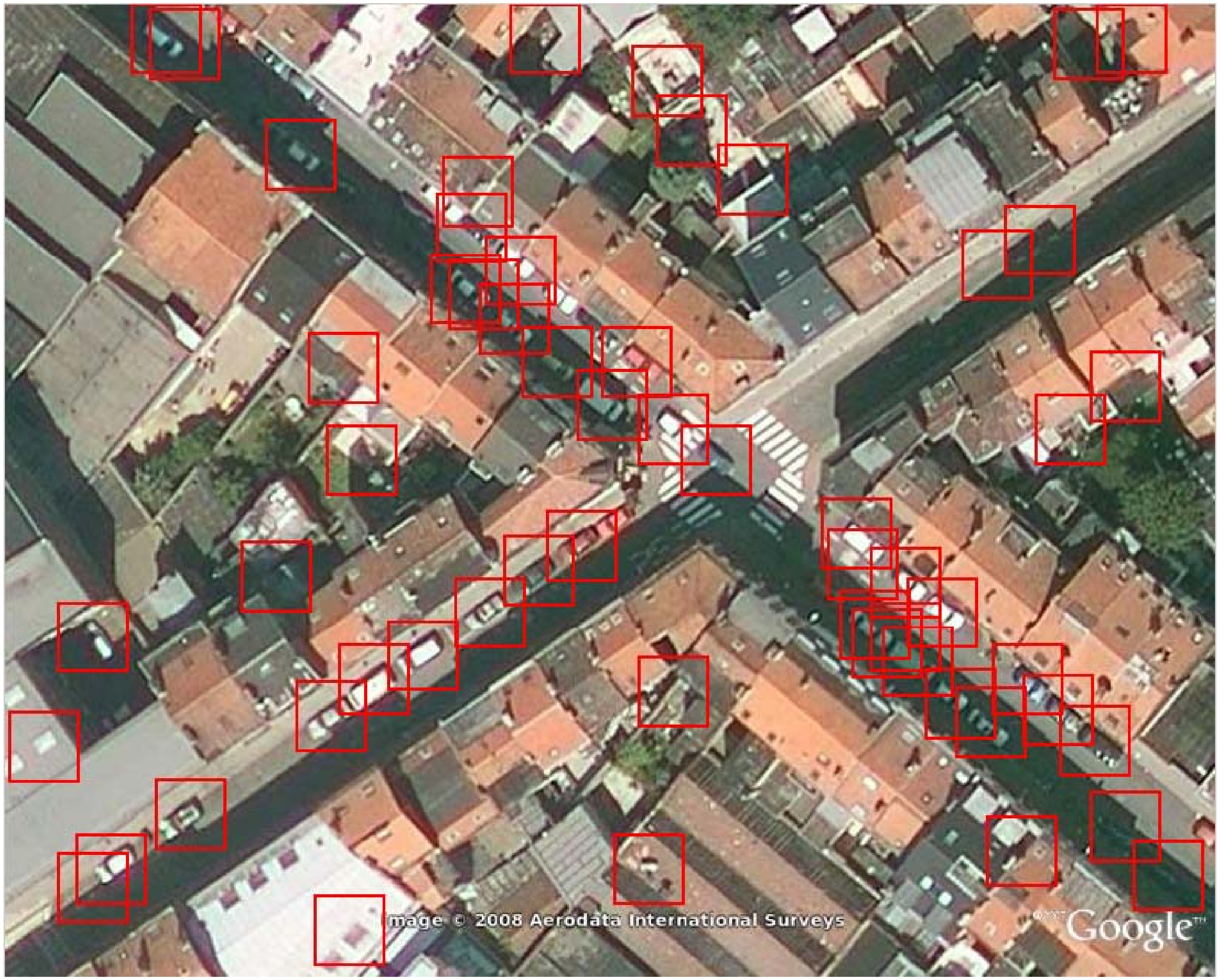


Satellite

- Th=0.95

Satellite

- Th = 0.99



Satellite

- Th=0.995

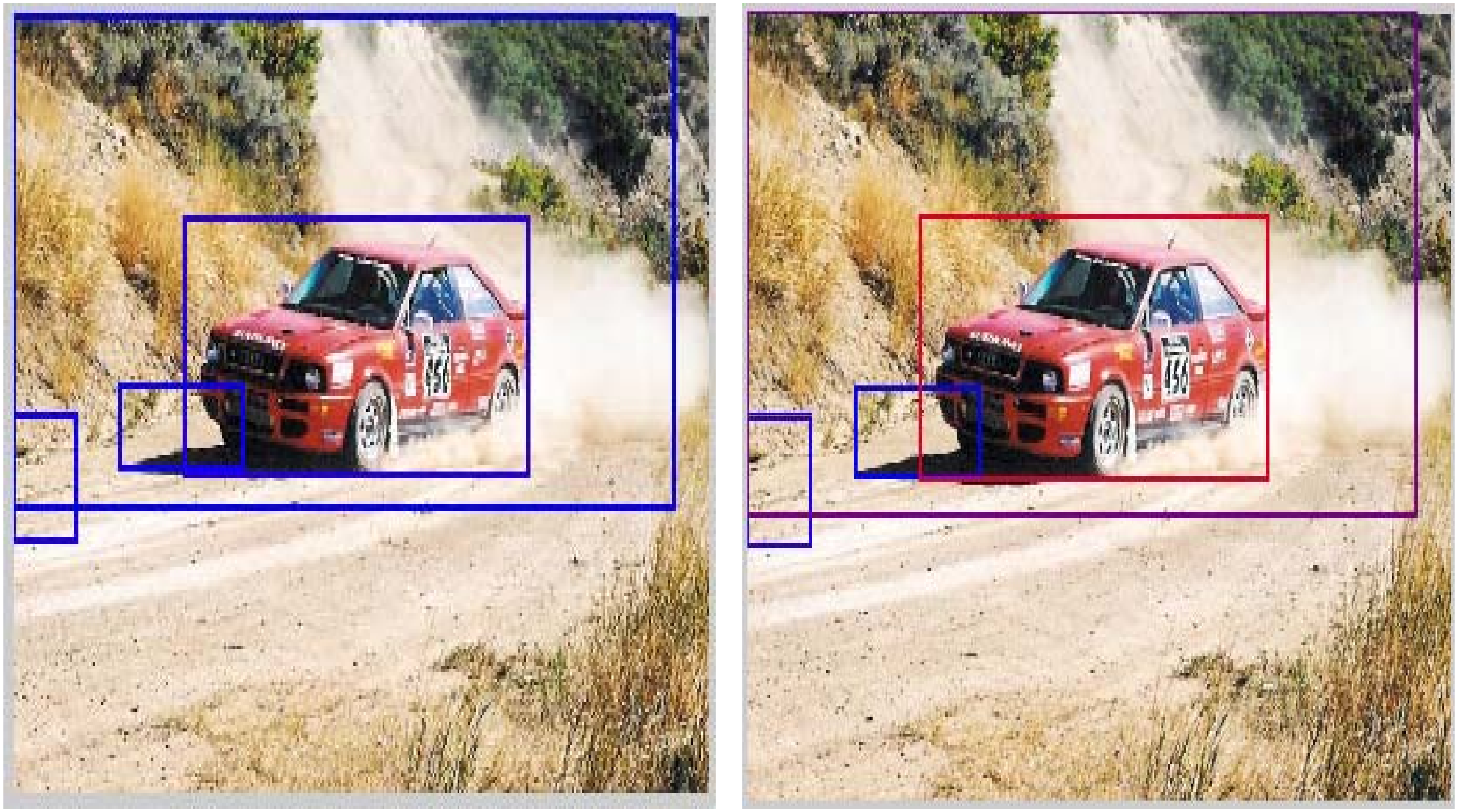
Outline

- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

Running TAS

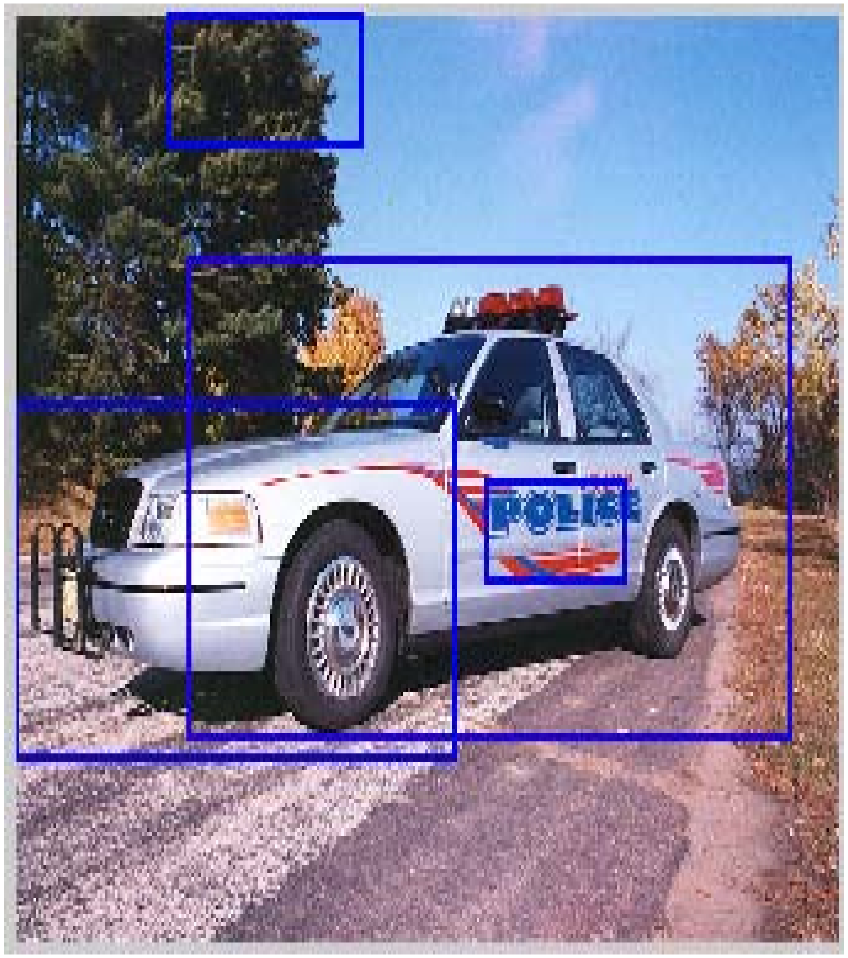
- Run TAS inference on all detected candidates
- False positives detected by the base detector will be filtered out
- Object not detected by the base detector could not be detected by TAS
- Data set: VOC2005, Google earth satellite images

