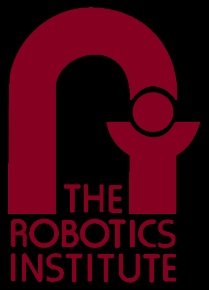
Direct Pose Estimation and Refinement

Hatem Alismail
halismai@cs.cmu.edu

Thesis Oral
July 28, 2016



Brett Browning (Co-chair)
Simon Lucey (Co-Chair)
Michael Kaess
Martial Hebert
Ian D. Reid, The University of Adelaide

Promised Completed Work

Chapters 5 and 6

Analysis of the proposed idea of direct alignment of binary descriptors (Bit-Planes)

Promised Completed Work

Chapters 5 and 6

Analysis of the proposed idea of direct alignment of binary descriptors (Bit-Planes)

Chapter 6

Applications to visual odometry in poorly lit underground mines

Promised Completed Work

Chapters 5 and 6

Analysis of the proposed idea of direct alignment of binary descriptors (Bit-Planes)

Chapter 6

Applications to visual odometry in poorly lit underground mines

Chapter 7

Direct (photometric) refinement of pose and structure jointly [*Bundle adjustment without correspondences*]

Promised Completed Work

Chapters 5 and 6

Analysis of the proposed idea of direct alignment of binary descriptors (Bit-Planes)

Chapter 6

Applications to visual odometry in poorly lit underground mines

Chapter 7

Direct (photometric) refinement of pose and structure jointly [*Bundle adjustment without correspondences*]

Chapter 4 (extra)

Evaluation of important details in direct visual odometry

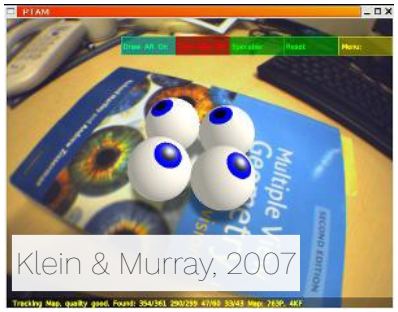
Contributions

- Robust and real-time pose estimation in challenging environments
  - Near darkness, blur, specular reflections, …
- First formulation of a direct (photometric) bundle adjustment for VSLAM
  - No correspondences required

Geometric Pose Estimation

One of the most successful applications of Computer Vision

AR/VR

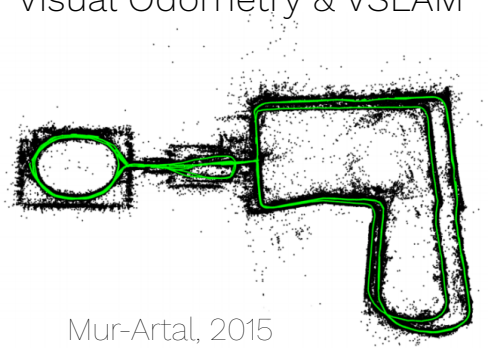


Klein & Murray, 2007

Structure-from-Motion (SFM)

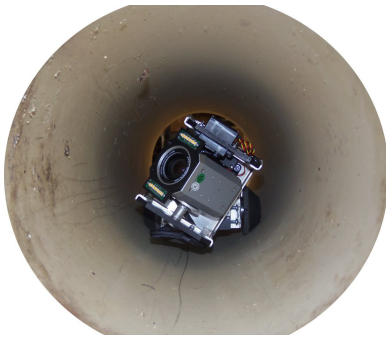
Agarwal, 2010

Visual Odometry & VSLAM



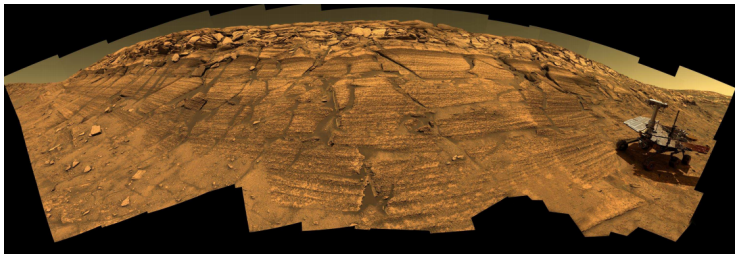
Mur-Artal, 2015

Robotic Inspection



http://www.superdroidrobots.com/

Space Exploration



Maimone, 2004

Autonomous Driving

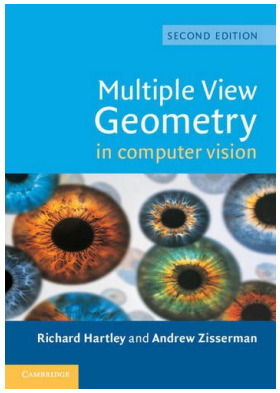
Google

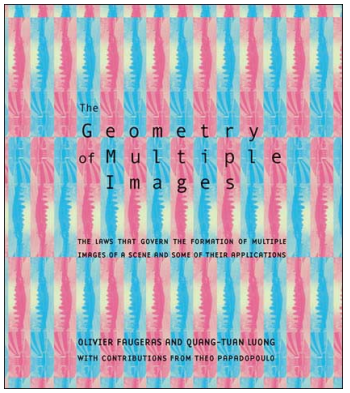
Pose Estimation too

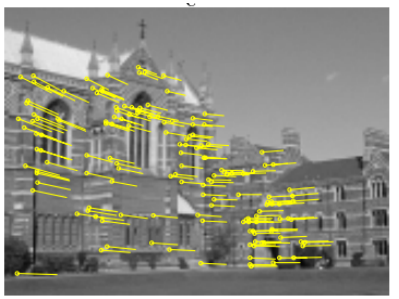
Image credit thewardrobedoor.com / Google

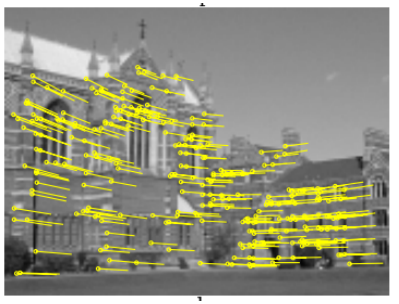
Feature-based Pipeline

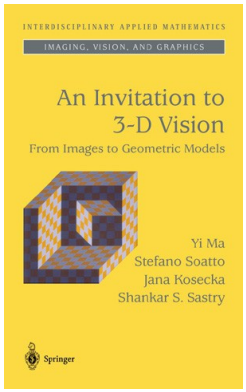
Major Contributor to the success of geometric pose estimation

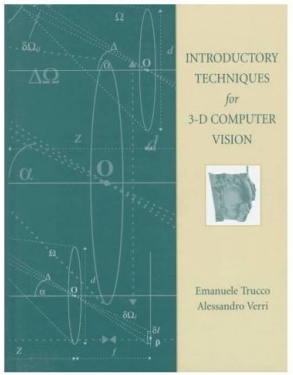


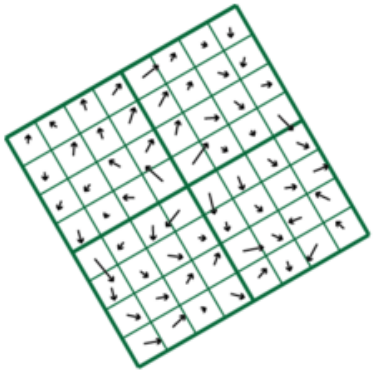


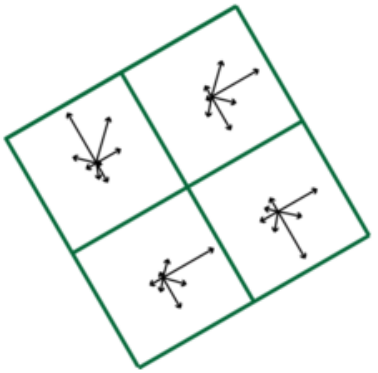




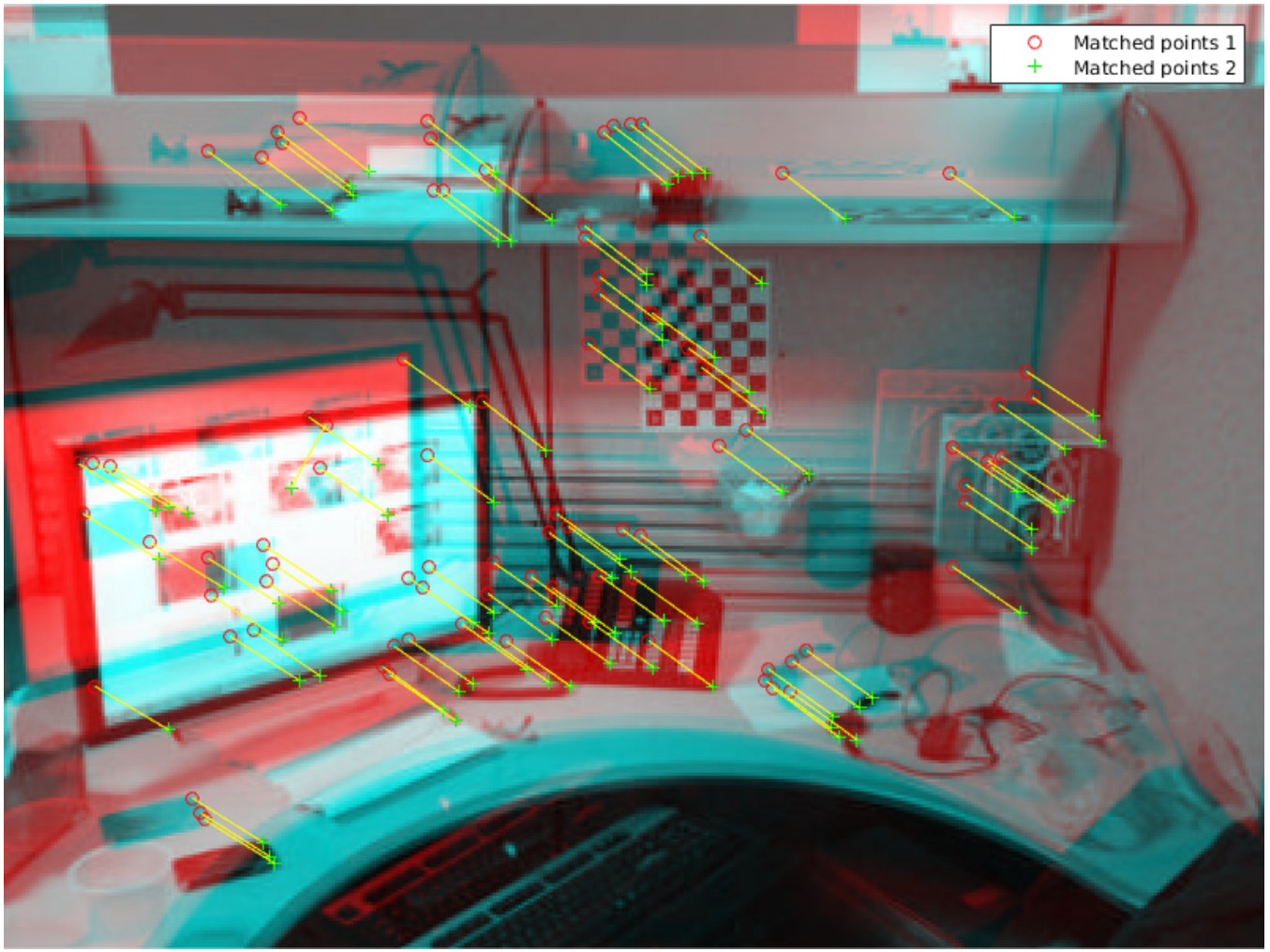




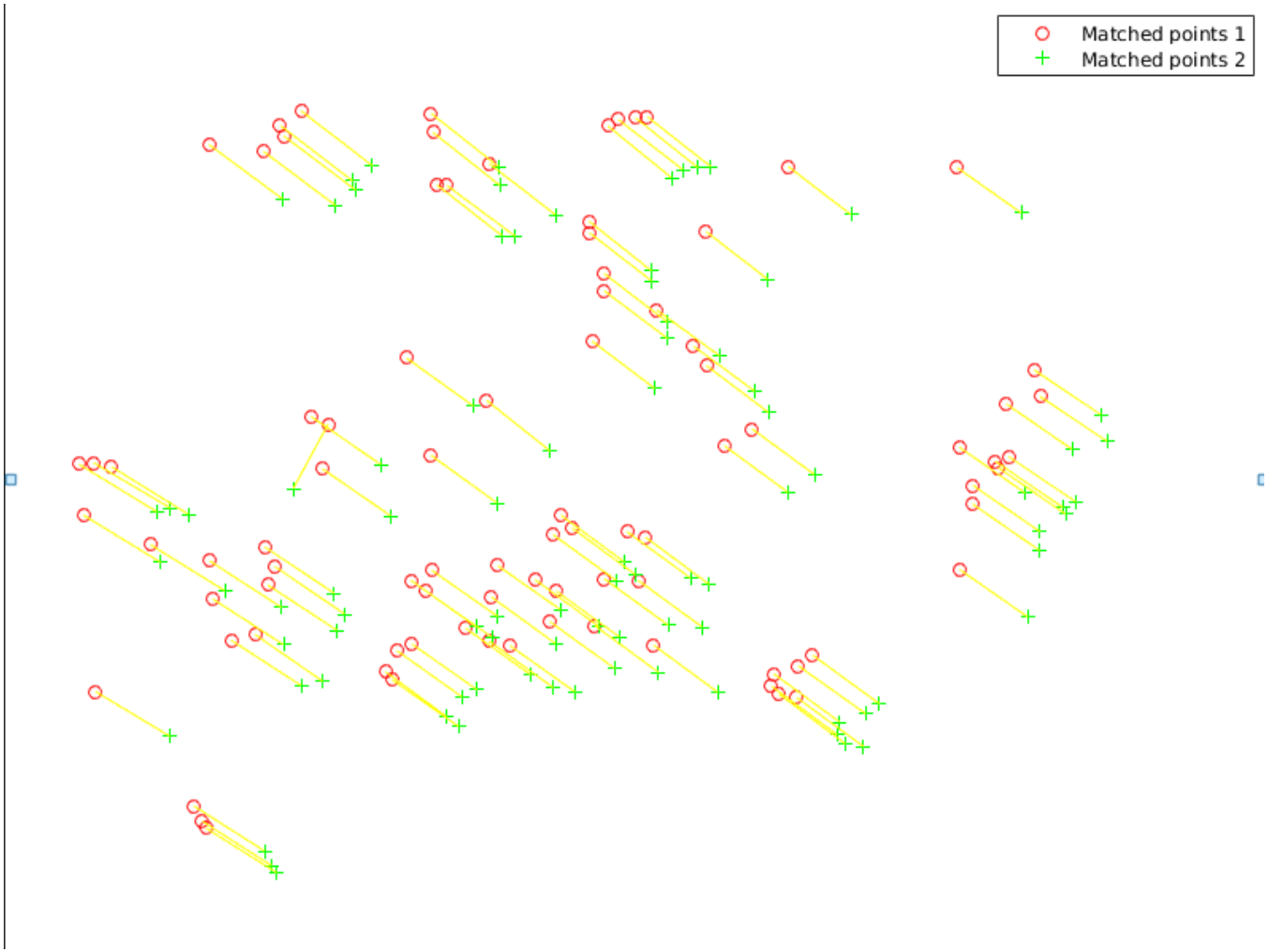




Feature-based Pipeline

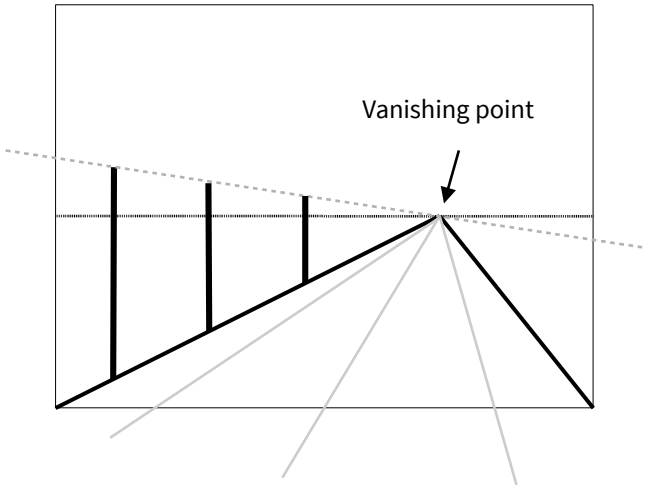


Feature-based Pipeline



Feature-based Pipeline

Natural transition from images to geometry



Viewpoint invariance

Mikolajczyk, 2007

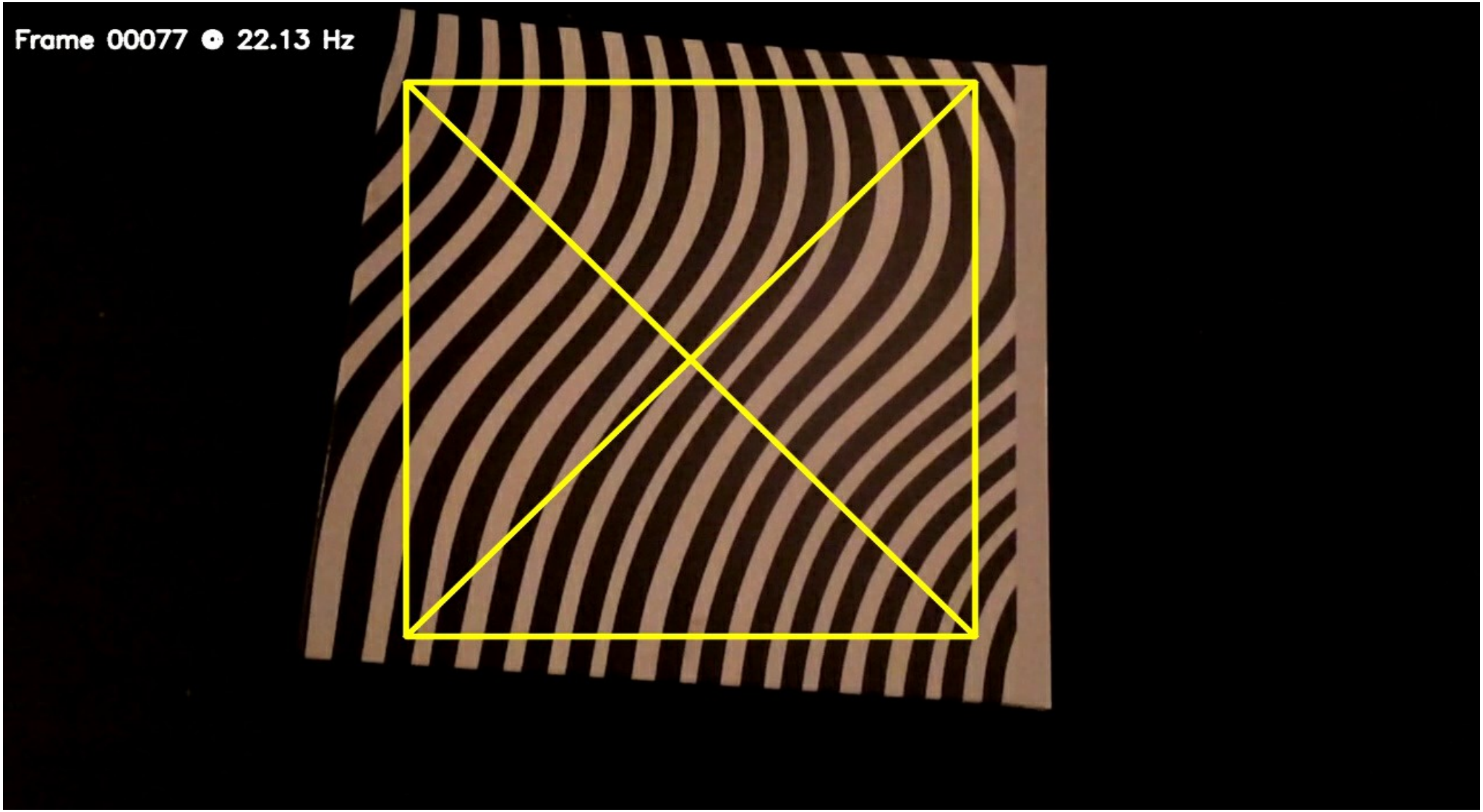
Illumination invariance

Mikolajczyk, 2007

Limitations of Keypoints

Limitations of Keypoints

Limitations of Keypoints



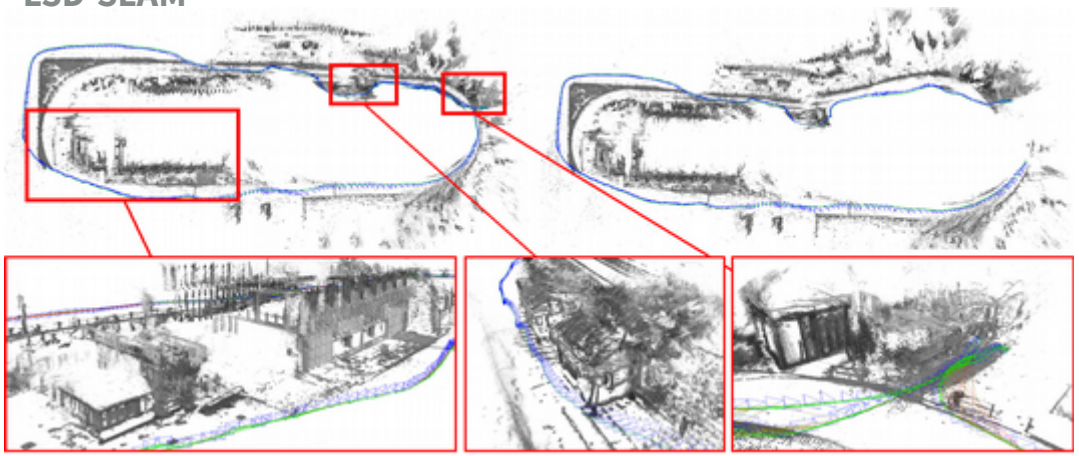
A step towards vision in challenging environments

http://hci.iwr.uni-heidelberg.de/benchmarks/Challenging_Data_for_Stereo_and_Optical_Flow

