Multi-view Traffic Signs Detection, Recognition, and 3D Localisation

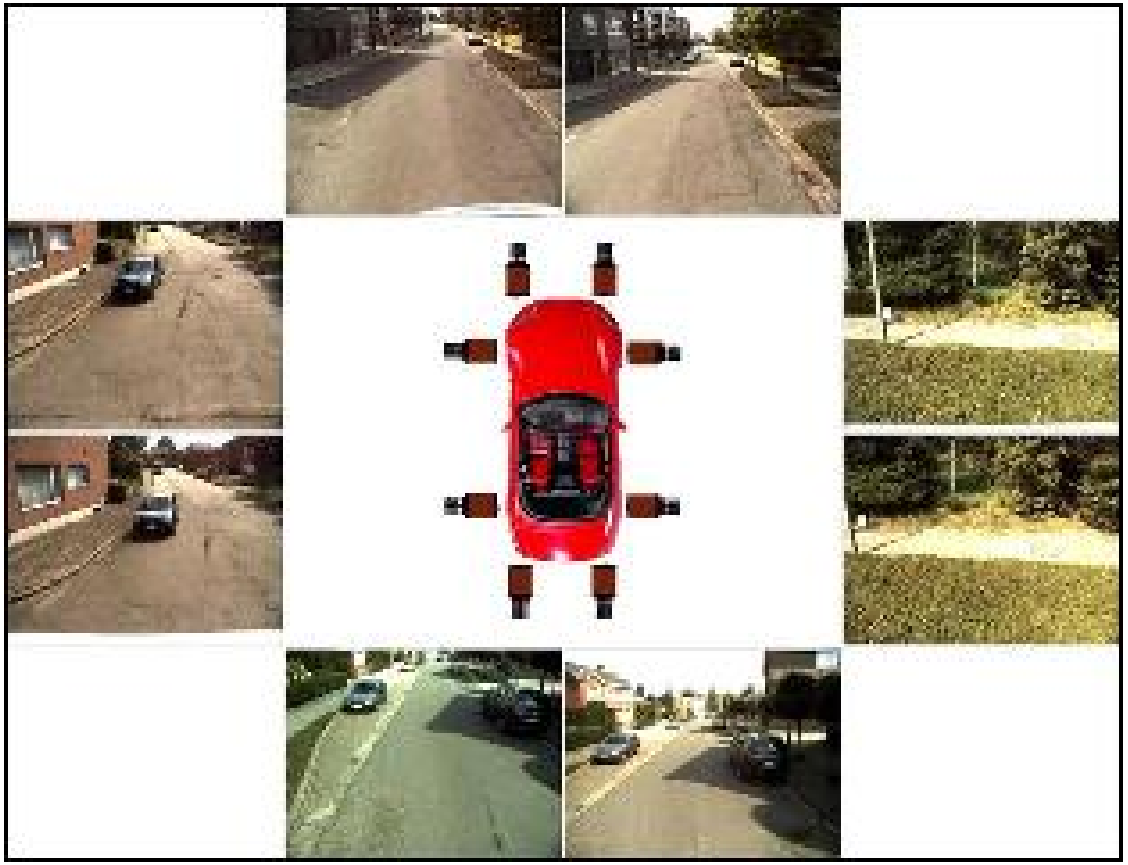
Radu Timofte[1], Karel Zimmermann[1], and Luc Van Gool[1,2]

[1]VISICS
Katholieke Universiteit Leuven, Belgium

[2]BIWI
ETH Zurich, Switzerland

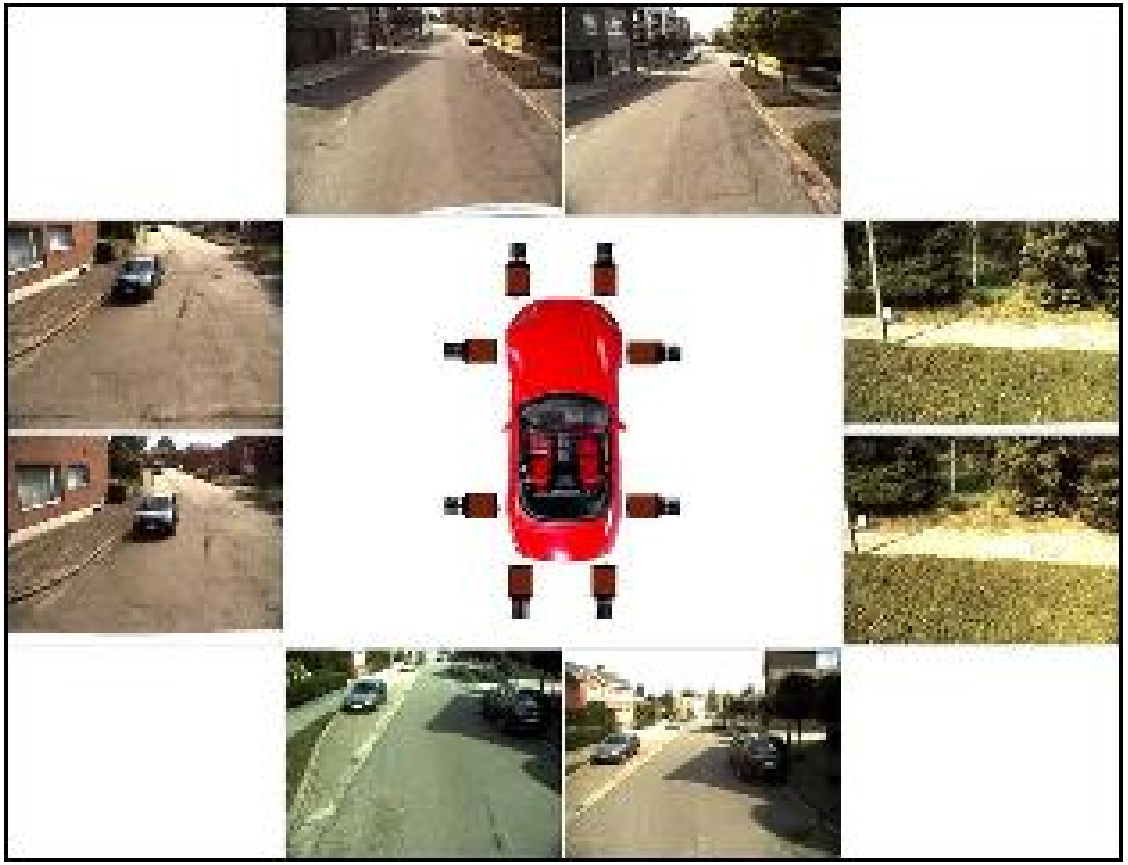
June 30, 2009

Problem definition



- **Input:** Large set of views and corresponding camera locations.
- **Output:** List of detected traffic signs.

Problem definition



- **Input:** Large set of views and corresponding camera locations.
- **Output:** List of detected traffic signs.

Outline

Single-view

- **Segmentation** - very fast bounding box selection process with $FN \longrightarrow 0$.
  - Traffic signs are designed to be well distinguishable from background $\Rightarrow$ have distinctive colors and shapes.
- **Detection** - Adaboost classifiers of bounding boxes.
- **Recognition** - based on SVM classifiers.

Multi-view

- **Global optimization** - over single-view detections constrained by 3D geometry.

Outline

Single-view

- **Segmentation** - very fast bounding box selection process with $FN \longrightarrow 0$.
  - Traffic signs are designed to be well distinguishable from background $\Rightarrow$ have distinctive colors and shapes.
- **Detection** - Adaboost classifiers of bounding boxes.
- **Recognition** - based on SVM classifiers.

