Context-based Reasoning for Object Detection and Object Pose Estimation.

José Oramas M.
VISICS, ESAT, KU Leuven
April 29th 2015

Outline

- Problem Statement
- Research Question
- Contributions
- Discussion

Thesis

Title:

Context-based Reasoning for Object Detection and Object Pose Estimation.

Supervisor:

- Prof. Tinne Tuytelaars.
- Prof. Luc De Raedt.

Examination Committee:

- Prof. Marie-Francine Moons.
- Prof. Luc Van Gool.
- Prof. Luc Van Eycken.
- Prof. Joseph Vandewalle
- Prof. Ales Leonardis.

Funding:

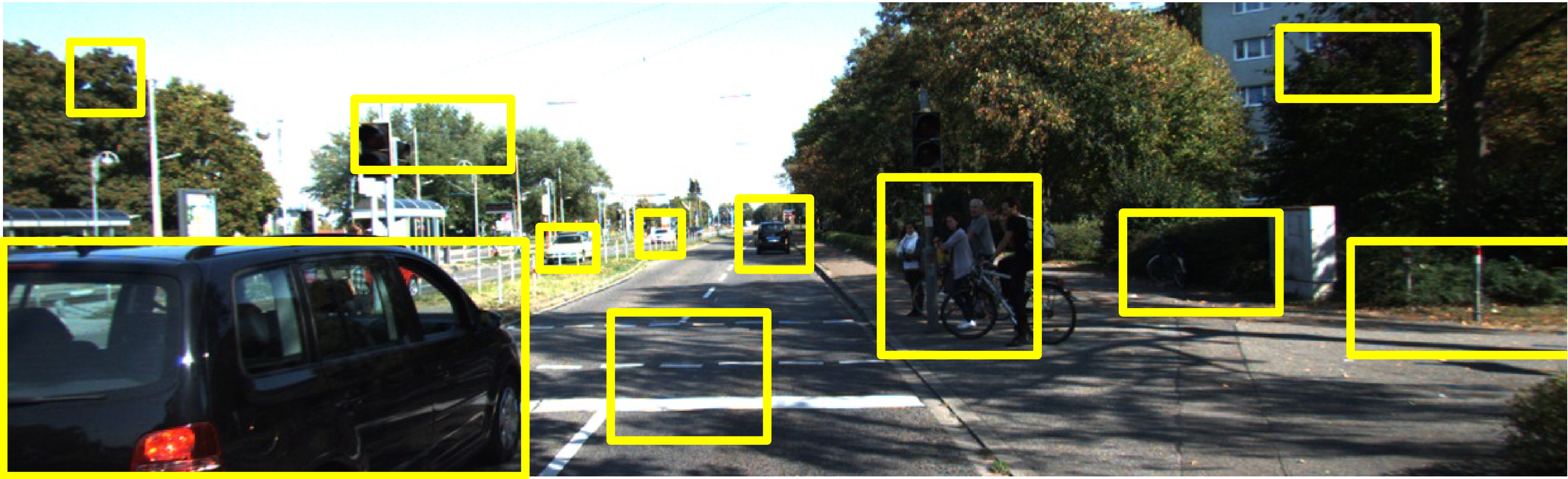
- DBOF Research Scholarship KUL 3E100864.
- FP7 ERC Grant 240530 COGNIMUND.
- KU Leuven OT Project VASI.

Object Detection

Introduction

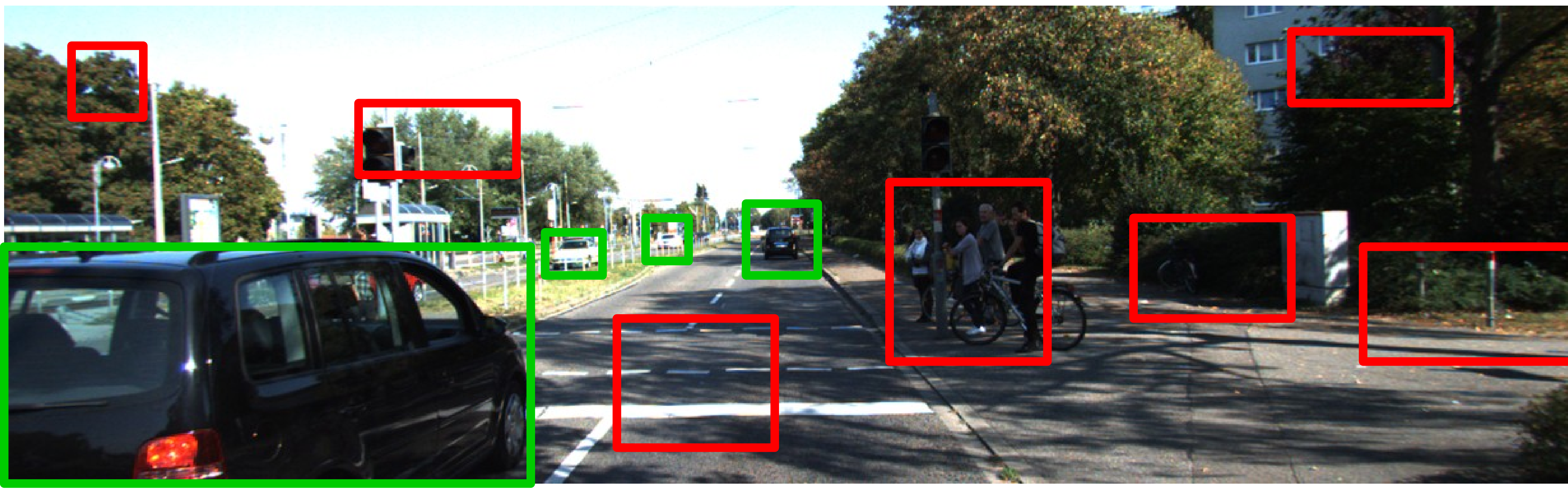
Object Detection

Sample regions (windows) over the image



Object Detection

Evaluate each region (window)



[There is a car | There is no car]

Introduction

Object Detection

Final prediction

Introduction

Object Detection

Driver assistance & Autonomous navigation

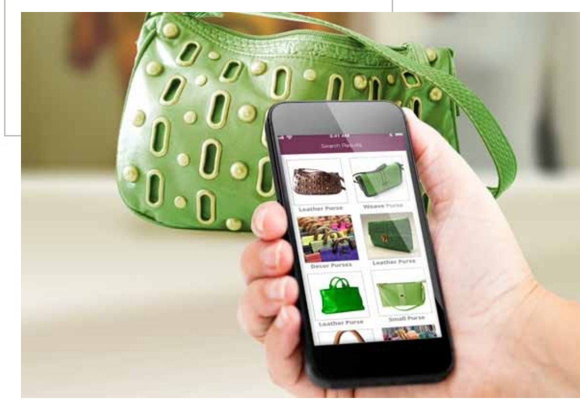
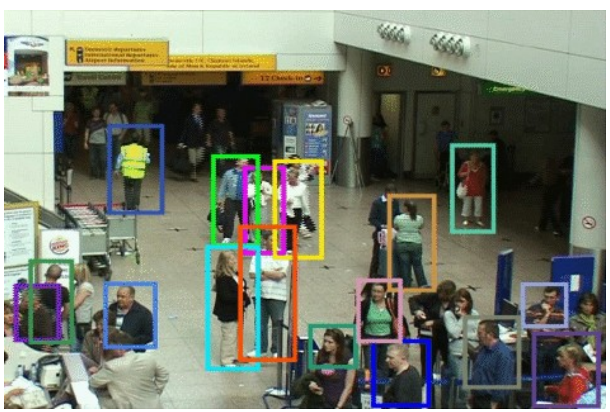
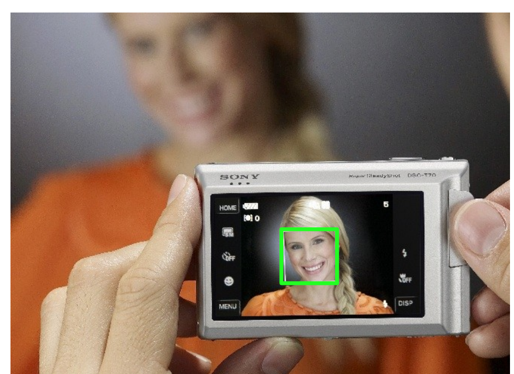


Image retrieval



Security & Surveillance



Automatic image acquisition + enhancement

Problem Statement

Object Detection

3D Appearance
- Murase et al., IJCV (1995) .
- Selinger et al., CVIU (1999) .
- Ponce et al., RFIA (2004) .
- Yan et al., ICCV (2007) .
- Song et al., ECCV (2014) .

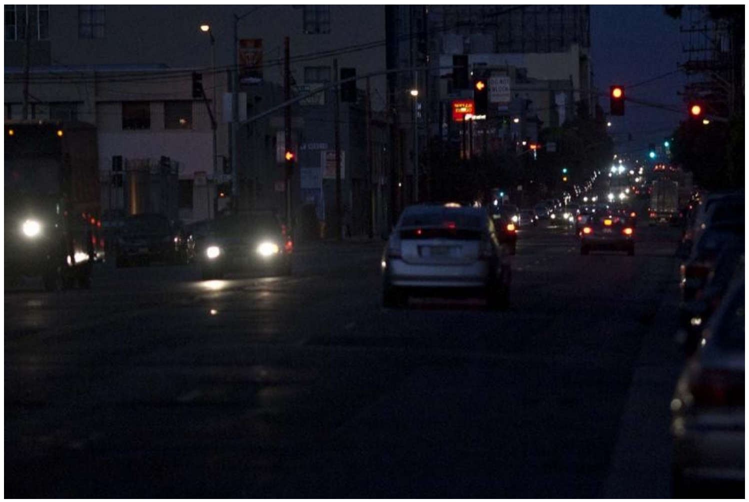
2D Appearance
- Dalal & Triggs., CVPR (2005) .
- Felzenszwalb et al., TPAMI (2010) .
- Fischler & Elschalager., TC (1973) .
- Viola and Jones, CVPR (2001) .

Object Pose / Viewpoint Estimation

Problem Statement

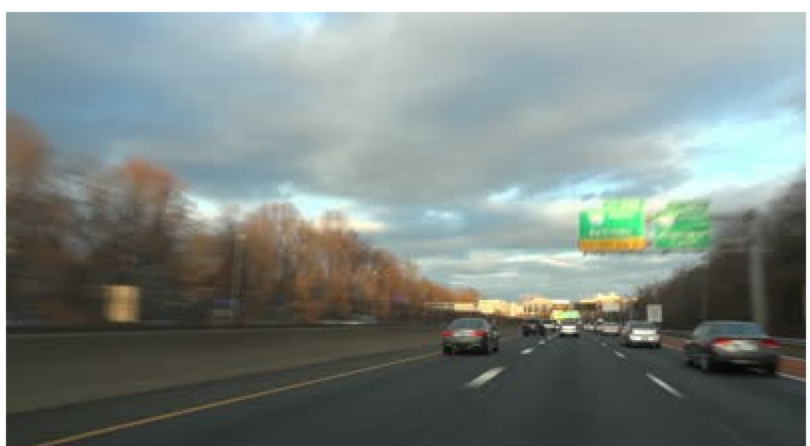
Challenges

Changes in Illumination



High Occlusions

Low resolution | objects at low scale



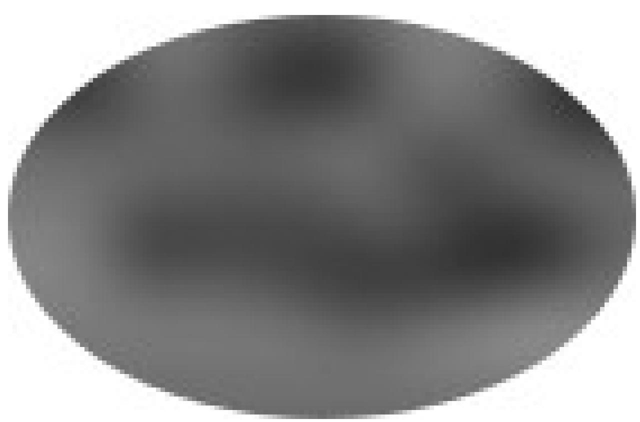
The Context Challenge

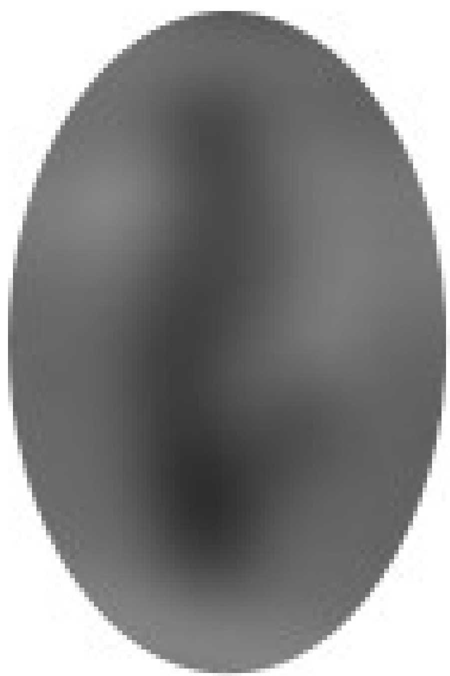
(Torralba & Oliva, IJCV'03)

Problem Statement

The Context Challenge (Torralba & Oliva, IJCV'03)

What is the category of the objects depicted in the following images?