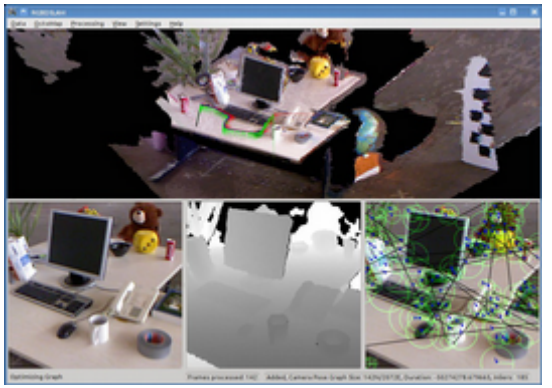
Direct Methods

Increasing availability of high frame-rate data

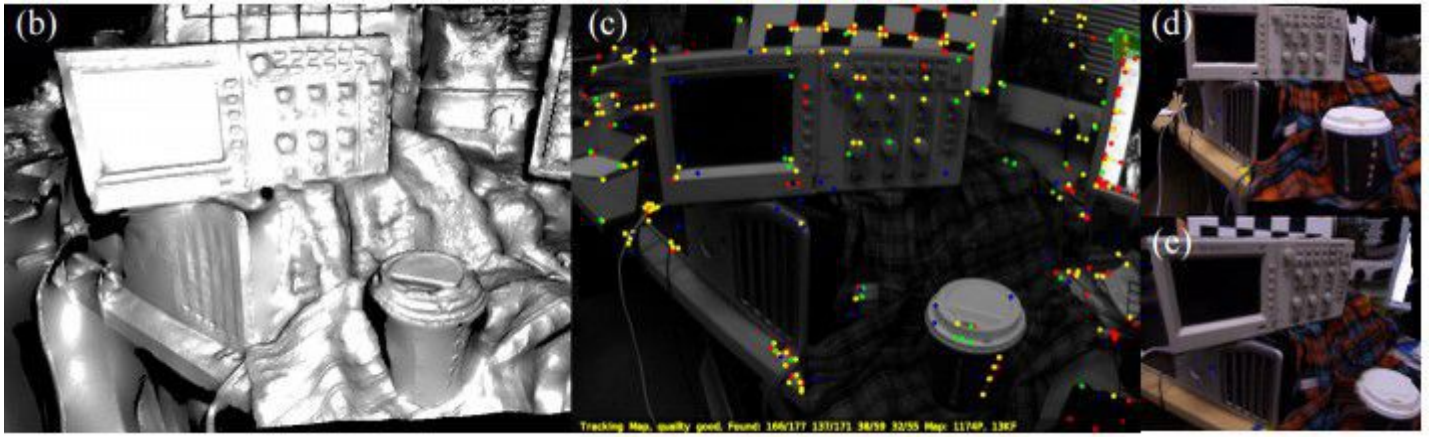
LSD-SLAM



RGBD-SLAM

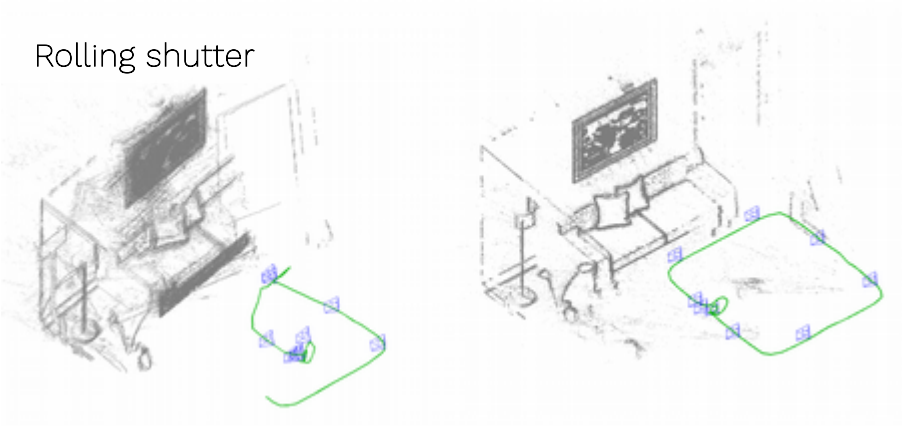


DTAM



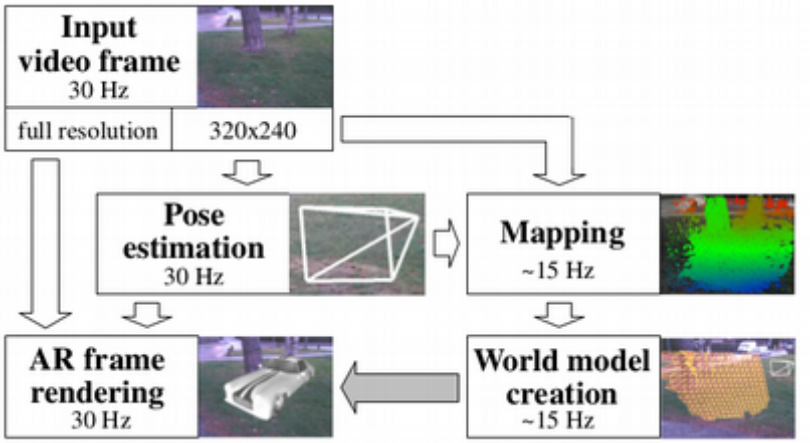
Increased Interest in Direct Methods

Rolling shutter



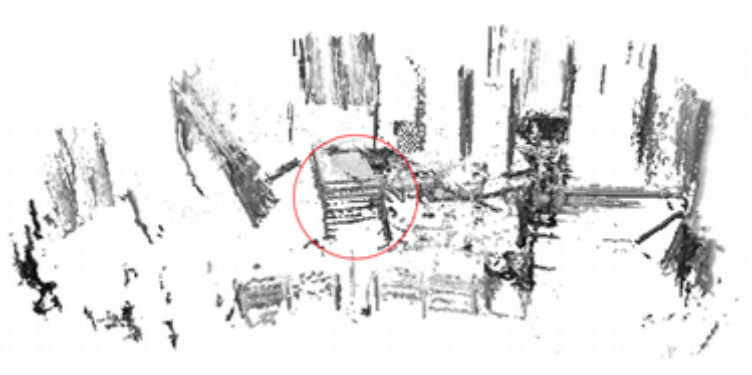
Kim, Cadena, Reid 2016

Mobile phones



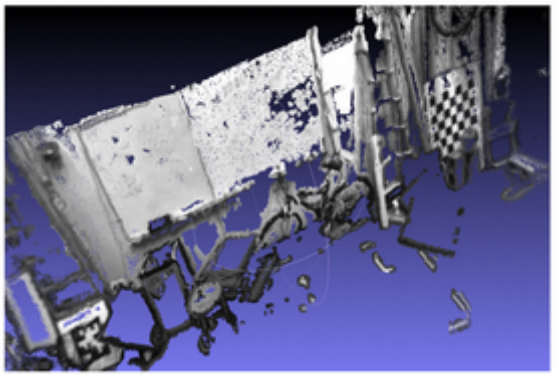
Schopes, Engel, Cremers, 2014

Visual-Inertial stereo SLAM



Usenko, Engel, Stuckler, Cremers, 2016

Visual-Inertial Mono SLAM



Concha, Loianno, Kumar, Civera, 2016

Direct Pose Estimation

The Lucas & Kanade Algorithm

Gradient-based search using image data directly to minimize a photometric error

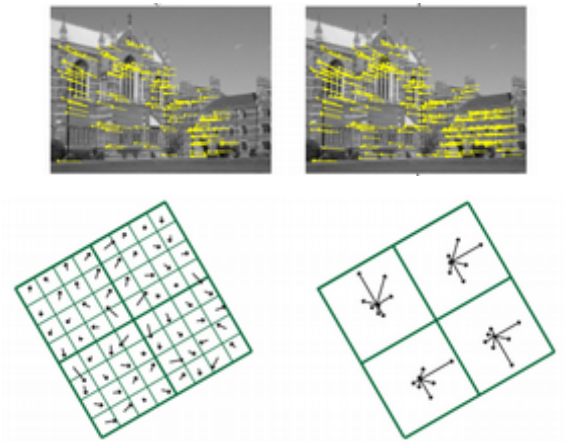
No keypoints necessary

More robust because we can use most of the image

The Correspondence Problem

Feature-based

Use pre-computed corrs.



Direct

Estimate corrs. with pose

Feature Based Methods for Structure and Motion Estimation

P. H. S. Torr[1] and A. Zisserman[2]

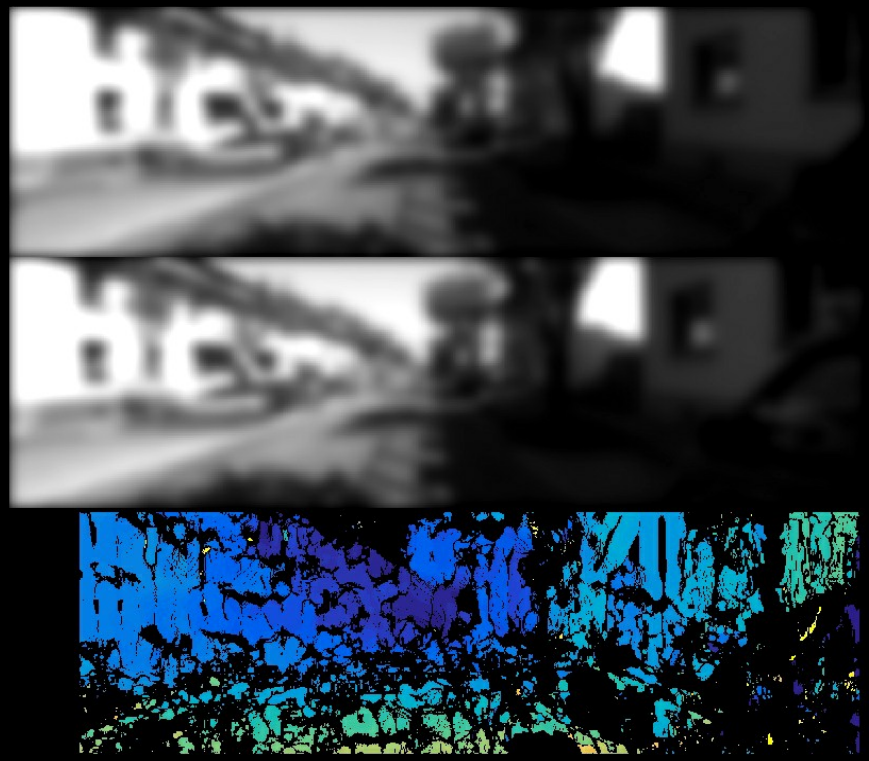
[1] Microsoft Research Ltd, 1 Guildhall St
Cambridge CB2 3NH, UK
philtorr@microsoft.com
[2] Department of Engineering Science, University of Oxford
Oxford, OX1 3PJ, UK
az@robots.ox.ac.uk

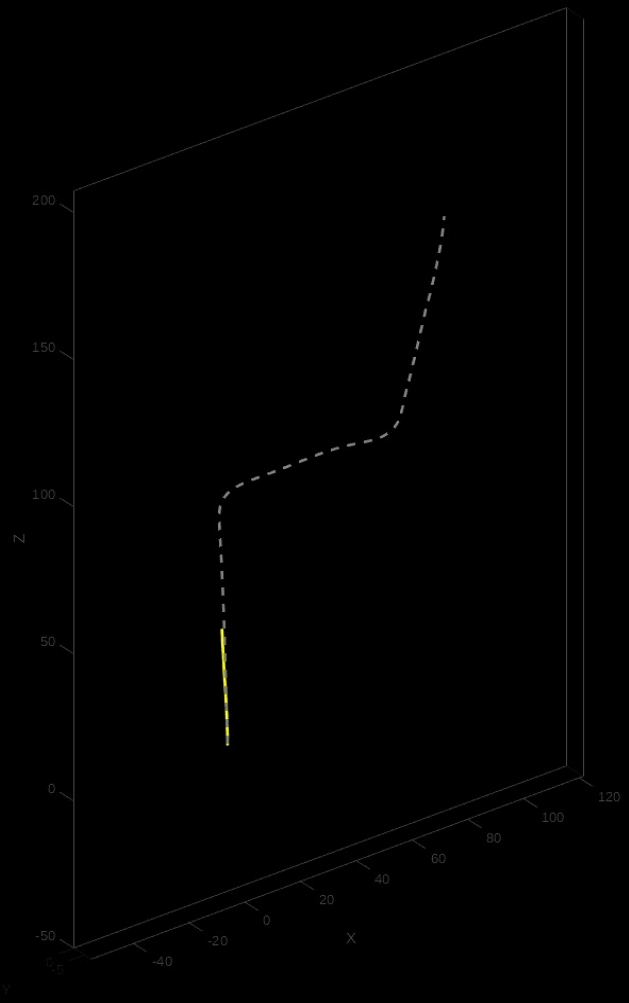
All About Direct Methods

M. Irani[1] and P. Anandan[2]

[1] Dept. of Computer Science and Applied Mathematics,
The Weizmann Inst. of Science, Rehovot, Israel.
irani@wisdom.weizmann.ac.il
[2] Microsoft Research, One Microsoft Way,
Redmond, WA 98052, USA.
anandan@microsoft.com

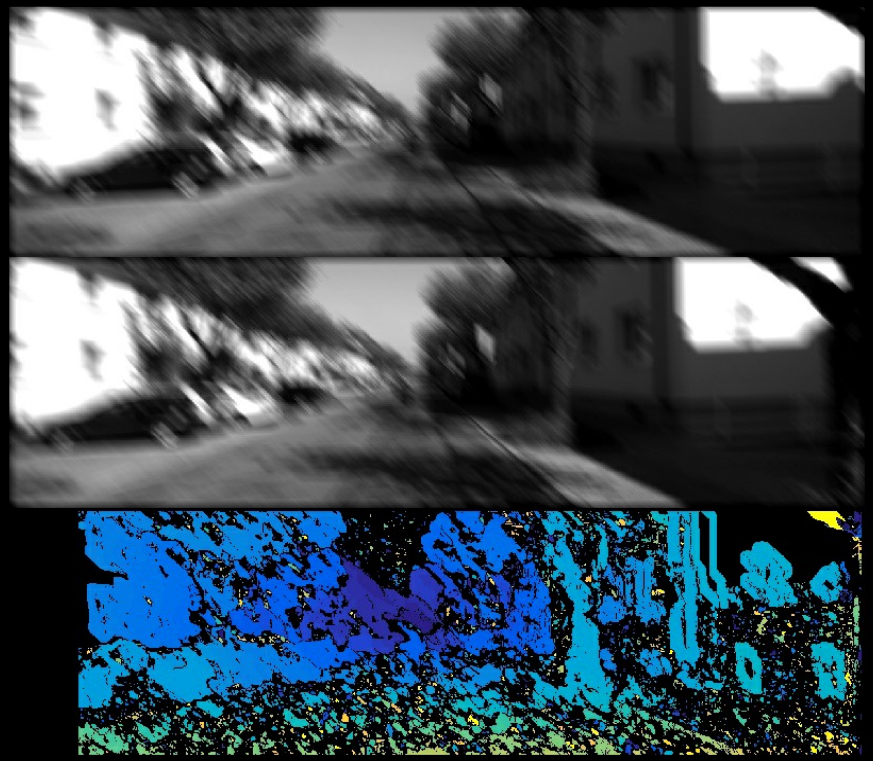
Robustness of Direct Pose Estimation

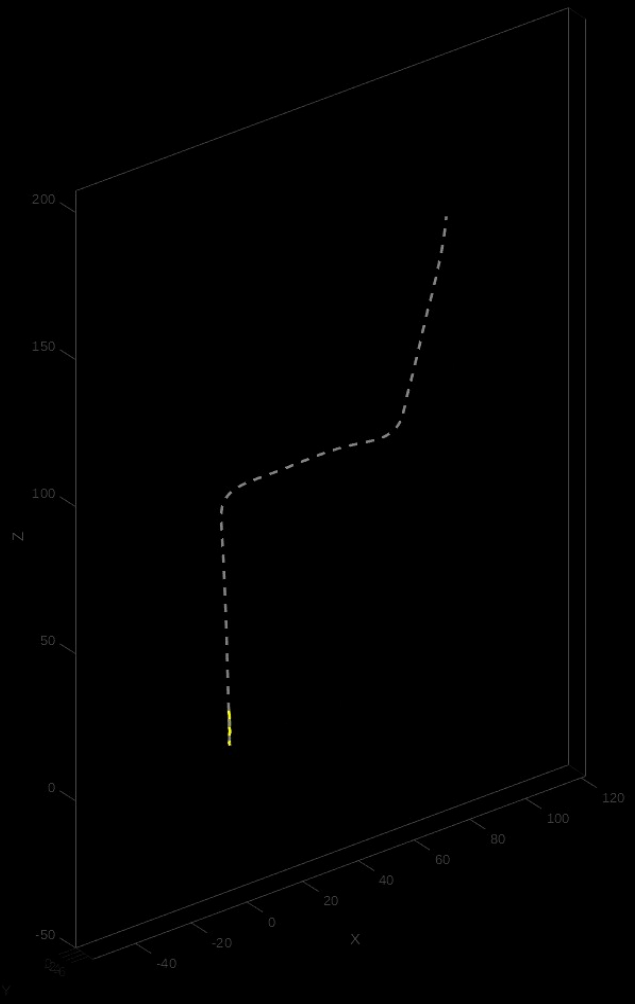




Using the Direct Disparity Space (DDS) algorithm in Chapter 3

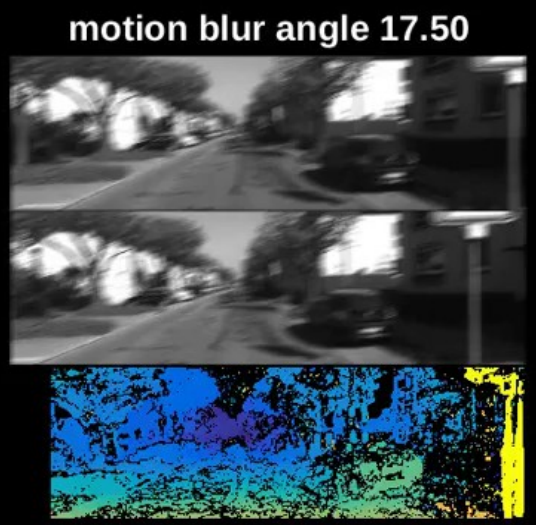
Robustness of Direct Pose Estimation

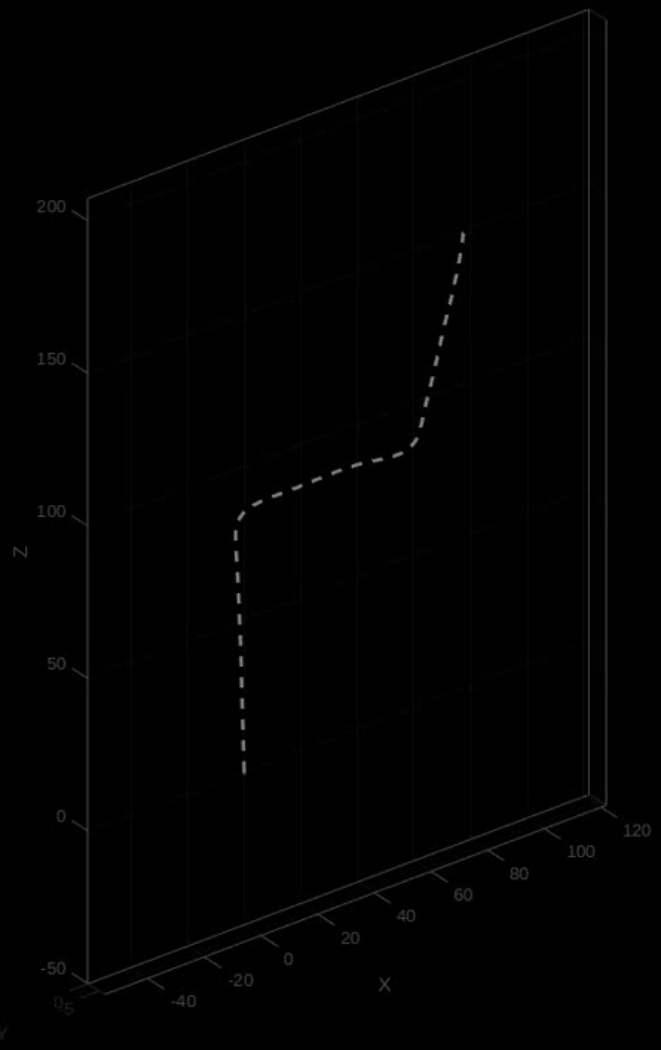




Using the Direct Disparity Space (DDS) algorithm in Chapter 3

Robustness of Direct Pose Estimation

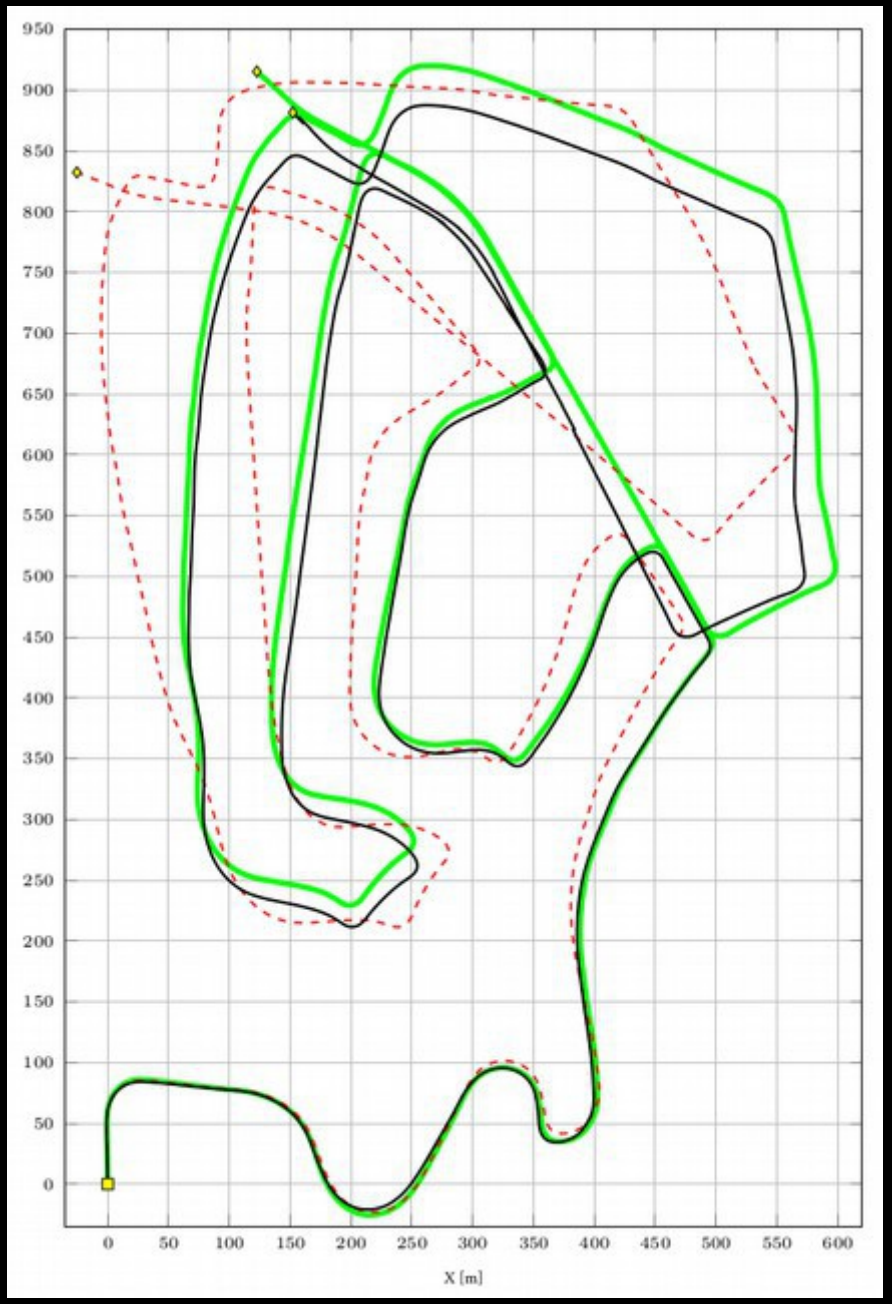




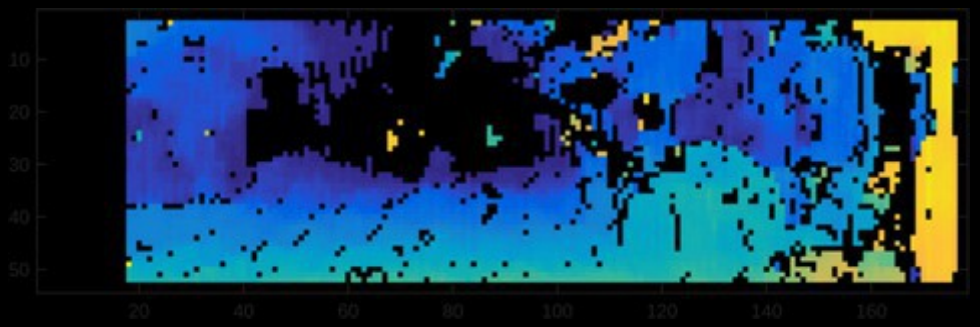
Using the Direct Disparity Space (DDS) algorithm in Chapter 3