Multi-view

- **Global optimization** - over single-view detections constrained by 3D geometry.

Outline

Single-view

- **Segmentation** - very fast bounding box selection process with $FN \longrightarrow 0$.
  - Traffic signs are designed to be well distinguishable from background $\Rightarrow$ have distinctive colors and shapes.
- **Detection** - Adaboost classifiers of bounding boxes.
- **Recognition** - based on SVM classifiers.

Multi-view

- **Global optimization** - over single-view detections constrained by 3D geometry.

Outline

Single-view

- **Segmentation** - very fast bounding box selection process with $FN \longrightarrow 0$.
  - Traffic signs are designed to be well distinguishable from background $\Rightarrow$ have distinctive colors and shapes.
- **Detection** - Adaboost classifiers of bounding boxes.
- **Recognition** - based on SVM classifiers.

Multi-view

- **Global optimization** - over single-view detections constrained by 3D geometry.

Color-based segmentation (thresholding)

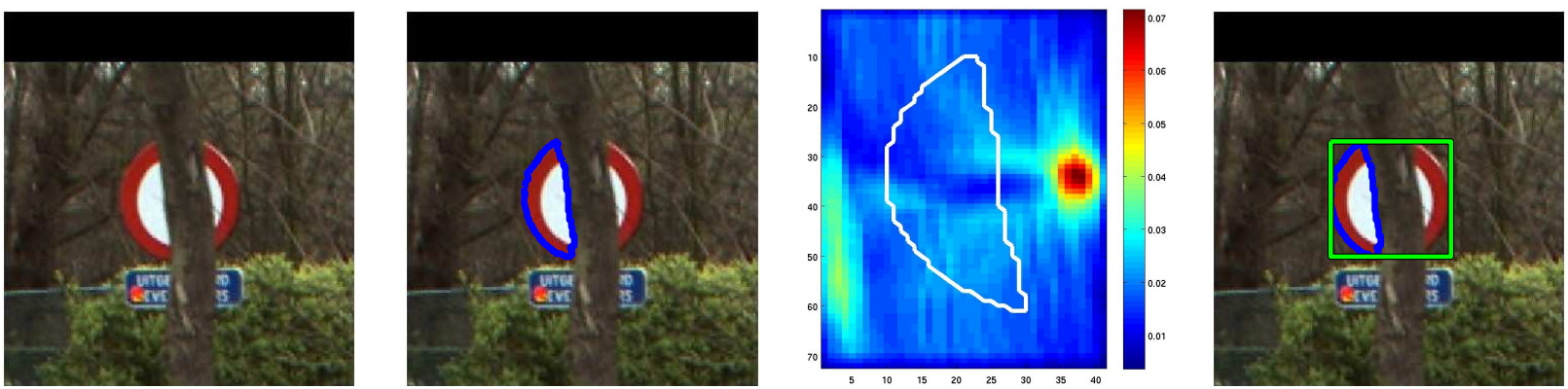
- Estimation of connected components of a thresholded image ($T = [0.5, 0.2, -0.4, 1.0]^\top$)

Original image	**Thresholded image $I(T)$**	**Connected components**	**Segmented bound. boxes**
			

Shape-based segmentation

- Searching for specific shapes (rectangles, circles, triangles).
- \+ Not all the traffic signs are locally threshold separable.
- \- More time consuming, many responses for small shapes (every small region is approximatelly some basic shape).

Original image | **Segmented region** | **Hough accumulator** | **Refined bound. box**



Learning segmentation

- There are thousands of possible settings of such methods e.g. different projections from color space.
- Learning is searching for a reasonable subset of these methods/settings.
- **Optimal trade-off among FN, FP and the number of methods.**

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} FP(\mathcal{T}) + K_1 \cdot FN(\mathcal{T}) + K_2 \cdot \mathrm{card}(\mathcal{T}))$$

- Boolean Linear Programming selects ≈ 50 methods out of 10000 in 2 hours.
- **Segmentation results** for example:

$$FN_{BB} = 1.5\%, FP = 3443/\ 2\mathrm{Mpxl\ image}, (FN_{TS} = 0.5\%)$$

Learning segmentation

- There are thousands of possible settings of such methods e.g. different projections from color space.
- Learning is searching for a reasonable subset of these methods/settings.
- **Optimal trade-off among FN, FP and the number of methods.**

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} FP(\mathcal{T}) + K_1 \cdot FN(\mathcal{T}) + K_2 \cdot \mathrm{card}(\mathcal{T}))$$

- Boolean Linear Programming selects ≈ 50 methods out of 10000 in 2 hours.
- **Segmentation results** for example:

$$FN_{BB} = 1.5\%, FP = 3443/\ 2\mathrm{Mpxl\ image}, (FN_{TS} = 0.5\%)$$

Learning segmentation

- There are thousands of possible settings of such methods e.g. different projections from color space.
- Learning is searching for a reasonable subset of these methods/settings.
- **Optimal trade-off among FN, FP and the number of methods.**

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} FP(\mathcal{T}) + K_1 \cdot FN(\mathcal{T}) + K_2 \cdot \mathrm{card}(\mathcal{T}))$$

- Boolean Linear Programming selects ≈ 50 methods out of 10000 in 2 hours.
- **Segmentation results** for example:

$$FN_{BB} = 1.5\%, FP = 3443/\ 2\mathrm{Mpxl\ image}, (FN_{TS} = 0.5\%)$$

Learning segmentation

- There are thousands of possible settings of such methods e.g. different projections from color space.
- Learning is searching for a reasonable subset of these methods/settings.
- **Optimal trade-off among FN, FP and the number of methods.**

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} FP(\mathcal{T}) + K_1 \cdot FN(\mathcal{T}) + K_2 \cdot \mathrm{card}(\mathcal{T}))$$

- Boolean Linear Programming selects ≈ 50 methods out of 10000 in 2 hours.
- **Segmentation results** for example:

$$FN_{BB} = 1.5\%, FP = 3443/\ 2\mathrm{Mpxl\ image}, (FN_{TS} = 0.5\%)$$

Learning segmentation

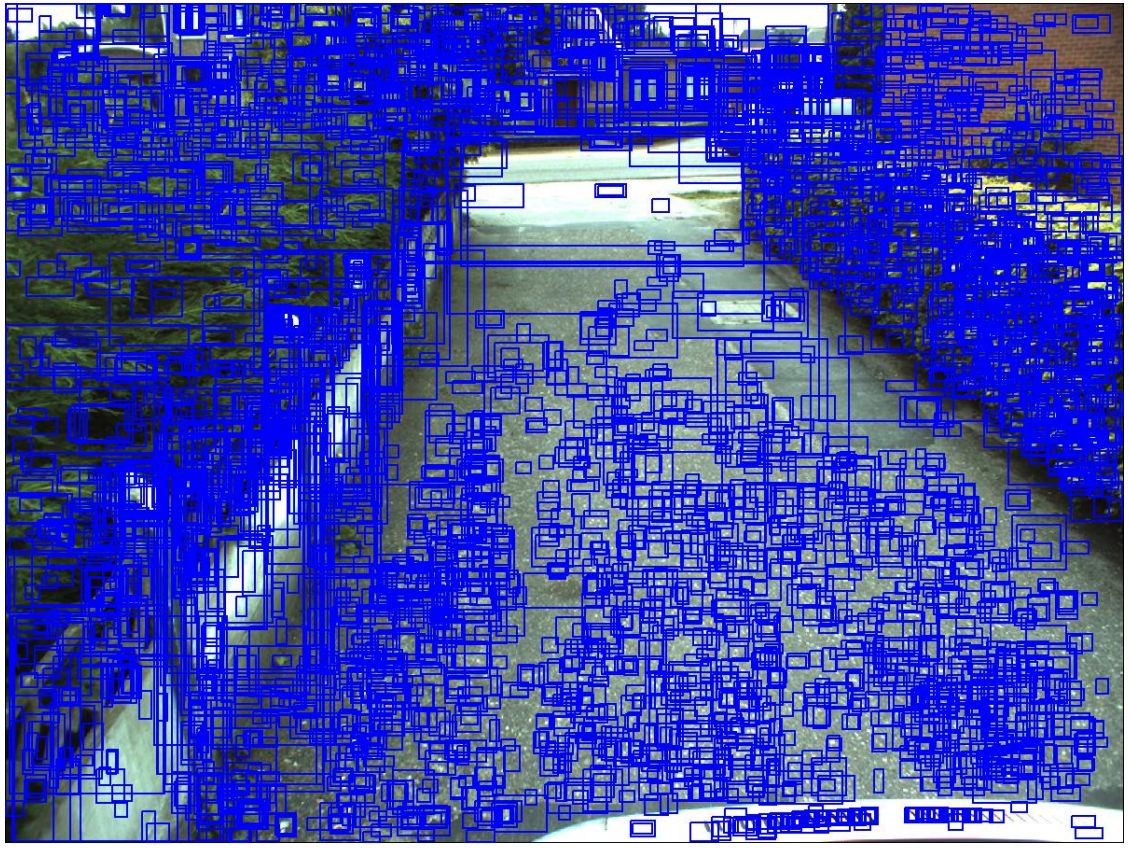
- There are thousands of possible settings of such methods e.g. different projections from color space.
- Learning is searching for a reasonable subset of these methods/settings.
- **Optimal trade-off among FN, FP and the number of methods.**

$$\mathcal{T}^* = \arg\min_{\mathcal{T}} FP(\mathcal{T}) + K_1 \cdot FN(\mathcal{T}) + K_2 \cdot \mathrm{card}(\mathcal{T}))$$

- Boolean Linear Programming selects ≈ 50 methods out of 10000 in 2 hours.
- **Segmentation results** for example:

$$FN_{BB} = 1.5\%, FP = 3443/\ 2\mathrm{Mpxl\ image}, (FN_{TS} = 0.5\%)$$

How does the output of **segmentation** look like?

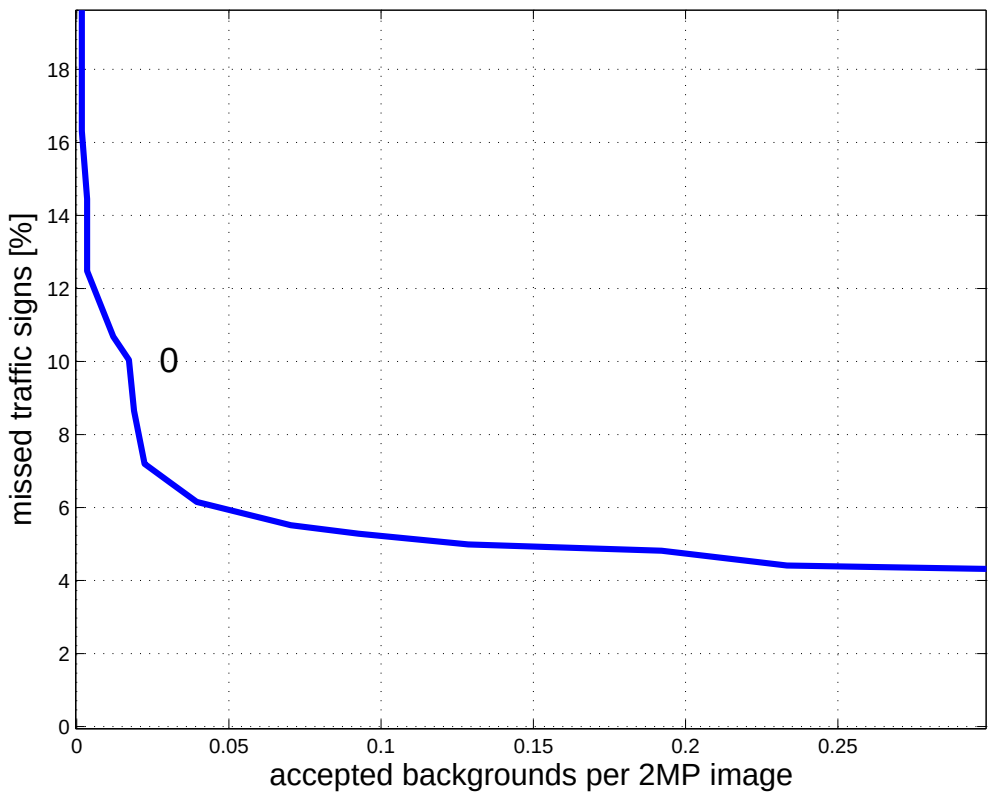


Detection

- **Detection:** suppresion of bounding boxes which does not look like a traffic sign.
  - Haar features computed on each channel of HSI space.
  - Separated shape-specific cascades of Adaboost classifiers.
- Detection (+segmentation) results:

Detection

- **Detection:** suppresion of bounding boxes which does not look like a traffic sign.
  - Haar features computed on each channel of HSI space.
  - Separated shape-specific cascades of Adaboost classifiers.
- **Detection (+segmentation) results**:



How does the output of **detection** look like?

3D optimization - introduction

- Single view detection and recognition is just preprocessing, the final decision is the subject of the global optimization over multiple views.
- The idea is based on Minimum Description Length, i.e. explaining detected bounding boxes by the lowest number of real world traffic signs.
- If detections satisfy some *visual and geometrical constraints*, then all of these detections are explainable by one real world traffic sign.

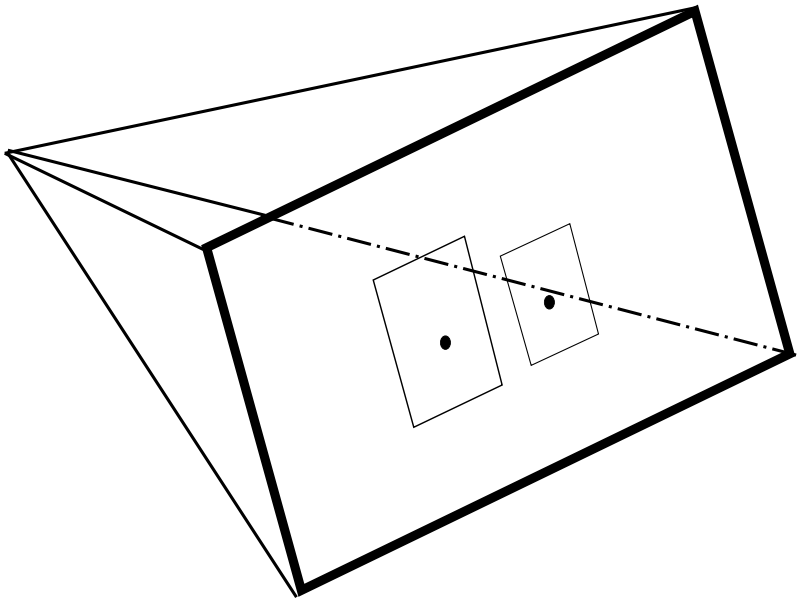
3D optimization - introduction

- Single view detection and recognition is just preprocessing, the final decision is the subject of the global optimization over multiple views.
- The idea is based on Minimum Description Length, i.e. explaining detected bounding boxes by the lowest number of real world traffic signs.
- If detections satisfy some *visual and geometrical constraints*, then all of these detections are explainable by one real world traffic sign.