Problem Statement

The Context Challenge (Torralba & Oliva, IJCV'03)

What is the category of the objects depicted in the following images?

car ?

pedestrian ?

Problem Statement

The Context Challenge (Torralba & Oliva, IJCV'03)

What is the object depicted in the following images?

Scene Context

Problem Statement

How to define the context of an object?

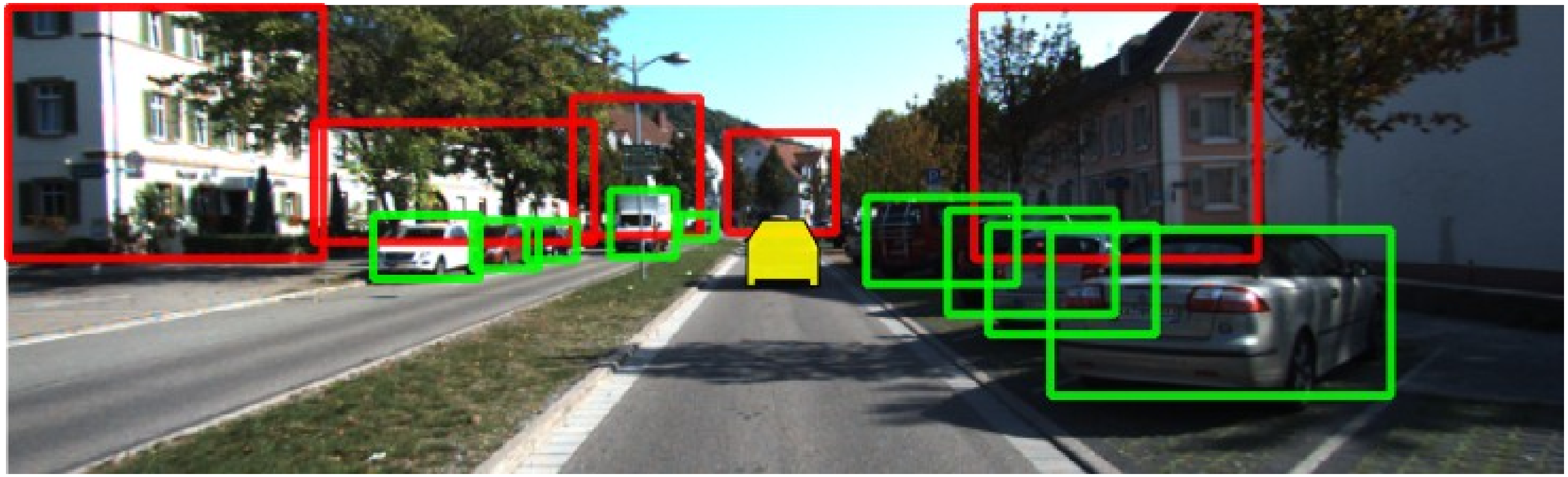
Geometric Context

Hoiem et al. ICCV'05

Problem Statement

How to define the context of an object?

Semantic Context



[Car | Building]

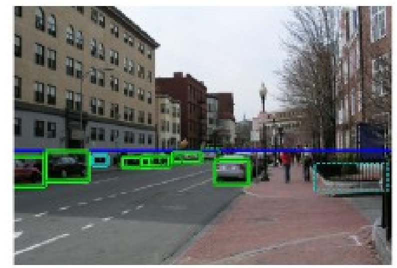
Desai et al. ICCV'11

Problem Statement

How to define the context of an object?

Scene Context

- Oliva & Torralba, ICJV'03.
- Russell et al., NIPS'07.
- Hoiem et al., IJCV'08

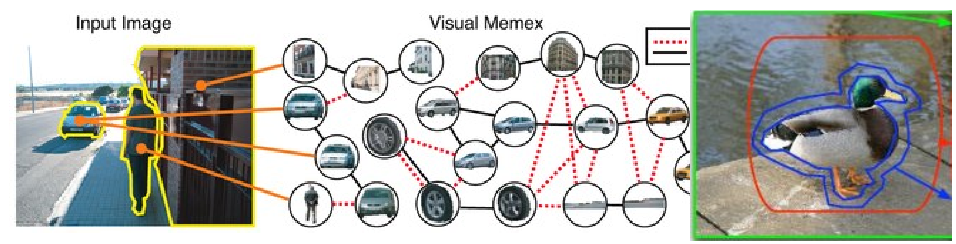


Geometric Context

- Hoiem et al., ICCV'05

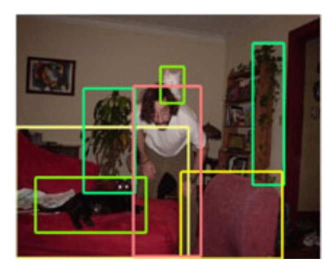
Local Context

- Perko & Leonardis., CVIU'10.
- Galleguillos et al., CVPR'10.
- Malisiewicz & Efros, NIPS'09.
- Bileschi et al., Ph.D. thesis.



Object Relations Context

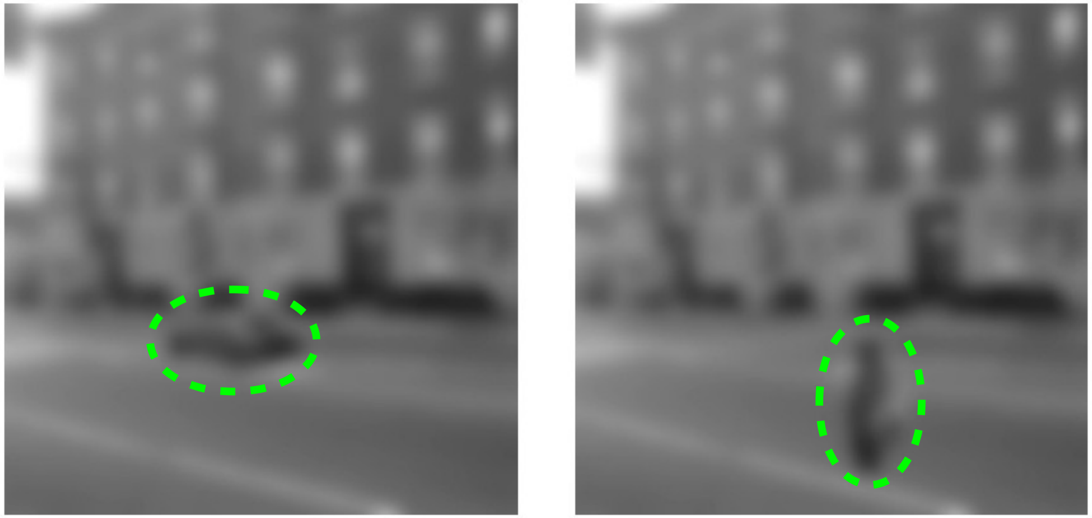
- Perko & Leonardis, CVIU'10.
- Desai et al., IJCV'11.
- Antanas et al., Neurocomputing'14



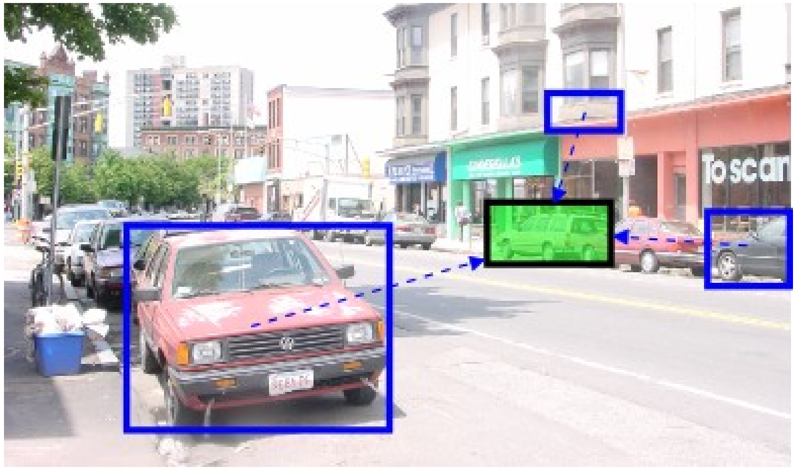
Problem Statement

In this study

Scene Context



Semantic Context (object relations)



Problem Statement

Relations between Objects

Natural group behaviors



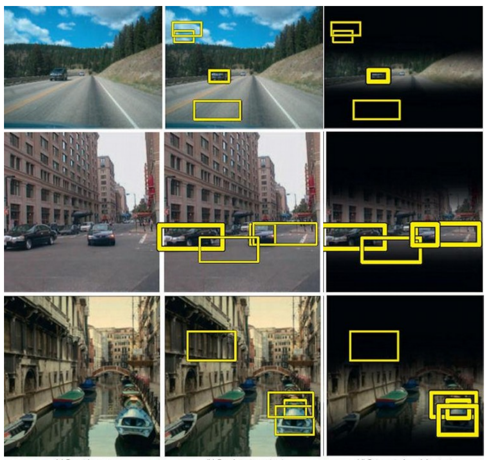
Man-Made objects in desired/permitted configurations



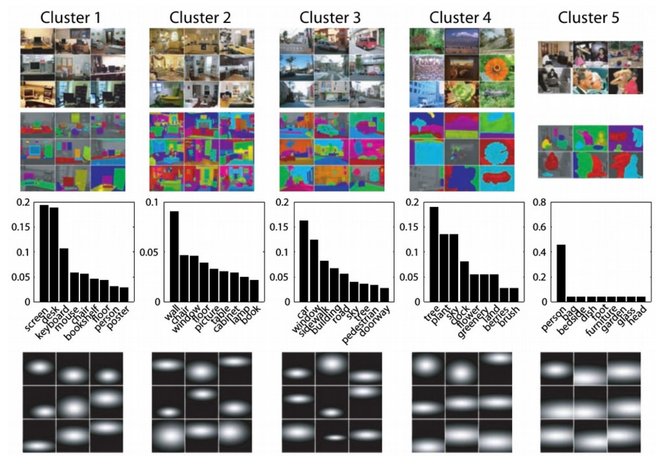

Problem Statement

Exploiting contextual information

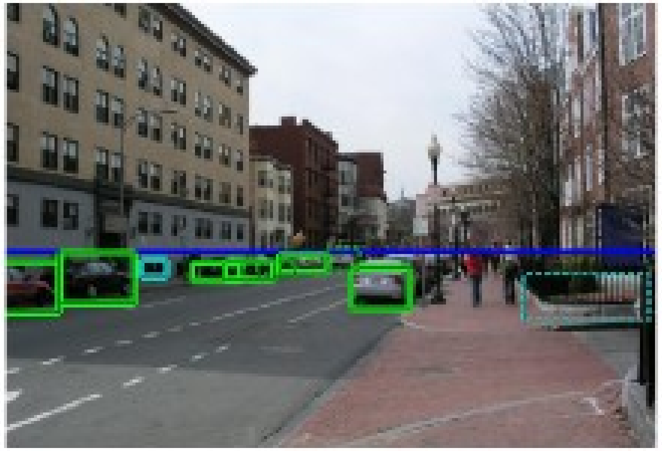
Scene Context



Oliva & Torralba., IJCV (2003).

