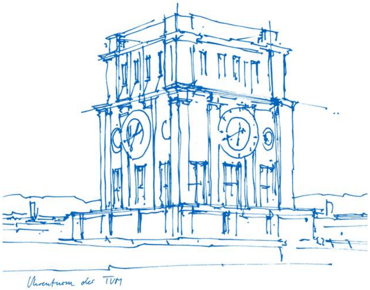
Computer Vision II: Multiple View Geometry (IN2228)

Chapter 11 Photometric Error and Direct SLAM

Dr. Haoang Li

05 July 2023 12:00-13:30



Announcements before Class

- ➢ Similarity between Codes of Assignment 4

✓ We have created a post on Moodle.
We do not want to be too harsh to students. Please refer to post for detailed information.

✓ Some students have sent emails to us to admit mistakes.
We promise that we will not take any further actions. The only loss is bonus.

✓ Some students have sent emails to us for explanation.
We will discuss your cases together later. Please give me some time because I have lots of things to handle.

✓ Some students still do not send us email. (IDs: 474, 270, 676, 388, 420, 247, 683)
I strongly recommend that you spontaneously contact us (please explicitly indicate your ID).

Announcements before Class

- Exam

- ✓ If a student fails the exam in the summer semester, he/she can take the repeat exam.

- ✓ If a student cannot take the exam in the summer semester (due to time conflict or sick), he/she can **directly** take the repeat exam.

- ✓ Currently, we do not receive new information about repeat exam, such as time and place.

- ✓ I have uploaded a new knowledge review document for Chapters 06—10. Please download it from course website or Moodle.

Announcements before Class

- Reminder of Exercise Session

✓ Today, we will hold the exercise 8 about direct method.
You will use the knowledge introduced in today's class to solve practical problems.

Wed 05.07.2023 Exercise 8: Direct Image Alignment

Wed 12.07.2023 Exercise 9: Direct Image Alignment

Explanation for Dominant Direction Association

➢ Use of Initial Pose Provided by Visual SLAM

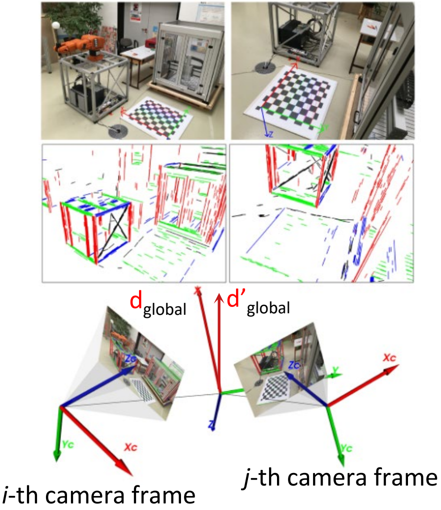
✓ Transform local dominant directions from the camera frame to the world frame based on initial rotation.

$d_{global} = R_i\, d_i \quad d_{global} = R_i\, d_i \quad d_{global} = R_i\, d_i$ (camera i)

$d'_{global} = R_j\, d_j \quad d'_{global} = R_j\, d_j \quad d'_{global} = R_j\, d_j$ (camera j)

where R_i and R_j are obtained based on constant velocity motion model and thus are not very accurate.

✓ Associate two dominant directions in the world frame
If a pair of directions has a small angle, two directions are associated.

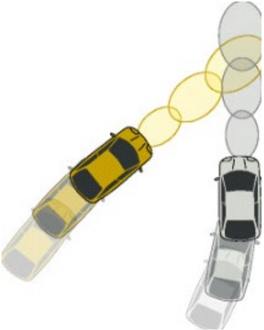


Explanation for Dominant Direction Association

➢ Three-level Camera Poses in Visual SLAM

✓ Initial pose

Directly obtain it by the constant velocity (acceleration) motion model.