Input to Our Algorithm

KITTI with thumbnail stereo

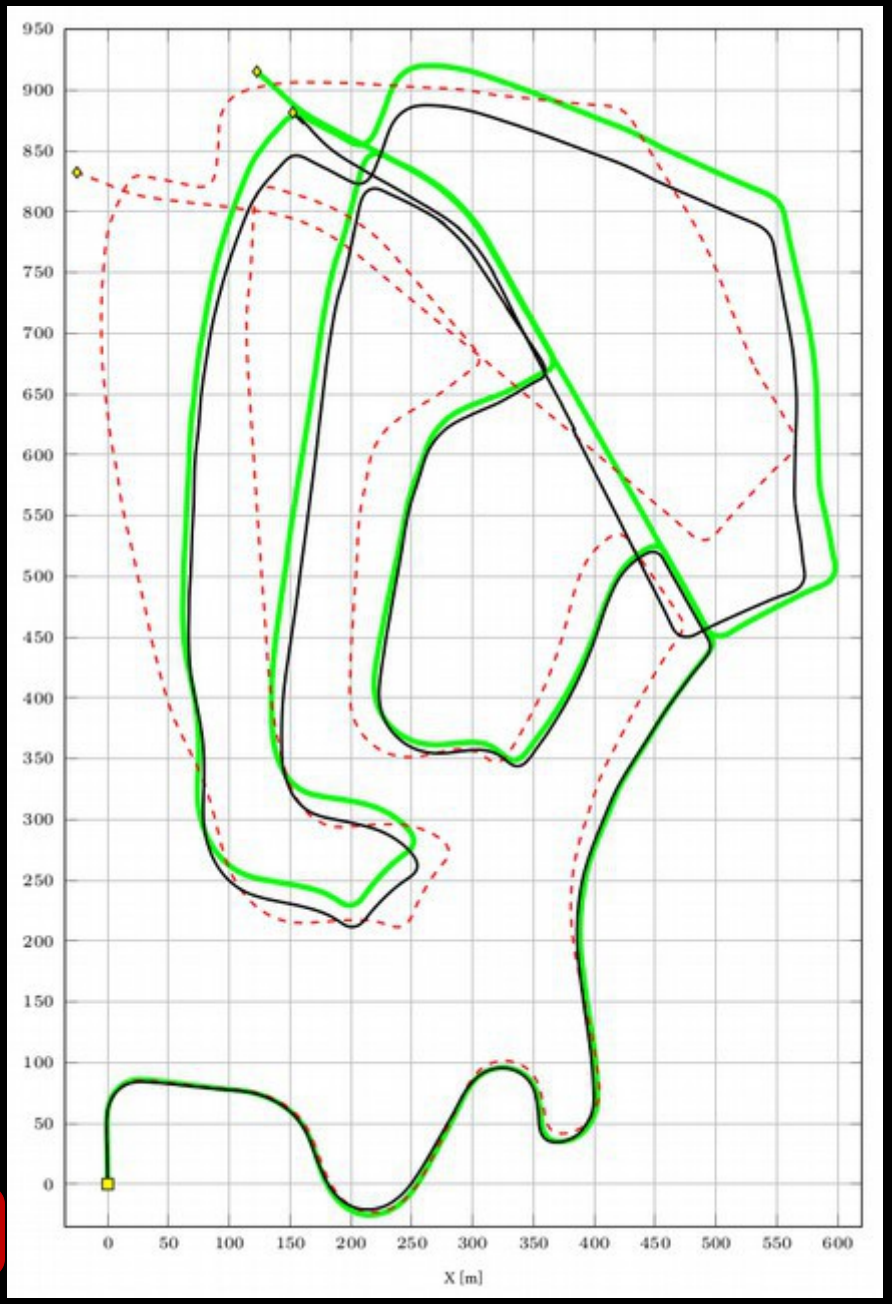


Green, DDS with 178x54 stereo images
Red, VISO with 1241x376 resolution
Black, GPS ground truth

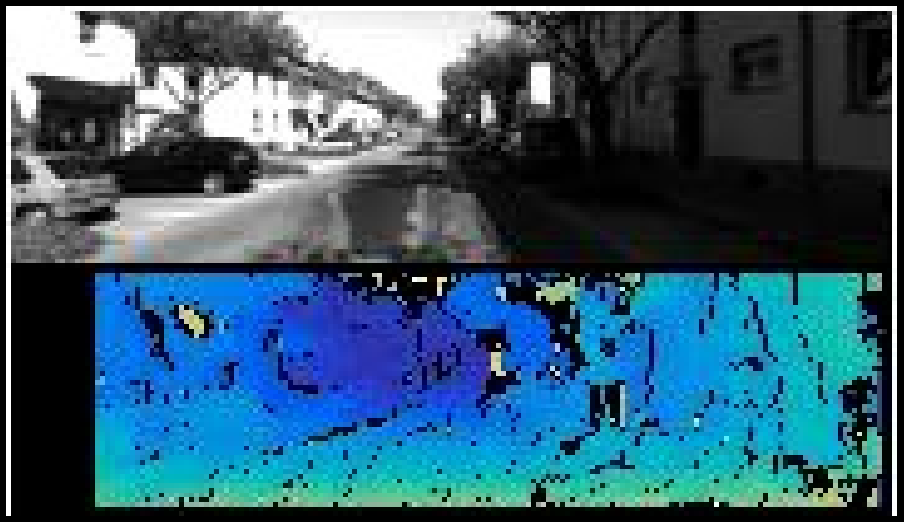


Stereo with opencv block matching
5x5 SAD window and 16 disparities search range

KITTI with thumbnail stereo

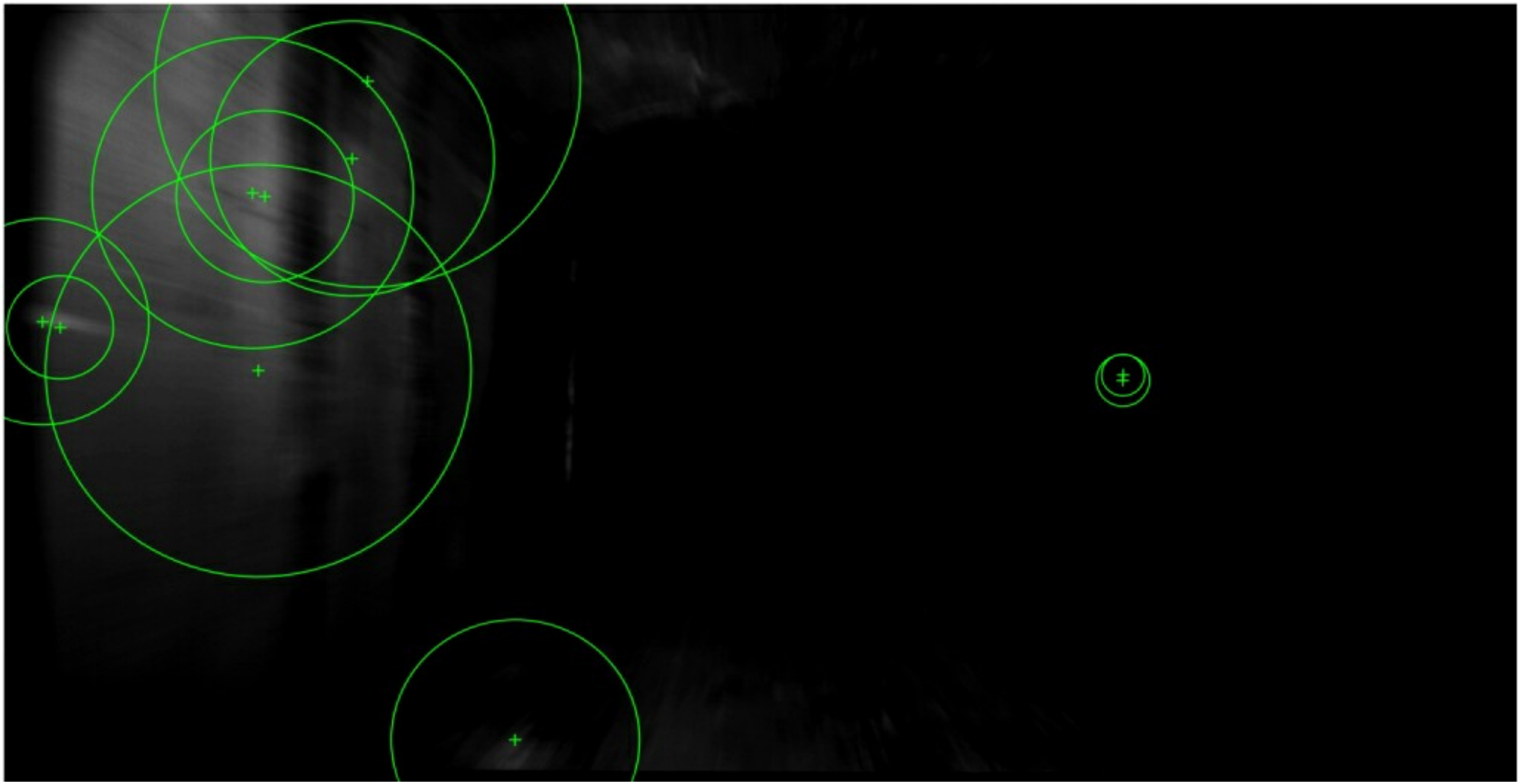


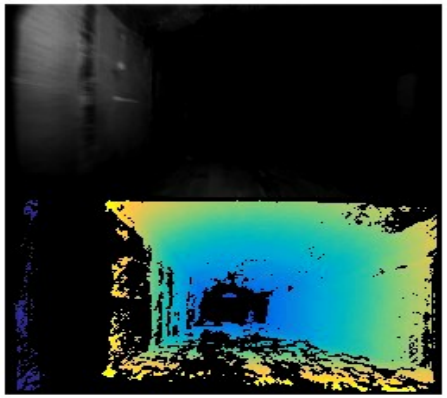
Green, DDS with 178x54 stereo images
Red, VISO with 1241x376 resolution
Black, GPS ground truth

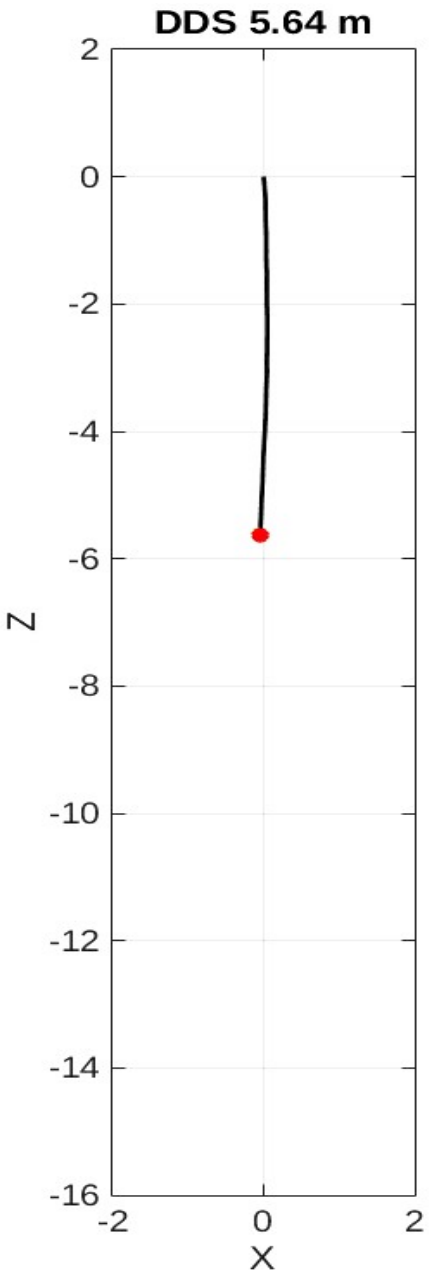


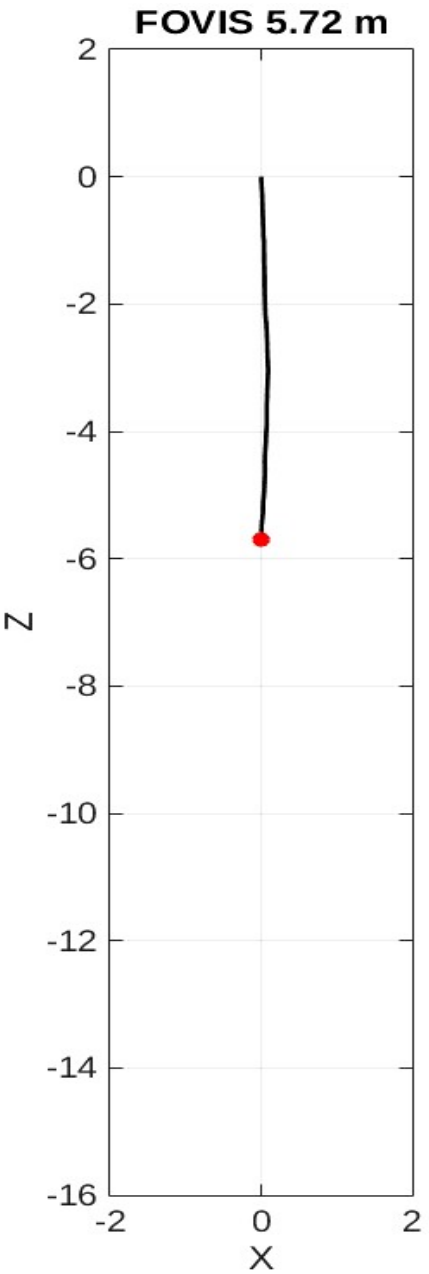
Stereo with opencv block matching
5x5 SAD window and 16 disparities search range

Real Data from Underground Mines





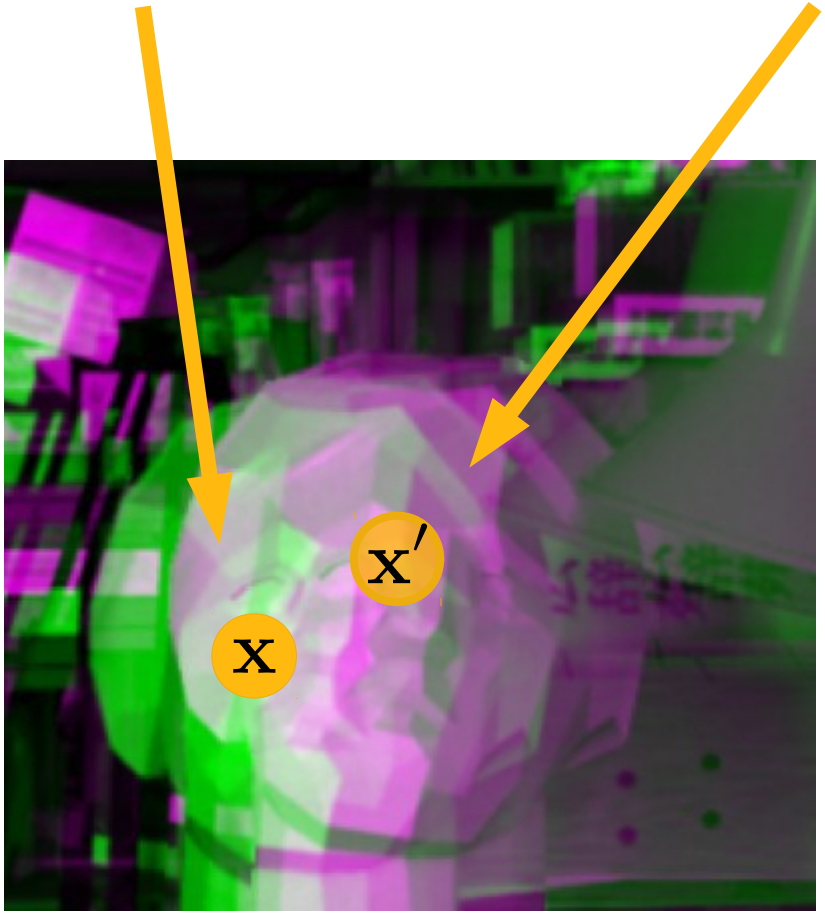

DDS 5.64 m
2
0
-2
-4
-6
-8
-10
-12
-14
-16
Z
-2
0
2
X


FOVIS 5.72 m
2
0
-2
-4
-6
-8
-10
-12
-14
-16
Z
-2
0
2
X

Limitations of Direct Methods

Pixel location in the reference image with intensity $\mathbf{I}_0(\mathbf{x})$

There exists an (unknown) **corresponding** location in input image with the same intensity



$$\mathbf{I}_0(\mathbf{x}) = \mathbf{I}_1(\mathbf{x}')$$

Brightness Constancy

The two views a related via a warp

$$\mathbf{x}' = \mathbf{w}(\mathbf{x}; \boldsymbol{\theta})$$

Brightness Constancy

$$\Delta\boldsymbol{\theta}^* = \underset{\Delta\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{\mathbf{x}} \|\mathbf{I}_0(\mathbf{x}) - \mathbf{I}_1(\mathbf{w}(\mathbf{x}; \boldsymbol{\theta} + \Delta\boldsymbol{\theta}))\|^2$$

Does not hold most of the time

Effective solution to the brightness constancy assumption

+

Direct Alignment
Lucas-Kanade

Robust Pose estimation

Effective solution to the brightness constancy assumption

+

Direct Alignment
Lucas-Kanade

Robust Pose estimation

Parameter-free

Invariant to arbitrary changes in illumination

Efficient to compute

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

	>	

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1		
	>	

1							

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	
	>	

1	1						

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
	>	

1	1	0					

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
	>	0

1	1	0	0				

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
	>	0
		1

1	1	0	0	1			

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
	>	0
	1	1

1	1	0	0	1	1		

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
	>	0
0	1	1

1	1	0	0	1	1	0	

(Zabih & Woofill, 94)

Binary Feature Descriptors

8	12	200
56	42	55
128	16	11

1	1	0
0	>	0
0	1	1

1	1	0	0	1	1	0	0

(Zabih & Woofill, 94)

Photometric Invariance

8	12	200
56	42	55
128	16	11

{1,1,0,0,1,1,0,0}

40	12	200
56	42	55
128	16	11

{1,1,0,0,1,1,0,0}

If using SSD, residual is
(40-8)^2 = **1024**

Must be Matching with a Binary Norm

8	12	200
56	42	55
128	16	11

255	12	200
56	42	55
128	16	11

{1,1,0,0,1,1,0,0} **{0,1,0,0,1,1,0,0}**

Hamming distance = 1
Descriptors remain close

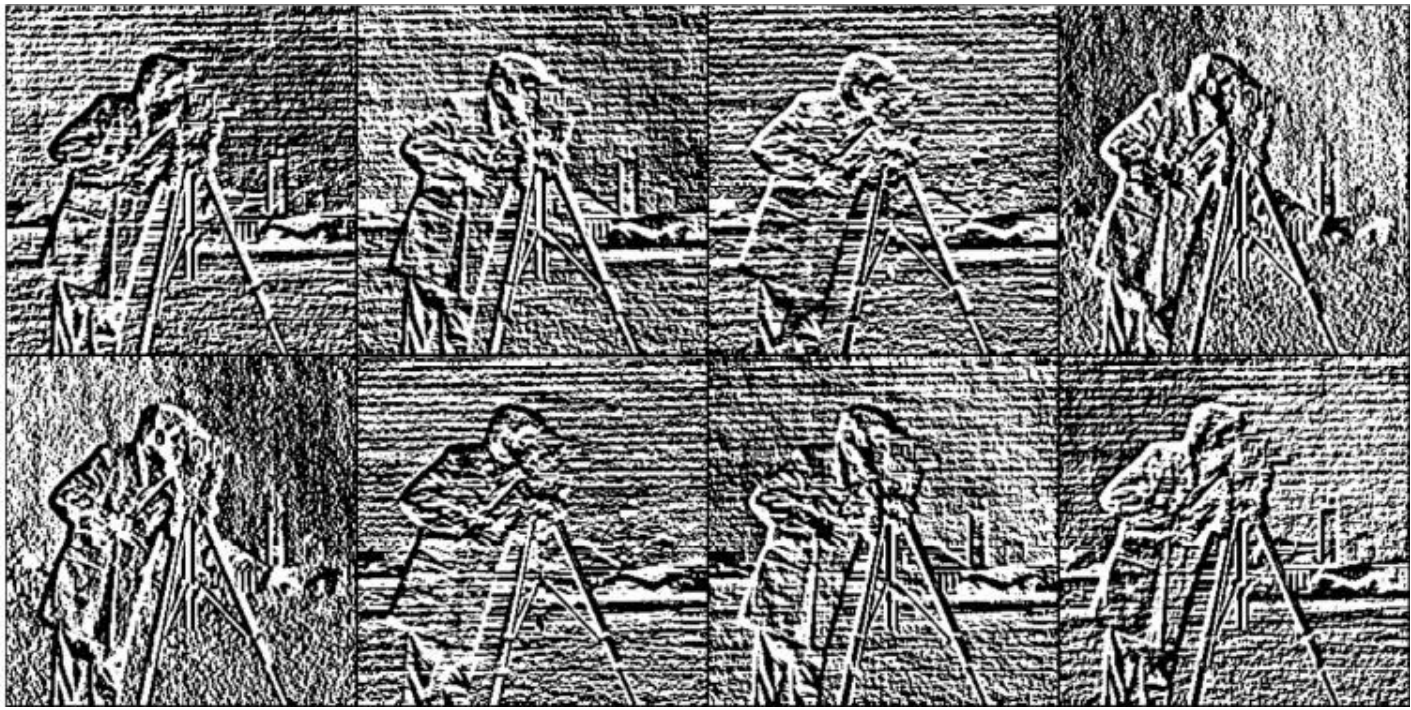
Using Binary Descriptors in LK

$$\Delta\boldsymbol{\theta}^* = \underset{\Delta\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{\mathbf{x}} \|\mathbf{I}_0(\mathbf{x}) - \mathbf{I}_1(\mathbf{w}(\mathbf{x};\boldsymbol{\theta}+\Delta\boldsymbol{\theta}))\|^2$$

We can approximate the Hamming distance, but lose invariance.