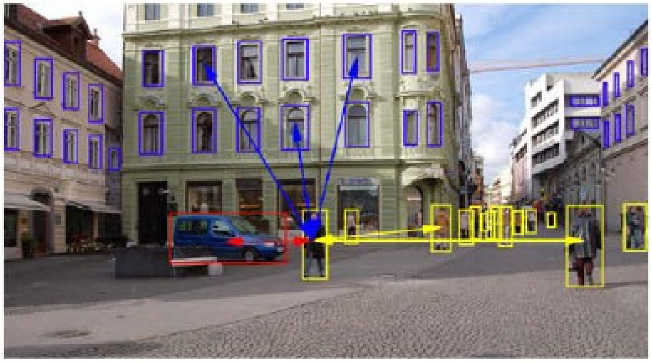


Russell et al. NIPS (2007) .

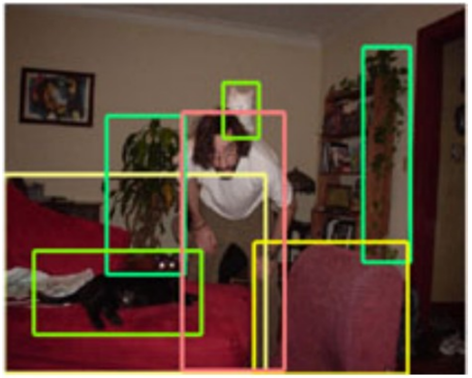


Hoiem et al., IJCV (2008).

Object Relations Context



Perko & Leonardis., CVIU (2010).



Desai et al., IJCV (2011).

Antanas et al., Neurocomputing (2014).

Problem Statement

Research Question:

Can contextual information improve performance of vision tasks?

Problem Statement

Main Research Question:

Can contextual information improve performance of vision tasks?

Research Question:

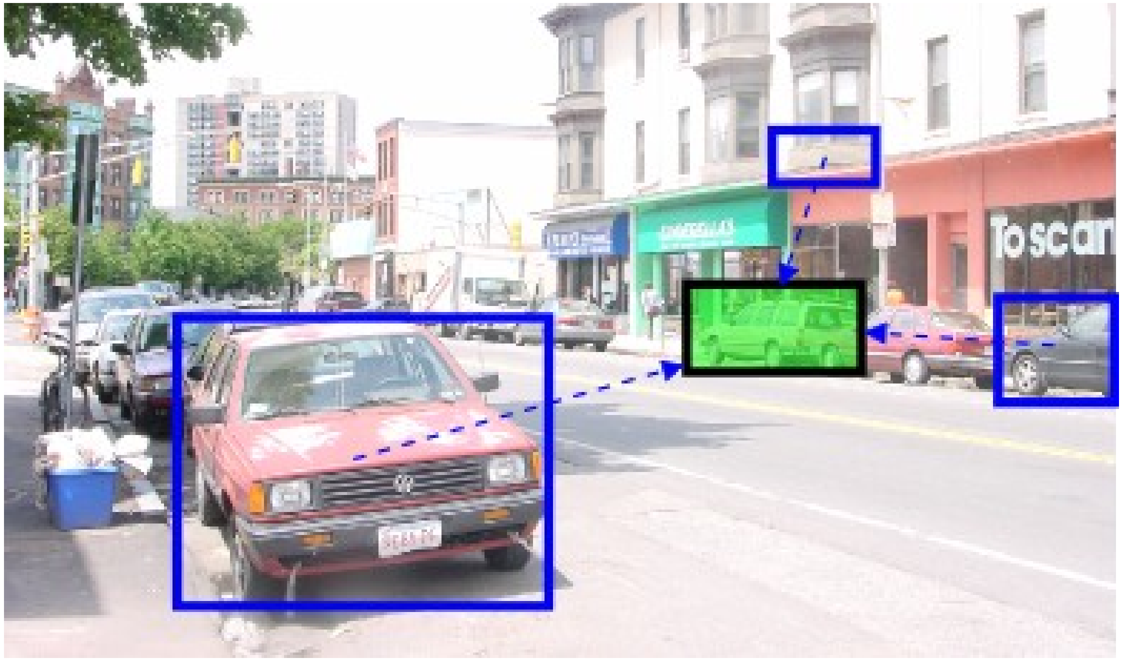
R1: ***Is contextual information, in the form of relations between objects, useful for object pose estimation?***

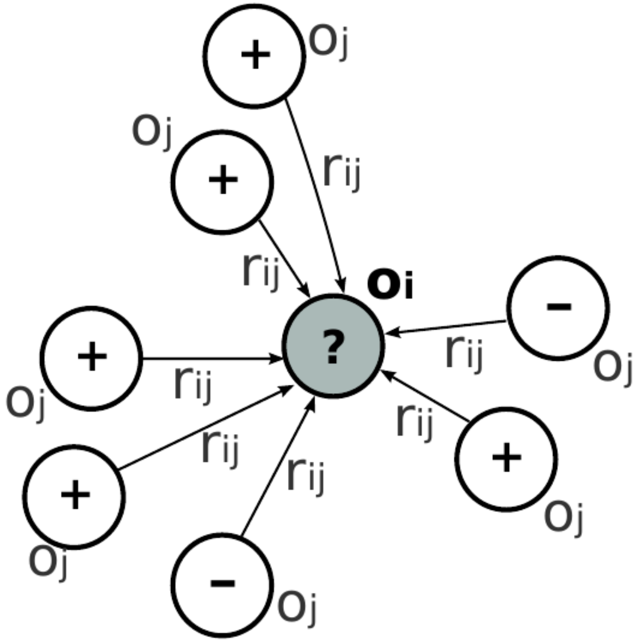
R2: ***Is contextual information, in the form of scene-driven cues, useful for the task of object viewpoint estimation?***

R3: ***To what extent does the nature of the association between objects affects the performance of using relations between objects to improve object detection?***

Context-based Pose Estimation

Relations between objects (ICCV'13)





Context-based object pose estimation

$$\theta_i^* = \arg \max_{(\theta_i \in o_i)} (wvRN(o_i|N_i))$$

Contextual [relational] classifier

$$wvRN(o_i|N_i) = \frac{1}{Z} \sum_{o_j \in N_i} v(o_i, o_j).w_j$$

(Mackassy et al. , JMLR 2007)

$$wvRN(\theta_i^+, o_i^+|N_i) = \frac{1}{Z} \sum_{o_j \in N_i} p(\theta_i^+, o_i^+|r_{ij}).w_j$$

Context-based Pose Estimation

Defining Relations between objects

Camera-centered (CC) **Object-centered (OC)**

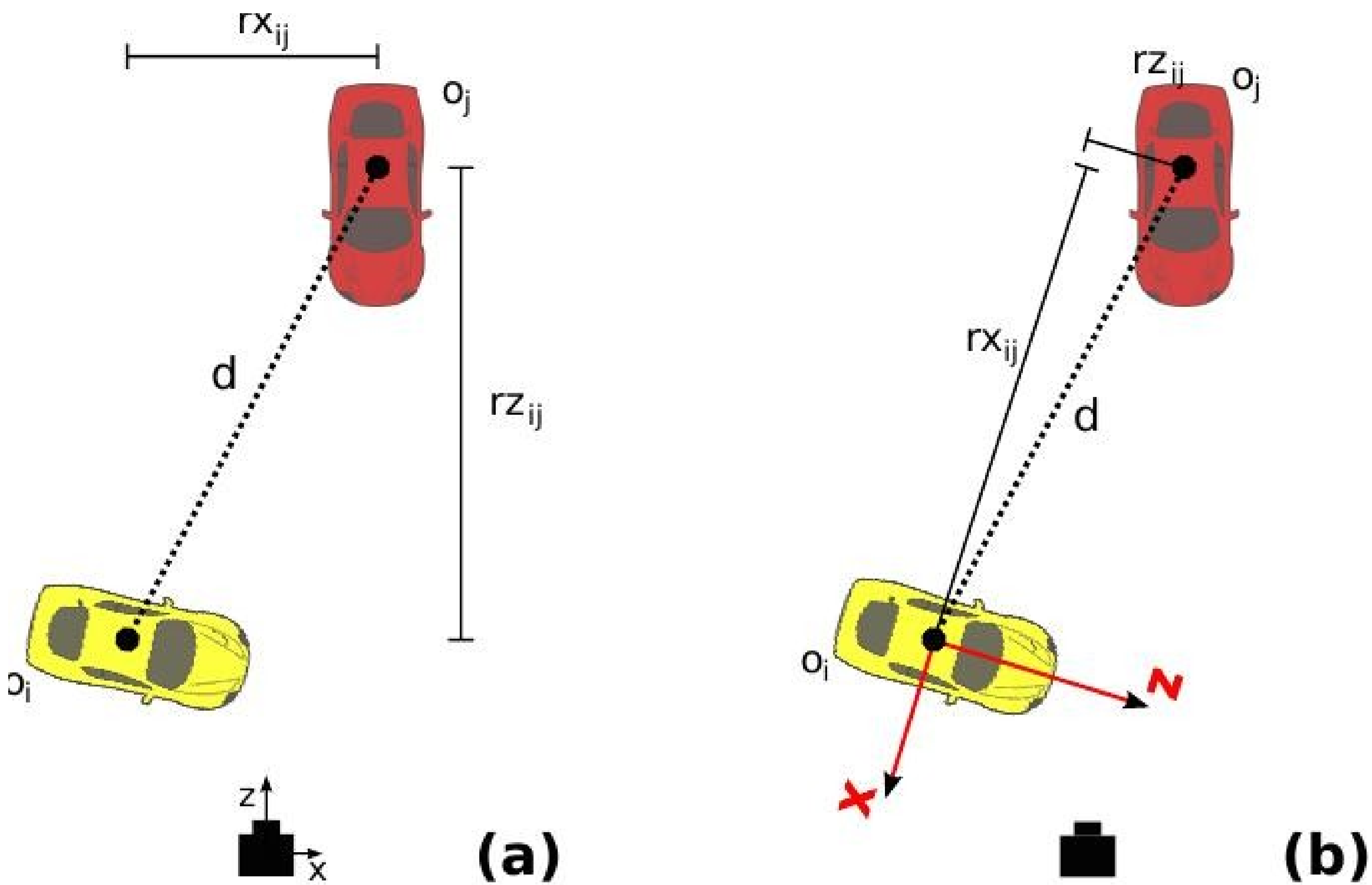


Figure: Pairwise relations between objects from (a) a camera-centered and from (b) an object-centered frame of reference.

Context-based Pose Estimation

Where can another car be located given the car in the center and the expected relative pose between them?