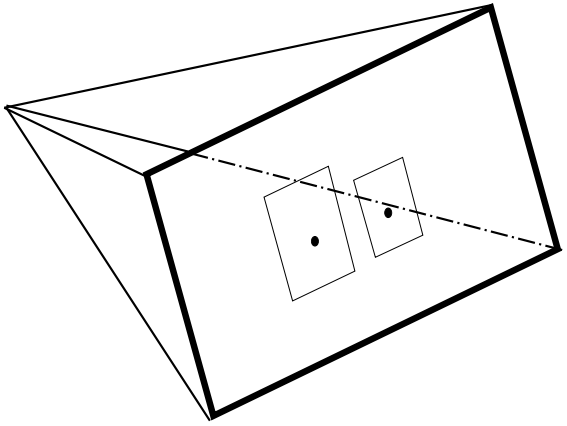
3D optimization - introduction

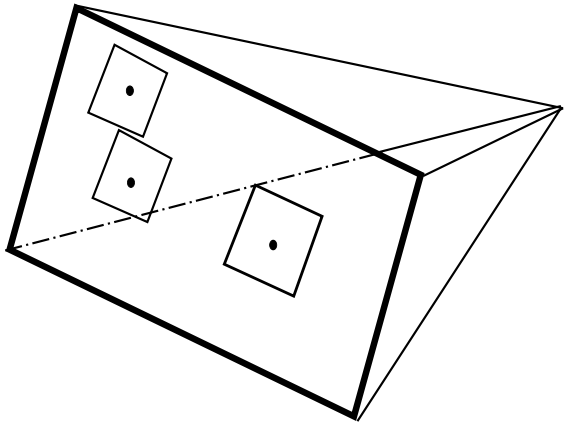
- Single view detection and recognition is just preprocessing, the final decision is the subject of the global optimization over multiple views.
- The idea is based on Minimum Description Length, i.e. explaining detected bounding boxes by the lowest number of real world traffic signs.
- If detections satisfy some *visual and geometrical constraints*, then all of these detections are explainable by one real world traffic sign.