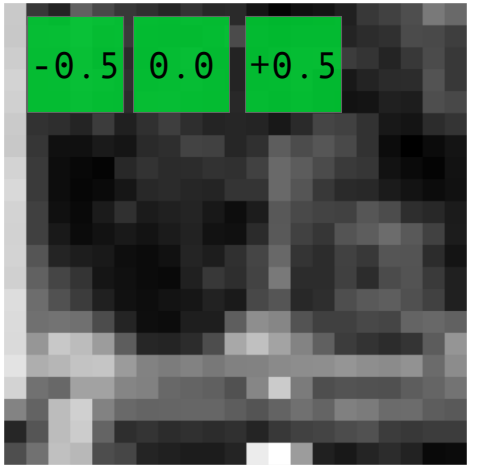
Bit-Planes Representation

Elegant Solution



Achieves two goals

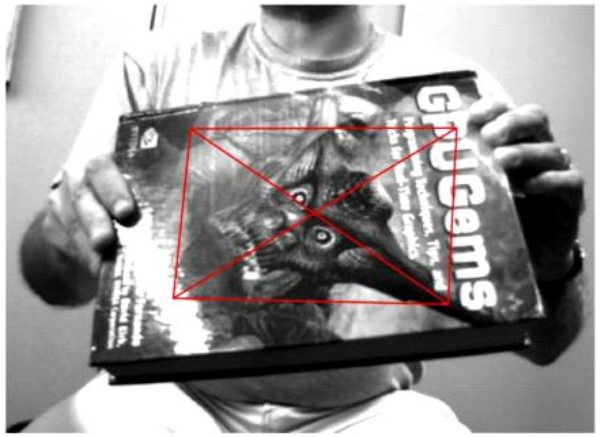
Gradients with simple convolution



Squared norm becomes equivalent to Hamming

```
  [1,1,0,0,1,0,0]
⊕ [1,1,0,0,0,1,0]
-----------------
  [0,0,0,0,1,1,0]
```

Does it work?



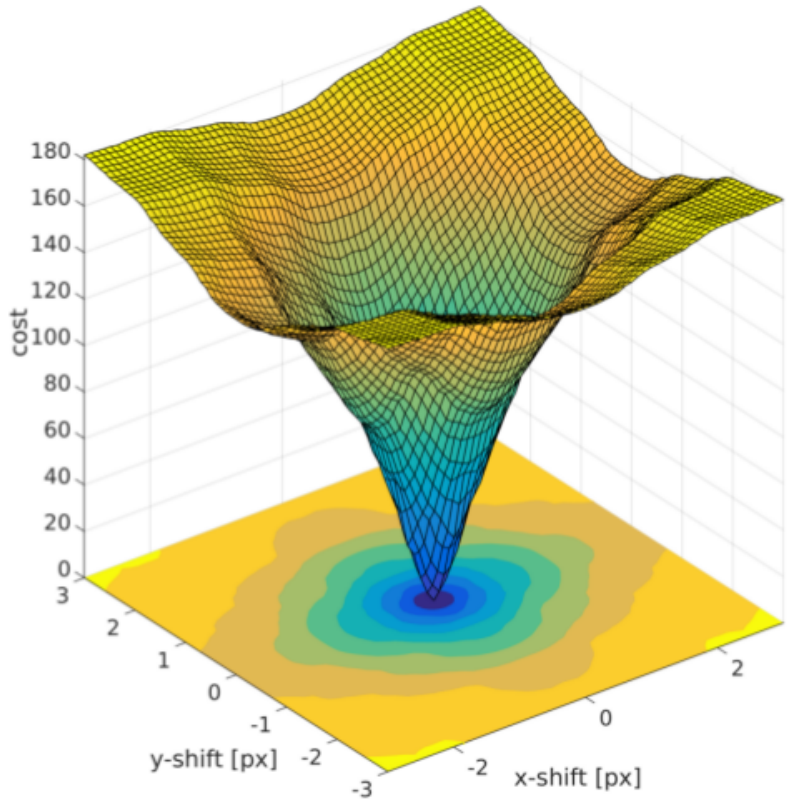
Intensity only

BitPlanes

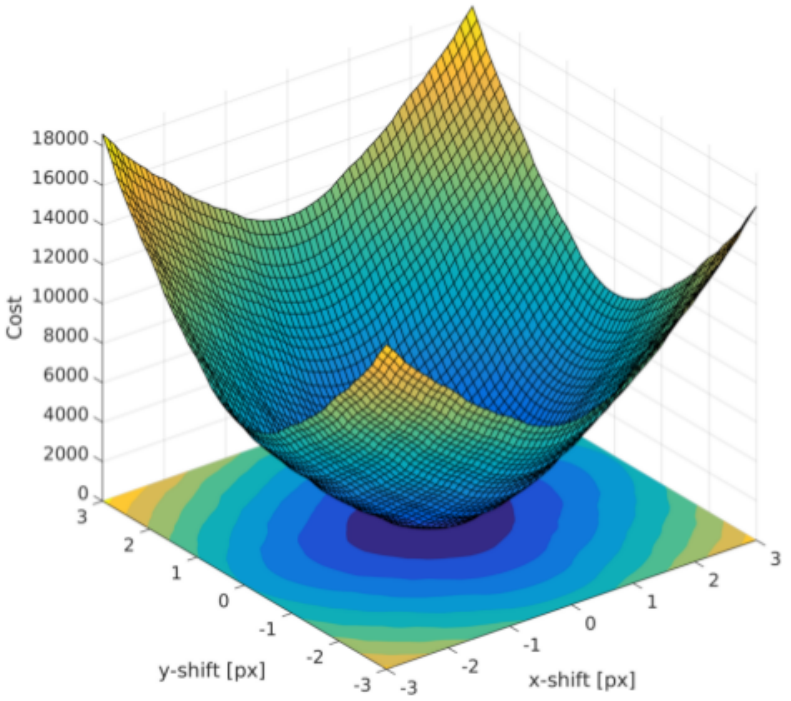
8DOF homography tracking

Why does it work?

Cost surface is quasi-quadratic



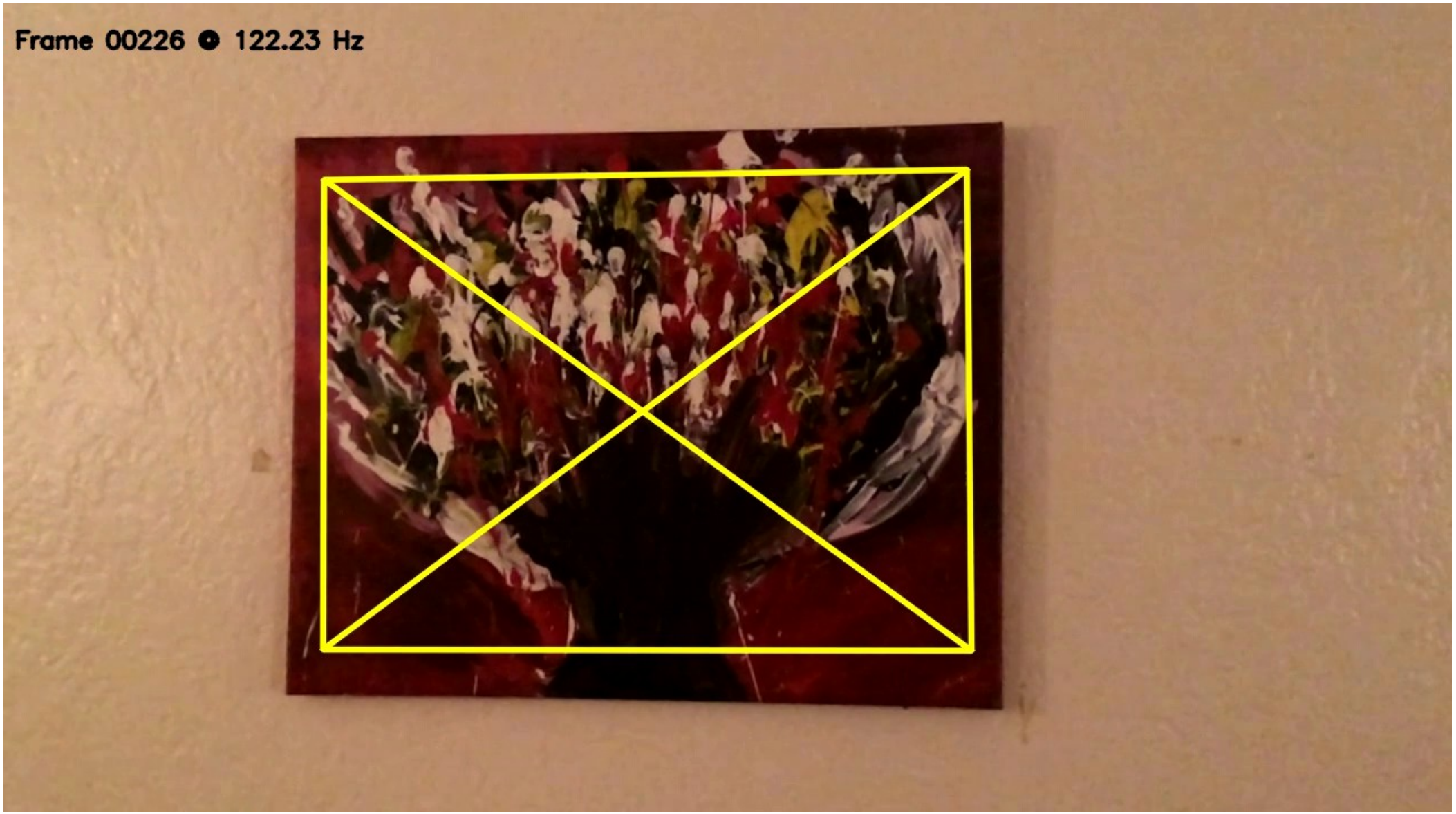
(a) Bit-Planes.



(b) Raw intensity.