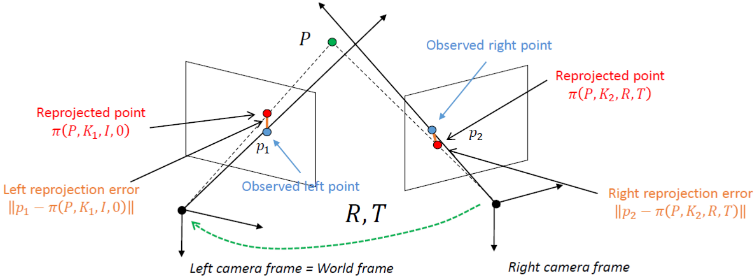
- Assume that we have three cameras. We have known the absolute rotation R_1 and R_2. We aim to initially guess the absolute pose R_3.
- We can first compute the relative pose R_{12} based on R_1 and R_2.
- Then we treat the relative pose R_{12} as the relative pose R_{23}.
- Finally, we combine the absolute pose R_2 and relative pose R_{23} to compute the absolute pose R_3.

Explanation for Dominant Direction Association

➢ Three-level Camera Poses in Visual SLAM

✓ Pose optimized based on local bundle adjustment
Re-projection error is the classical loss. (Details will be introduced tomorrow)
This loss is a general loss suitable for both structured and non-structured scenes.

$$P = argmin_P \ \|p_1 - \pi(P, K_1, I, 0)\|^2 + \|p_2 - \pi(P, K_2, R, T)\|^2$$



Explanation for Dominant Direction Association

➢ Three-level Camera Poses in Visual SLAM

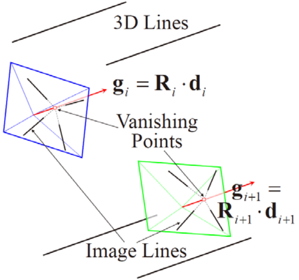
✓ Pose further optimized based on dominant direction alignment
This knowledge is introduced in our previous class.
This is only applicable to the structured environment.

Global Local

$$\mathbf{g}_i = \mathbf{R}_i \mathbf{d}_i$$

i is the ID of camera

Rotation can be used to align the dominant directions from the **camera frame** to the **world frame**



Today's Outline

- Overview and Motivation
- Photometric Error
- Direct SLAM Methods
- Photometric Calibration

Overview and Motivation

➢ Two Strategies in Multi-view Geometry

✓ Two representative papers published in ECCV 1999.

All About Direct Methods

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

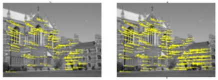
[1] Dept. of Computer Science and Applied Mathematics,
The Weizmann Inst. of Science, Rehovot, Israel.
irani@wisdom.weizmann.ac.il
[2] Microsoft Research, One Microsoft Way,
Redmond, WA 98052, USA.
anandan@microsoft.com

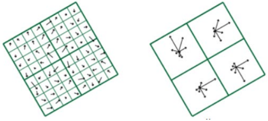
Feature Based Methods for Structure and Motion Estimation

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

Direct method





Feature-based method

Overview and Motivation

- ➢ Two Dominant SLAM Derived from The Above Two Strategies

- Indirect Method (Feature-based Method)

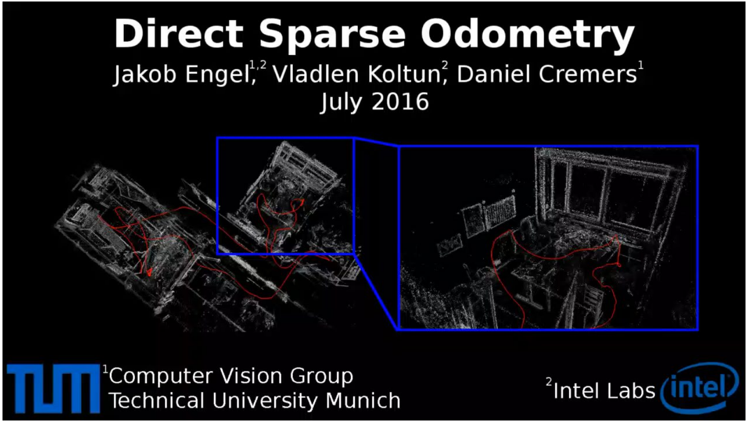
Demo video of SLAM from University of Zaragoza