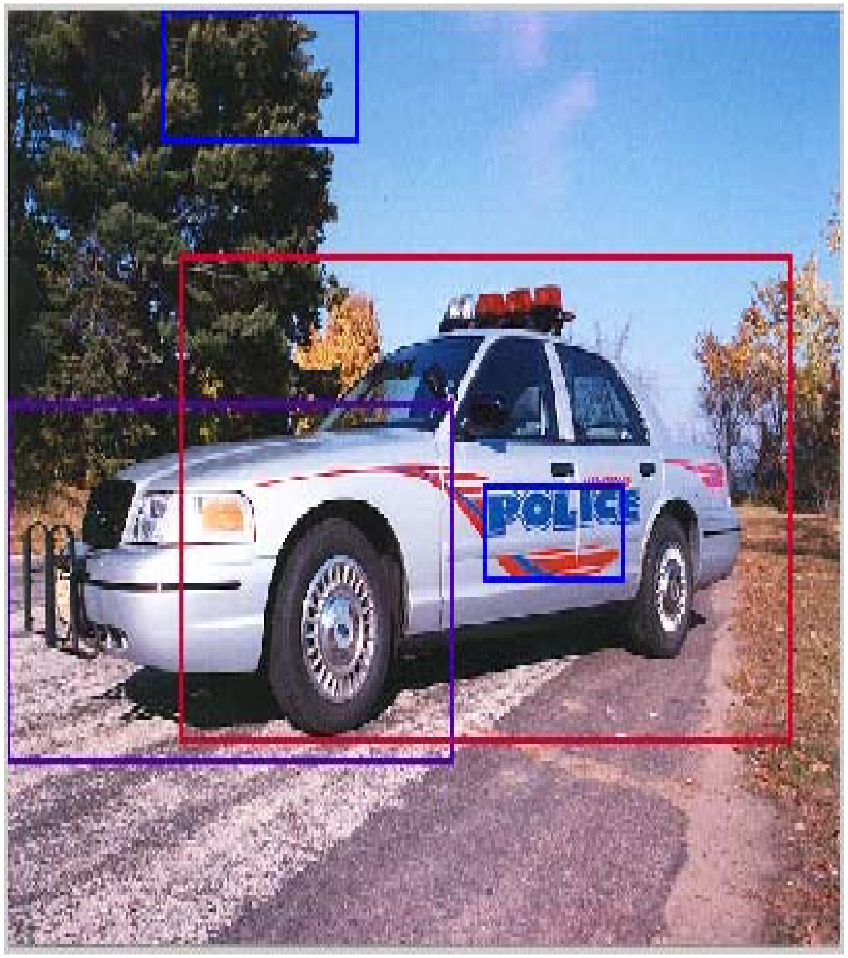
Base Detector vs TAS



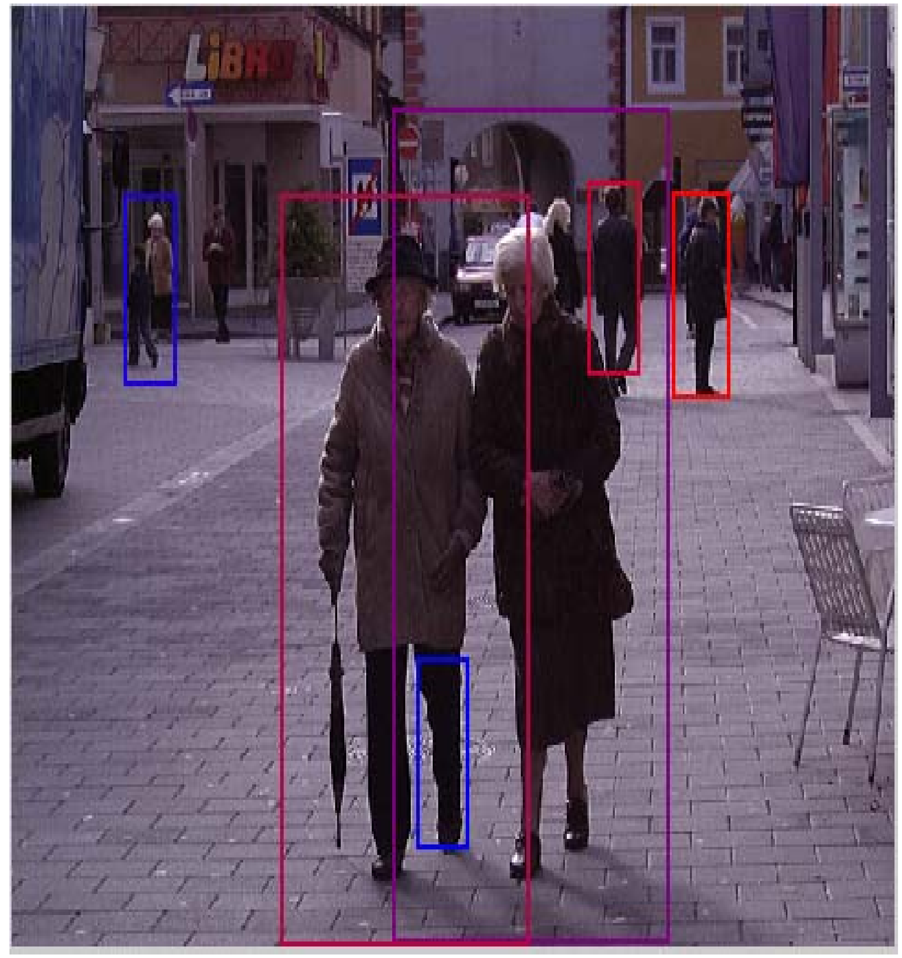
Left: base detector result. Right: TAS result