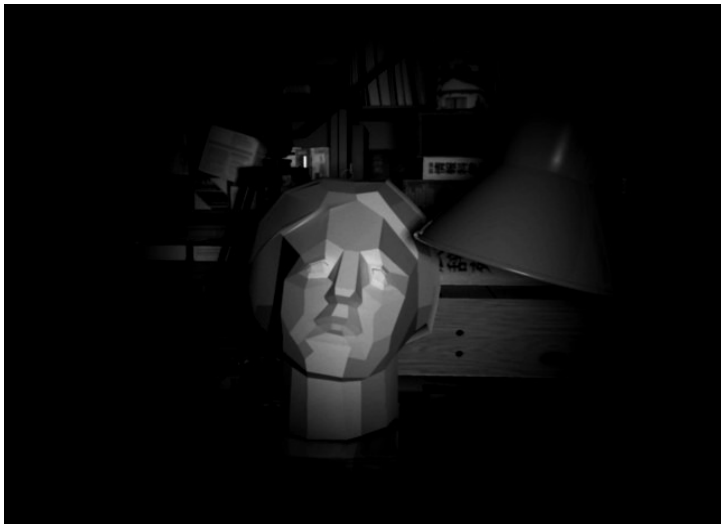
Real time Tracking

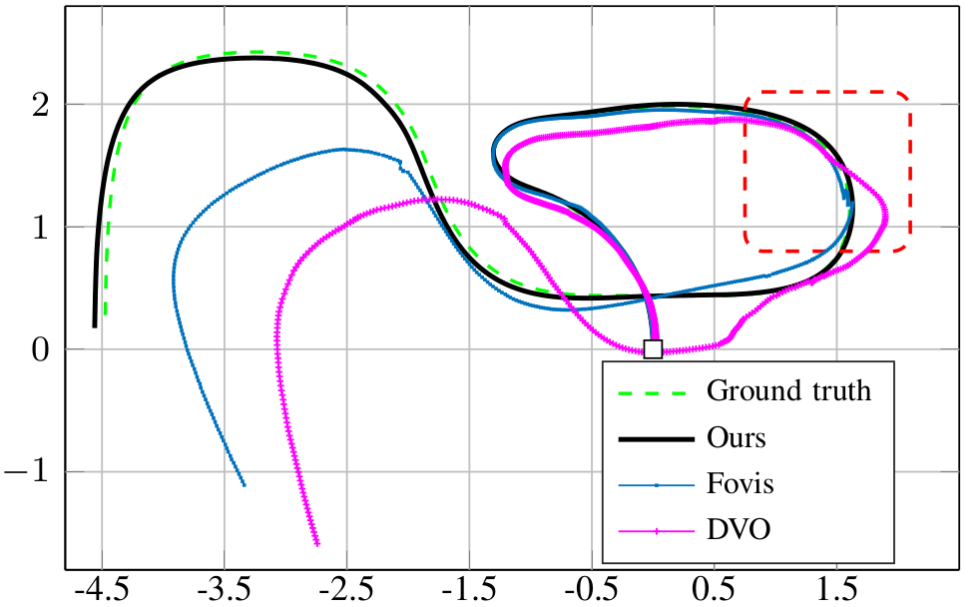
Real time Tracking



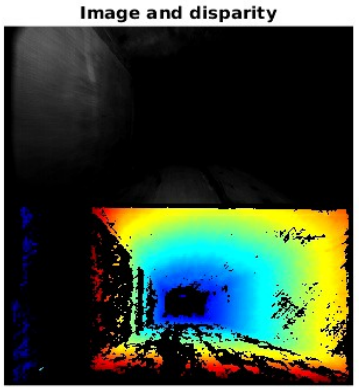
Application to Robust Visual Odometry

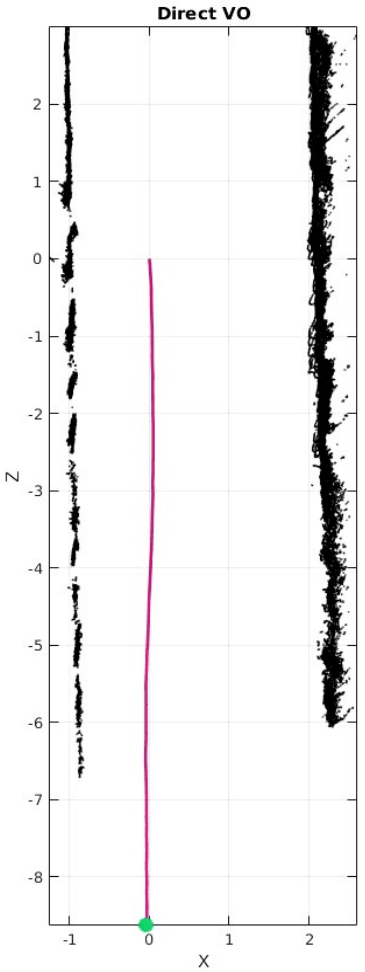
https://github.com/halismai/bpvo

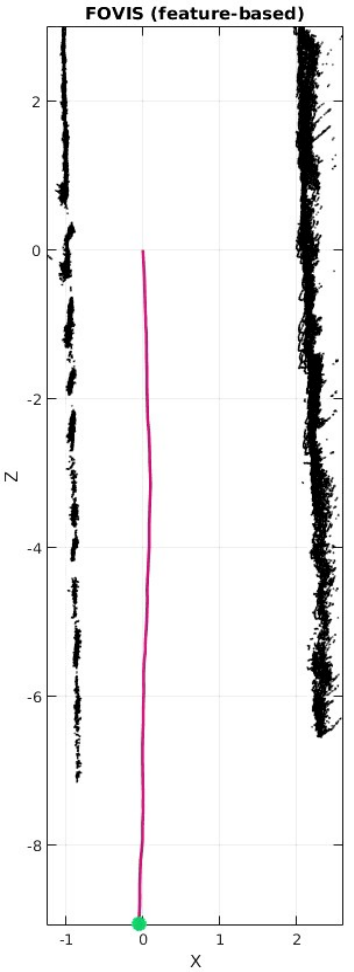


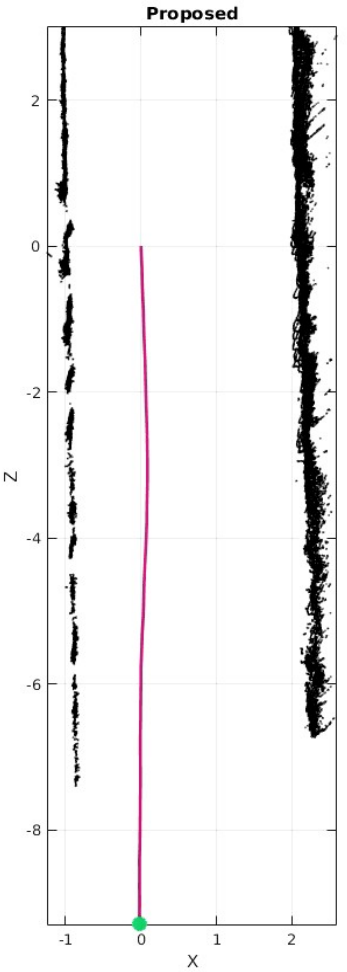


Application to Underground Mines



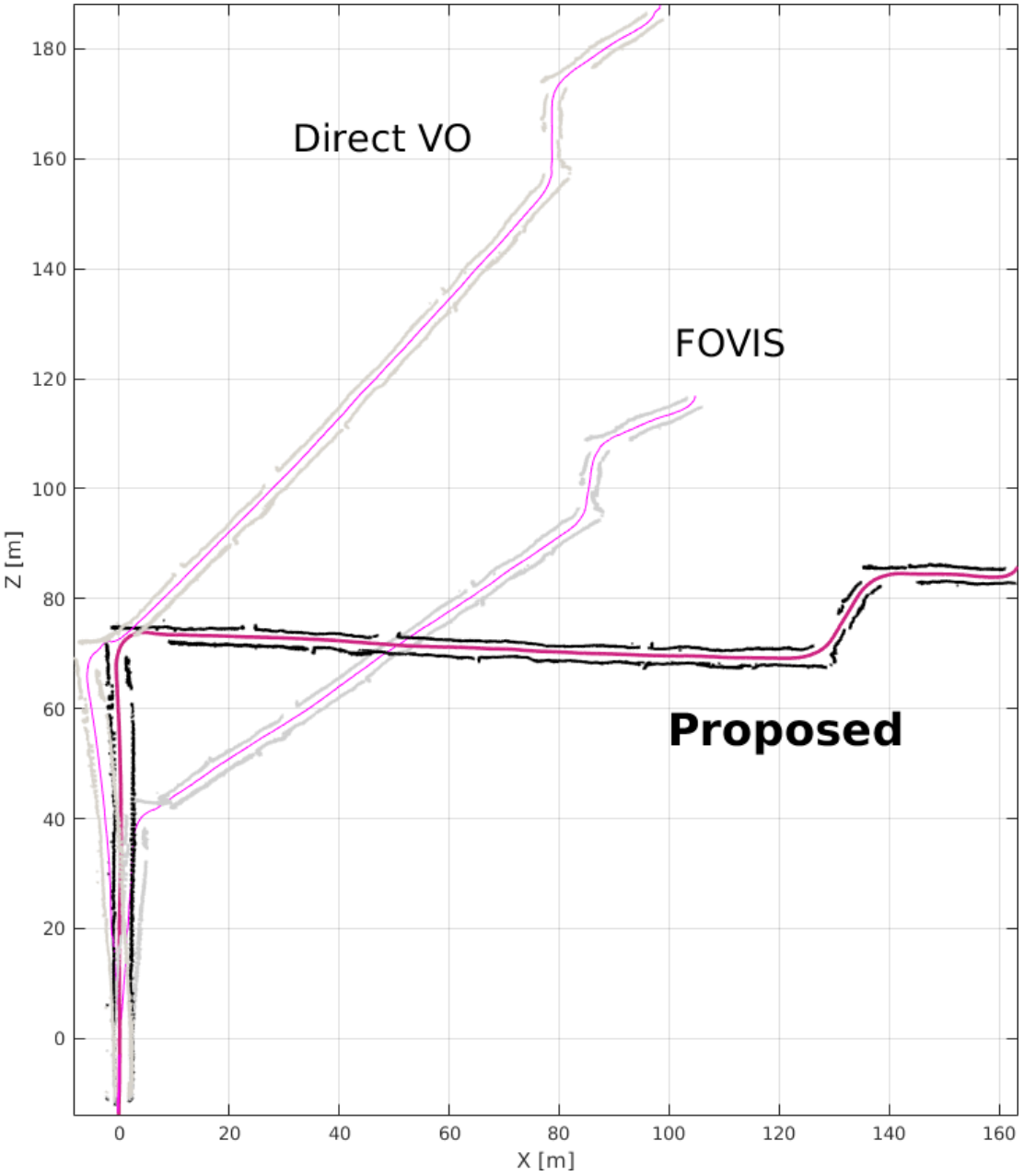




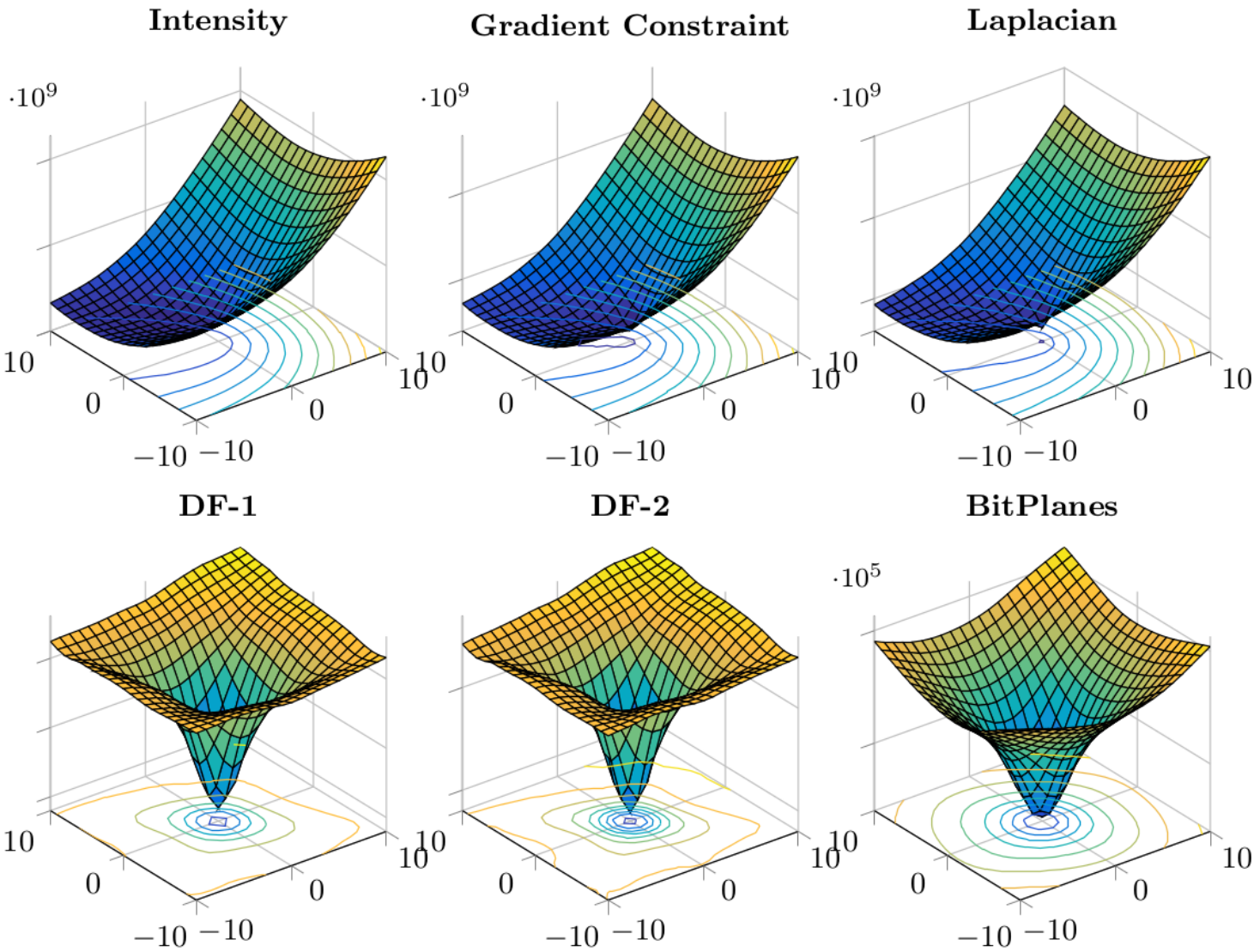


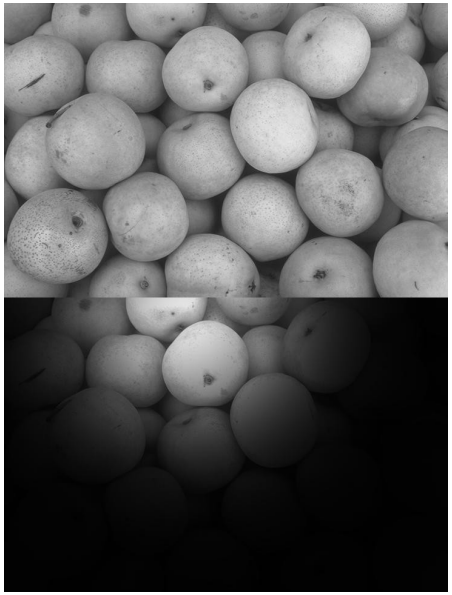
Sped up 2X

Application to Underground Mines



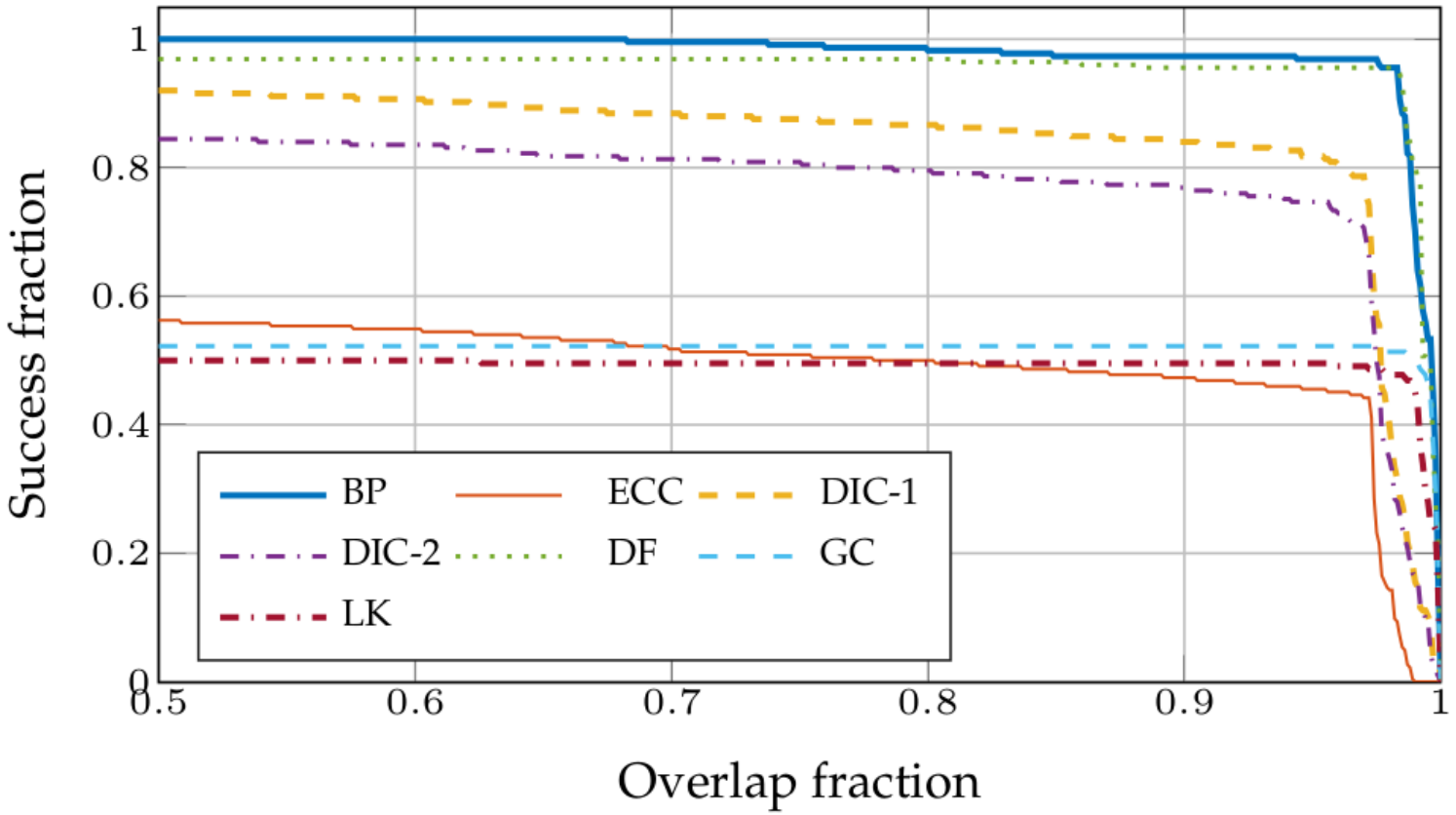
Comparison with other descriptors





DF: Descriptor Fields, Crivellaro et al. CVPR 2014

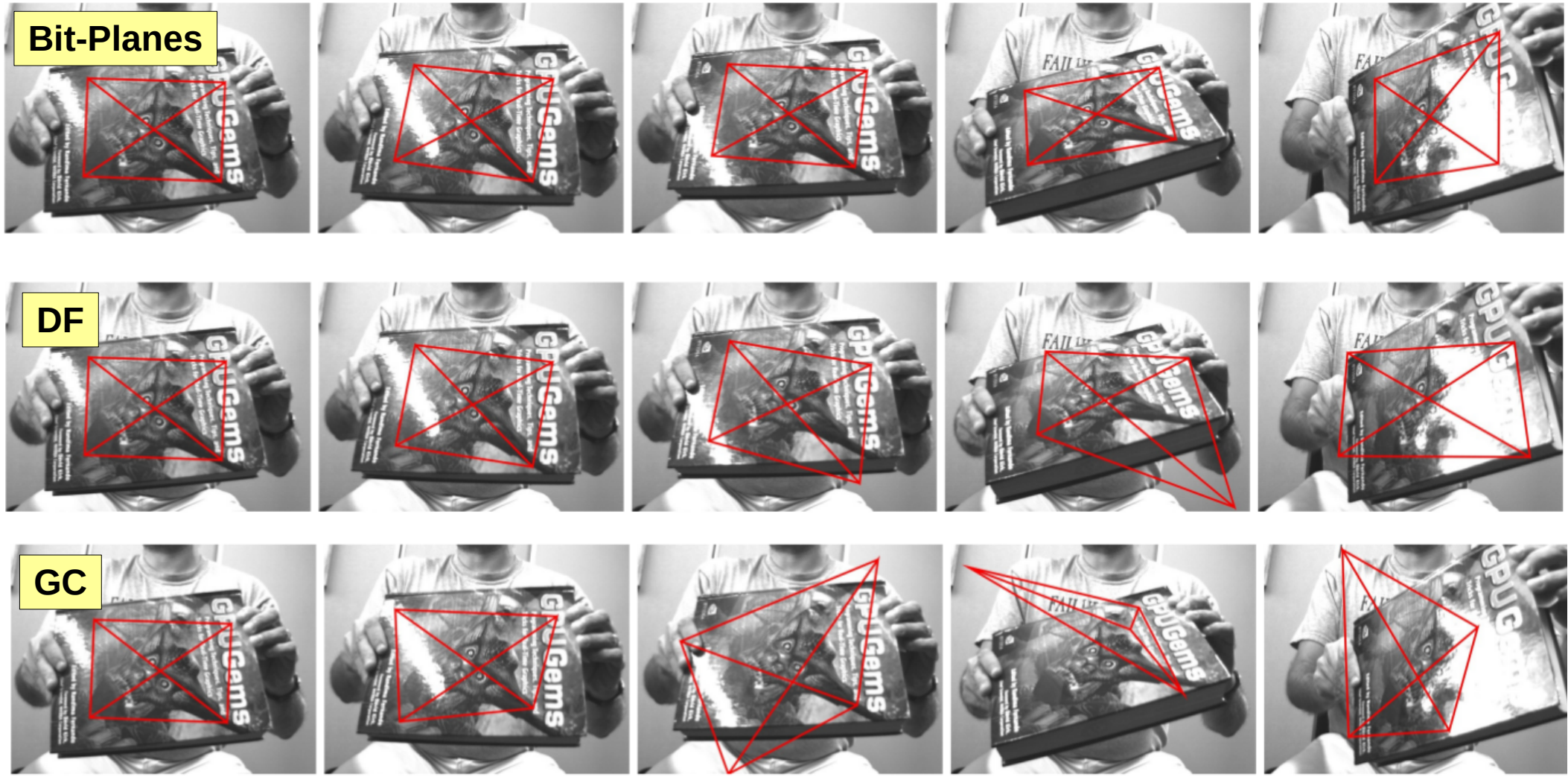
Quantitative Benchmark



Chapters 5 and 6 in dissertation include additional evaluation and quantitative results with applications to VO and parametric image alignment

Benchmark by:
Gauglitz, et al 2011

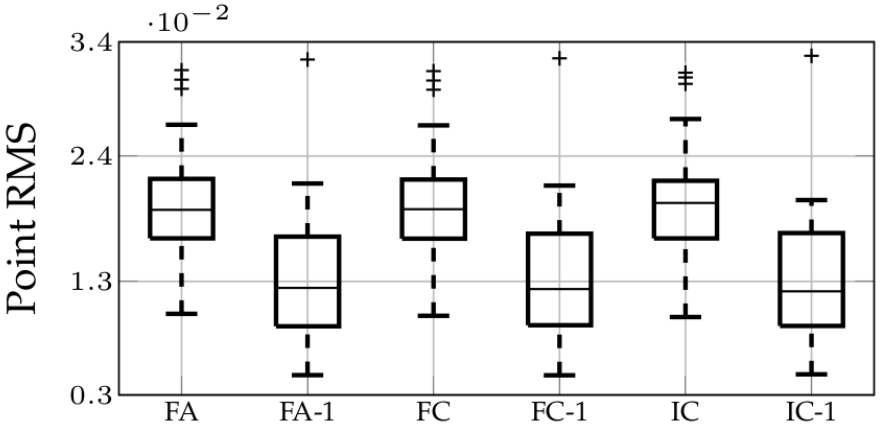
Comparison with other descriptors



DF: Descriptor Fields, Crivellaro et al. CVPR 2014

Additional Details

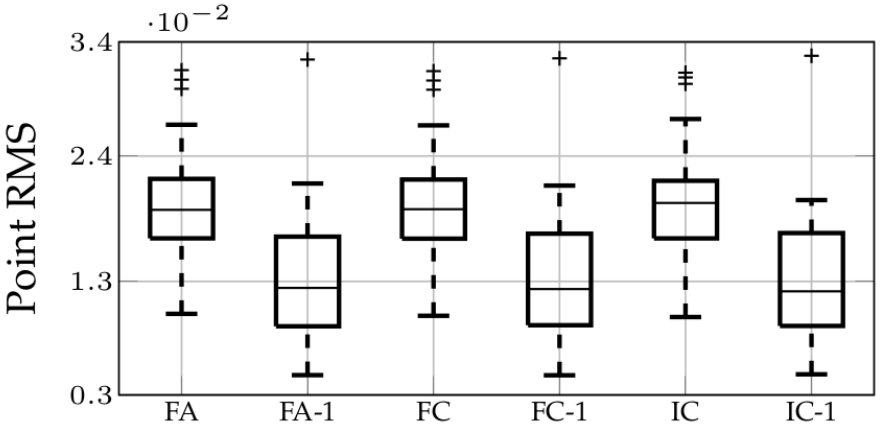
- Warp descriptor images, or warp image then recompute descriptors?



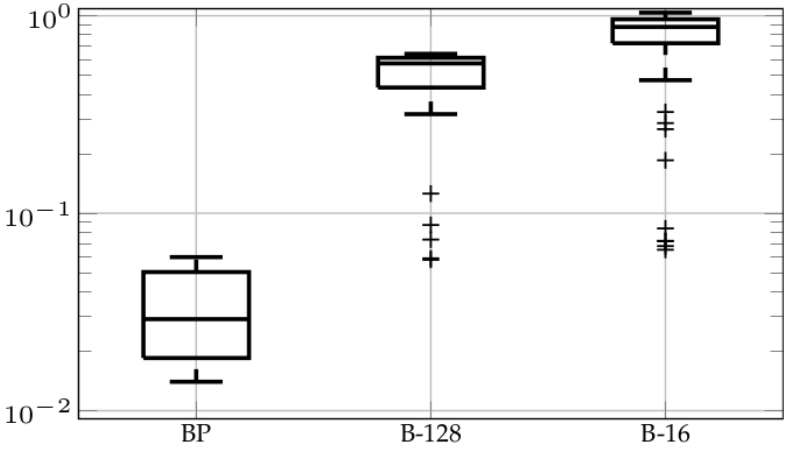
Chapters 5 and 6

Additional Details

- Warp descriptor images, or warp image then recompute descriptors?



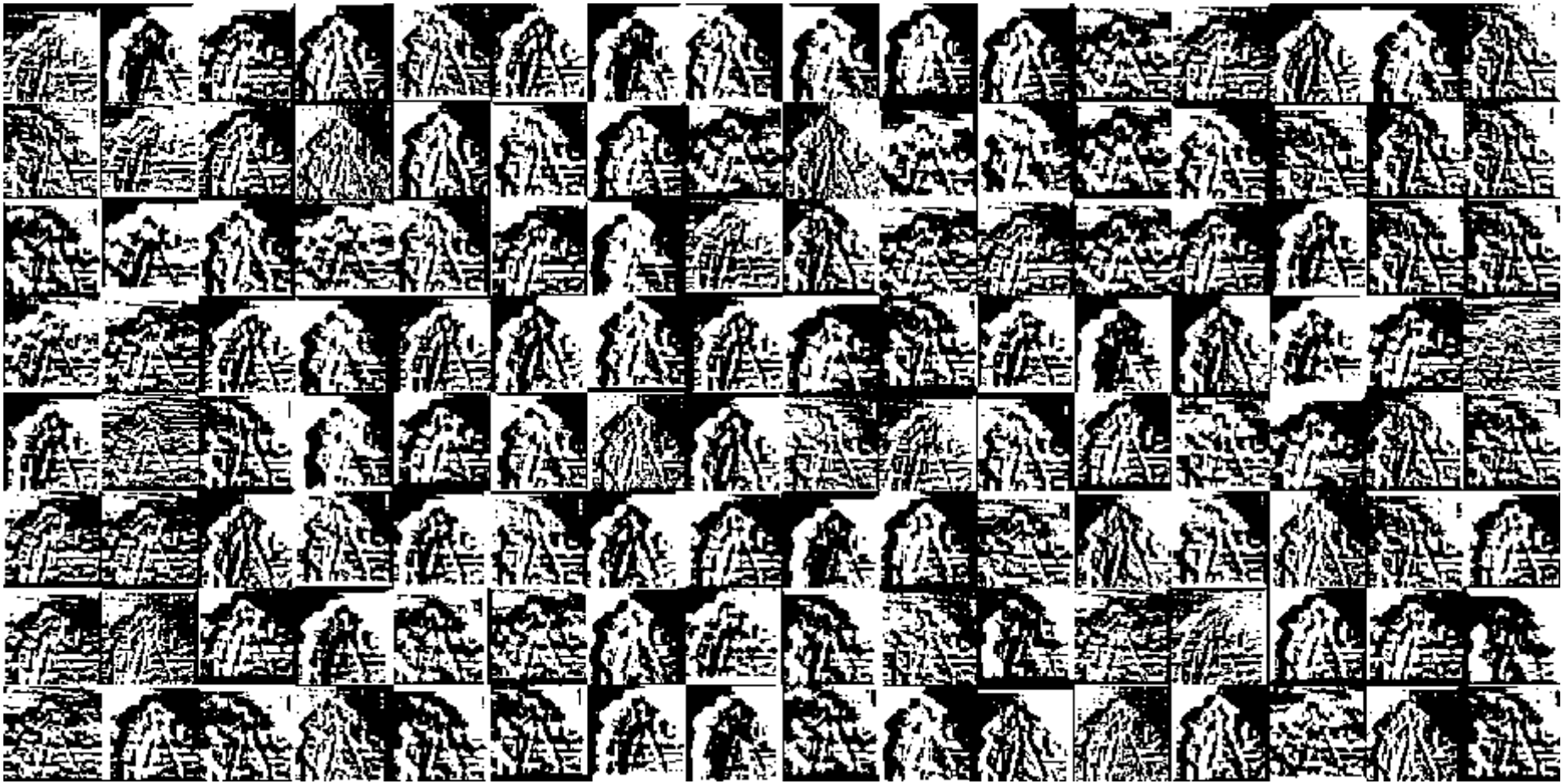
- Does it work with more sophisticated binary descriptors?