Minimum Description Length in 3D

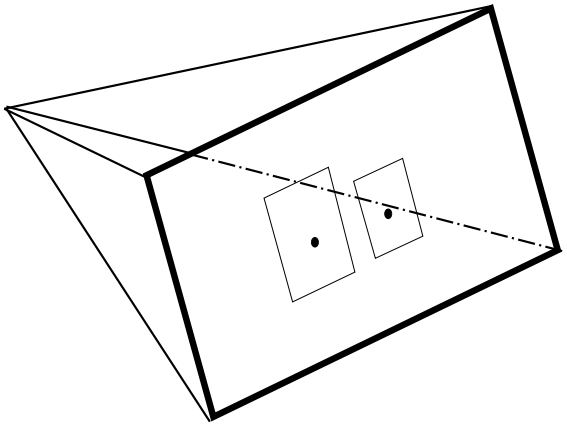


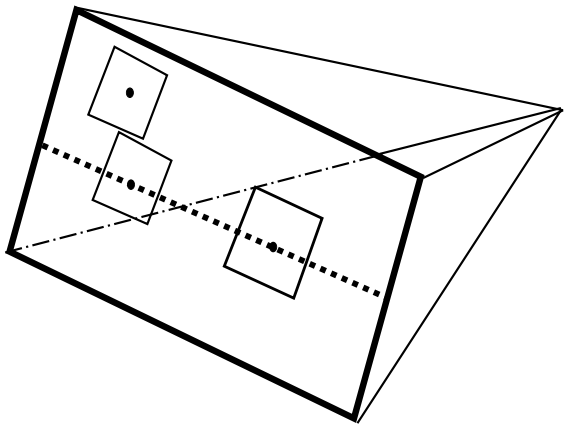
Minimum Description Length in 3D



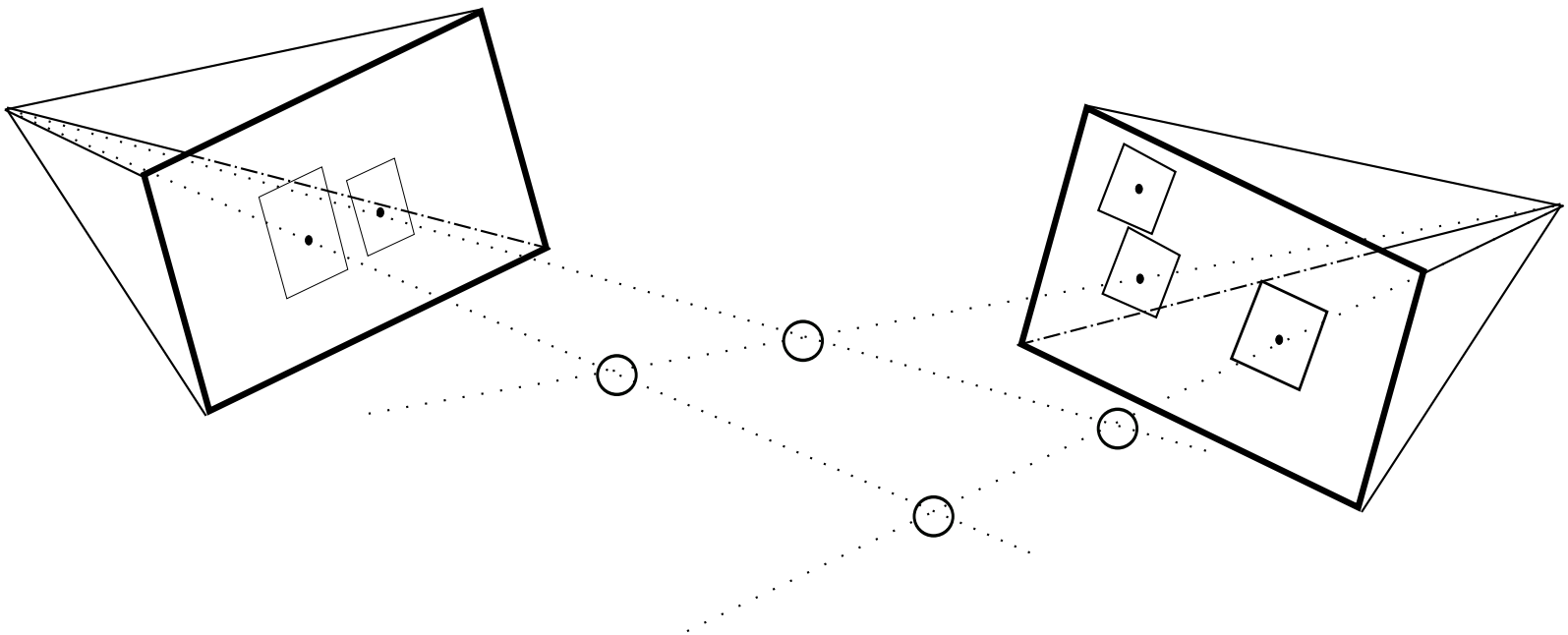


Minimum Description Length in 3D

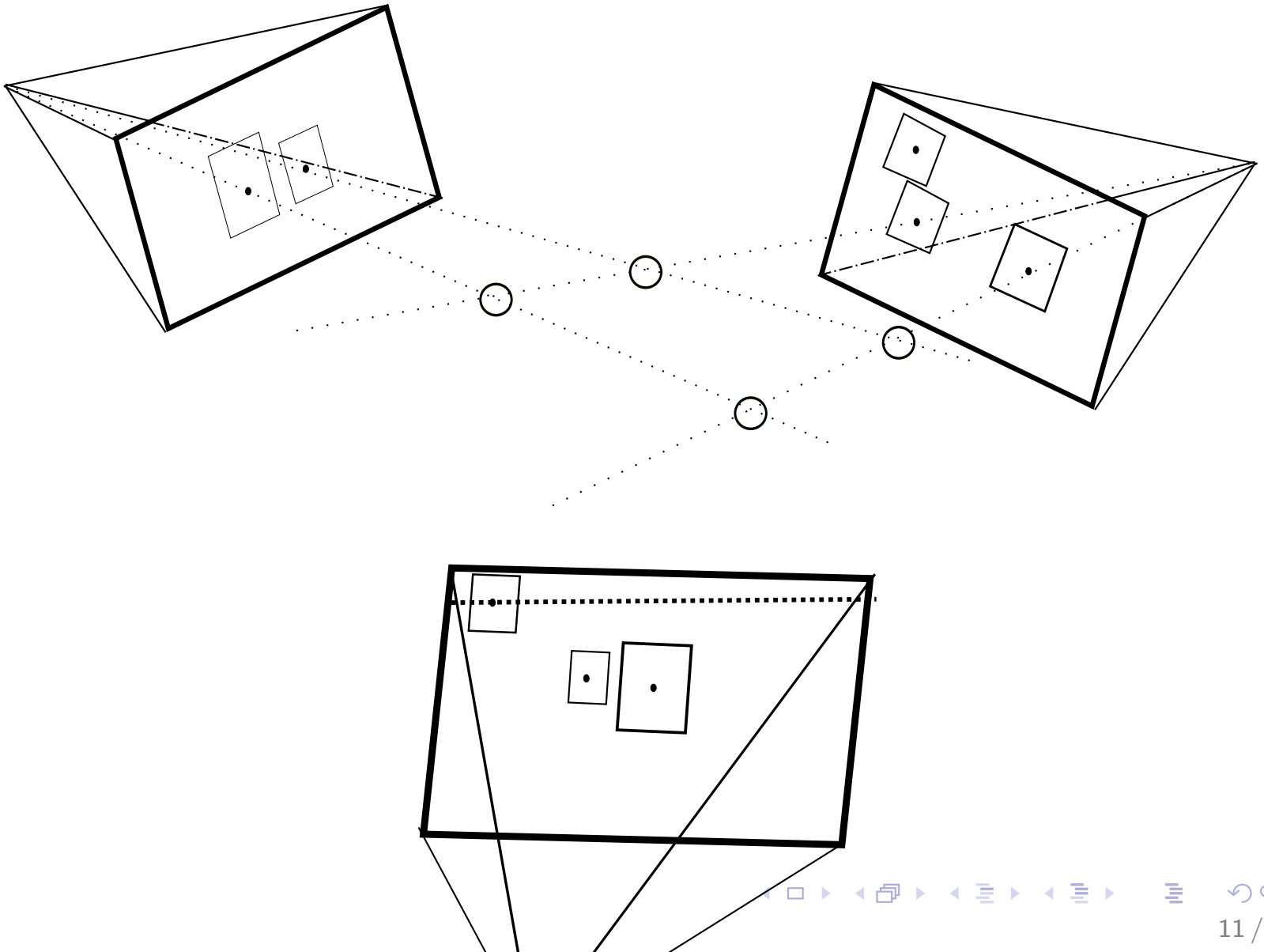




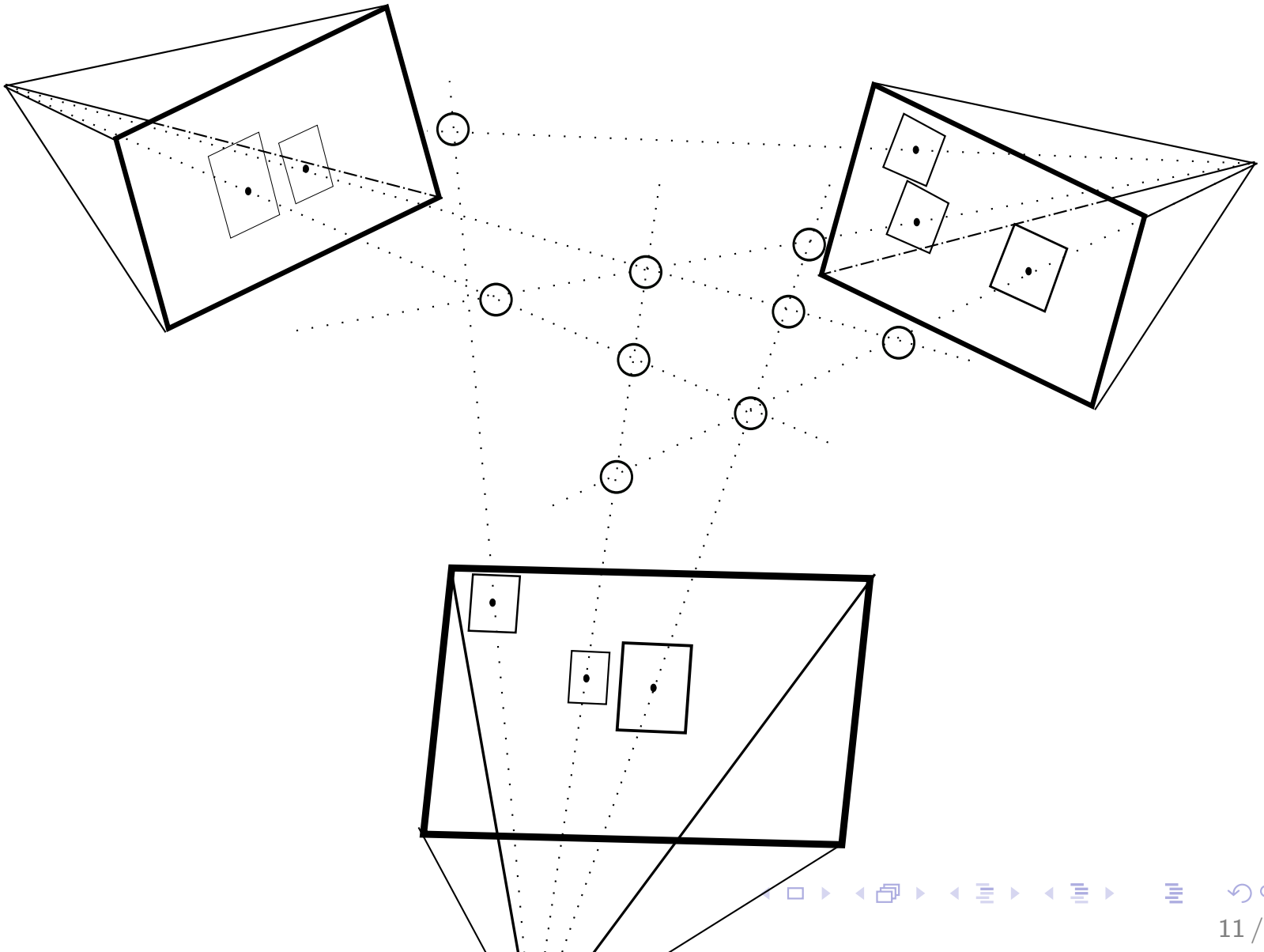
Minimum Description Length in 3D



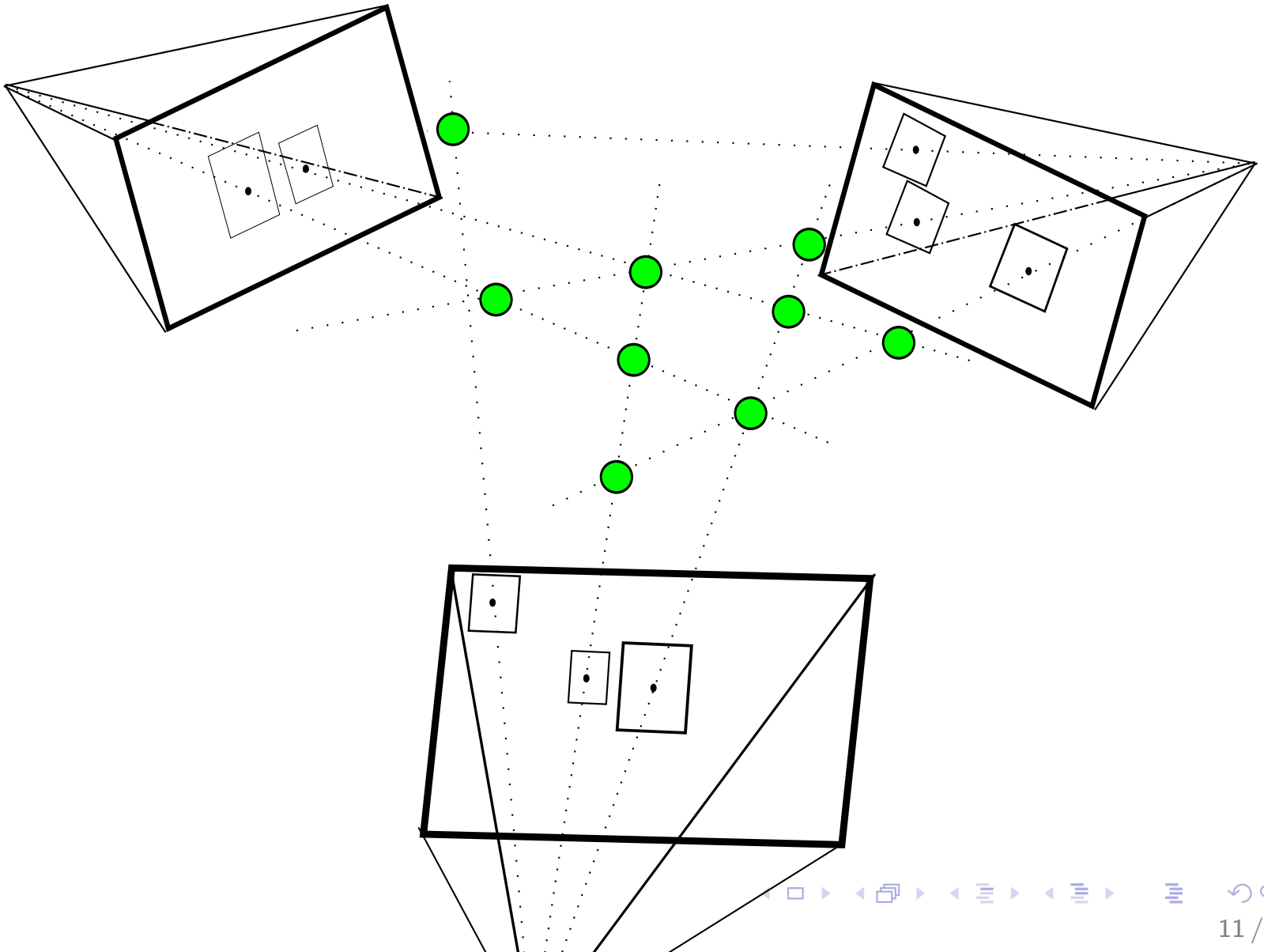
Minimum Description Length in 3D



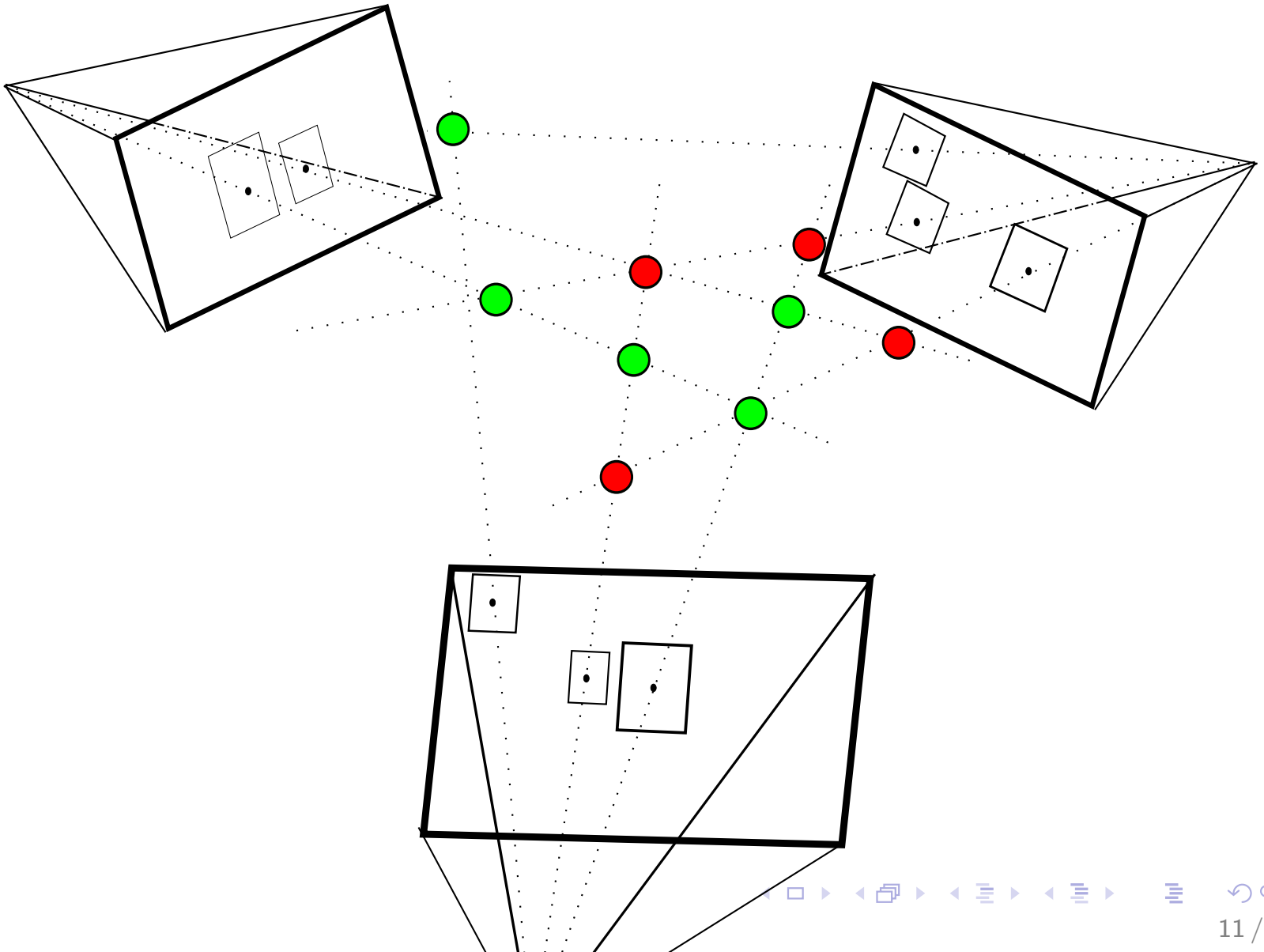
Minimum Description Length in 3D