Base Detector vs TAS

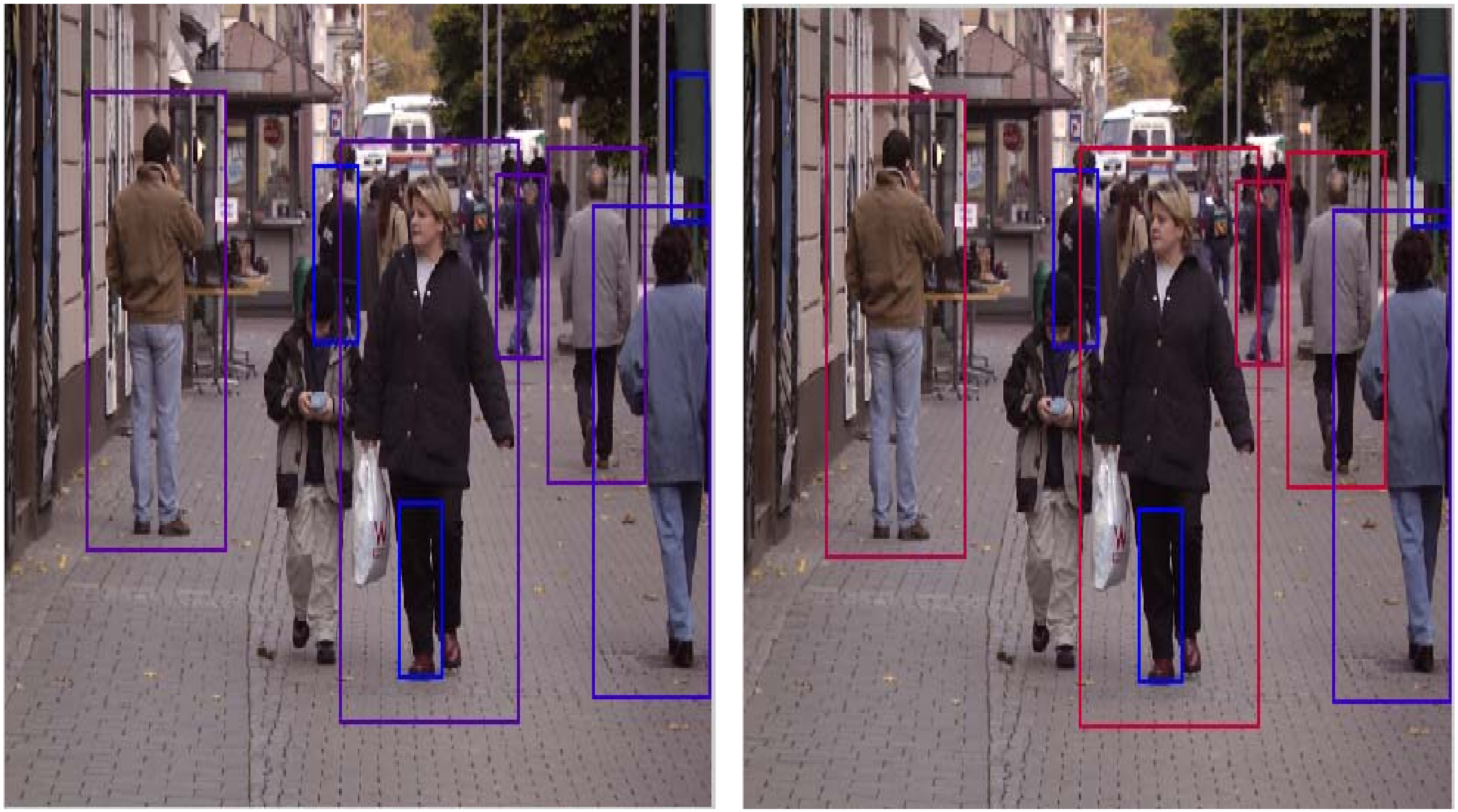




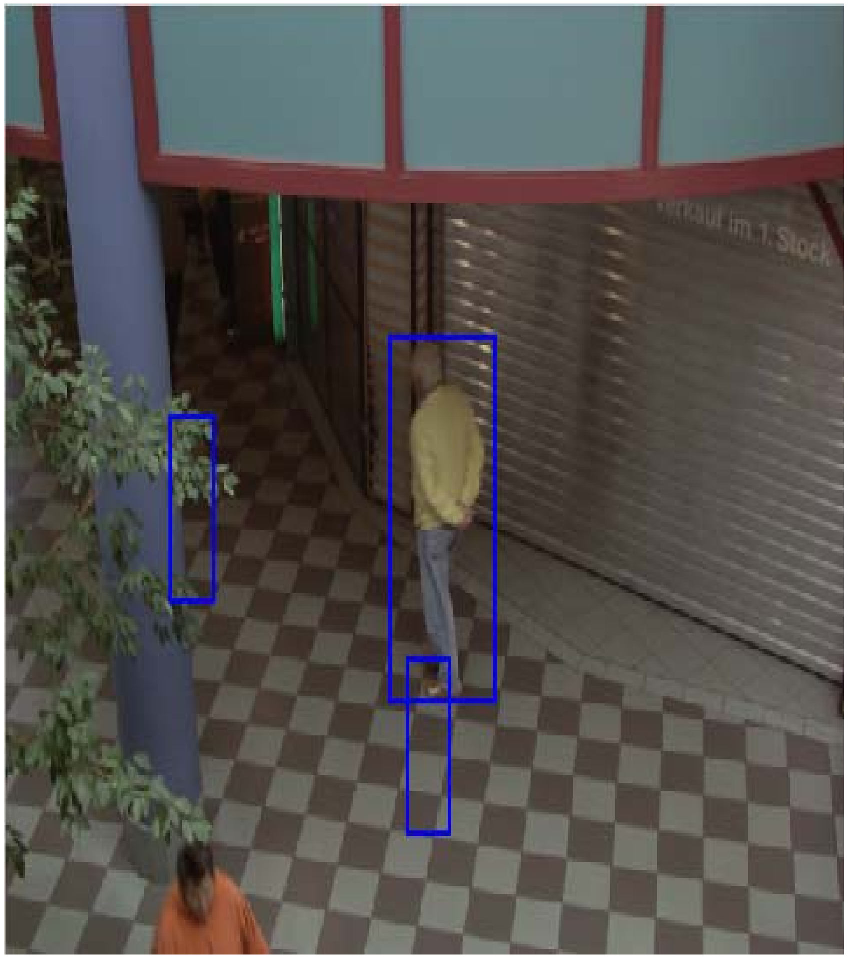
Base Detector vs TAS

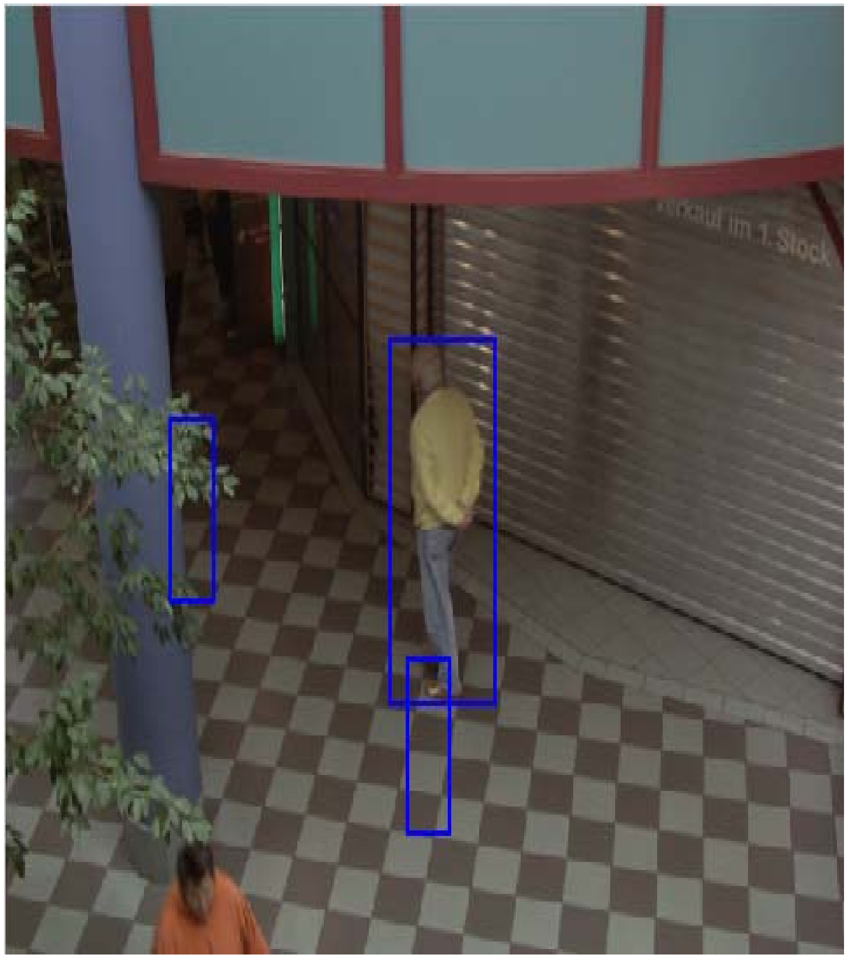


Base Detector vs TAS



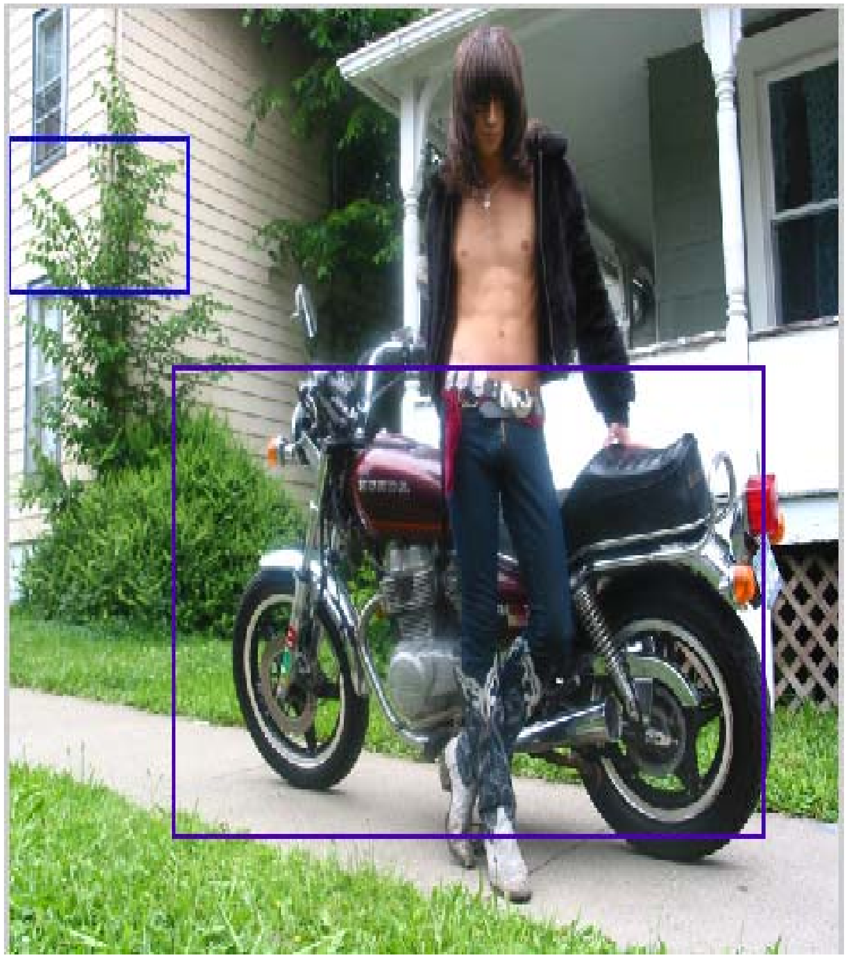
Base Detector vs TAS

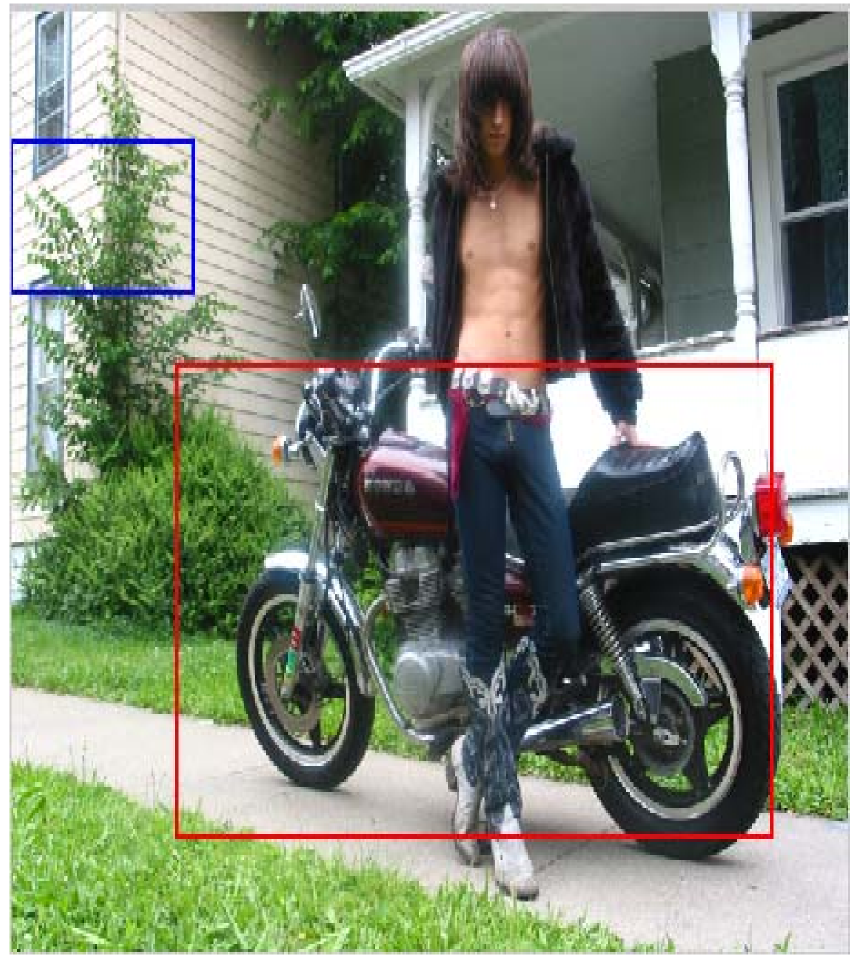




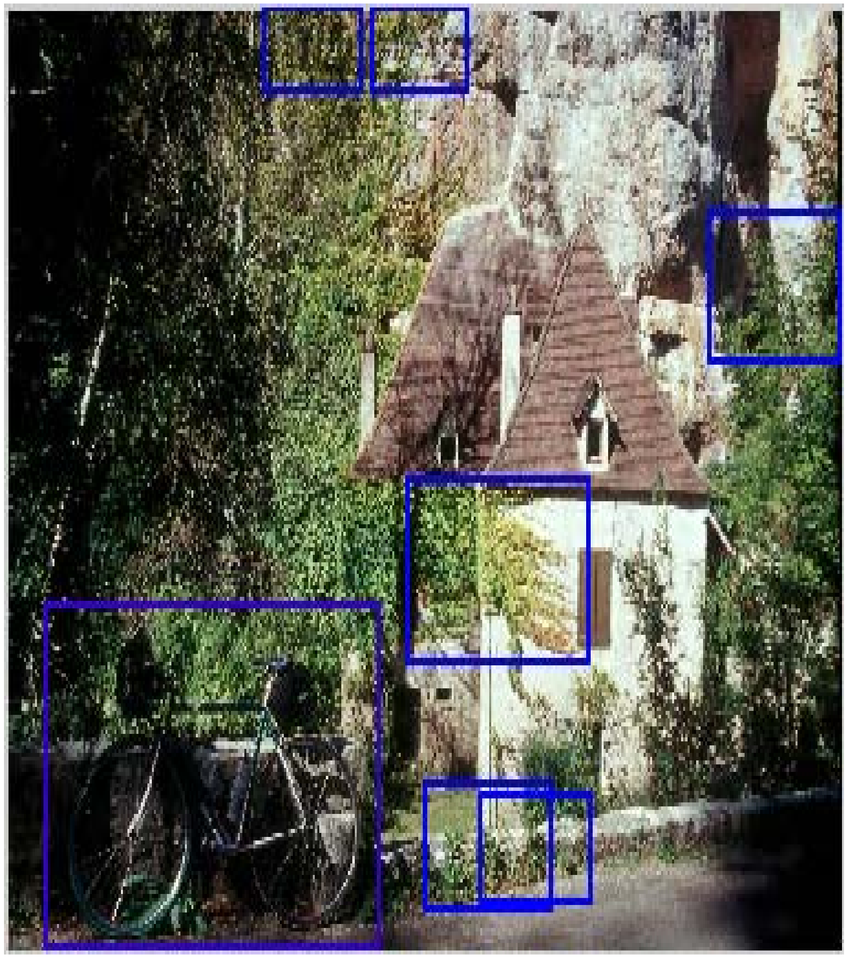
Base Detector vs TAS

Base Detector vs TAS