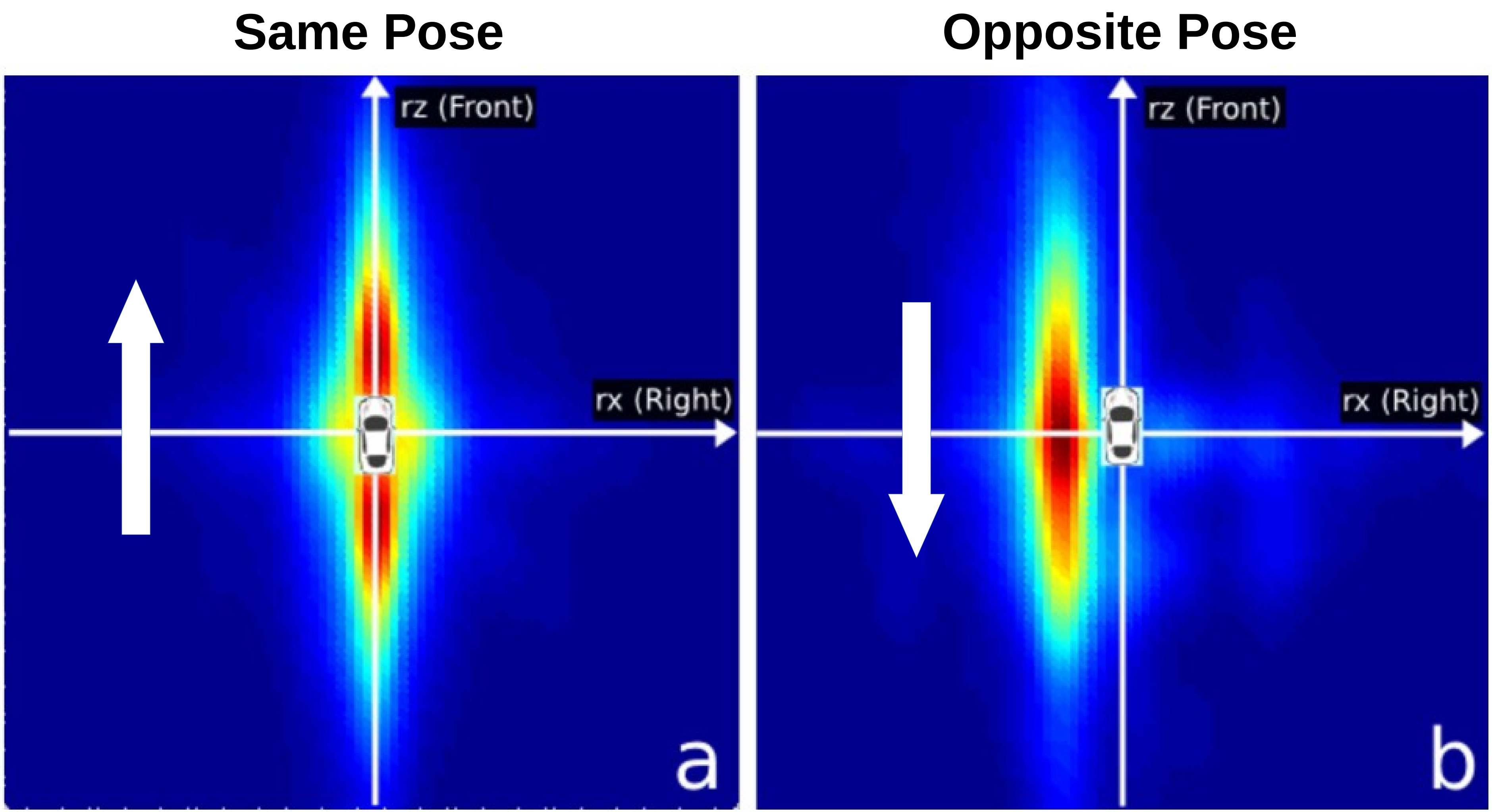


Figure: Top-view of the distribution of object-centered relations for cars with (a) the same and (b) opposite pose, respectively.

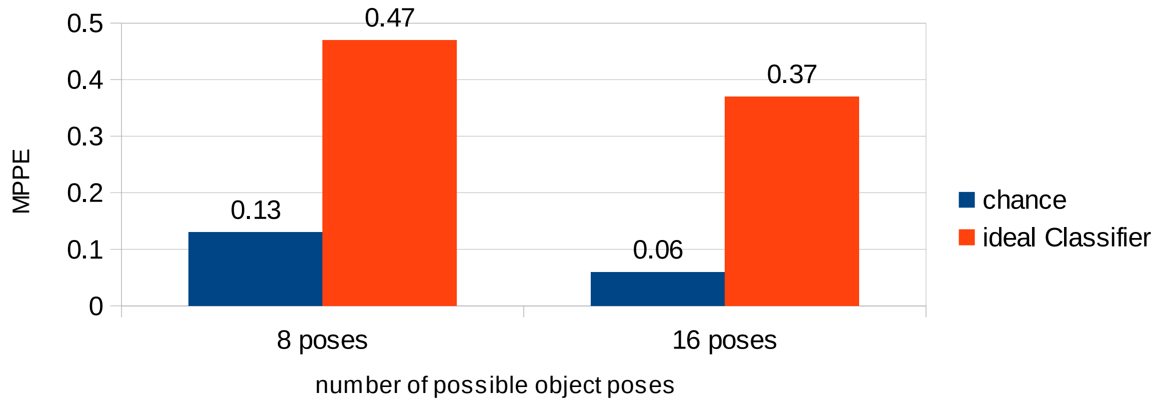
Context-based Pose Estimation

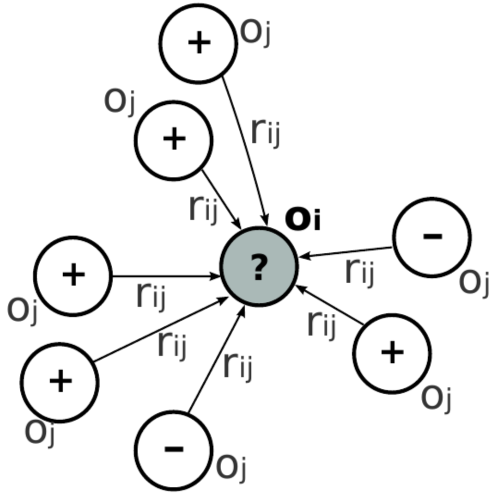
Some results

Ideal Setting

(purely contextual method)

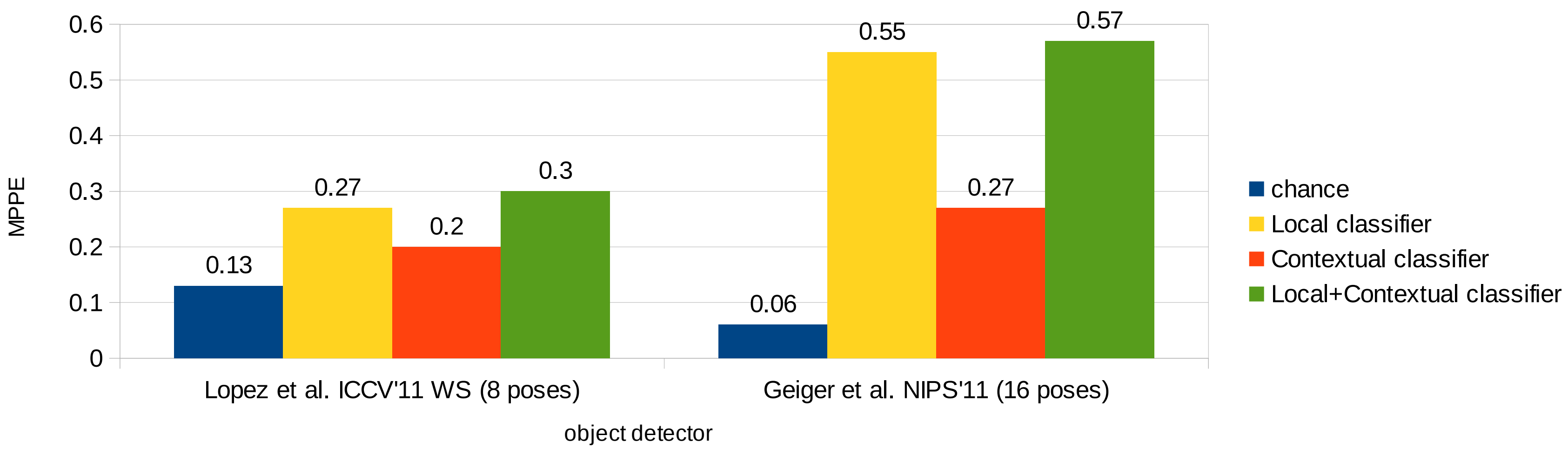
Figure: Mean Precision in Pose Estimation (MPPE) on the KITTI dataset.





Realistic Setting

Figure: Mean Precision in Pose Estimation (MPPE) on the KITTI dataset.



Context-based Pose Estimation

Qualitative Results

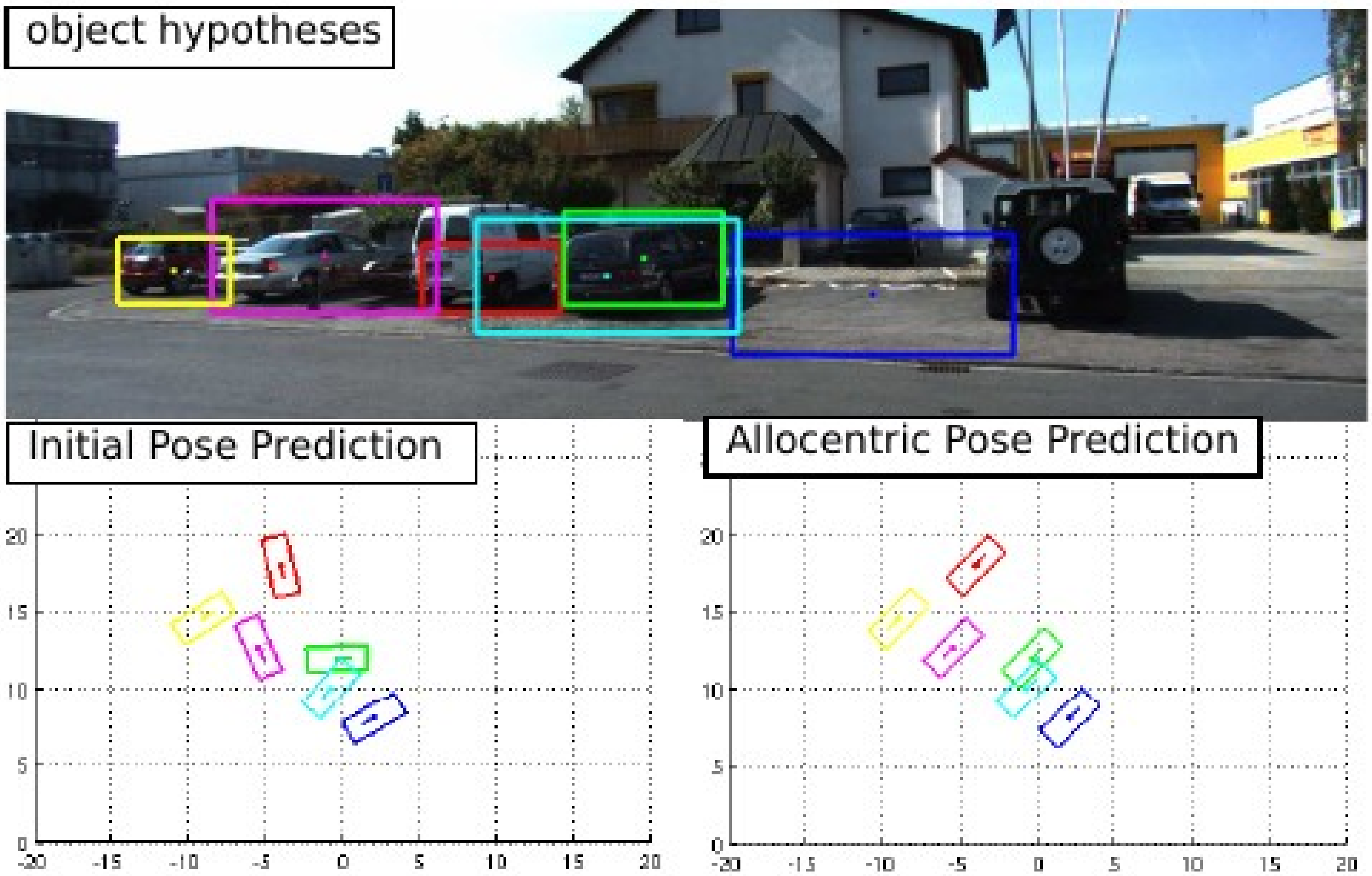


Figure: Object bounding box is color coded. Notice the difference between the initial pose prediction given by the detector and the context-based prediction.