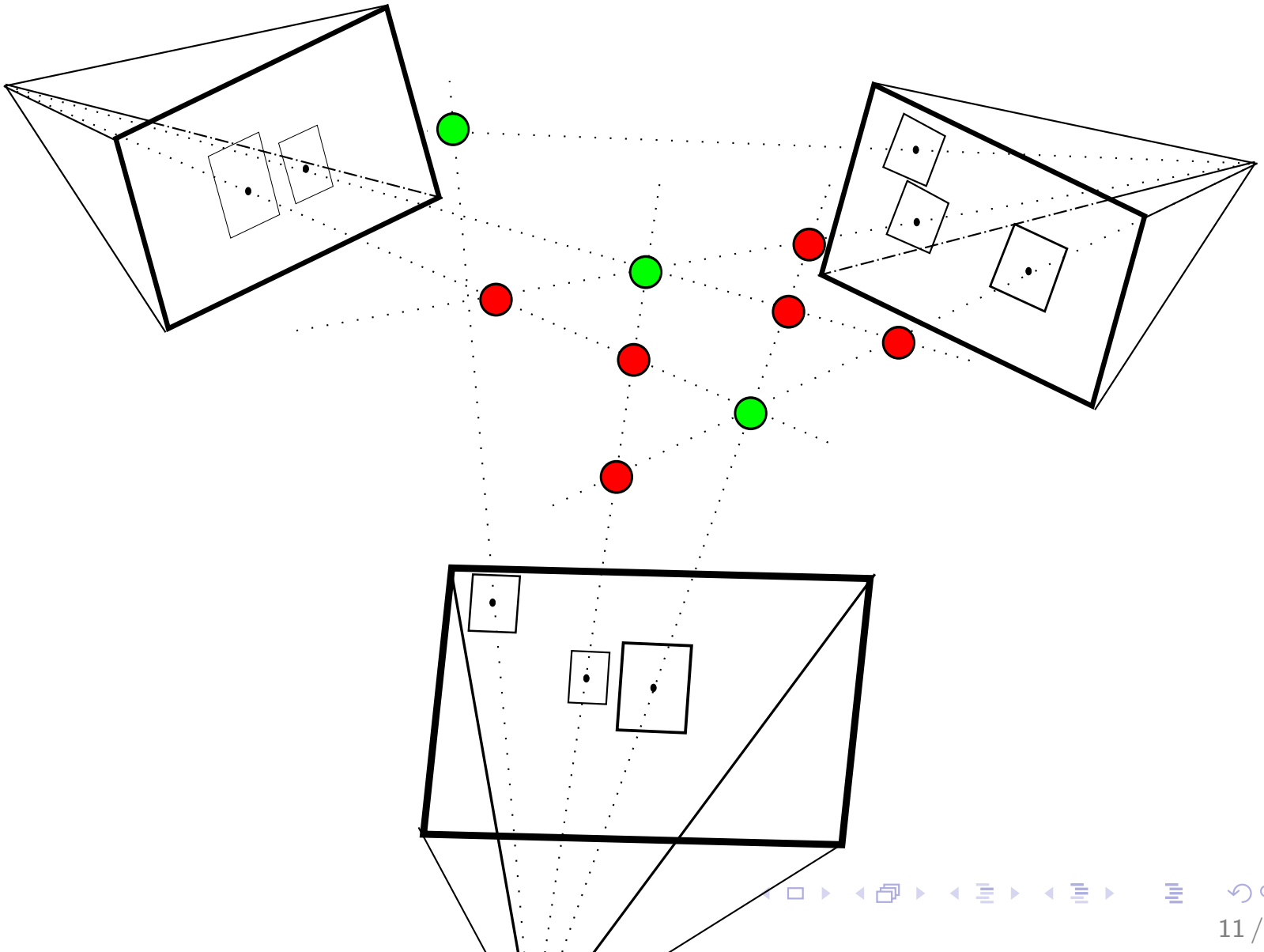
Minimum Description Length in 3D



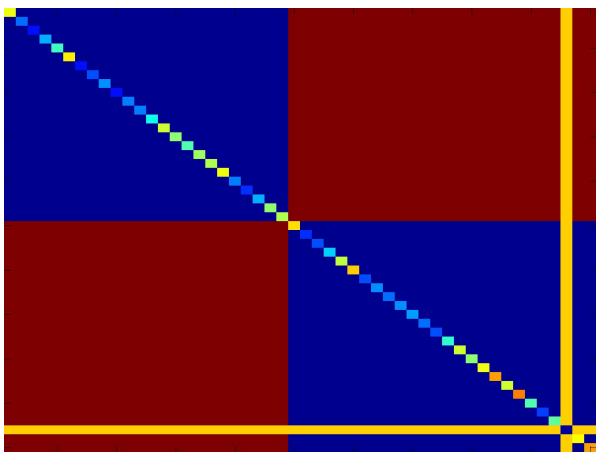
Minimum Description Length in 3D



Minimum Description Length in 3D



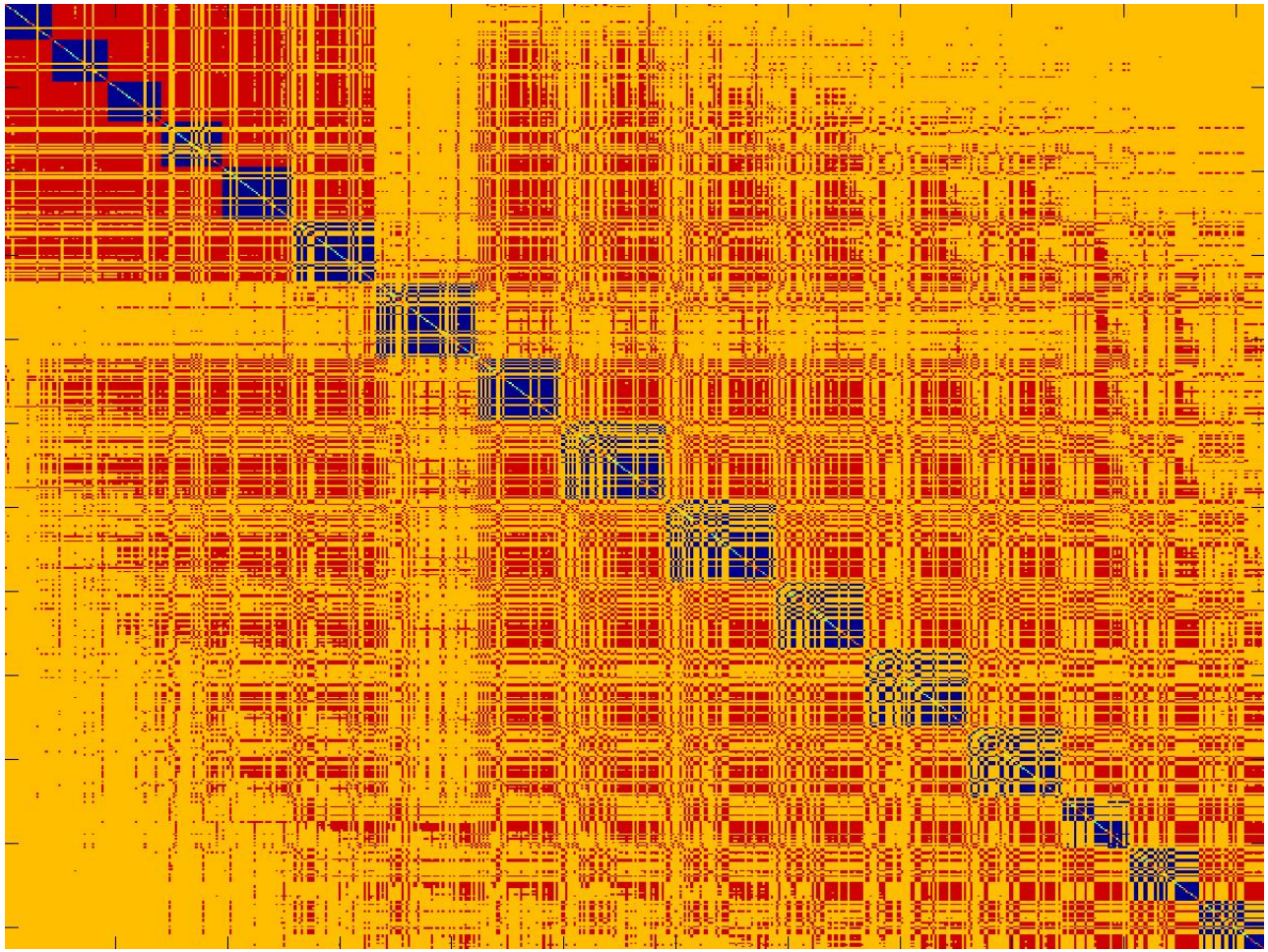
Problem formulation



$$\max_{\mathbf{x} \in \{0,1\}} \mathbf{x}^\top \cdot \cdot \mathbf{x}$$

Example with 16 views

Example with 16 views



Results

- The summary of 3D results:

#	No.frames/TSs	3D Localised TS	FP	Recognised TS
1	8×3001 /78	75(96.2%)	3	74(98.7%)
2	8×6201 /71	68(95.8%)	7	65(95.6%)