Chapters 5 and 6

Additional Details

e.g.: BRIEF (128 Bit-Planes)

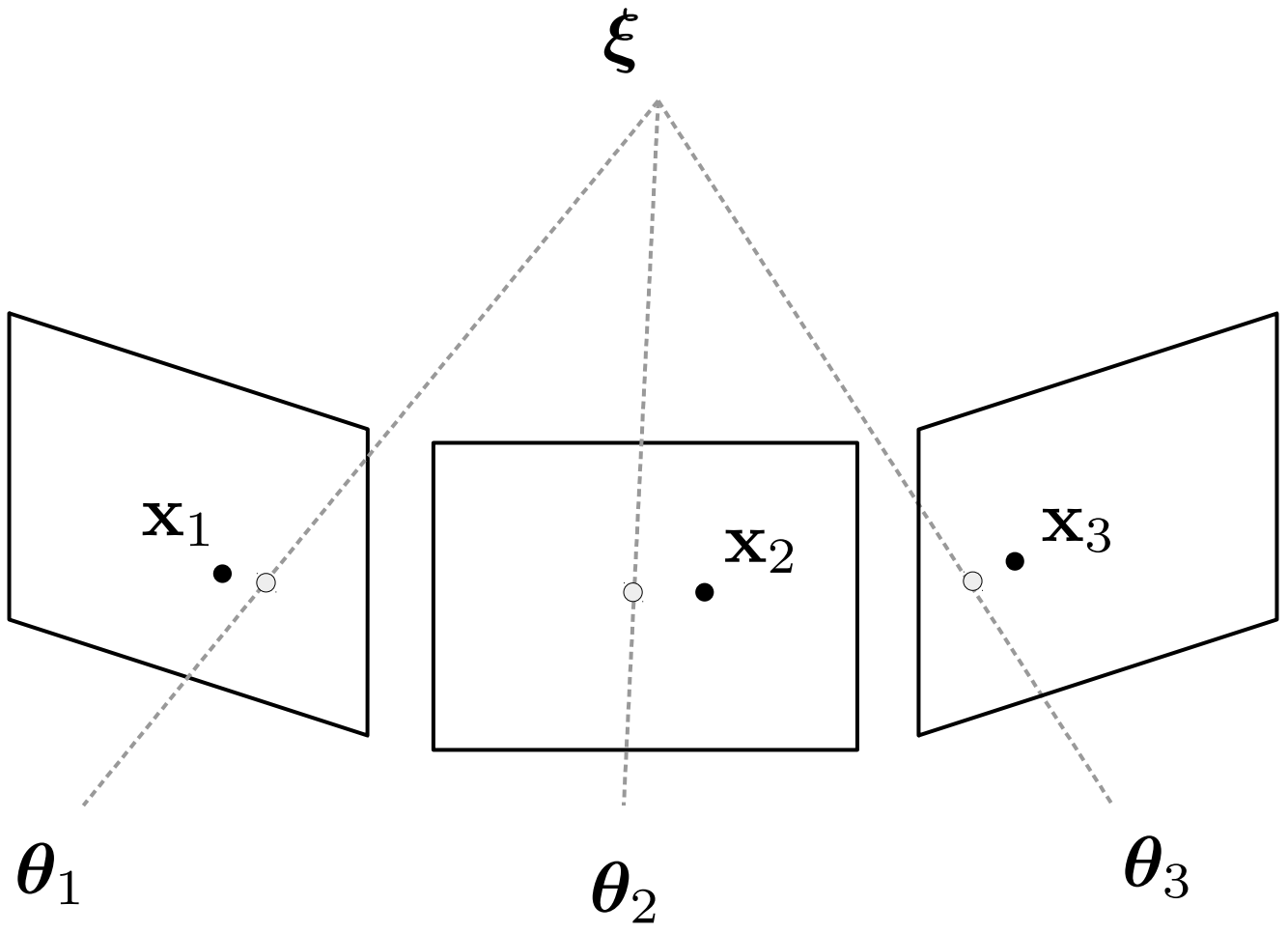


Multi-view Refinement

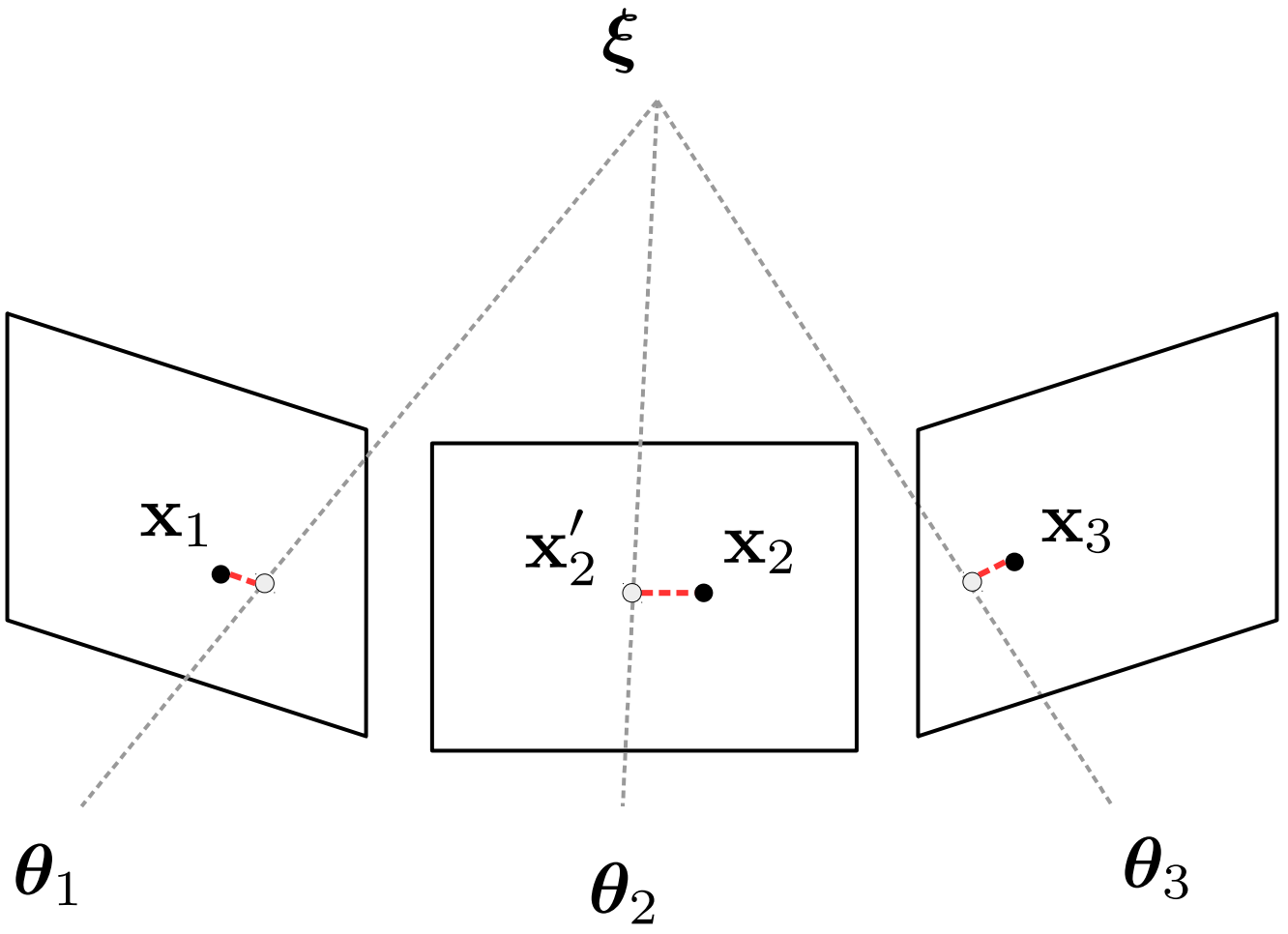
- Features remain the method of choice in VSLAM
  - Due to joint refinement of pose and structure over multiple views using **bundle adjustment**

- Can we formulate a *Direct Bundle Adjustment*?
  - Bundle adjustment without correspondences
  - No longer need corner/edge structures

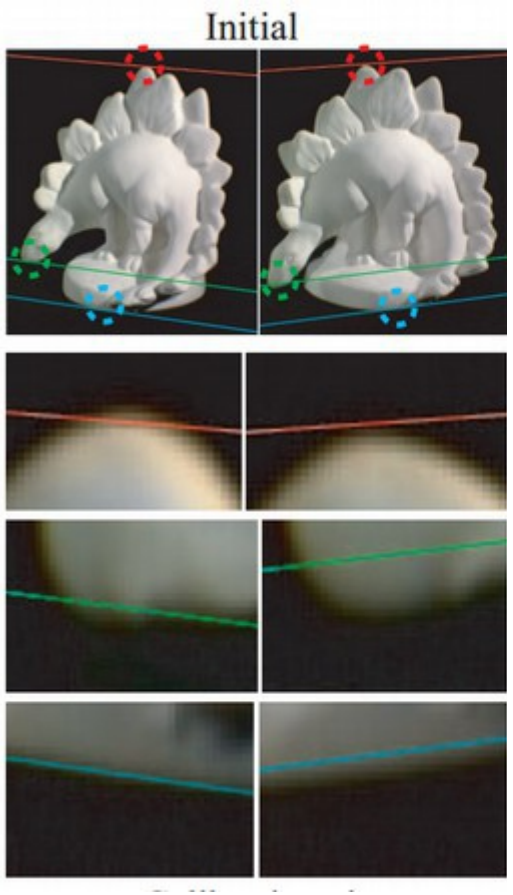
Geometric Bundle Adjustment



Geometric Bundle Adjustment



Problems with the Reprojection Error

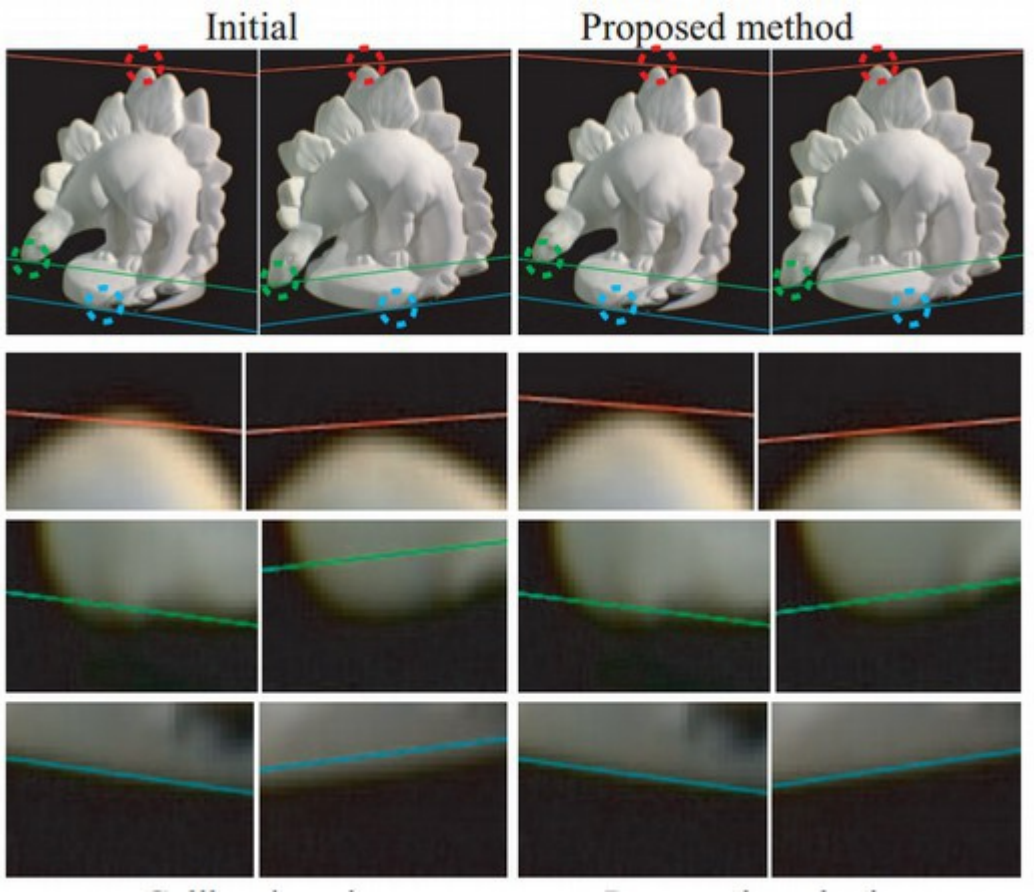


Furukawa & Ponce, CVPR 2008

Attribute errors in feature localization to slight miscalibration

Interleave correspondence refinement with geometric BA

Problems with the Reprojection Error

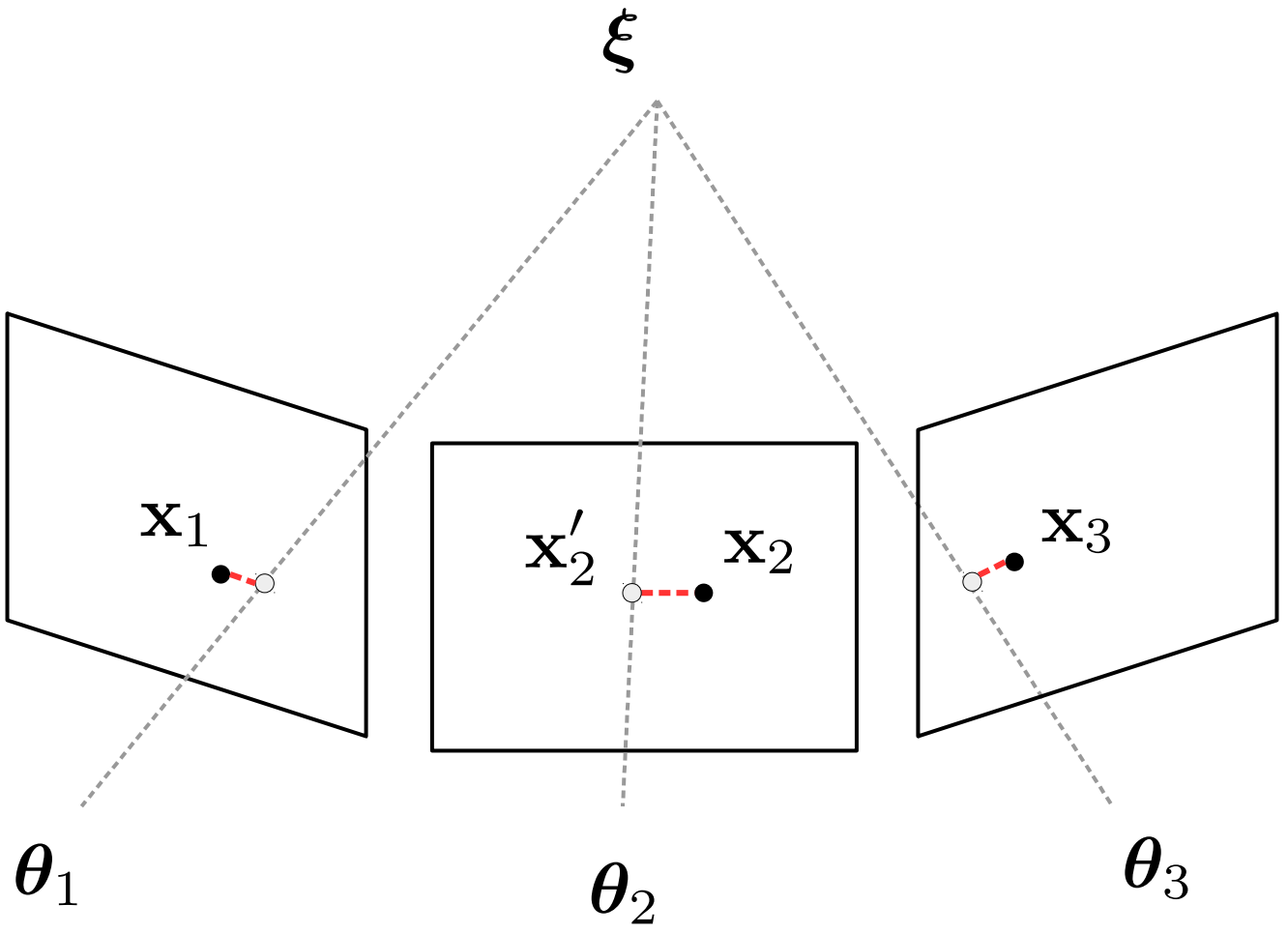


Furukawa & Ponce, CVPR 2008

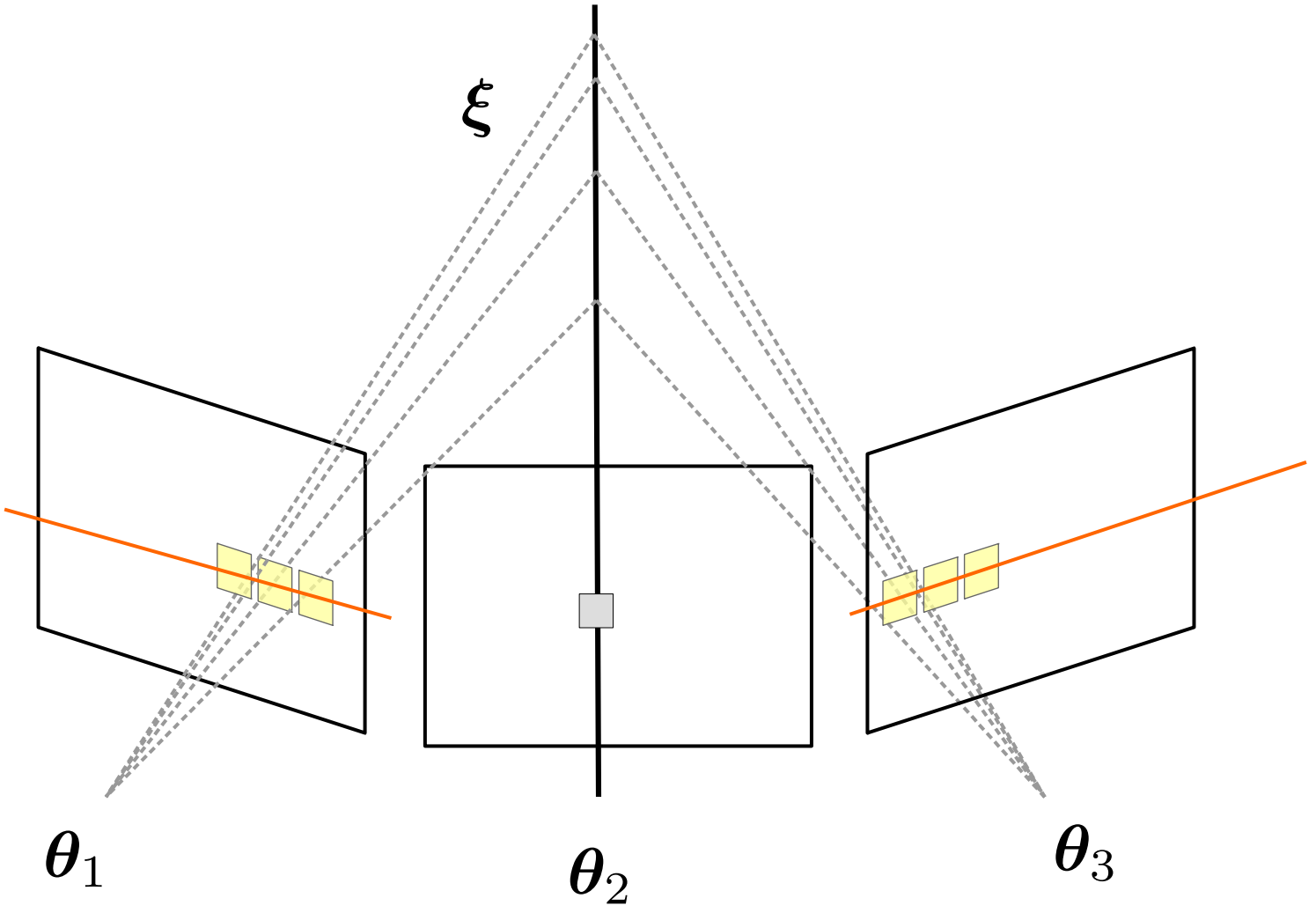
Attribute errors in feature localization to slight miscalibration