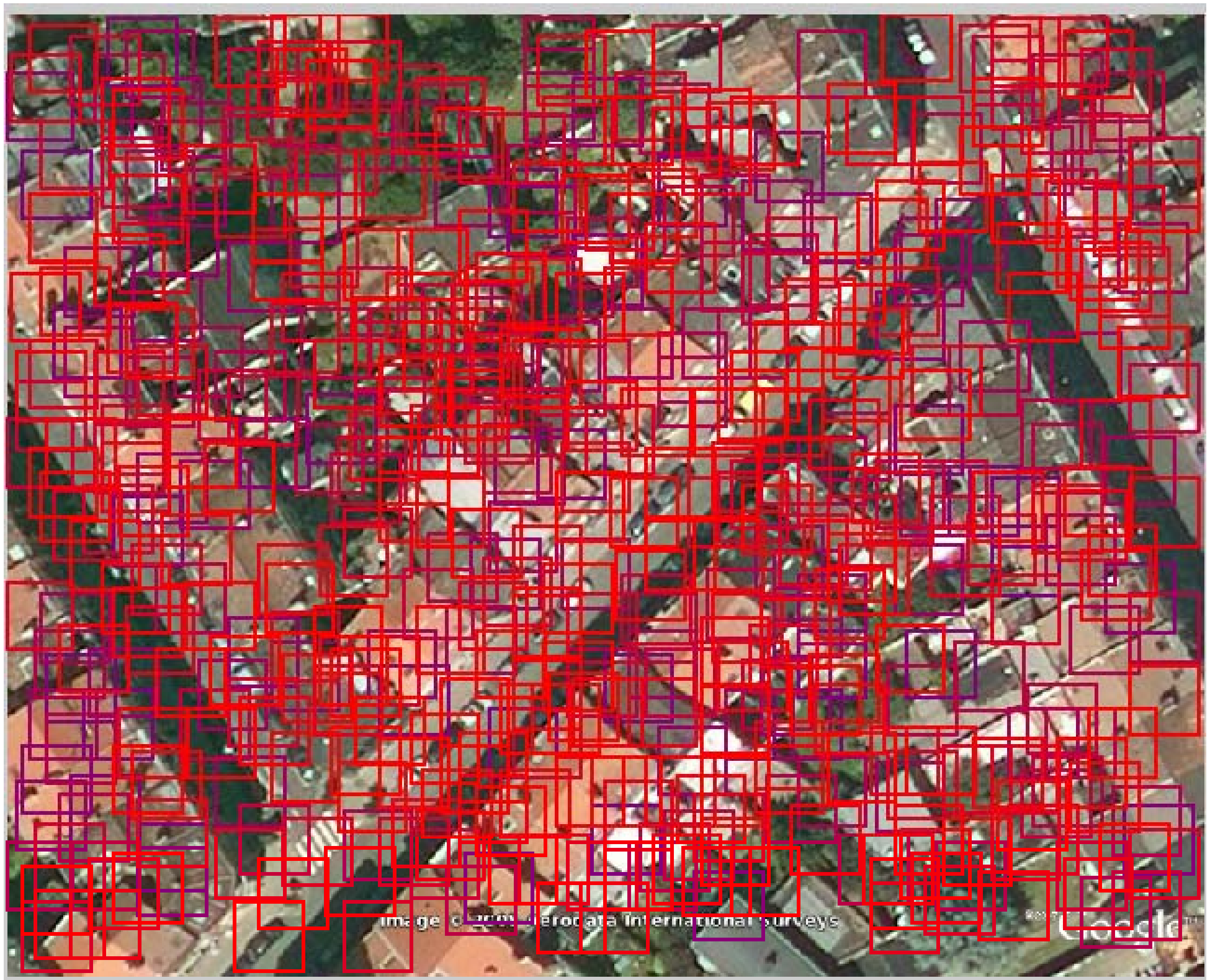




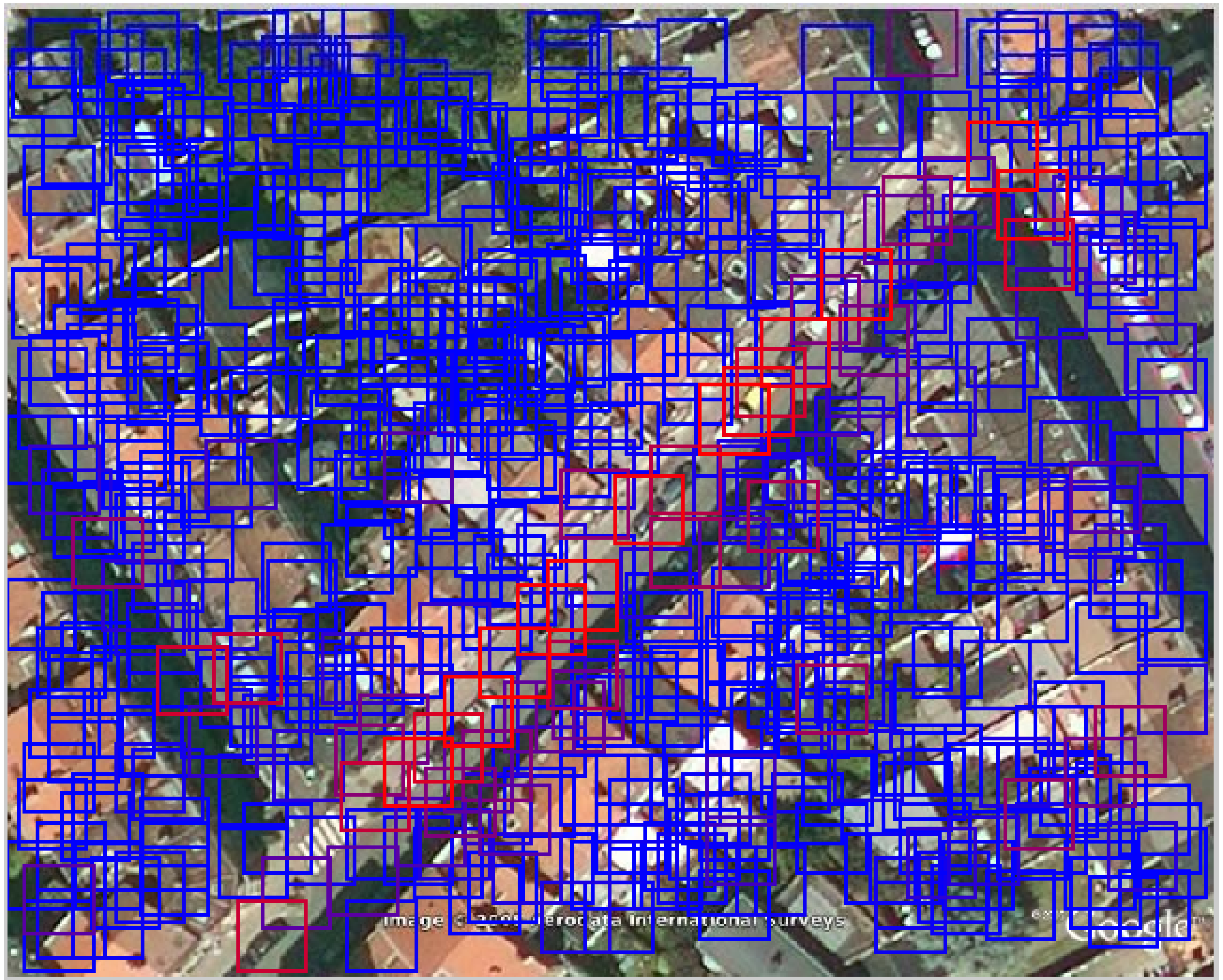
Base Detector vs TAS



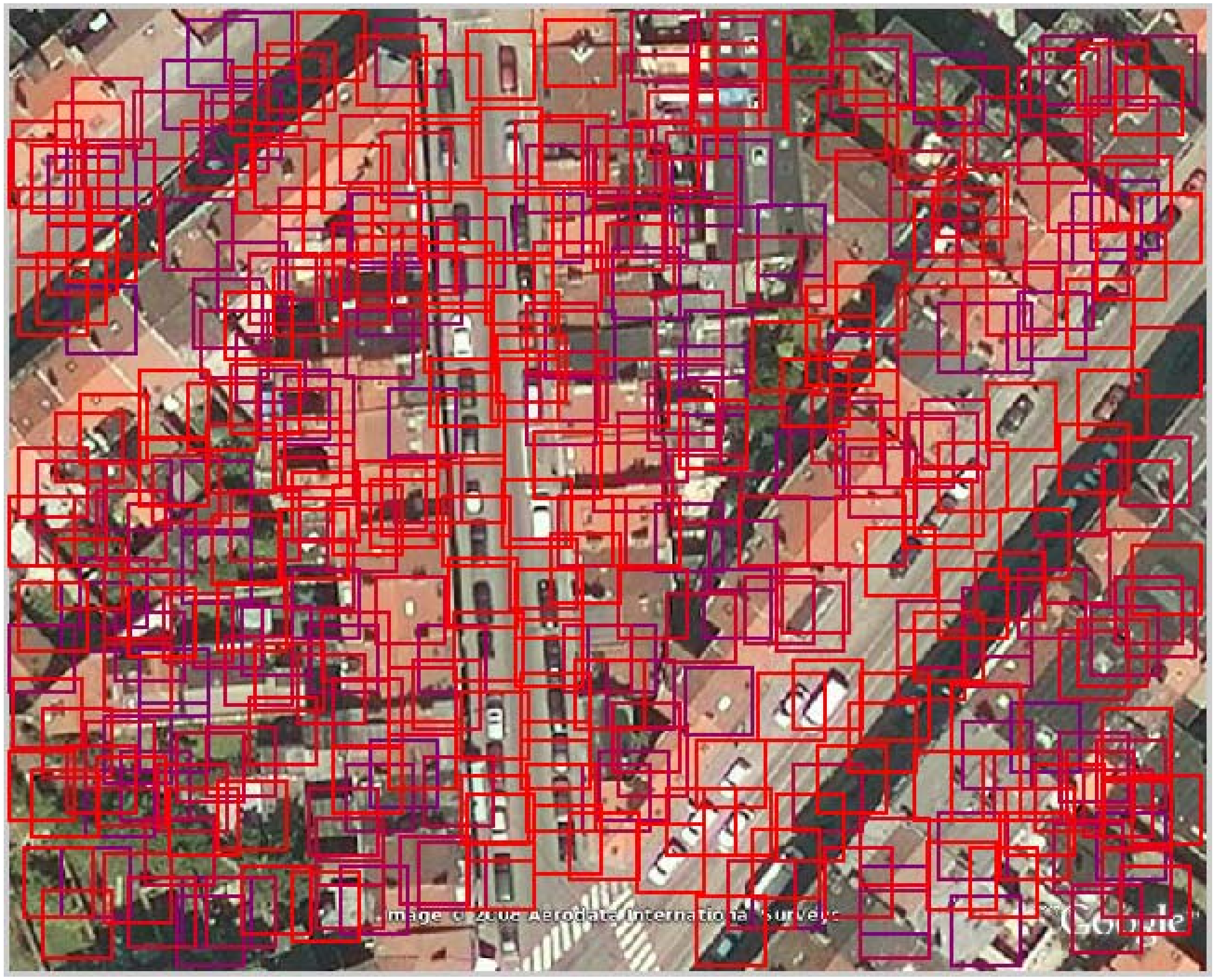

Base Detector



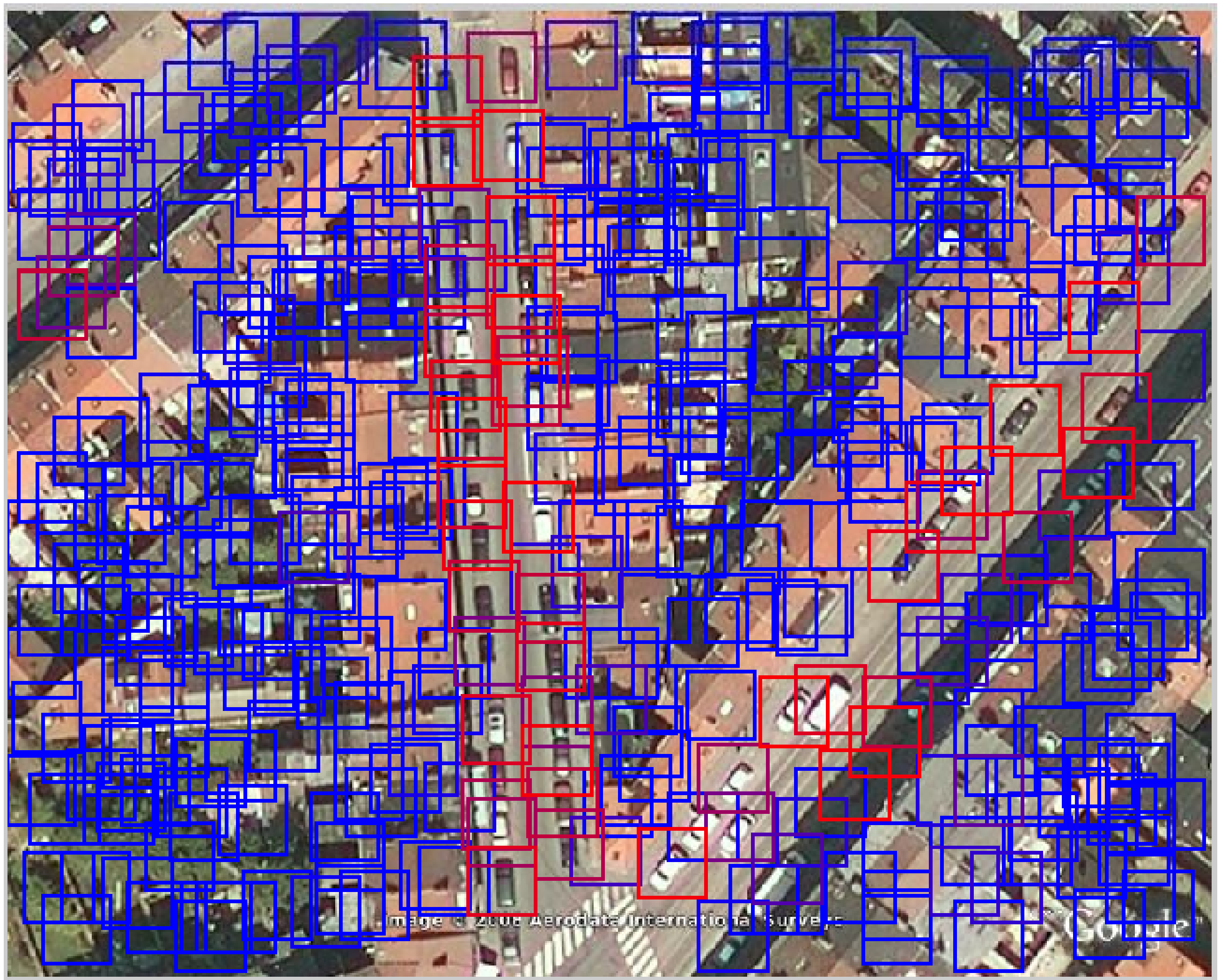
TAS



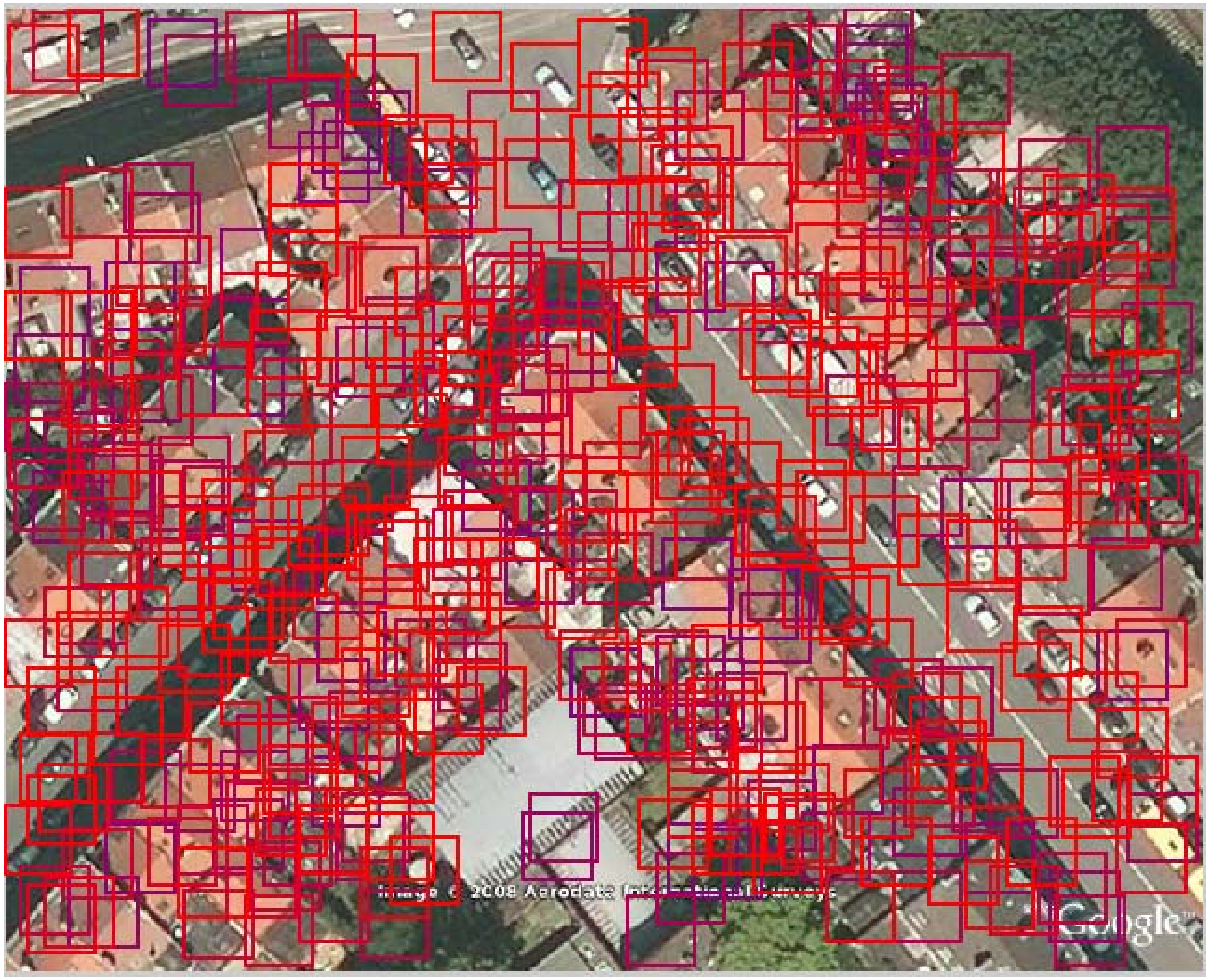
Base Detector



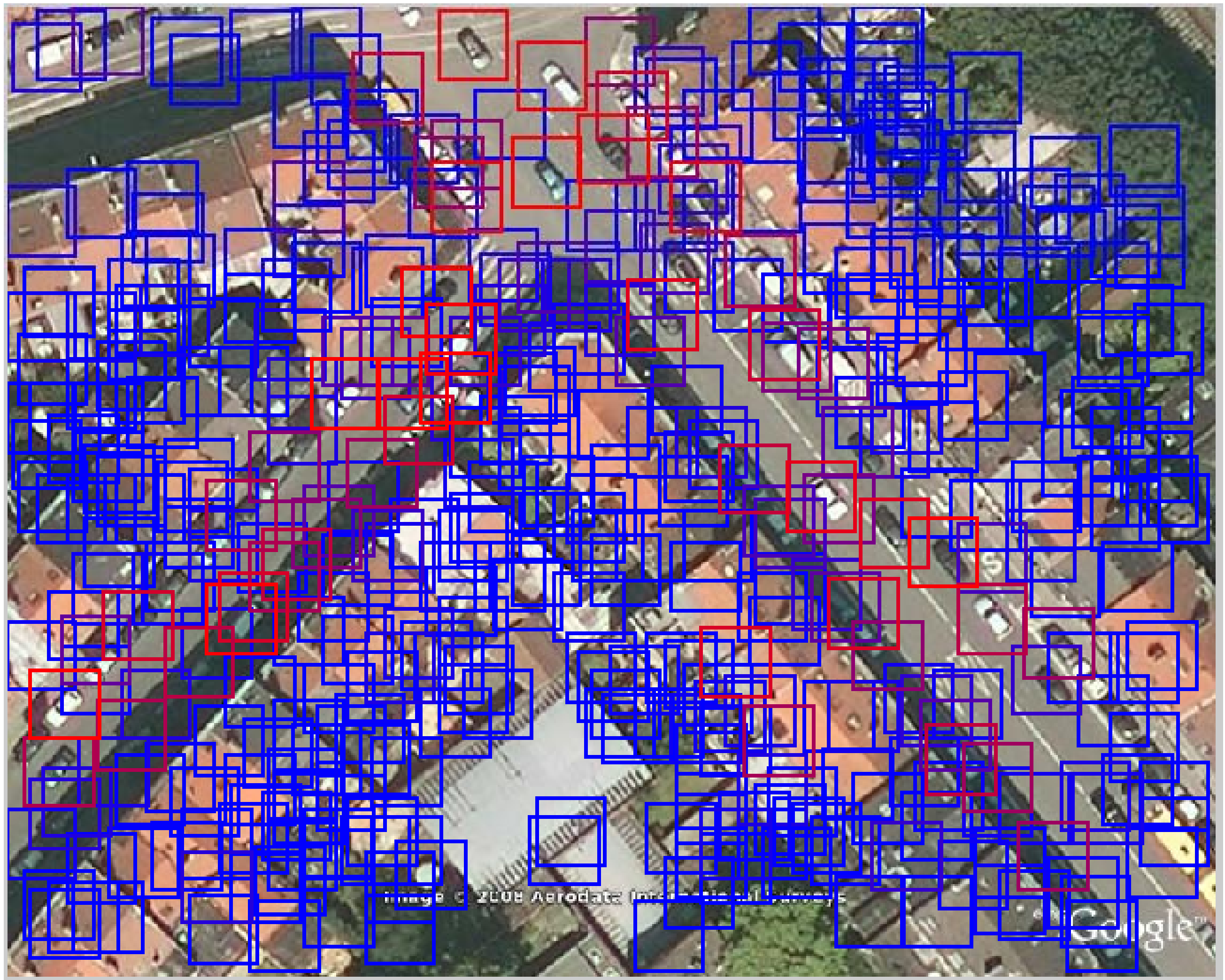
TAS



Base Detector



TAS



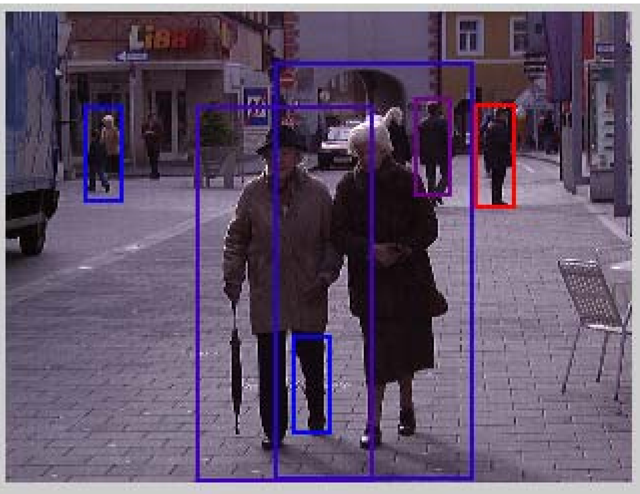