This method relies on the point features

Overview and Motivation

➢ Two Dominant SLAM Derived from The Above Two Strategies

- Direct SLAM



Demo video of VO from our Computer Vision Group, TUM

This method uses the photometirc loss

Overview and Motivation

➢ Recap on Feature-based Method

✓ Abstract image as a set of keypoints

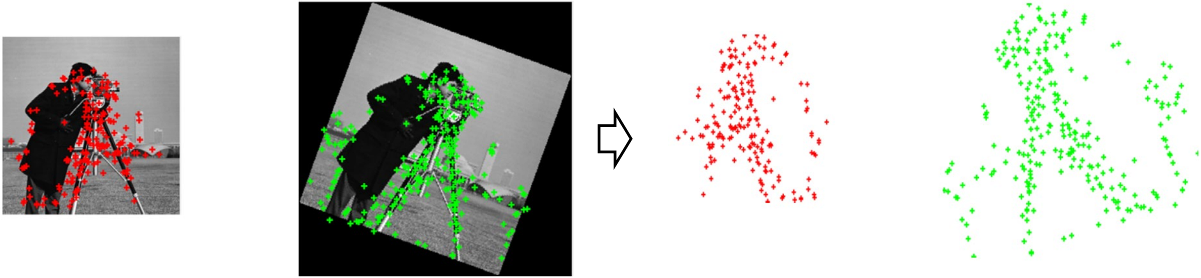


Image is reduced to a sparse set of keypoints. They are usually matched with feature descriptors.

Overview and Motivation

➢ Recap on Feature-based Method

✓ Advantages of feature-based methods

Relatively robust to viewpoint change and illumination variation

Wide-baseline matching

Illumination invariance

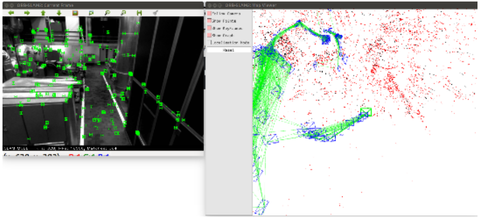
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\text{Using invariant descriptors introduced before}}$$

Overview and Motivation

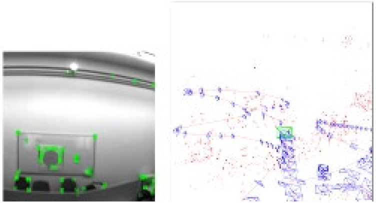
➢ Recap on Feature-based Method

✓ Disadvantages of feature-based methods

- Creates only a sparse map of the world.



- Does not sample across all available image data, e.g., discard the information around edges & weak intensities.

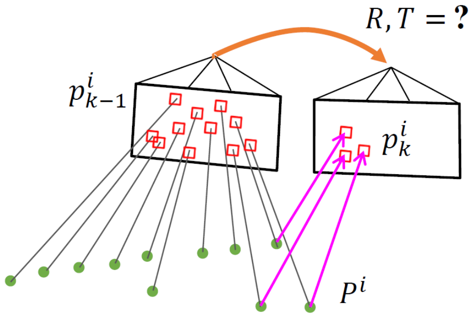


Overview and Motivation

➢ Recap on Feature-based Method

✓ Estimate the relative pose between two cameras

- Extract & match features
- Epipolar geometry constraint
- Bundle Adjust by minimizing the Re-projection Error



✓ Pros and cons

- Can cope with large frame-to-frame motions
- Slow due to costly feature extraction, matching, and outlier removal (e.g., RANSAC)

Overview and Motivation

➢ Motivation of Direct Method

✓ From two-step to one-step method to estimate the **relative pose**

Feature-based method is a **two-step** method: we will first track the features, and then determine the camera movement based on these features. Such a two-step strategy is difficult to guarantee overall optimality due to noise propagation.

Can we simultaneously determine the camera motion and feature correspondence?
We can use direct method based on the brightness consistency assumption.

Photometric Error

➢ Problem Formulation

✓ p1 and p2 are the perspective projections of 3D point P. They are associated by the unknown relative pose and depth. 3D point P is the bridge.