KU LEUVEN

PSI VISICS

Research question 1:

Is contextual information, in the form of relations between objects, useful for object pose estimation?

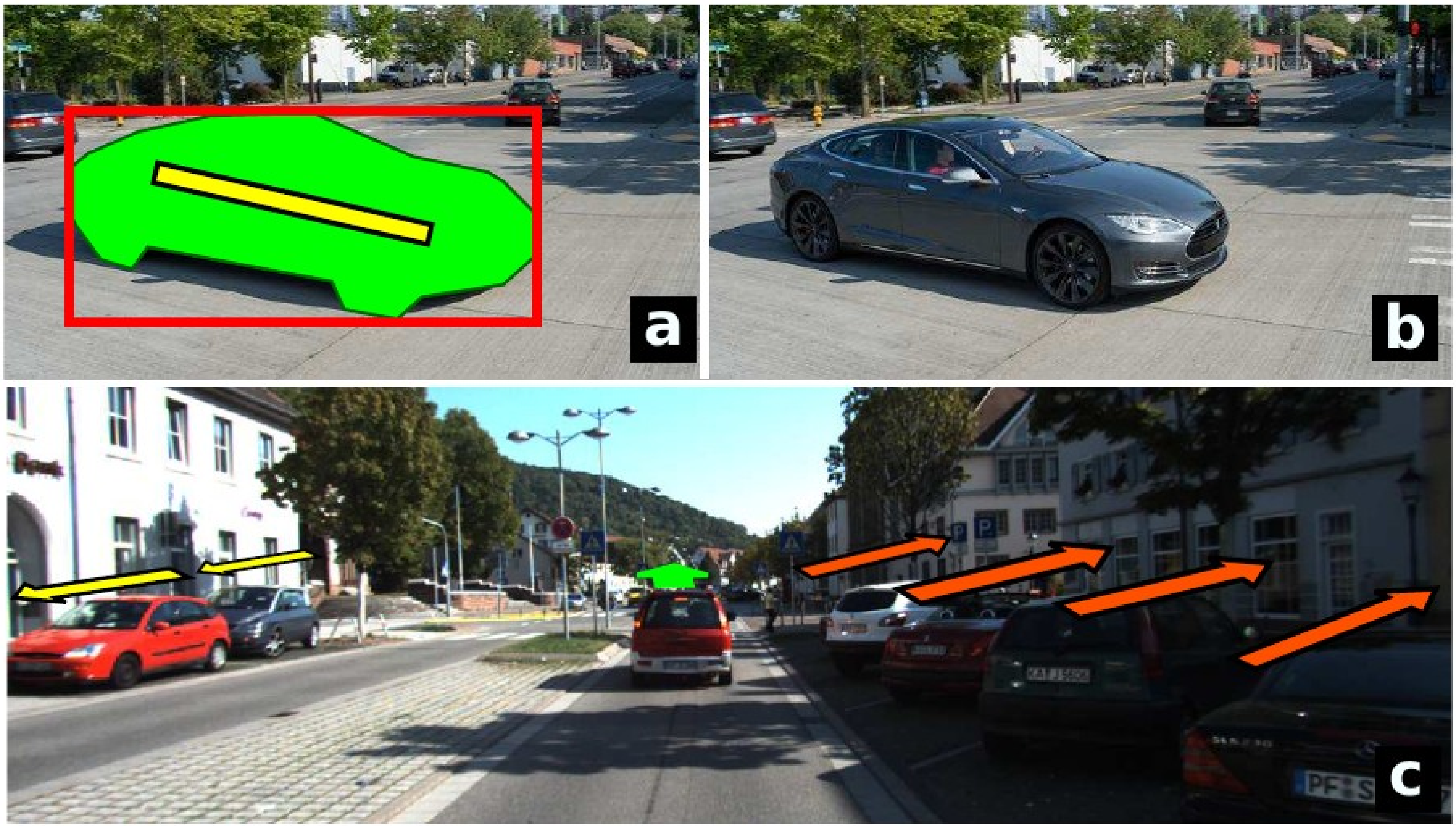
- Purely contextual experiments show that the proposed methods are able to encode information about the orientation of participant objects.

- Combination of local and contextual methods improves initial pose estimation performance.

Scene-driven Viewpoint

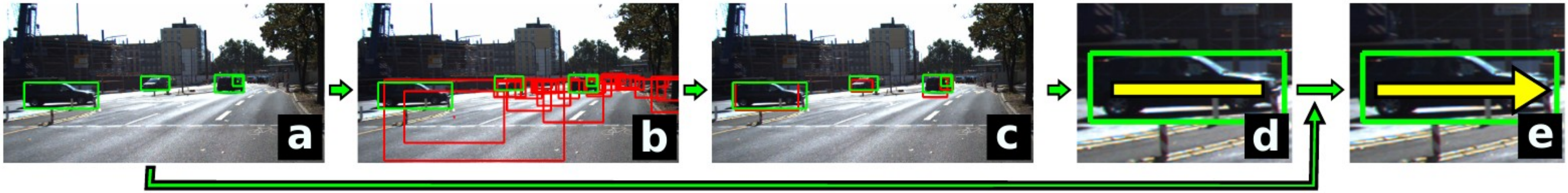
Scene context (BMVC'14)

- Exploit physical extent of elongated objects (a,b).
- Regions of the scene tend to host objects with particular features (c).



Scene-driven Viewpoint

Algorithm pipeline



a) Object detection.

b) Scene-driven object proposal generation.

c) Object hypotheses – proposals matching.

d) Object elongation orientation classification.

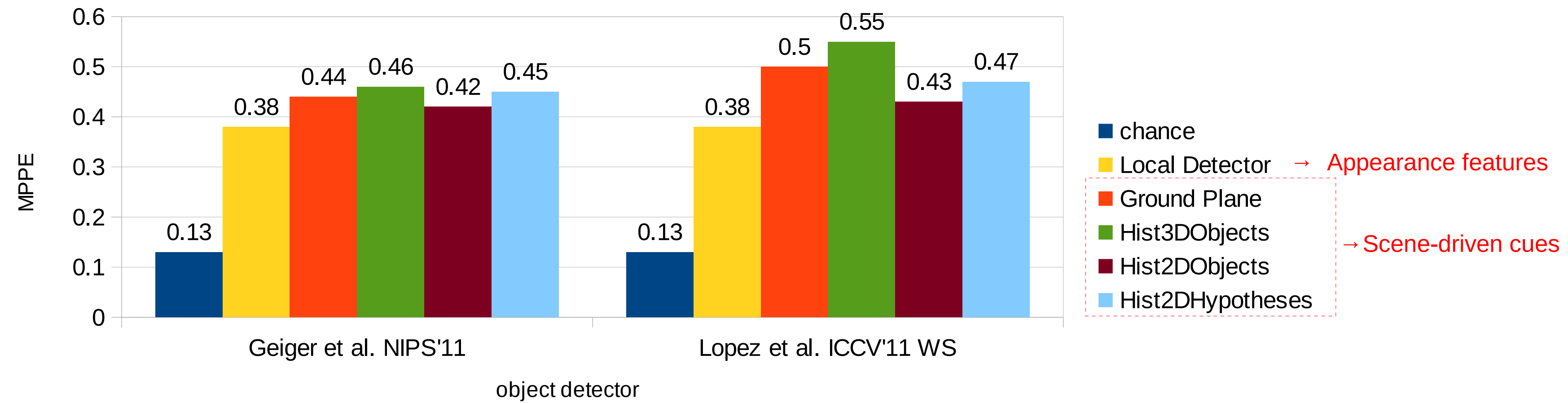
e) Object viewpoint classification.

Scene-driven Viewpoint

Quantitative Results (8 viewpoints)

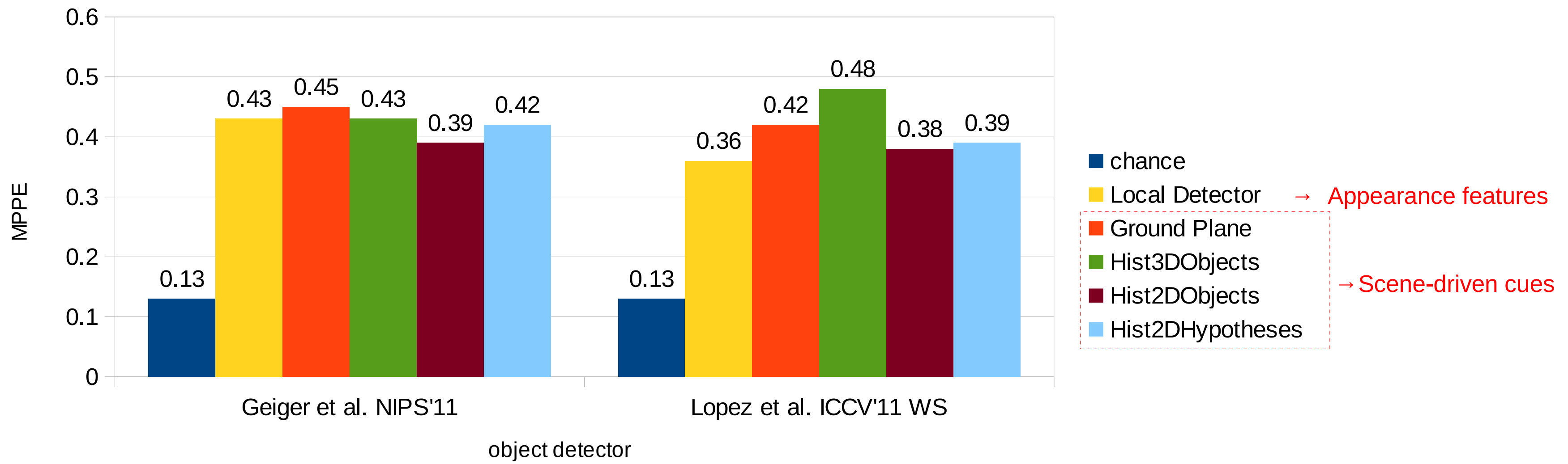
Easy image set (object height>50px)

Figure: Mean Precision in Pose Estimation (MPPE) on the KITTI dataset .



Full image set (all the objects)

Figure: Mean Precision in Pose Estimation (MPPE) on the KITTI dataset .



Scene-driven Viewpoint

Quantitative Results (8 viewpoints)

- **Continuous Line:** object detector prediction.
- **Dashed Line:** scene-driven object proposals.
- **Circle:** ground-truth viewpoint.