3	8×2001 /44	41(93.2%)	2	41(100%)
4	8×4001 /76	73(96.1%)	8	71(97.3%)
$\sum$	8×15204 /269	257(95.6%)	20	251(97.7%)

- The average accuracy of 3D localisation is of 24.54 cm.

Results

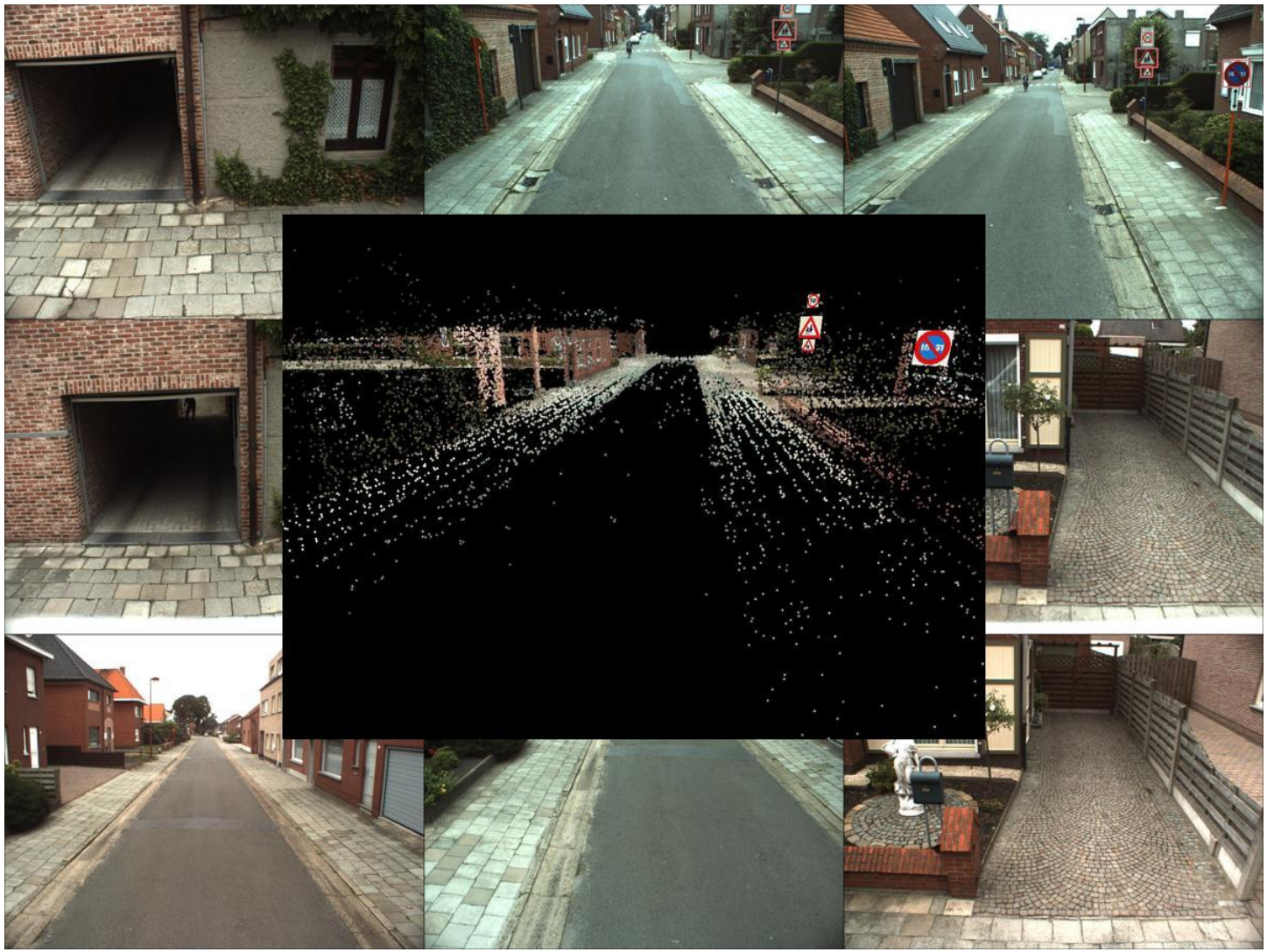
- The summary of 3D results:

#	No.frames/TSs	3D Localised TS	FP	Recognised TS
1	8×3001 /78	75(96.2%)	3	74(98.7%)
2	8×6201 /71	68(95.8%)	7	65(95.6%)
3	8×2001 /44	41(93.2%)	2	41(100%)
4	8×4001 /76	73(96.1%)	8	71(97.3%)
$\sum$	8×15204 /269	257(95.6%)	20	251(97.7%)

- The average accuracy of 3D localisation is of 24.54 cm.

Visualisation of 3D results in one camera

3D visualisation



Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

Conclusions

- Traffic Sign Detection, Recognition and 3D Localisation is a challenging problem
- We propose a multi-view scheme, which combines 2D and 3D analysis
- The main contributions are:
  - Boolean Linear Programming formulation for fast candidate extraction in 2D
  - Minimum Description Length formulation for best 3D hypothesis selection
- Work in progress...

- Questions?