Normalized image points

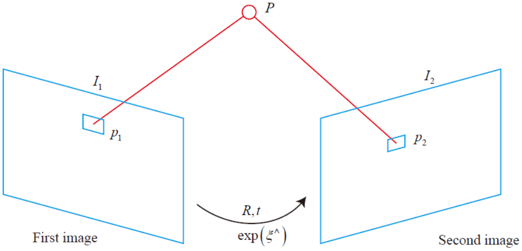
$$\mathbf{p}_1 = \begin{bmatrix} \bar{u} \\ \bar{v} \\ 1 \end{bmatrix}_1 = \underbrace{\frac{1}{Z_1}}_{\text{depth}} \boxed{\mathbf{P}},$$

Left camera frame

$$\mathbf{p}_2 = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}_2 = \mathbf{K}\left(\mathbf{R}\boxed{\mathbf{P}} + \mathbf{t}\right)$$

Right camera frame

Ordinary point



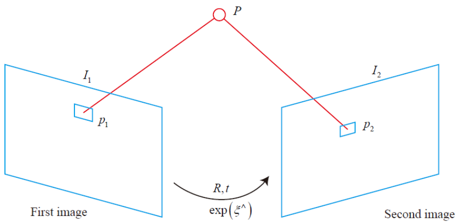
Photometric Error

➢ Problem Formulation

✓ Brightness Consistency Assumption

- Given an **arbitrary** camera pose and depth of p1, we can estimate the position of p2.
- If the camera pose and depth are not good enough, the appearance of the **estimated** p2 and the **extracted** p1 will be significantly different.
- We have prior information that correspondence should have the same brightness, i.e., brightness consistency assumption.
- Therefore, we aim to find the optimal relative camera pose and depth to minimize the brightness difference, i.e., find the optimal p2 that is more similar to p1.
- The optimal pose and correspondence are obtained simultaneously.

Photometric Error



➢ Definition

✓ Photometric error of a single pixel

Images, i.e., 2D matrix composed of intensity values

$$e = \mathbf{I}_1(\mathbf{p}_1) - \mathbf{I}_2(\mathbf{p}_2)$$

Computed position w.r.t. relative pose and depth of p1

Known coordinates of p1

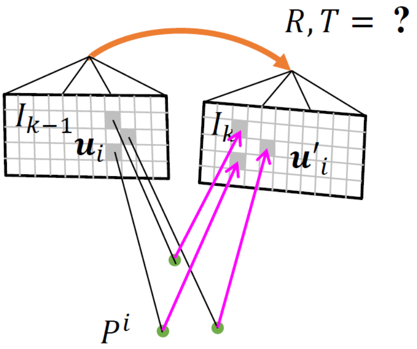
✓ Extension to multiple pixels

$$\min_{\mathbf{T}} J(\mathbf{T}) = \sum_{i=1}^{N} e_i^T e_i, \quad e_i = \mathbf{I}_1(\mathbf{p}_{1,i}) - \mathbf{I}_2(\mathbf{p}_{2,i})$$

A least-squares error

Photometric Error

➢ Definition



✓ Practical setup

- No feature extraction, no matching, no RANSAC needed
- Instead, we directly minimize Photometric Error

3D point back-projected by unknown depth

$$P^i, R, T = \arg\min_{P^i, R, T} \sum_{i=1}^{N} \left(I_{k-1}(p_{k-1}^i) - I_k\left(\pi(P^i, K, R, T)\right)\right)$$

Known

Unknown pose

✓ Pros and cons

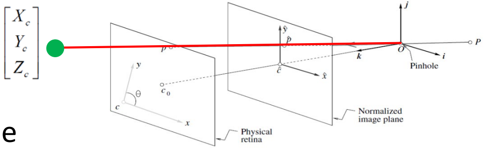
- All image pixels can in principle be used (higher accuracy, higher robustness to motion blur and weak texture (i.e., weak gradients))
- Increasing the camera frame rate reduces computational cost per frame (no RANSAC needed)
- Very sensitive to initial value limited frame to frame motion

Photometric Error

➢ Discussion about Depth

✓ Role of Depth

Based on the depth, we can back project any pixel into the three-dimensional space and then project it into the next image.



Camera frame

✓ Depth can be obtained in different ways.