Outline

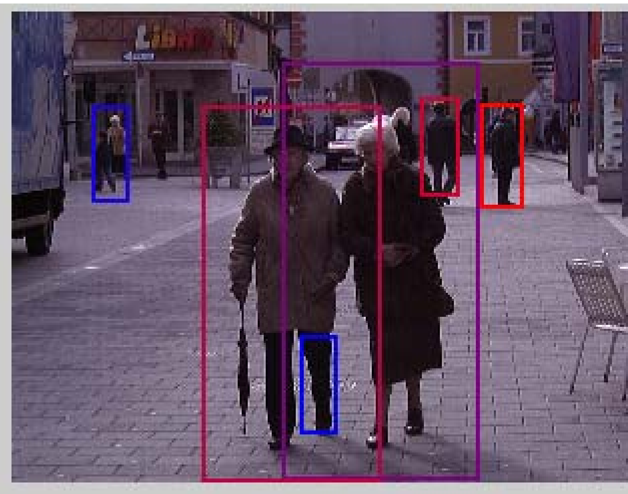
- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

Things-and-Stuff Relationships

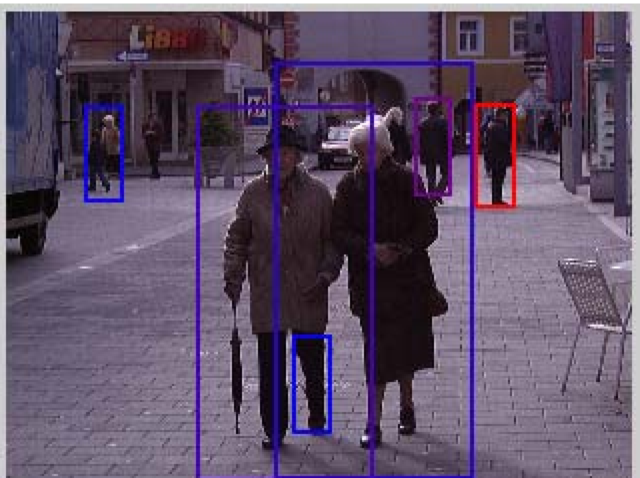
- Feature description: 44 features, including color, texture, shape
- The relationships are learnt during training
- The relationships change the score of a candidate
- 25 relationship candidates