Interleave correspondence refinement with geometric BA

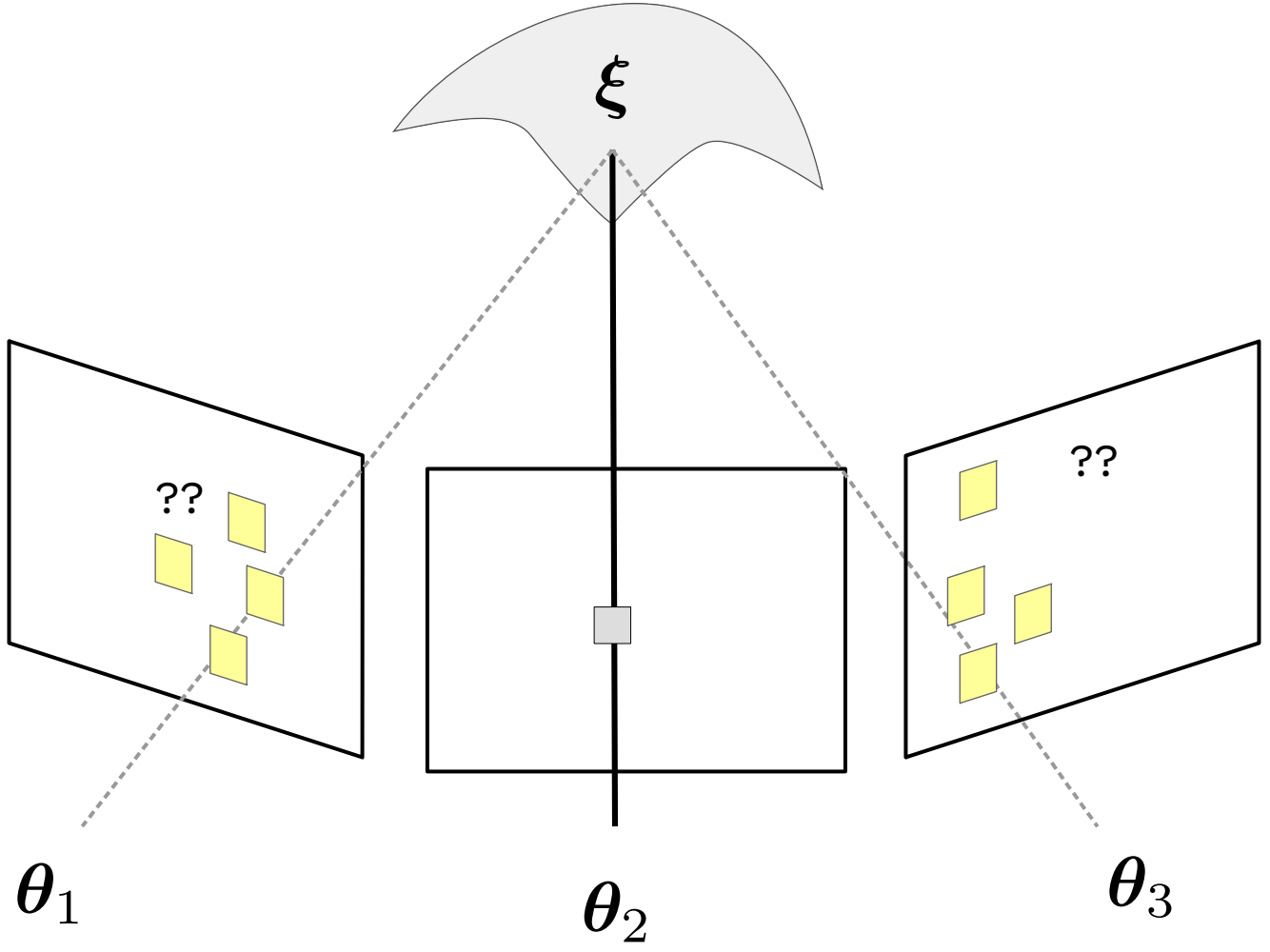
Geometric Bundle Adjustment



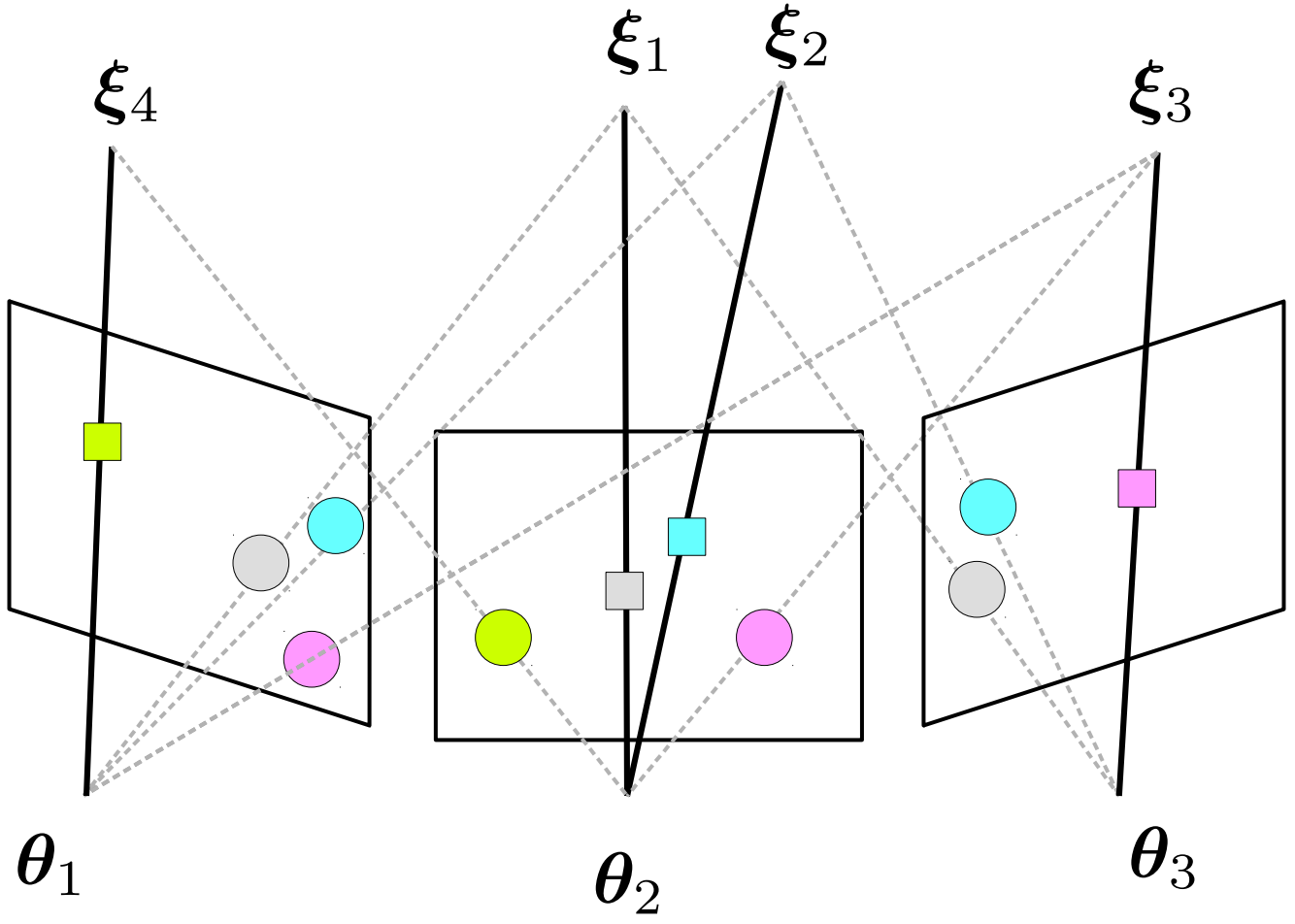
Multi-view Stereo



Photometric Bundle Adjustment



Photometric Bundle Adjustment

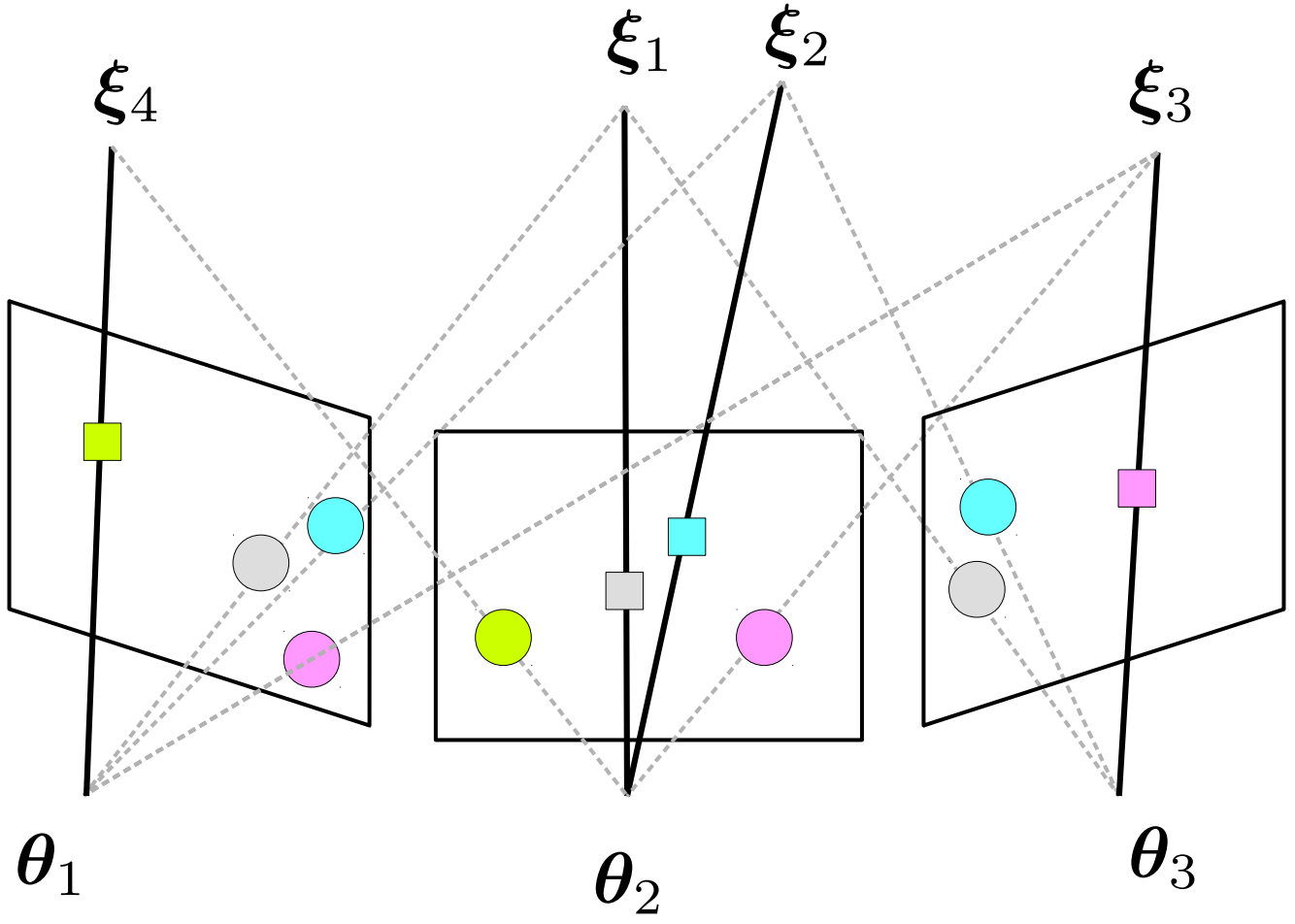


Patch at reference frame

Potential search area

Need to determine visibility info

Photometric Bundle Adjustment

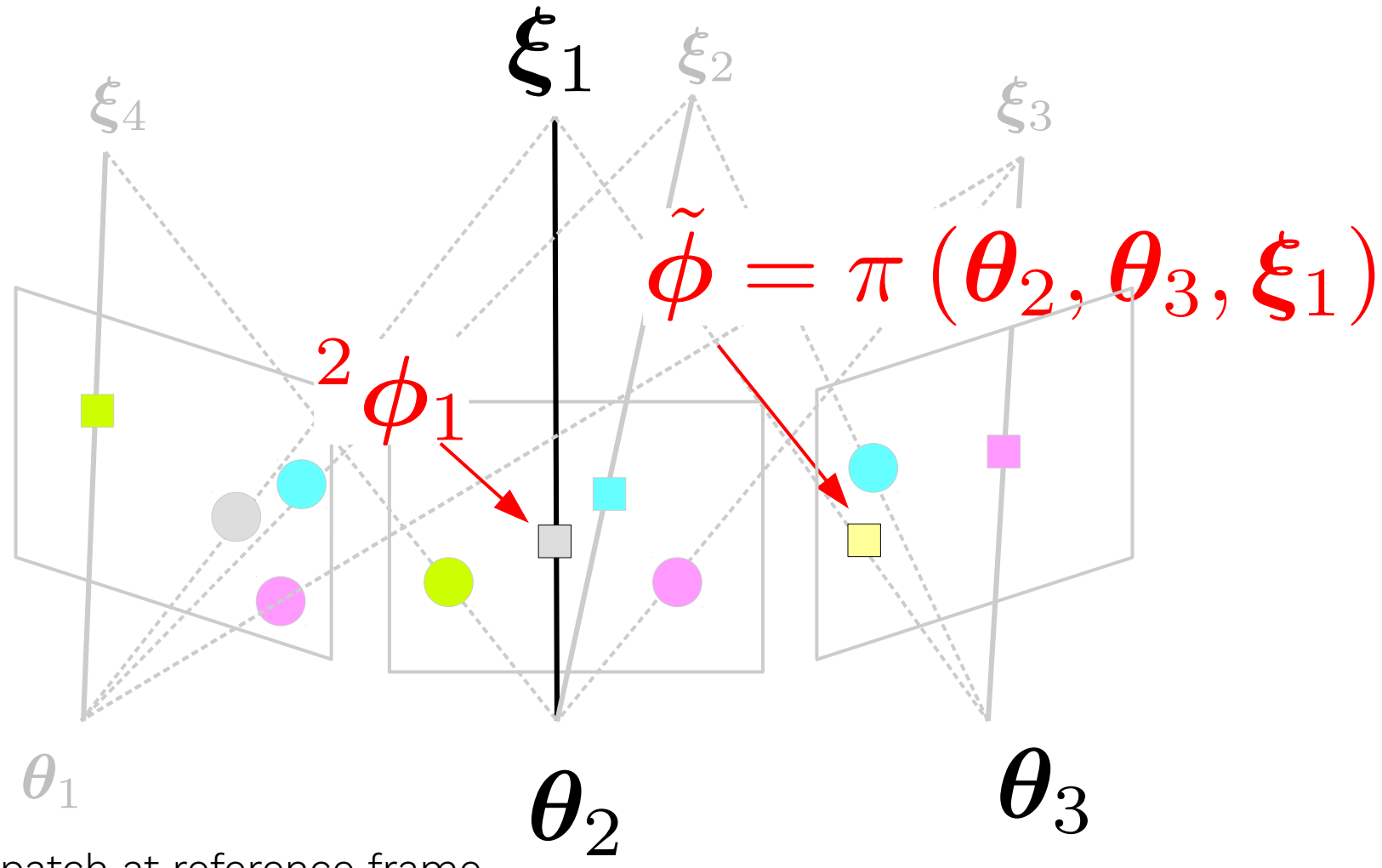


Patch at reference frame

Potential search area

Need to determine visibility info

Photometric Bundle Adjustment



Fixed patch at reference frame

Potential search area

Objective Per Point

$$\| {}^{r}\boldsymbol{\phi}_j - \tilde{\boldsymbol{\phi}}\left(\boldsymbol{\theta}_r, \boldsymbol{\theta}_i, \boldsymbol{\xi}_j\right) \|_2^2$$

Point ‘j’ with reference frame ‘r’ projected at frame ‘i’

Objective Per Point

$$\| {}^{r}\boldsymbol{\phi}_j - \tilde{\boldsymbol{\phi}}\left(\boldsymbol{\theta}_r, \boldsymbol{\theta}_i, \boldsymbol{\xi}_j\right) \|_2^2$$

Point ‘j’ with reference frame ‘r’ projected at frame ‘i’

Objective now depends on 2 poses
Creates additional terms in the Hessian

Objective Per Point

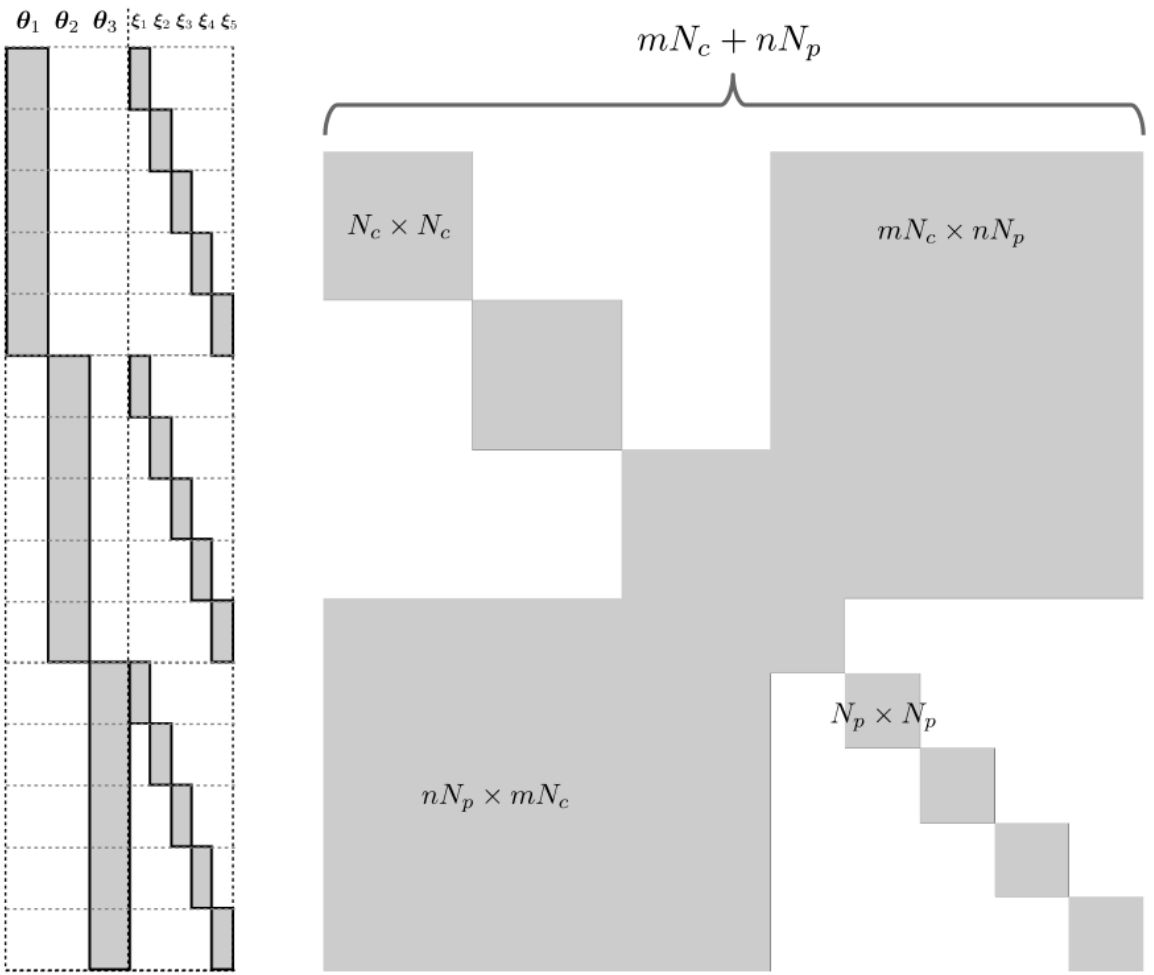
$$\| {}^{r}\boldsymbol{\phi}_j - \tilde{\boldsymbol{\phi}}\left(\boldsymbol{\theta}_r, \boldsymbol{\theta}_i, \boldsymbol{\xi}_j\right) \|_2^2$$

Point ‘j’ with reference frame ‘r’ projected at frame ‘i’

Objective now depends on 2 poses
Creates additional terms in the Hessian

Make the reference patch **fixed**
(store it in a buffer)

Sparsity Pattern



Hessian is Identical to geometric BA

Jacobian is slightly denser

For example, using a 3x3 patch results in 9 instead of 2 residuals

$$\mathbf{J}(\boldsymbol{\theta}) = \nabla \mathbf{I}(\mathbf{u}' + \mathbf{u}) \frac{\partial \mathbf{u}'}{\partial \boldsymbol{\theta}}$$

$$\mathbf{J}(\boldsymbol{\xi}) = \nabla \mathbf{I}(\mathbf{u}' + \mathbf{u}) \frac{\partial \mathbf{u}'}{\partial \boldsymbol{\xi}}$$

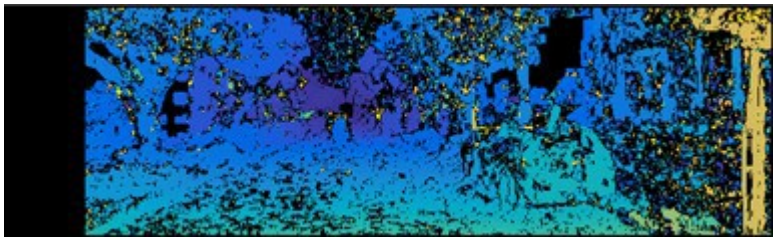
The same Jacobian as in geometric BA, but multiplied by warped image gradient

Point/Pixel Selection

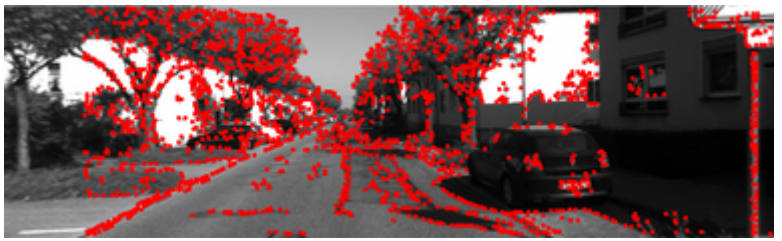
- Has depth and is local max of gradient magnitude

Input image

Gradient magnitude



Depth initialization



Selected pixels

Pixels are selected as **integer** positions

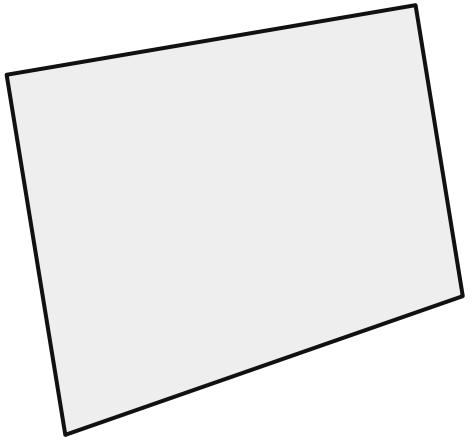
~4096 points per frame on KITTI data

Visibility

- Project the current set of 3D points onto new frame

$$\{\boldsymbol{\xi}_j, \boldsymbol{\phi}_j\}_{j=1}^{N}$$

Initialization of the scene points thus far, each with its photometric data and a 5x5 intensity patch



$\text{Frame } (m+1)$

Visibility

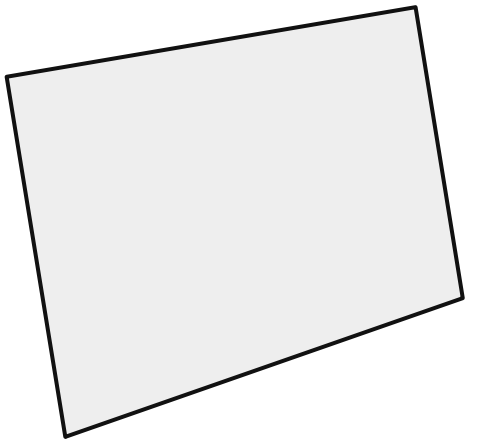
- Project the current set of 3D points onto new frame

$$\{\boldsymbol{\xi}_j, \boldsymbol{\phi}_j\}_{j=1}^{N}$$

Initialization of the scene points thus far, each with its photometric data and a 5x5 intensity patch

$$\boldsymbol{\theta}_{m+1}$$

Pose initialization for frame (m+1)



$\text{Frame } (m+1)$

Visibility

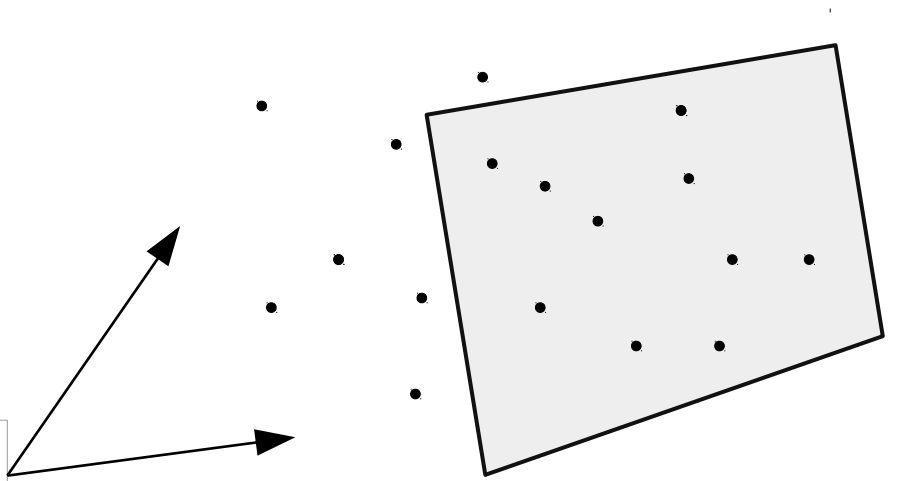
- Project the current set of 3D points onto new frame

$$\boldsymbol{\theta}_{m+1}$$

Pose initialization for frame (m+1)

$$\{\boldsymbol{\xi}_j, \boldsymbol{\phi}_j\}_{j=1}^{N}$$

Initialization of the scene points thus far, each with its photometric data and a 5x5 intensity patch