- Depth can be directly obtained by RGB-D camera.
- If we have a binocular (stereo) camera, the pixel depth can also be calculated based on the disparity.
- If we only have a monocular camera, we have to treat the depth of P as a unknown variable and optimize it along with camera pose.

$$\text{Depth } \lambda \boxed{K^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}} = \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix}$$

Image normalization

Photometric Error

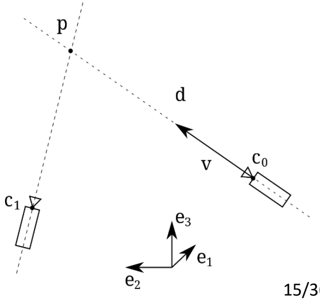
- ➢ Discussion about Depth

✓ Inverse Depth Parametrization

Some features in the environment (like clouds) are far off, leading to the distance estimate of infinity. This can cause some problems of **numerical stability**.

To get around it, the inverse of the distance is introduced. All of the infinite values become zeros which tend to cause fewer problems.

$$\rho = \frac{1}{\|\mathbf{p} - \mathbf{c}_0\|}$$



For more scientific and systematic illustration, please refer to [1].

[1] Javier Civera, Andrew J. Davison, and J. M. Martınez Montiel, "Inverse Depth Parametrization for Monocular SLAM," IEEE TRO, 2008

Photometric Error

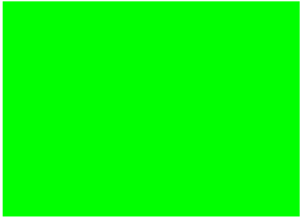
➢ Discussion about Pixels to Track

✓ Three types of pixel densities

We can track all the pixels, which is called the **dense direct method**.

- In an image, there are millions of pixels, so we cannot calculate the photometric errors for all the pixels in real-time on the existing CPU and require GPU acceleration.
- In addition, by analogy with the DLT tracker, the points with non-obvious pixel gradients will not contribute much to motion estimation
- It will be difficult to estimate the 3D position during reconstruction (some 2D points may not be tracked).

Dense methods
track every pixel



In a VGA image: 300’000+ pixels

Photometric Error

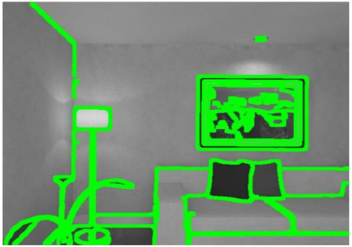
➢ Discussion about Pixels to Track

✓ Three types of pixel densities

We can track partial pixels with significant gradients. This is called a **semi-dense direct method**.

- If the pixel gradient is zero, the entire Jacobian is zero, which will not contribute to the problem.
- Therefore, we can only use pixels with high gradients, i.e., discard areas where the pixel gradients are not obvious.
- We use the tracked pixels to reconstruct a semi-dense structure.

Semi-Dense methods track only edges



In a VGA image: ~10,000 pixels

Photometric Error

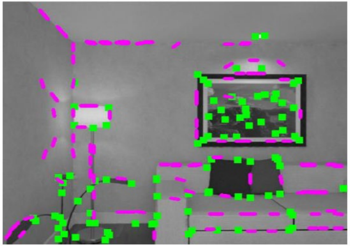
➢ Discussion about Pixels to Track

✓ Three types of pixel densities

We can track sparse key points, which we call the **sparse direct method**.

- Usually, we can obtain hundreds to thousands of key points (based on Harris detector).
- This sparse direct method does not need to calculate descriptors (like SIFT) and only uses hundreds of pixels.
- This method is the fastest, but it can only calculate sparse reconstruction.

Sparse methods
track sparse pixels



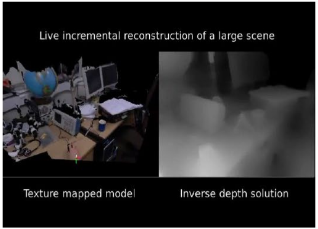
In a VGA image: ~2,000 pixels

Photometric Error

➢ Discussion about Pixels to Track

✓ Some representative methods (more information will be provided later)