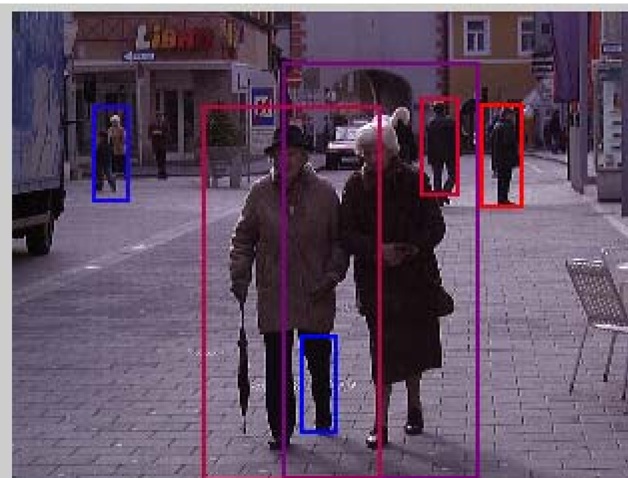
Relationships



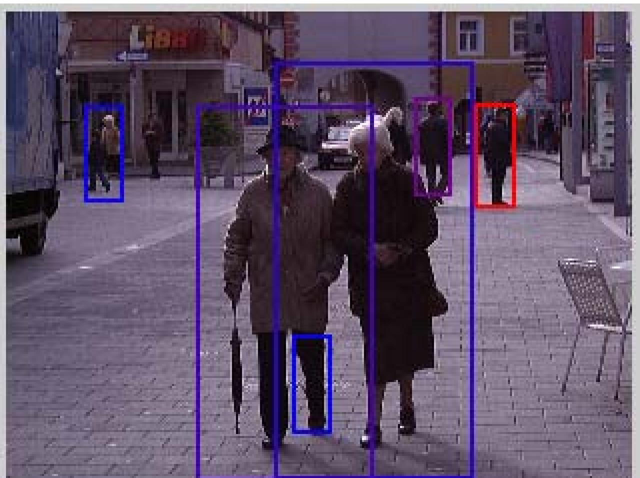


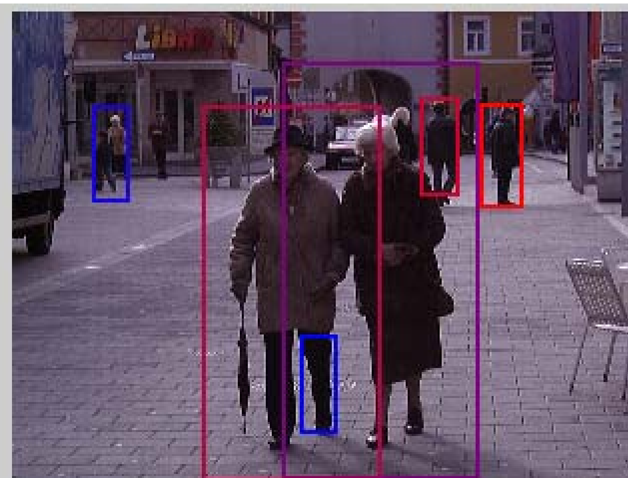
Relationships



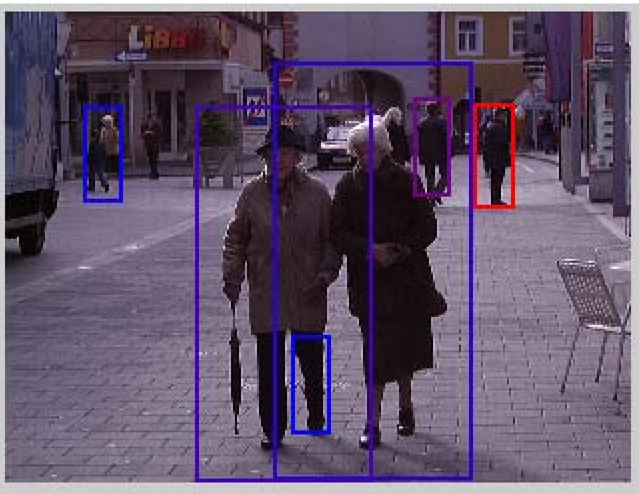


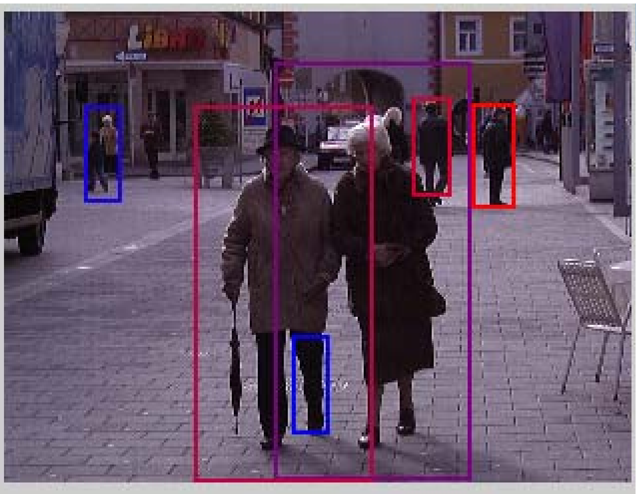
Relationships



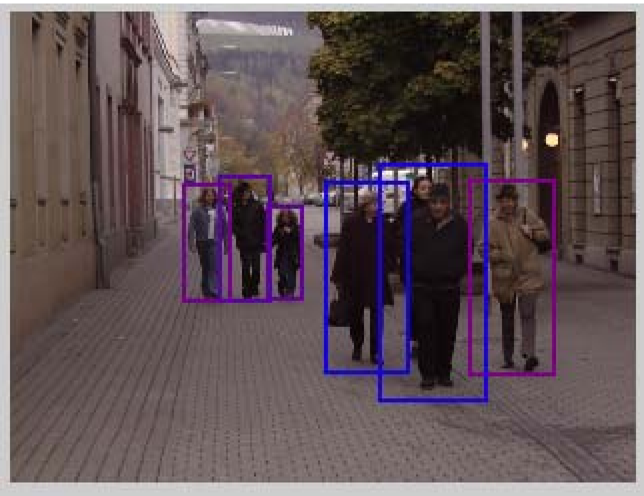


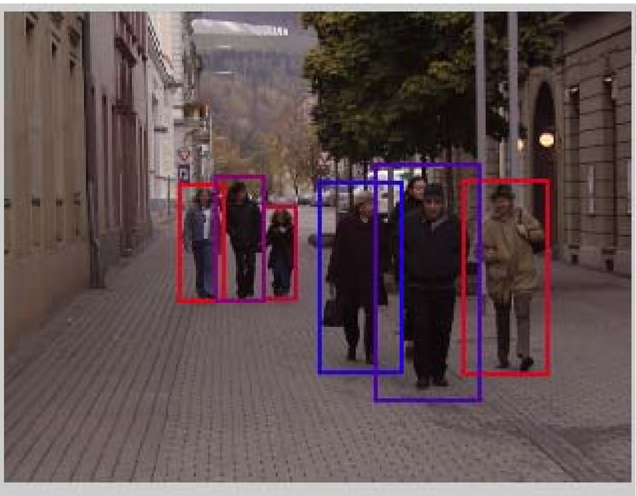
Relationships





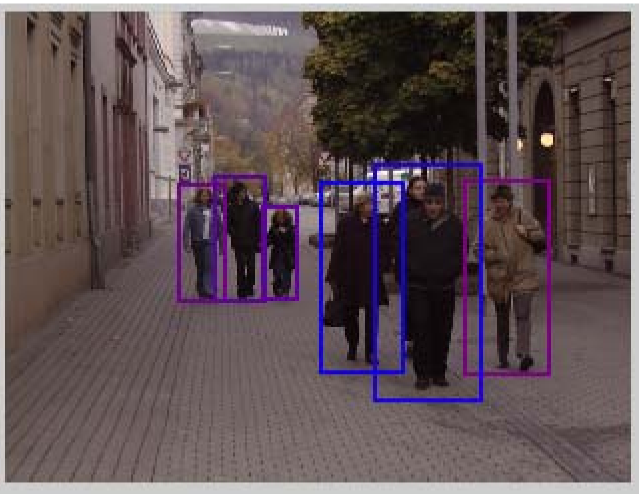
Relationships

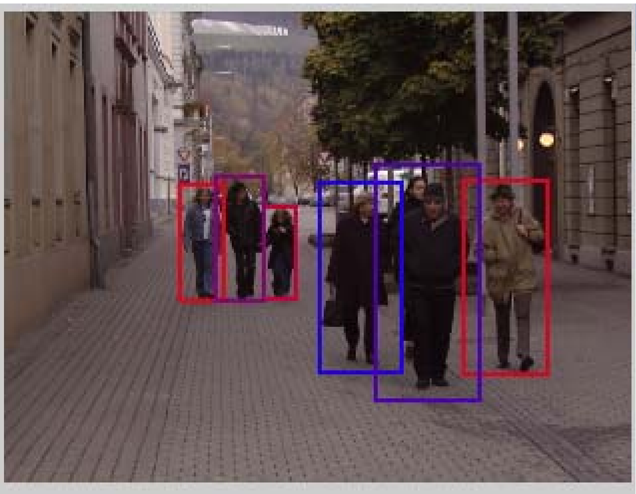




Relationships

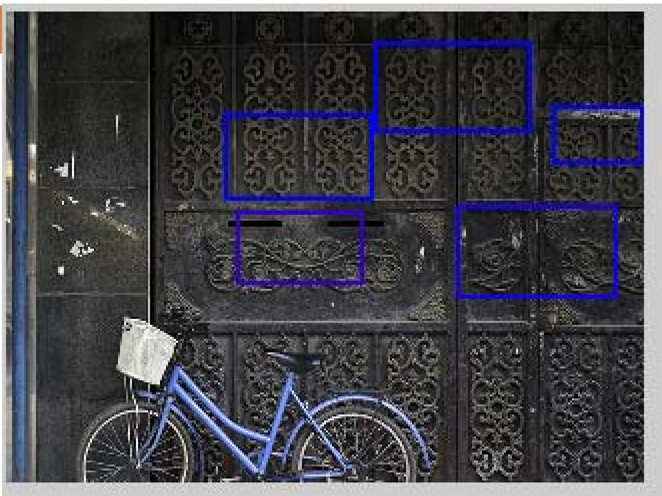
Relationships

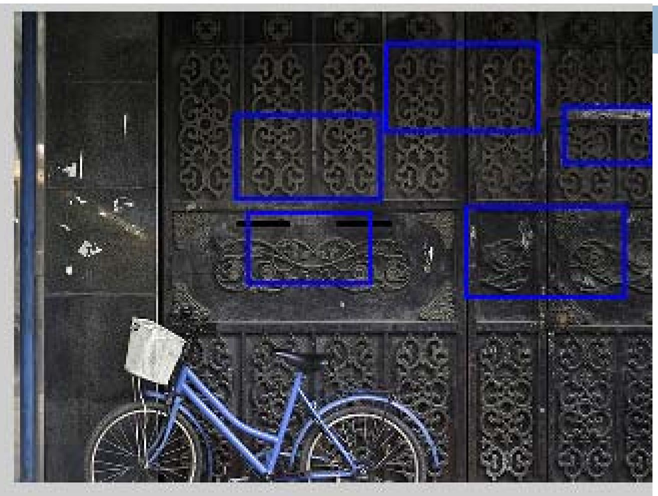




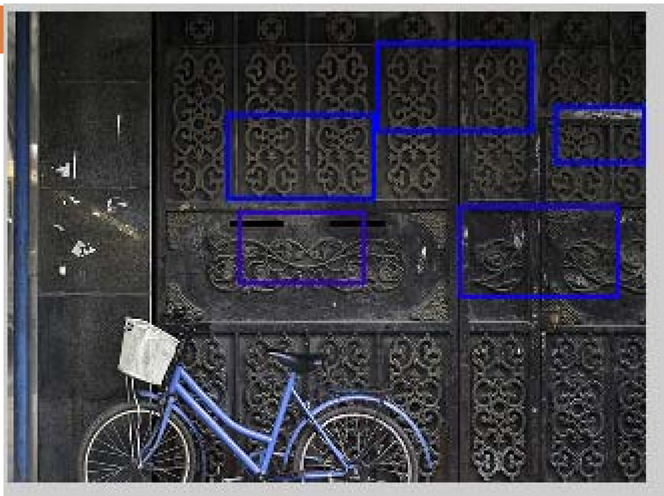
Relationships

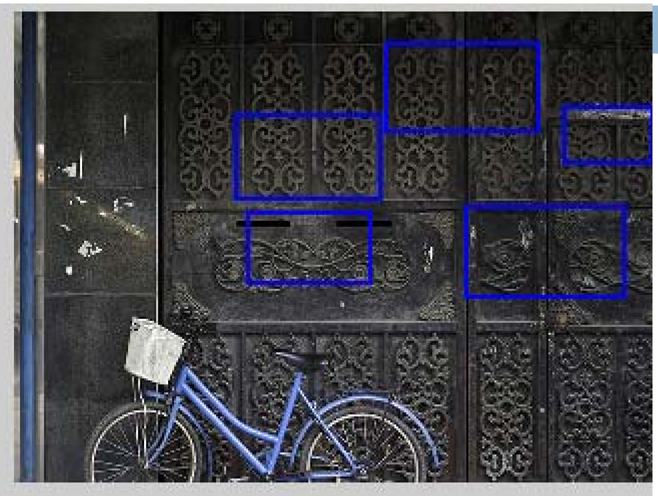
Relationships



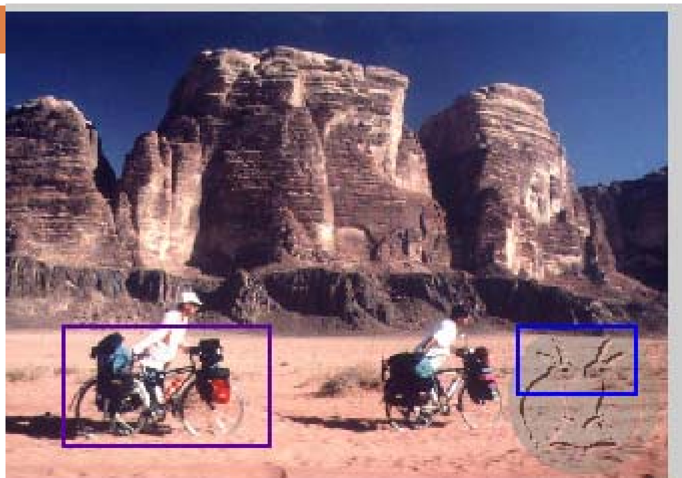


Relationships

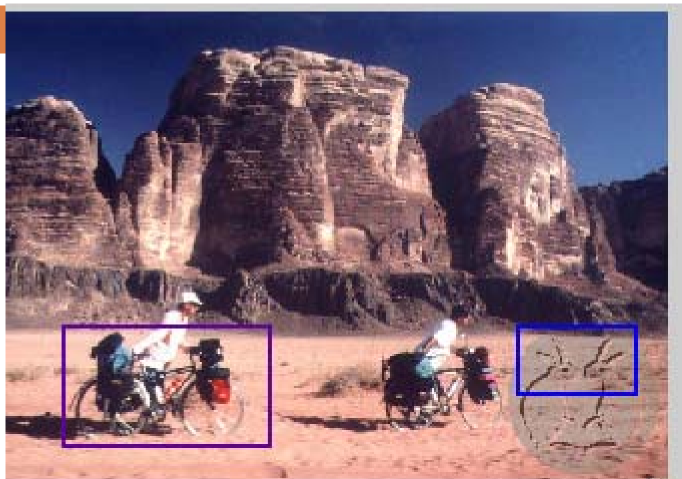




Relationships

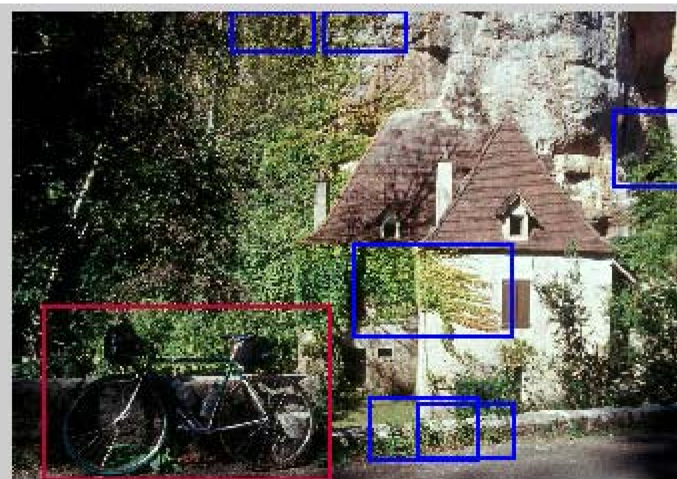


Relationships



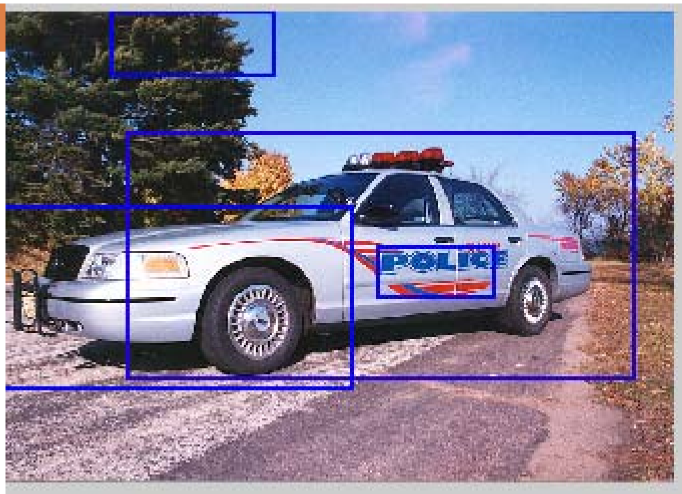
Relationships

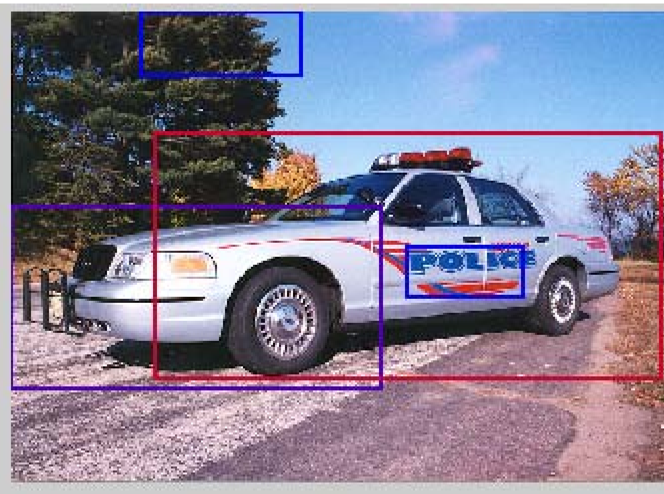
Relationships



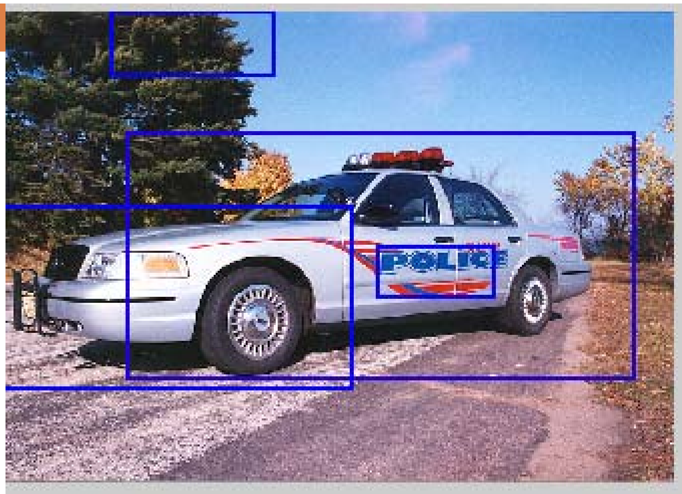
Relationships

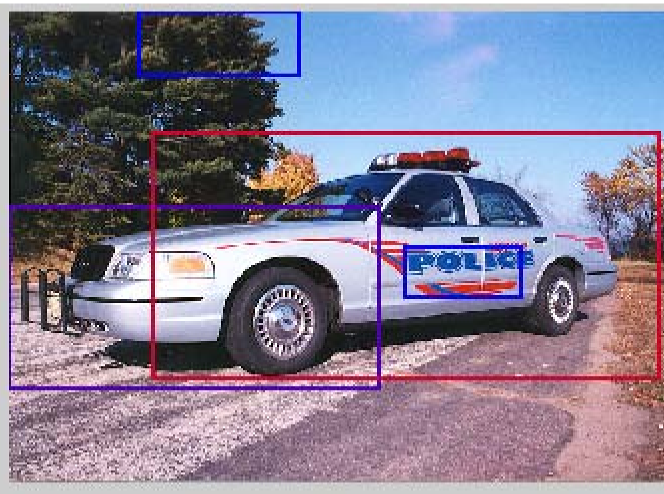
Relationships



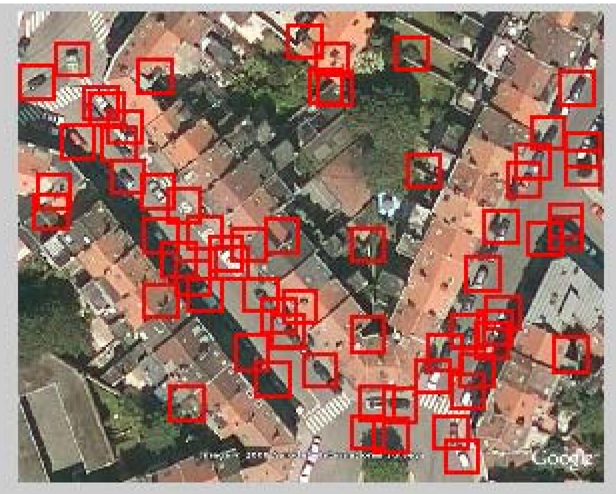


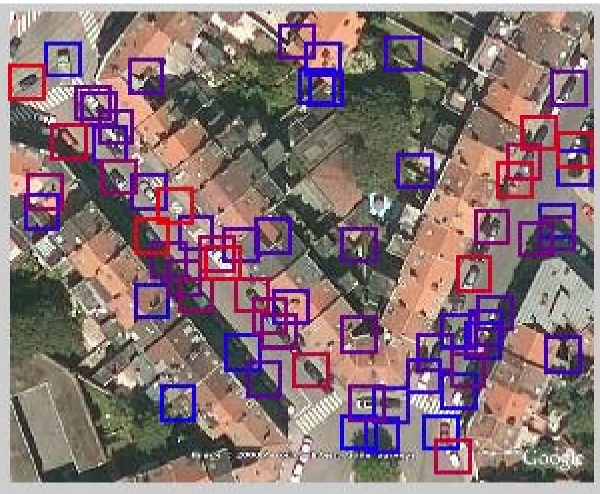
Relationships