Revisiting the Research Questions

Research question 2:

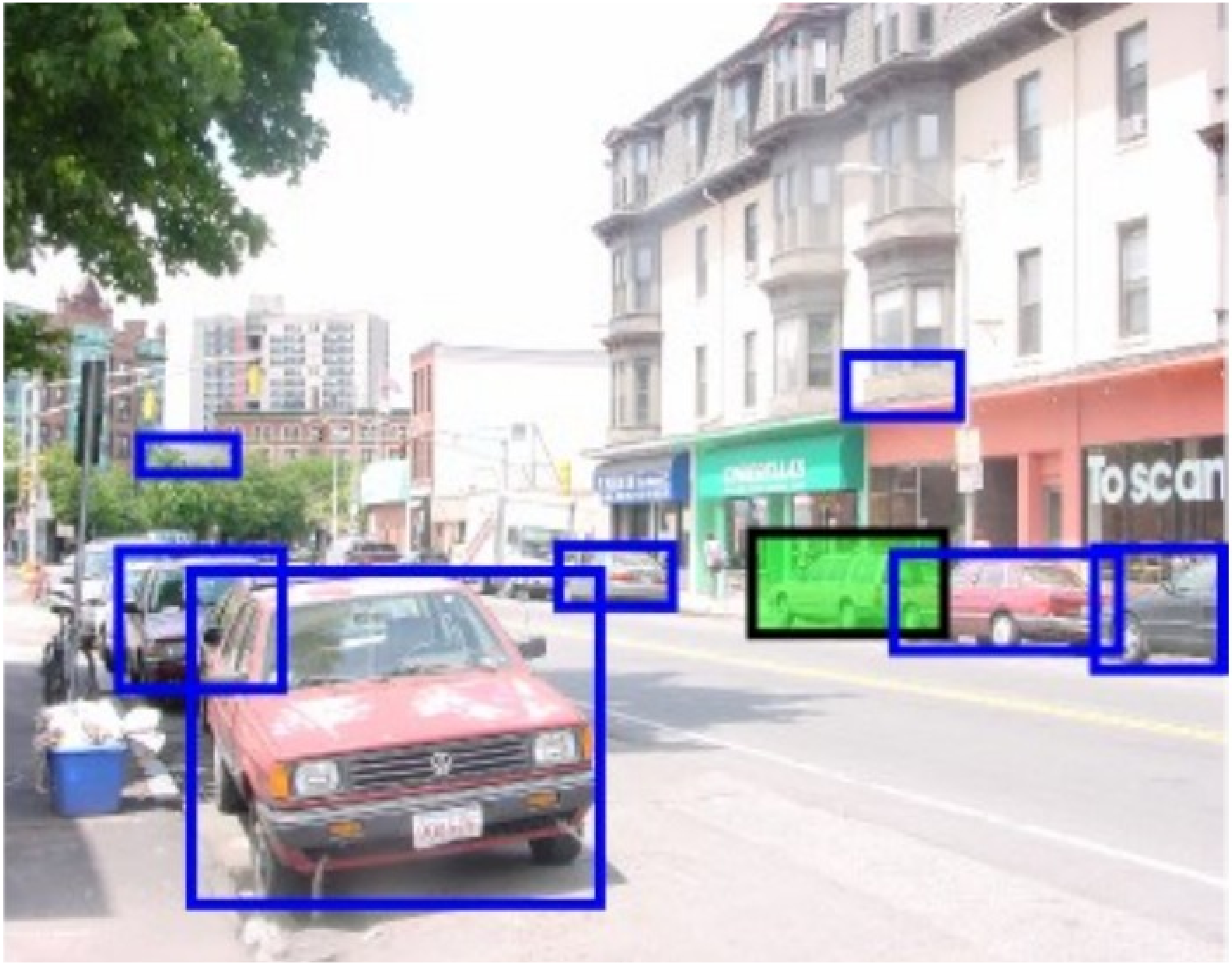
Is contextual information, in the form of scene-driven cues, useful for object viewpoint estimation?

- Experiments suggest that scene can effectively serve as a source of contextual information for object viewpoint estimation.

- Combination of scene-driven cues and methods based on intrinsic features produces improvements on object viewpoint estimation performance.

Object Association

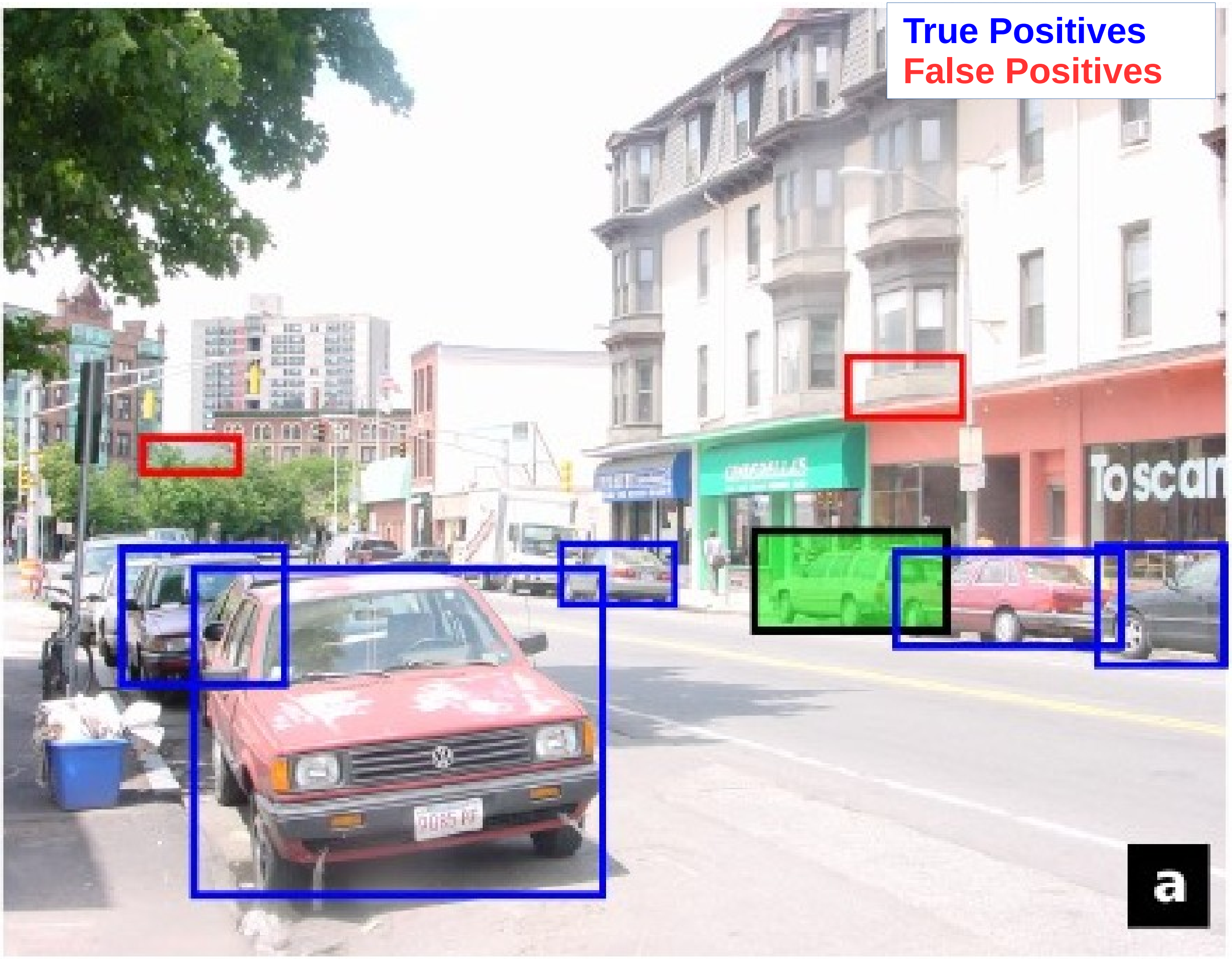
Context-based Object Detection (WACV'14)



Aggressive Inference

Object Association

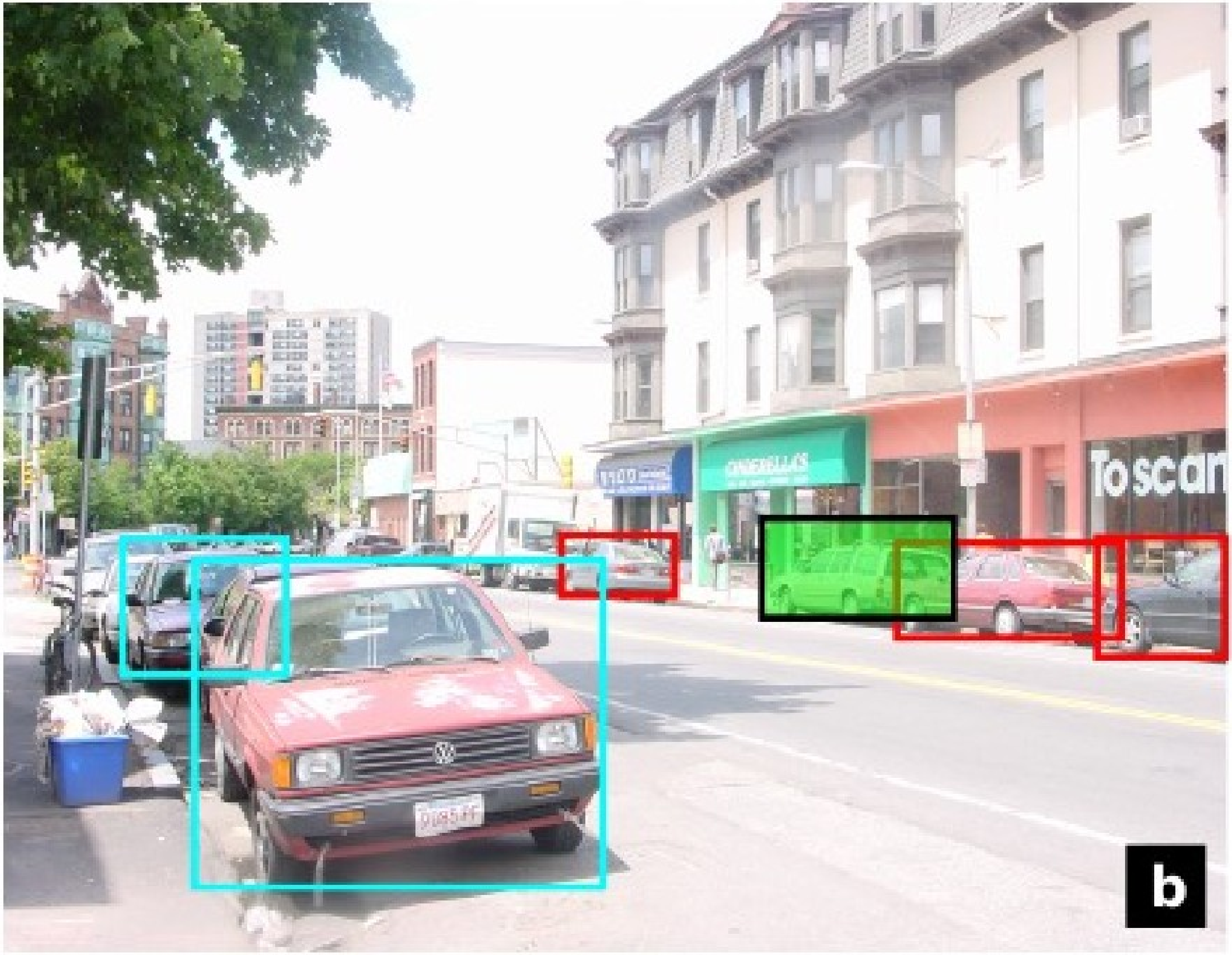
How to properly use relations between objects?



Cautious Inference

Object Association

How to properly use relations between objects?



Relationship-driven association

Proposed Method

How objects associate to each other ?