Projections using pose initialization onto new frame

$$\text{Frame } (m+1)$$

Visibility

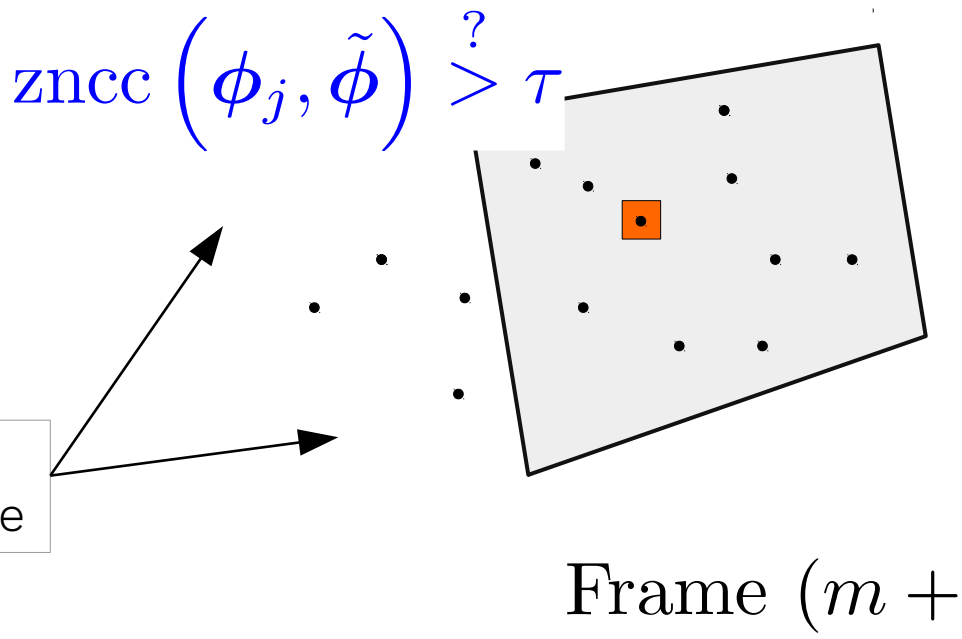
- Project the current set of 3D points onto new frame

$$\{\boldsymbol{\xi}_j, \boldsymbol{\phi}_j\}_{j=1}^{N}$$

Initialization of the scene points thus far, each with its photometric data and a 5x5 intensity patch

$$\boldsymbol{\theta}_{m+1}$$

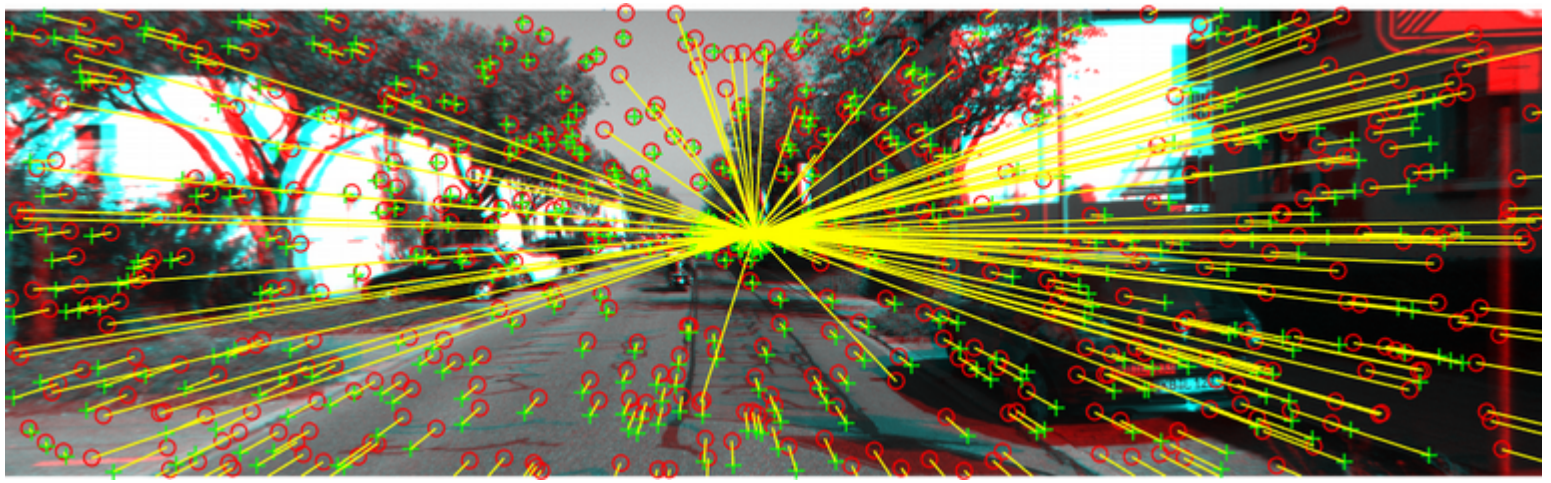
Pose initialization for frame (m+1)



Visibility

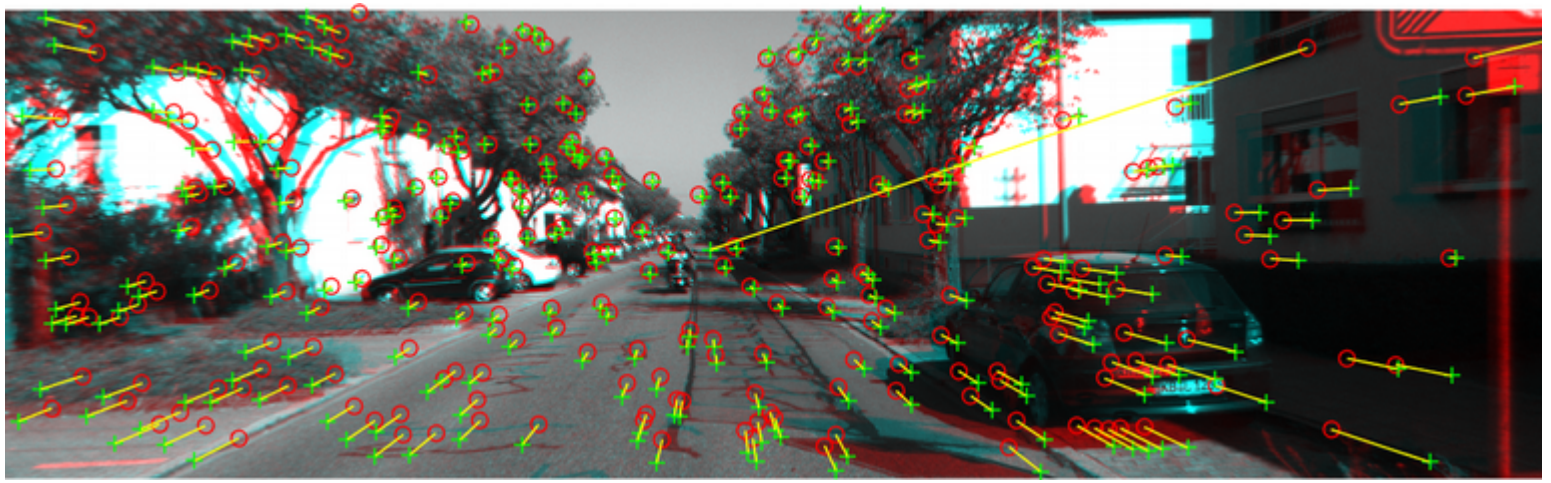
- Project the set of 3D points onto new frame

Depends on accuracy of pose and structure initialization



Visibility

- Project the set of 3D points onto new frame
- Outliers handed with robust minimization
  - Huber loss

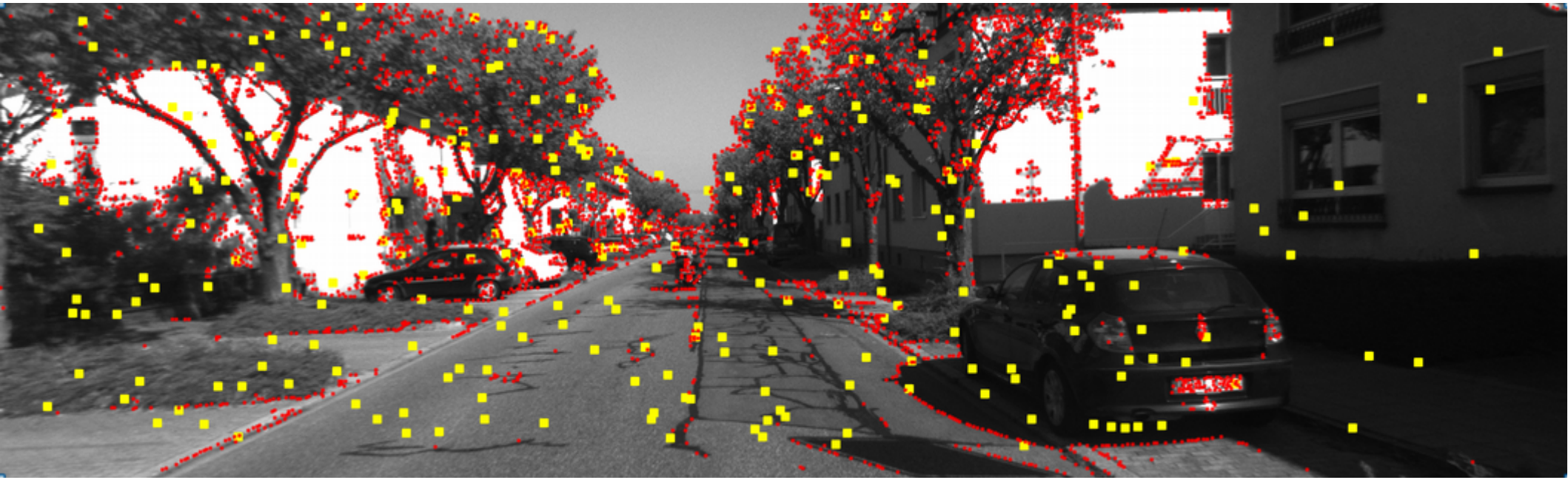


Implementation details

- 3x3 patch photometric error
  - Experimented with center-weighted error
    - No significant difference
- 5x5 patch for zncc (with threshold = 0.6)
- Sliding window with 5 frames
- Depth initialization from block matching stereo

One Last Detail

- Avoid adding duplicate points

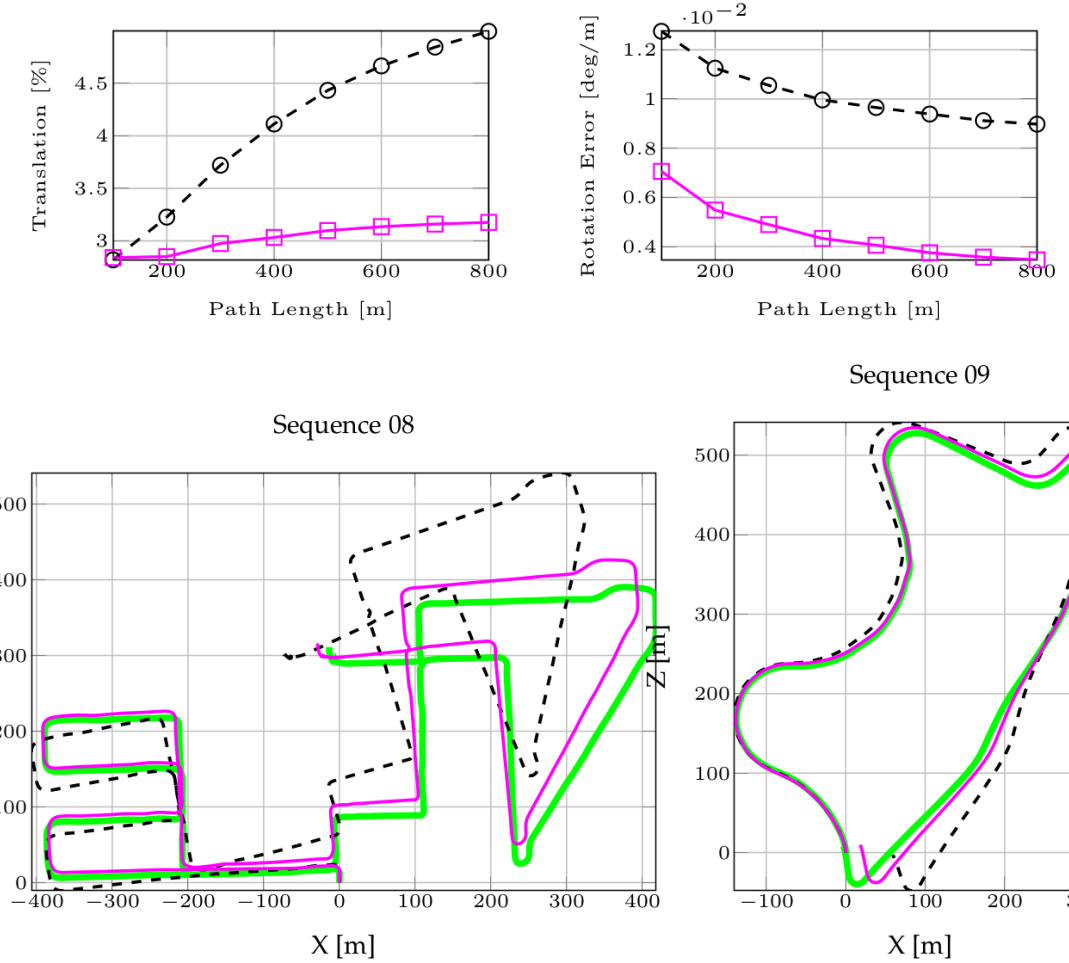


Location of a projected point from the previous frame

Candidates to initialize new scene points

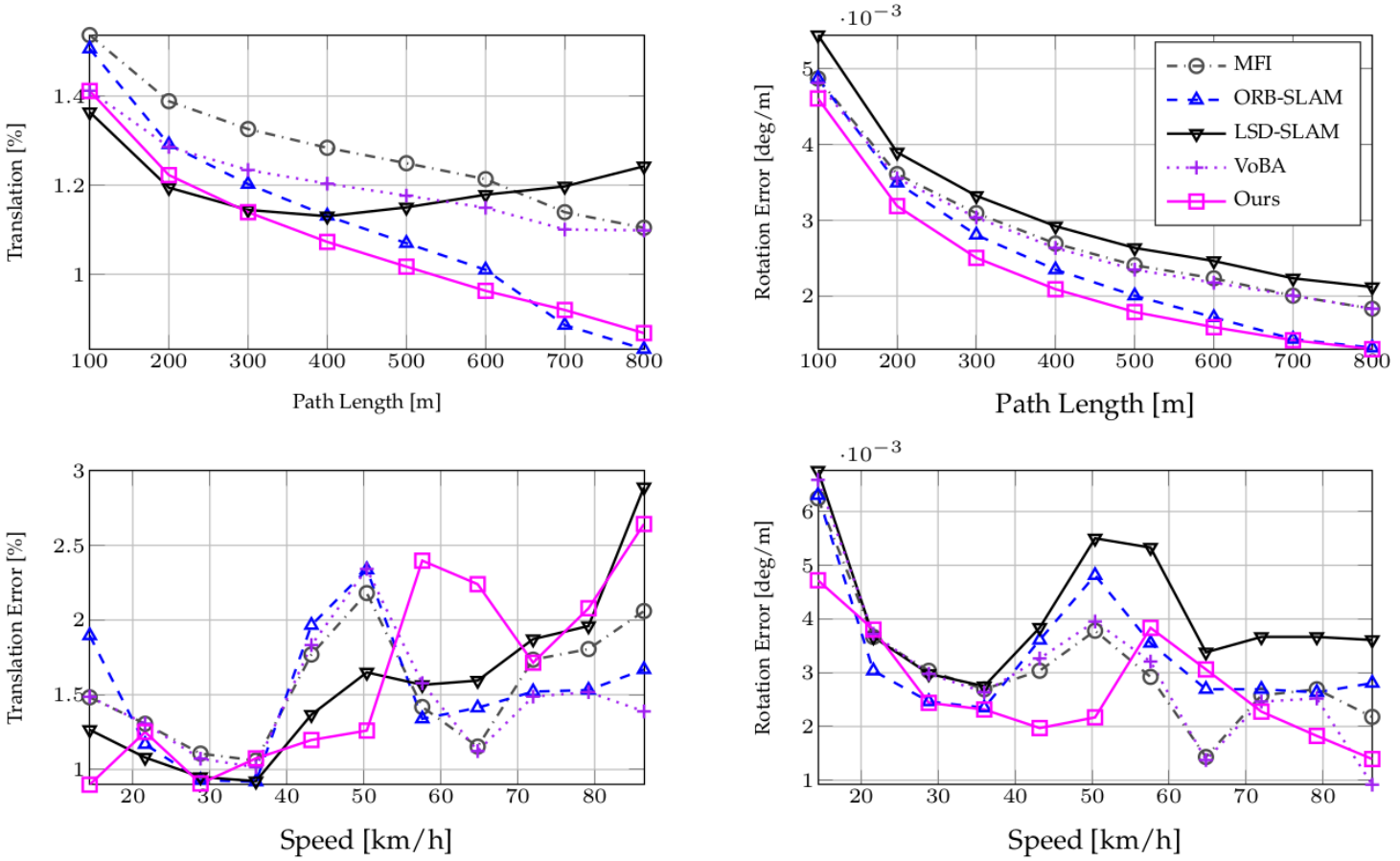
Improvement on poor initialization

Pose init from direct VO; Depth init from Block Matching stereo



Improvement on good initialization

Pose init from ORB-SLAM; Depth init from Block Matching stereo



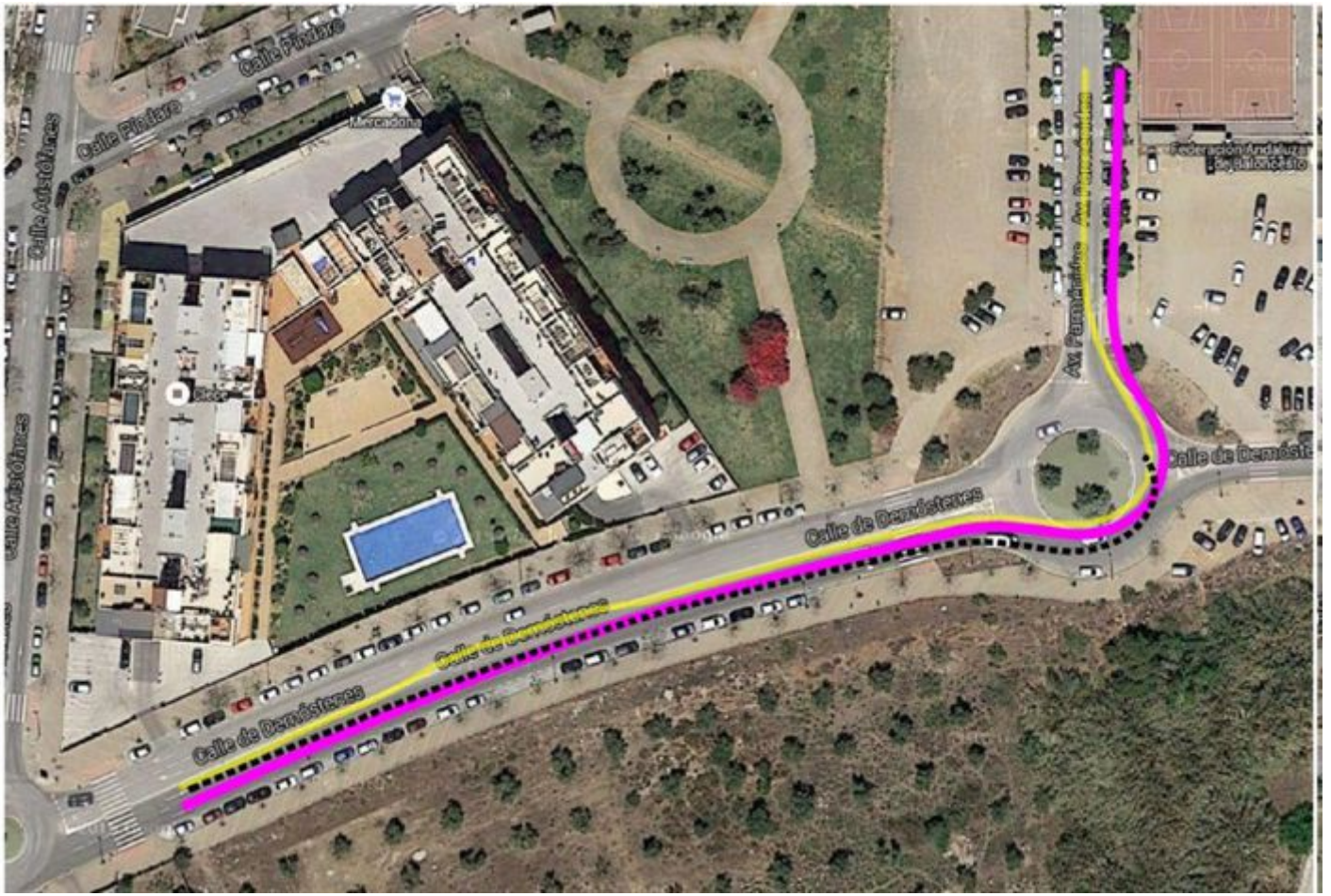
ORB-SLAM does traditional bundle adjustment + loop closure

Results with no initialization

Higher frame-rate data

ORB-SLAM

Ours



Open Questions

- Theoretical characterization of photometric BA
  - Convergence basin and conditions, motion…
- Theoretical/experimental understanding of why it can improve on geometric BA
  - Could improve current geometric BA
- Good results with simple techniques
  - Other options for: pixel selection, visibility, descriptors, …

Summary

- Extended the state-of-the-art in pose estimation to challenging domains
  - Low light, sudden & drastic appearance change, specular reflections, …
- Demonstrated faster than real-time results for template tracking and visual odometry
- Presented the 1st formulation of direct bundle adjustment for VSLAM
  - Improves results obtained with geometric BA

Conclusions

- Direct alignment using binary descriptors is a robust solution to pose estimation
  - When the inter-frame displacement is small
  - Use Bit-Planes for correct objective in least-squares
- Direct (photometric) bundle adjustment is a feasible solution for VSLAM and can improve accuracy

Thank you