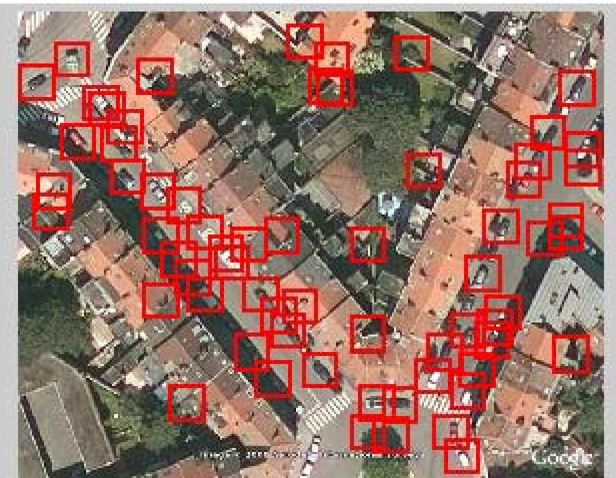


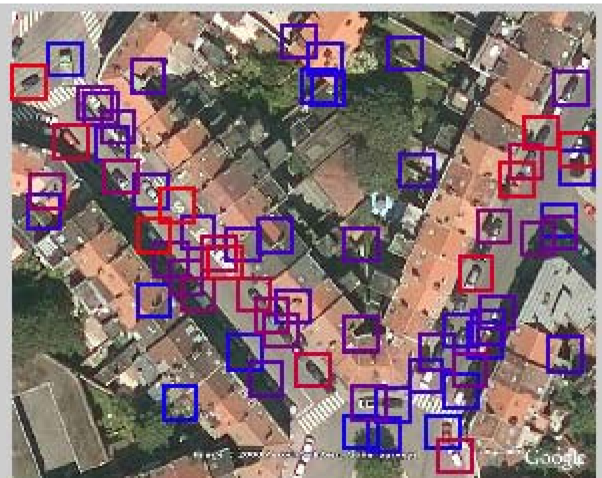
Relationships





Relationships

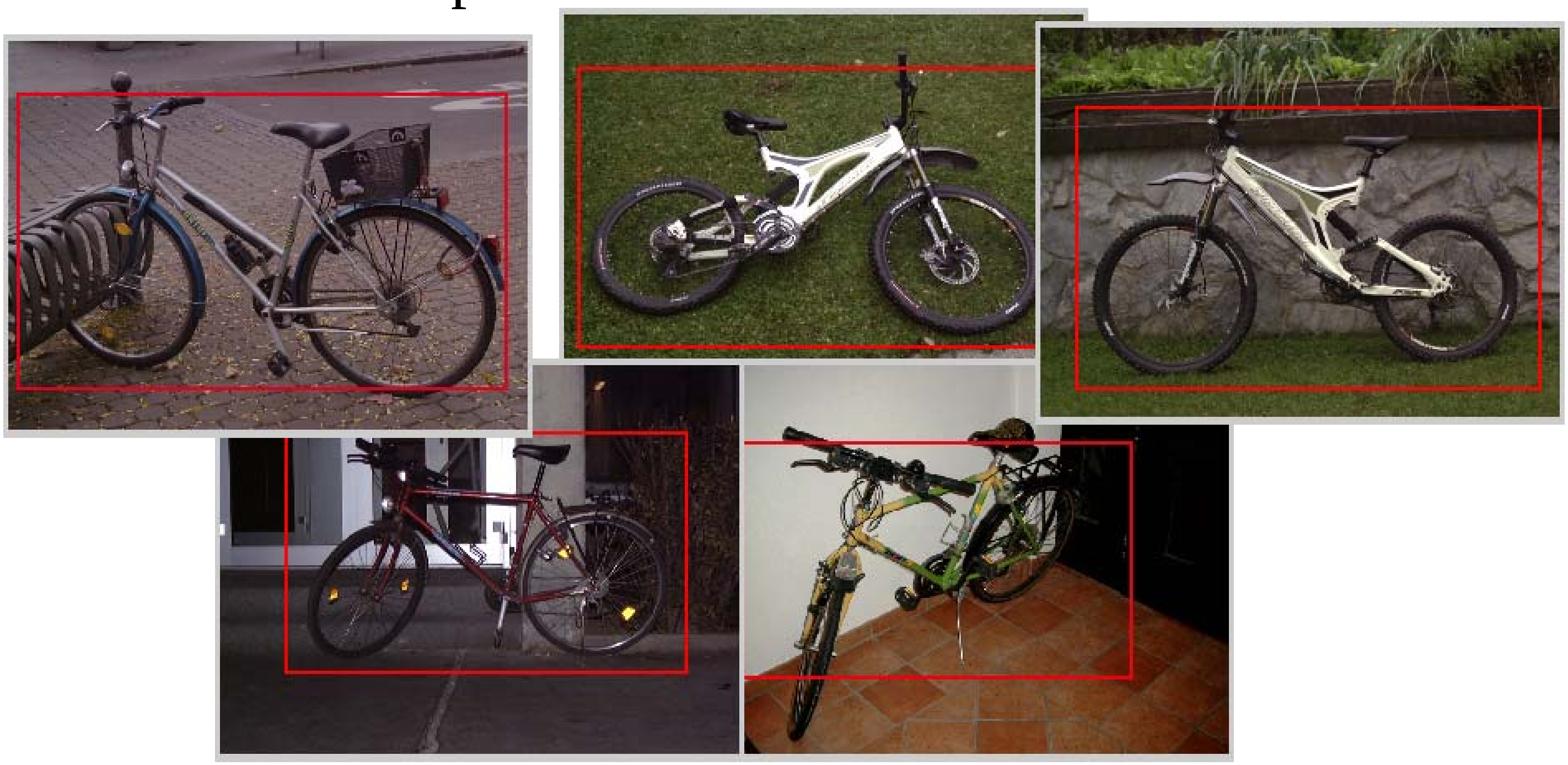




Relationships

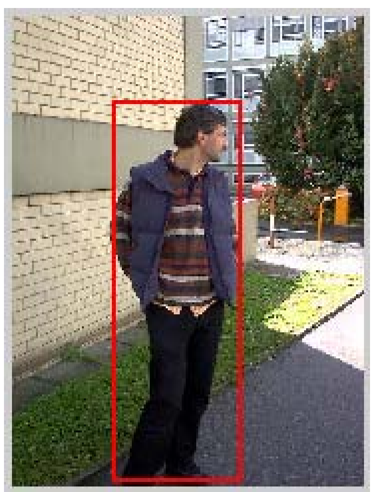
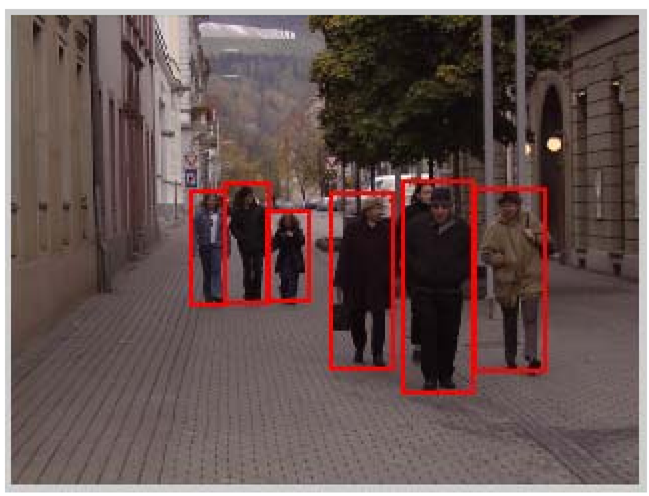
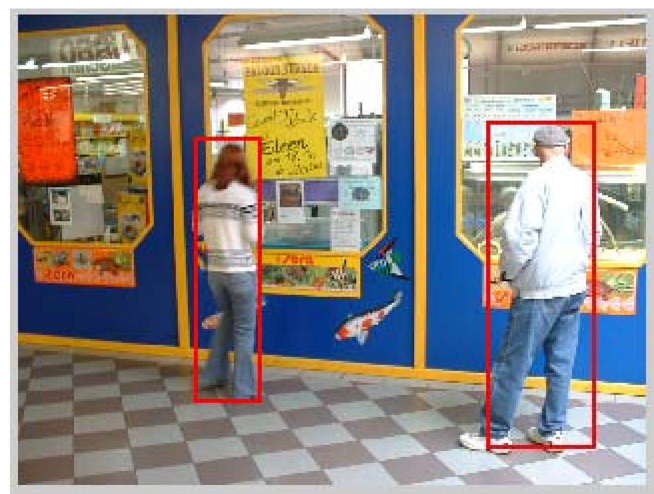
Relationships

- Some regions inside the bounding box have relationships with the candidate



Relationships

- View point.
  - Different viewpoints generate different relationships
- Region features might be misleading

Relationships

- The diversities of the backgrounds
- The region features inside the bounding box might be a complementary cue to the features used by the base detector



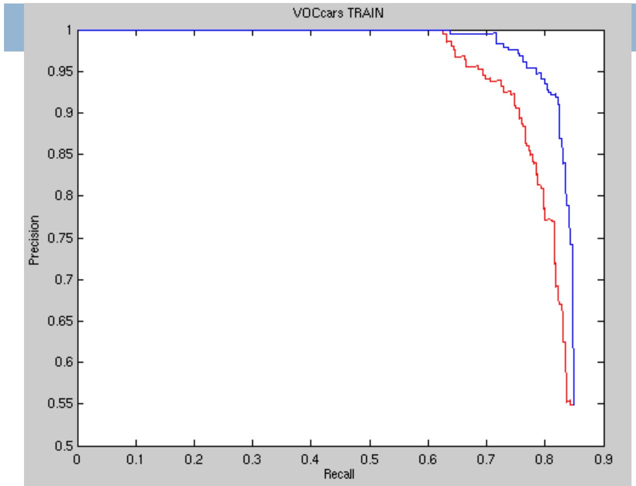

Outline

- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

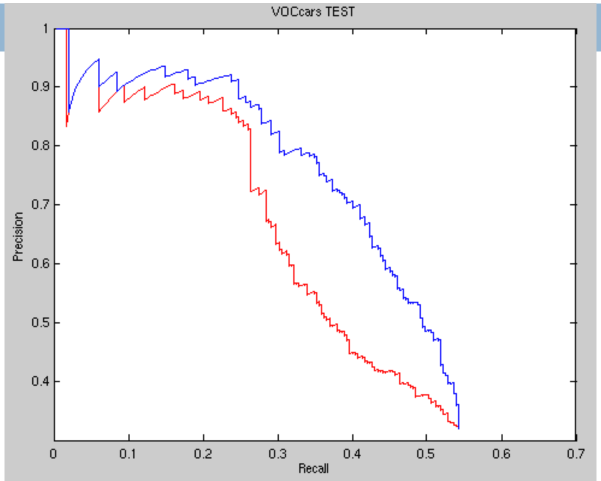
Performance Analysis

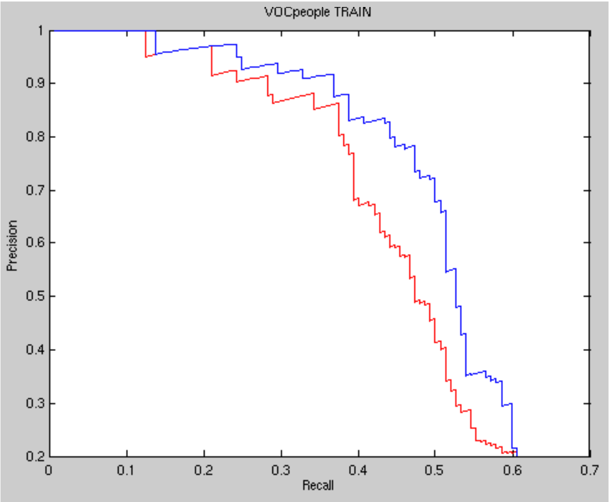
- Training samples: 15
- Test samples: 15
- Image size: 792x636
- Test machine: Core(TM)2 Quad@2.40GHz, 8G RAM
- Implemented in Matlab
- Detection and segmentation are not included
- Required computing power
  - Learning – 2141.67 seconds of CPU time
  - Inferring – 63.89 seconds of CPU time