Category-driven association

Relationship-driven association

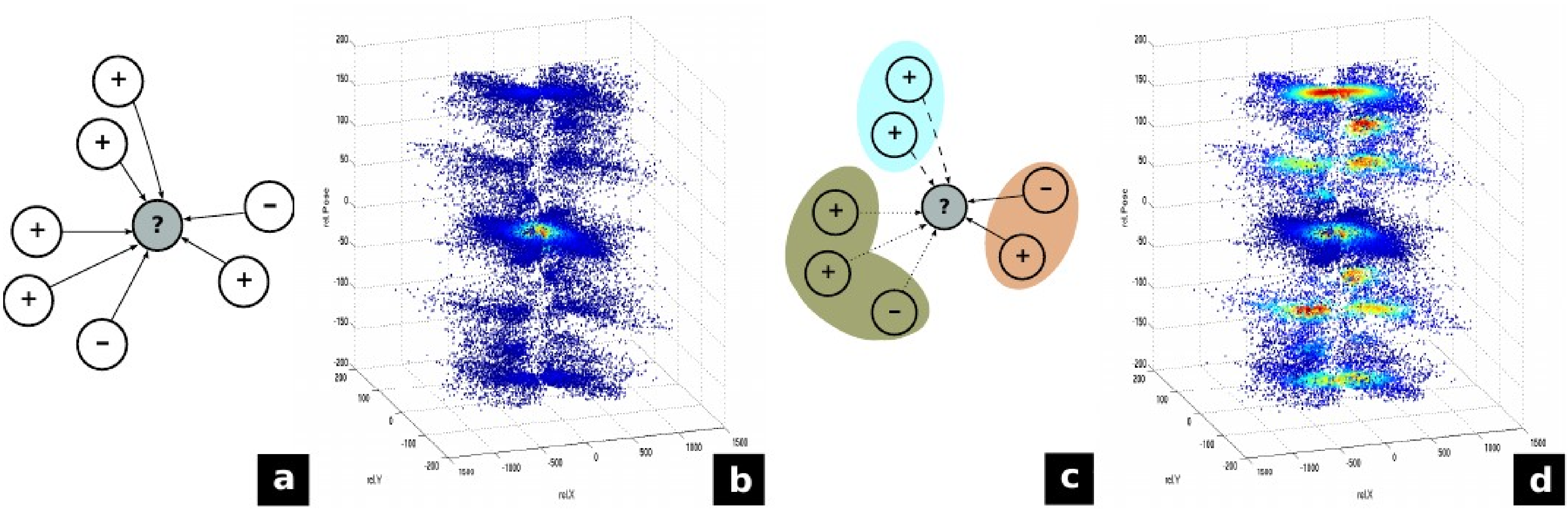


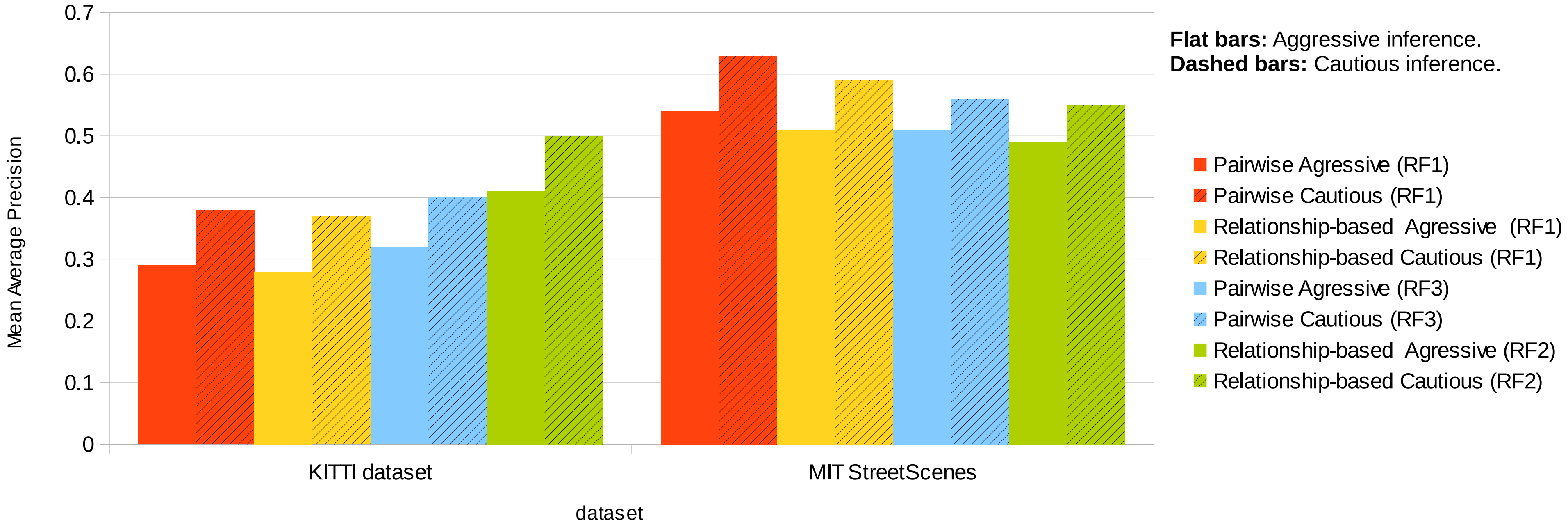
Figure: Category-based association: a) voting, b) density distribution; and Relationship-based association c) voting, d) density distribution. Density distributions from cars on the KITTI dataset.

$$wvRN(o_i|N_i) = \frac{1}{Z} \sum_{o_j \in N_i} v(o_i, o_j).w_j$$

Object Association

Only using contextual information

Figure: Mean average precision performance using the detector from [1] to collect object hypotheses.



Flat bars: Aggressive inference.
Dashed bars: Cautious inference.

Object Association

Combination of Local and Contextual Information

Collecting Hypotheses using [1]

Dataset **KITTI benchmark**		**RF1** Class-based Homophily Global		**RF2** Relation-based Homophily Global	
Set	Detector [1]	aggre.	caut.	aggre.	caut.
all	0.61±0.011	0.61±0.009	0.63±0.007	0.65±0.011	***0.68±0.003***

Dataset **MIT StreetScenes**		**RF3** Class-based Homophily Global		**RF2** Class-based Homophily Global	
Set	Detector [1]	aggre.	caut.	aggre.	caut.
all	0.69±0.006	0.77±0.001	***0.80±0.028***	0.73±0.011	0.76±0.014

Table: Mean average precision performance using the detector from [1] to collected object hypotheses.

[1] López et al. ,ICCV WS 2011.
[2] Felzenszwalb et al. ,TPAMI 2010.

Collecting Hypotheses using DPM [2]

Dataset **KITTI benchmark**		**RF3** Class-based Homophily Global		**RF2** Relation-based Homophily Global	
Set	Detector [2]	aggre.	caut.	aggre.	caut.
all	0.65±0.003	0.68±0.007	0.71±0.007	0.72±0.009	**0.75±0.003**

Dataset **MIT StreetScenes**		**RF3** Class-based Homophily Global		**RF2** Class-based Homophily Global	
Set	Detector [2]	aggre.	caut.	aggre.	caut.
all	0.62±0.004	0.66±0.011	**0.71±0.012**	0.65±0.026	0.69±0.014

Table: Mean average precision performance using the detector from [2] to collected object hypotheses.

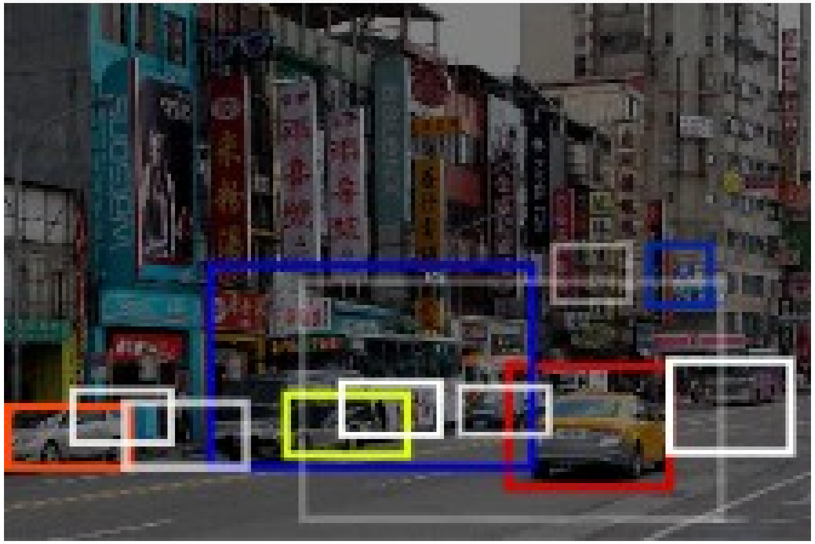
However...

Is there something that can be done to improve recall?

Recovering missed detections

In summary

a) Perform object detection.



b) Recover missed object instances by generating object proposals.

Recovering missed detections

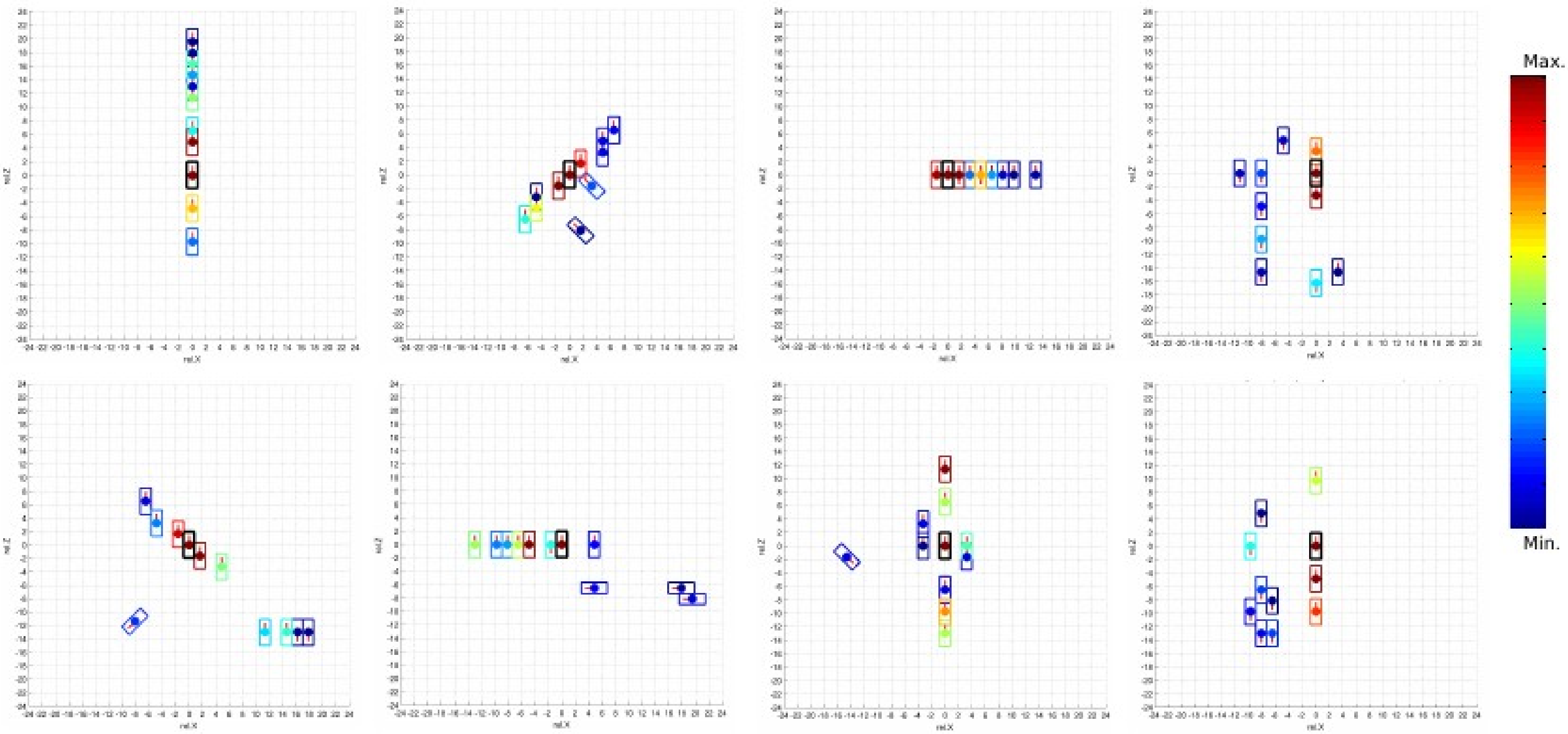
Contribution

- A method to discover higher-order relations between objects.
- Use the modeled relations to recover missed object instances.

Higher order relations between cars marked by color codes

Recovering missed detections

Discovered higher-order relations.

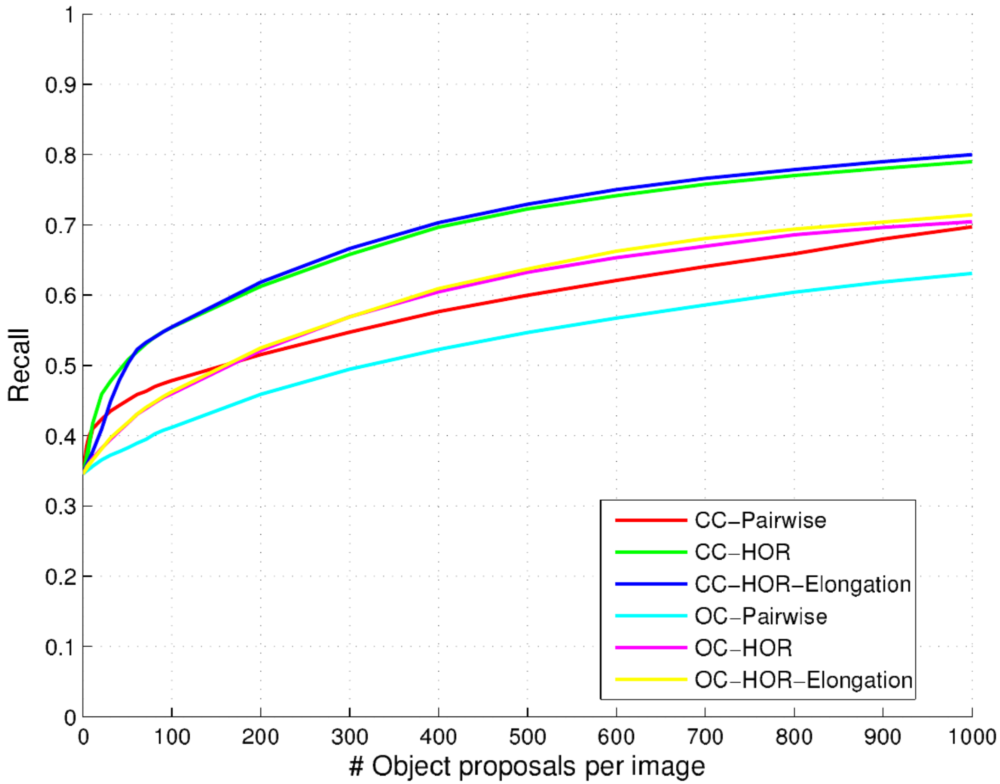


- Top-view of the discovered Higher-order Relations (HOR) between cars in the KITTI dataset.
- Relations are defined from an object-centered perspective.
- Reference object is in the center and colored in black.
- The occurrence likelihood of the related objects is color-coded in jet scale.

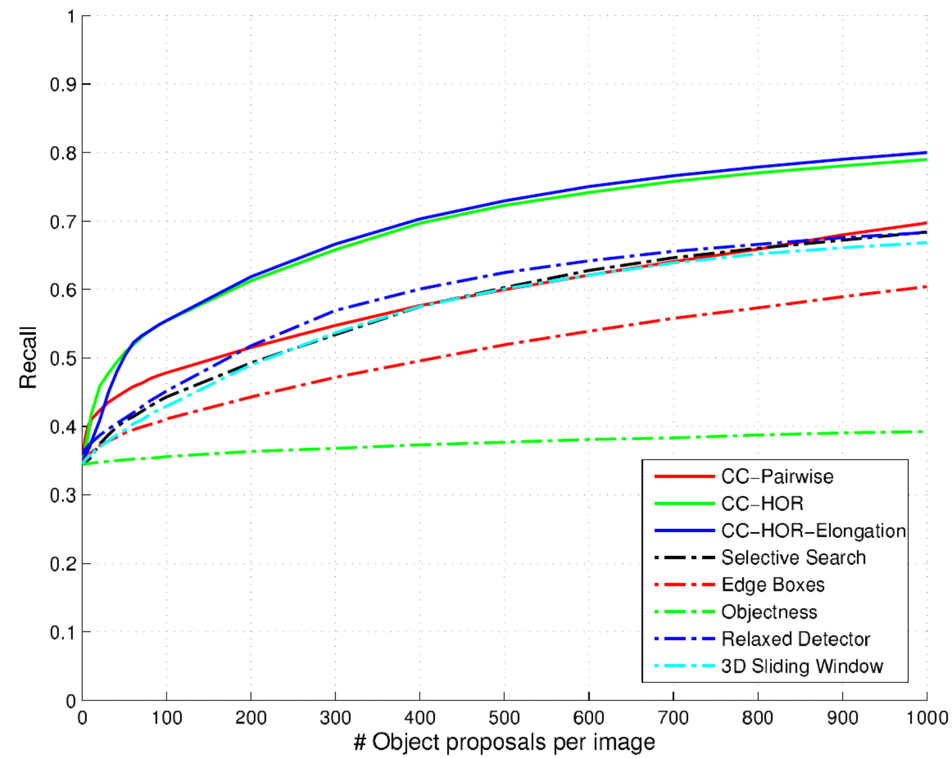
Recovering missed detections

Some results

Comparison w.r.t. relation-based methods



Comparison w.r.t. to other methods



Recall vs. number of generated object proposals on the KITTI dataset (IoU=0.5)

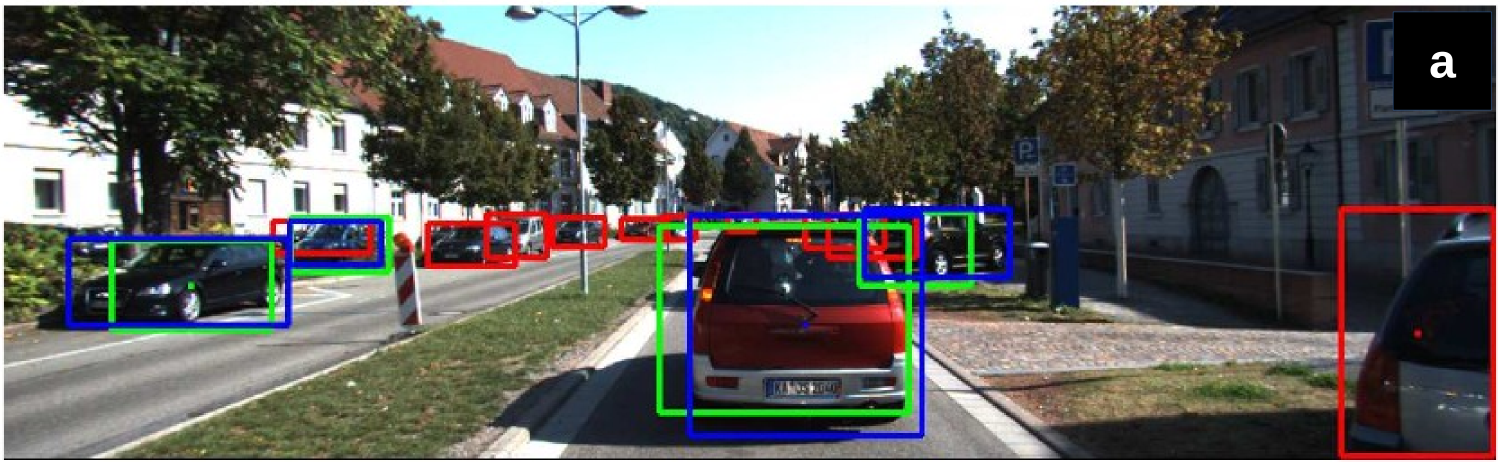
CC: camera centered frame of reference
OC: object centered frame of reference
HOR:Higher-order relations

Recovering missed detections

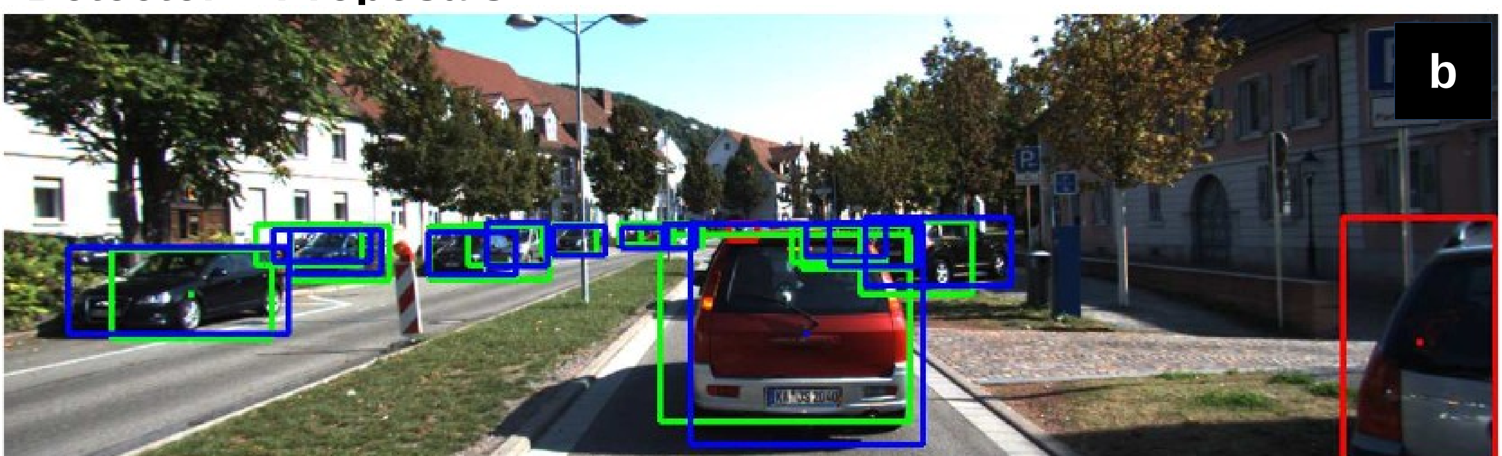
Some results

Qualitative results

Detector alone



Detector + Proposals



Object annotations | matched object instances | unmatched object instances

Research question 3:

To what extend does the nature of the association between object affects the performance of using relations between objects to improve object detection?

- Using most certain objects as source of contextual information increases the gains in object detection precision brought by contextual information.

- Assuming that objects are associated by underlying relationships increases the performance of relations-based methods.

- Methods that reason about object relations can be effectively used to recover miss detected object instances. As a result, this improves object detection performance in terms or recall.

Conclusions

Lessons Learned

- **Collective classification should be used cautiously in vision problems** (Chapter 4) .
- **Object pose / viewpoint estimation is not a purely local problem** (Chapter 3 & 5).
- **Object relations can be used to improve object detection recall** (Chapter 6).

Conclusions

Future Work

- **Integration of detailed local models for object categories.** (e.g. Xiang et al. 3Ddr'13 , Zia et al., CVPR'14, Girshick et al., CVPR'14)
- **Perform the prediction of continuous object pose/viewpoint angles.**
- **Integrate more advanced methods for Collective Classification.** (e.g. Statistical Relational Learning (SRL))

Publications

- Fernando B., Gavves, E., Oramas M., J., Ghodrati, A., Tuytelaars, T.
  Modeling video evolution for action recognition. CVPR 2015*.*
- Oramas M. J., Tuytelaars T.
  Scene-driven Cues for Viewpoint Classification of Elongated Object Classes. BMVC 2014.
- Oramas M. J., De Raedt L., Tuytelaars T.
  Reasoning about object relations for object pose classification. NCCV 2014.
- Oramas M. J., De Raedt L., Tuytelaars T.
  Towards cautious collective inference for object verification. WACV 2014.
- Antanas L., van Otterlo M., Oramas Mogrovejo J., Tuytelaars T., De Raedt L.
  There are plenty of places like home: Using relational representations in hierarchies for distance-based image understanding. Neurocomputing 2014.
- Billiet L., Oramas M. J., Hoffmann M., Meert W., Antanas L.
  Rule-based hand posture recognition using qualitative finger configurations acquired with the Kinect. ICPRAM 2013.
- Oramas M. J., De Raedt L., Tuytelaars T.
  Allocentric pose estimation. ICCV 2013.
- Antanas L., van Otterlo M., Oramas M. J., Tuytelaars T., De Raedt L.
  A relational distance-based framework for hierarchical image understanding. ICPRAM 2012*.*
- Antanas L., van Otterlo M., Oramas M. J., Tuytelaars T., De Raedt L.
  Not far away from home: A relational distance-based approach to understand images of houses. IPL 2010.
- Oramas M., J., Tuytelaars, T.
  Recovering hard-to-find object instances by sampling context-based object proposals. Submitted to ICCV 2015.
- Martinez-Camarena, M., Oramas M., J., Tuytelaars, T.
  Towards sign language recognition based on body parts relations. Submitted to ICIP 2015.

Thank you for your attention

Context-based Reasoning for Object Detection and Object Pose Estimation.

José Oramas M.
VISICS, ESAT, KU Leuven
April 29th 2015