Base Detector vs TAS

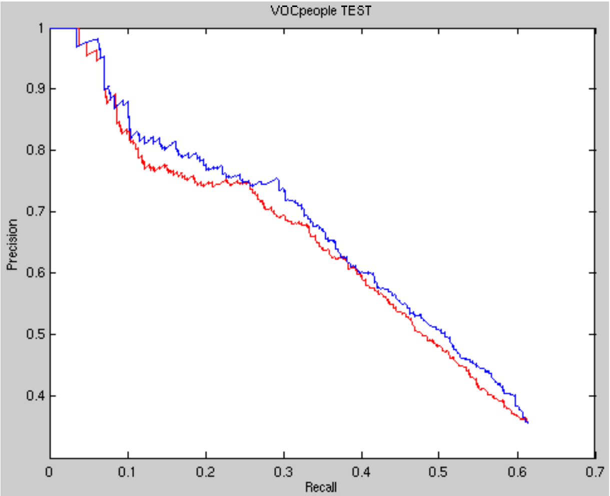


Cars



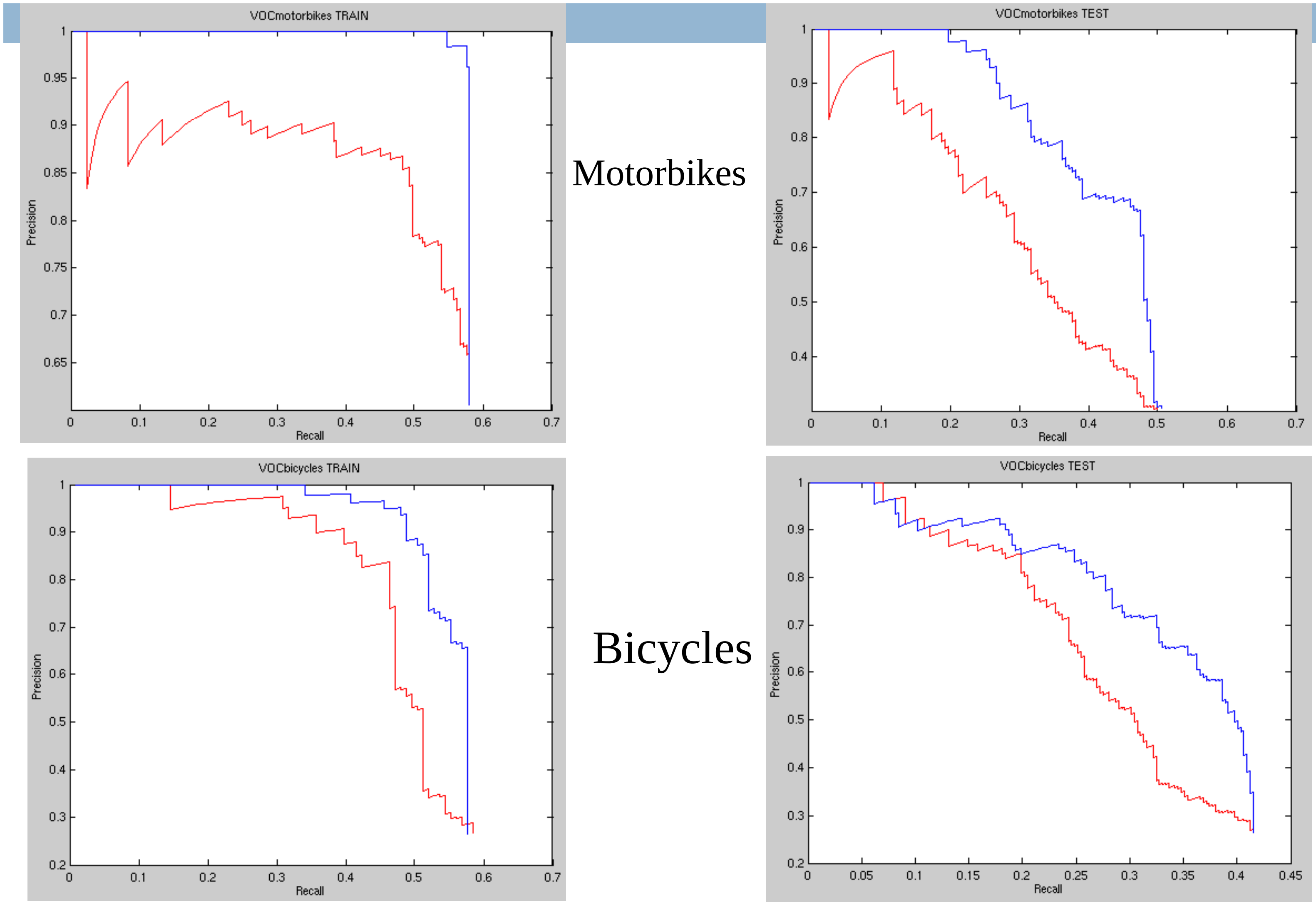


People



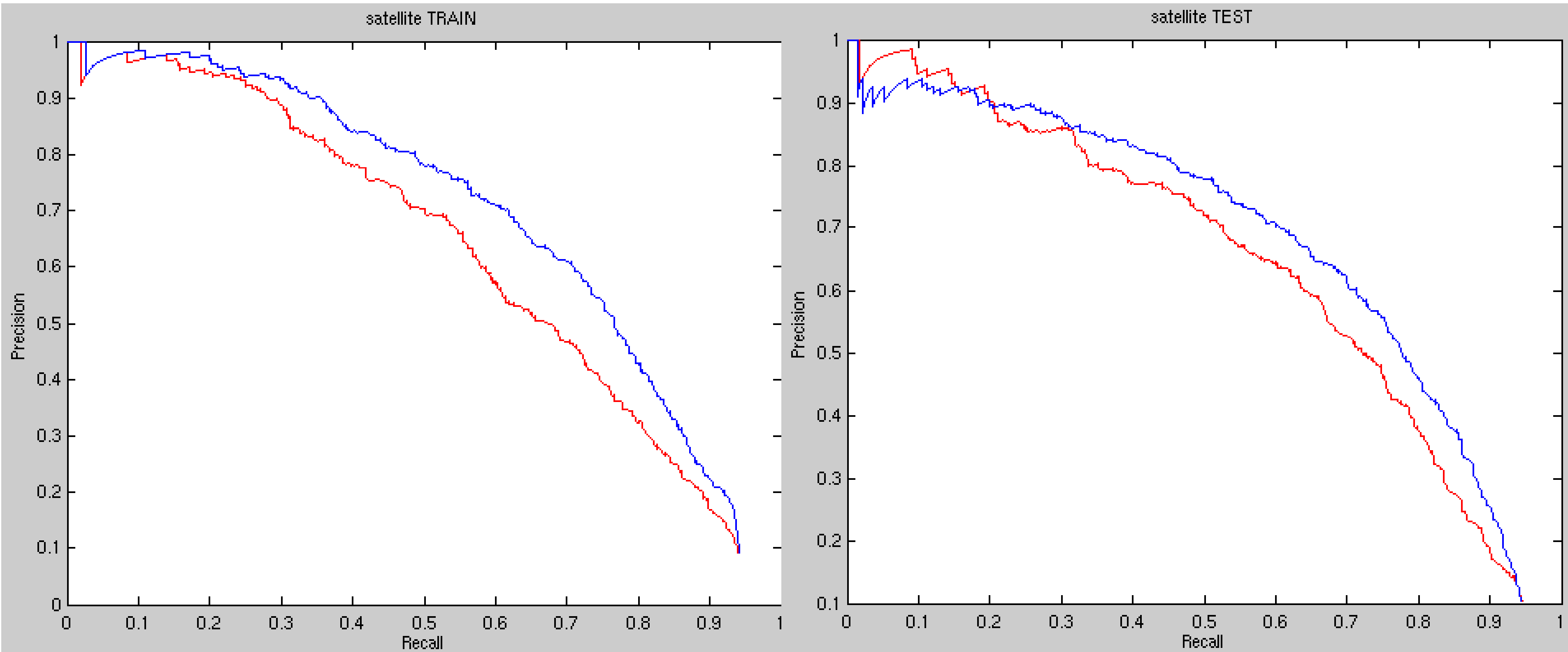
Red: base detector. Blue: TAS

Base Detector vs TAS - Motorbikes



Red: base detector. Blue: TAS

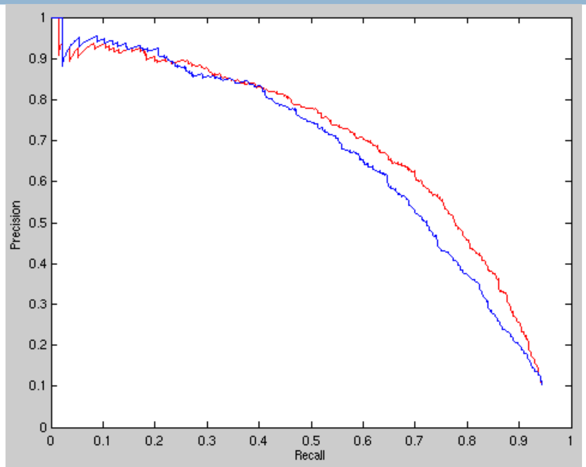
Base Detector vs TAS - Satellite



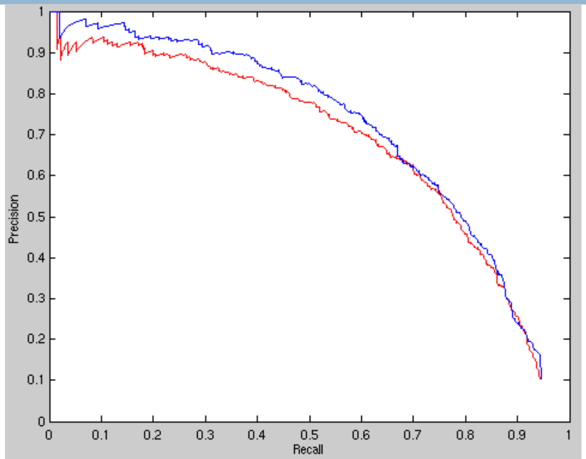
Outline

- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

Number of Region Clusters

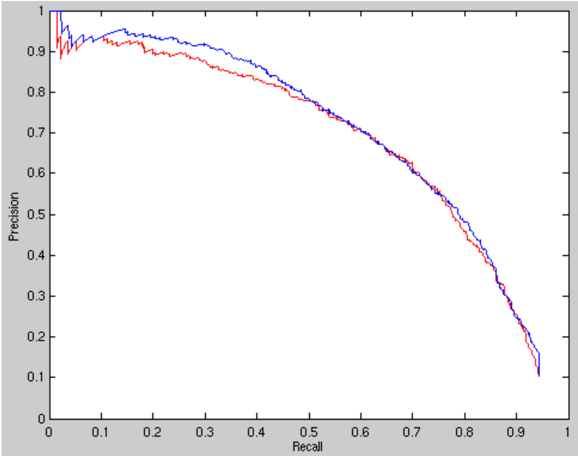


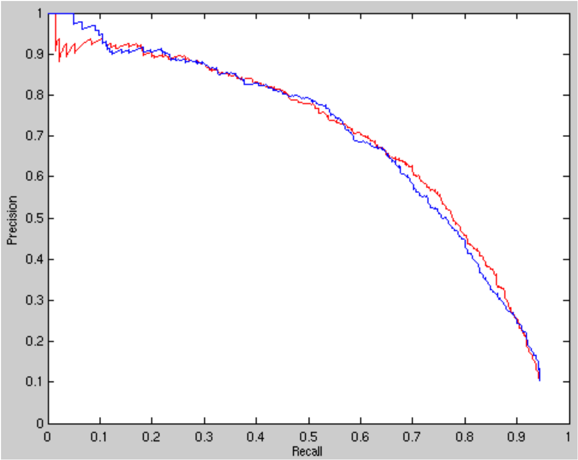
Red: 10



Blue: 3

Blue: 5



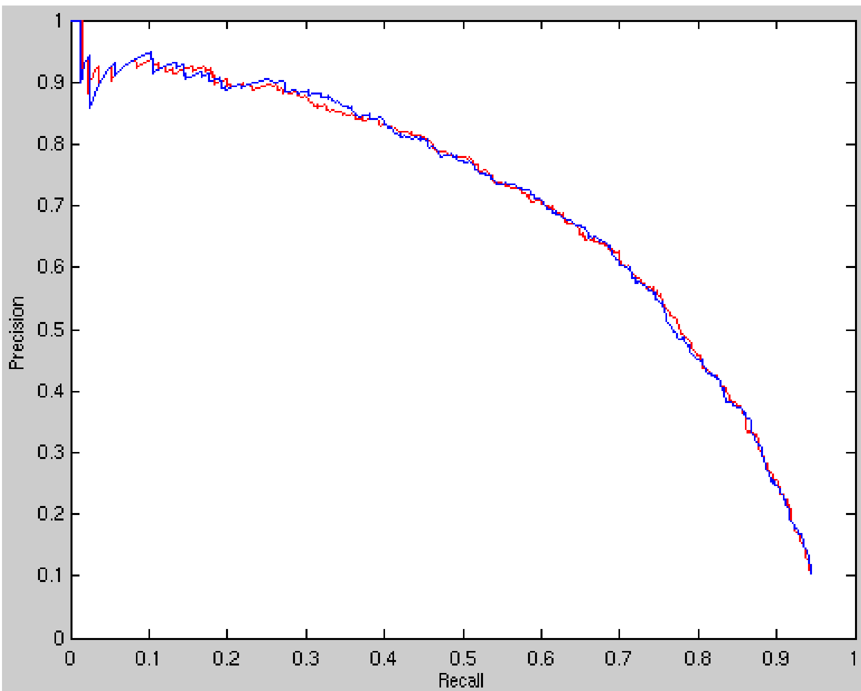


Blue: 20

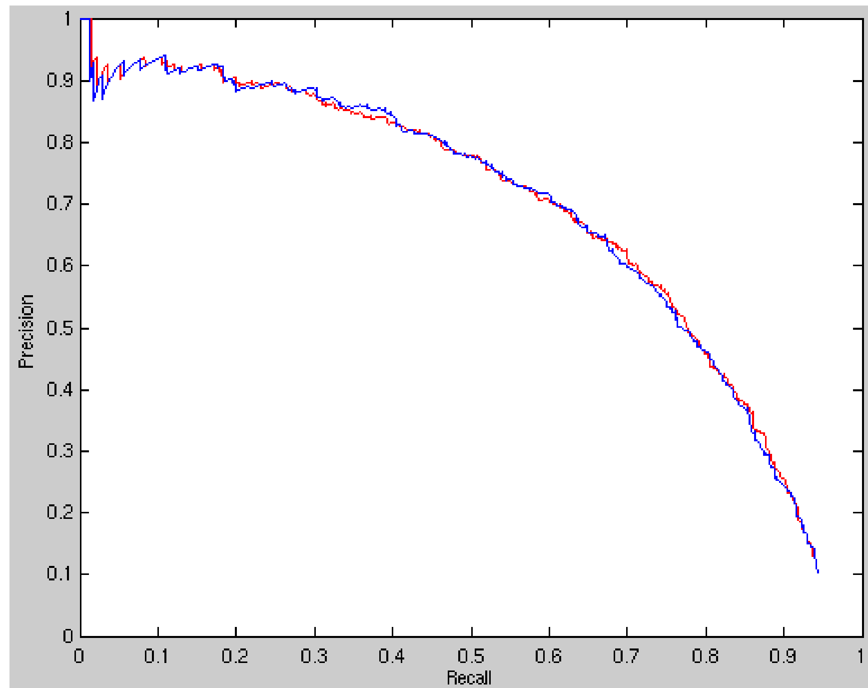
Blue: 30

Number of Gibbs Iterations

Red: 10



Blue: 20



Blue: 100

Outline

- Training and inferring
- Preprocessing
- Experimental results
- Things-and-stuff relationships
- Performance
- Effect of parameters
- Conclusion

Conclusion

- Can be easily integrated with detectors
- The performance is dependent on the detector
- The “stuff” can come from the context as well as the object itself
- Especially suitable for background consistent and view point consistent datasets, ex: aerial images
- 3D information could be used to improve the performance

Reference

- **Learning Spatial Context: Using Stuff to Find Things**, Geremy Heitz and Daphne Koller. European Conference on Computer Vision (ECCV), 2008
- TAS http://ai.stanford.edu/~gaheitz/Research/TAS/
- Superpixel http://www.cs.sfu.ca/~mori/research/superpixels
- HOG implemetation http://pascal.inrialpes.fr/soft/olt
- PASCAL VOC2005 http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2005/index.html