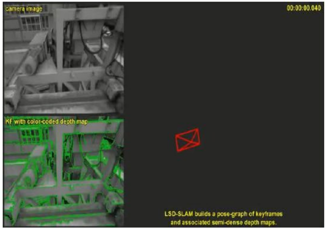
Dense methods
track every pixel



In a VGA image: 300'000+ pixels

DTAM [Newcombe '11], REMODE [Pizzoli'14]

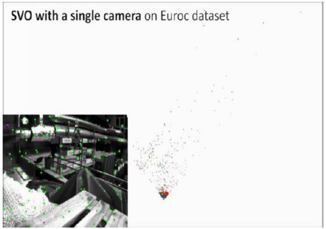
Semi-Dense methods
track only edges



In a VGA image: ~10,000 pixels

LSD-SLAM [Engel'14]

Sparse methods
track sparse pixels



In a VGA image: ~2,000 pixels
e.g., 120 feature patches × (4×4 pixels per patch)

SVO [Forster'14], DSO [Engel'17]

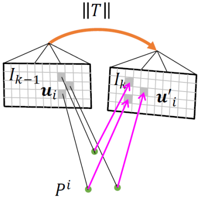
Photometric Error

➢ Discussion about Baseline (Relative Pose)

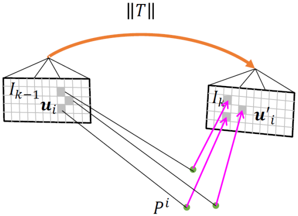
✓ What is the influence of the motion baseline on the convergence rate of direct methods?

Intuitively, direct SLAM is not suitable for large baselines for two reasons:

- Initial pose may be unreliable, which leads to the local minimum.
- Photometric consistency assumption is not satisfied.



For **small motion** baselines, $\|T\|$,
the **photometric error is usually small**



For **large motion** baselines, $\|T\|$,
the **photometric error is usually large**
(due to large geometric and illumination changes)

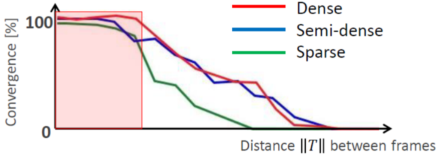
Photometric Error

➢ Discussion about Baseline (Relative Pose)

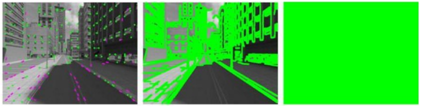
✓ What is the influence of the motion baseline on the convergence rate of direct methods?

We had the following empirical findings [1]:

- Dense and Semi-dense behave similarly for both large and small baselines.
- Sparse methods behave equally well as dense or semi dense methods for small motion baselines.



(when the two images have 0 motion baseline, we say their overlap is 100%; as the baseline increases, their overlap decreases)



(c) Sparse (d) Semi-Dense (e) Dense

[1] Forster, Zhang, Gassner, Werlberger, Scaramuzza, SVO: Semi Direct Visual Odometry for Monocular and Multi Camera Systems, IEEE Transactions on Robotics (T-RO), 2017.

Photometric Error

➢ A Short Summary

A systematic comparison between feature-based and direct method

Feature-Based	Direct
can only use & reconstruct corners	can use & reconstruct whole image
faster	slower (but good for parallelism)
flexible: outliers can be removed retroactively.	inflexible: difficult to remove outliers retroactively.
robust to inconsistencies in the model/system (rolling shutter).	not robust to inconsistencies in the model/system (rolling shutter).
decisions (KP detection) based on less complete information.	decision (ordinary point) based on more complete information.
no need for good initialization.	needs good initialization.

Key point

Direct SLAM Methods

➢ Representative Direct SLAM Methods

LSD-SLAM [1]

DSO [2]

SVO [3]

- PTAM
- ORB-SLAM
- SVO
- LSD-SLAM
- DSO

Indirect methods: Minimize the feature reprojection error

Direct methods: Minimize the feature photometric error

[1] Engel, Schoeps, Cremers, LSD SLAM: Large scale Semi Dense SLAM , European Conference on Computer Vision (ECCV), 2014.

[2] Engel, Koltun, Cremers, DSO: Direct Sparse Odometry , IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2017.

[3] Forster, Zhang, Gassner, Werlberger, Scaramuzza, SVO: Semi Direct Visual Odometry for Monocular and Multi Camera Systems , IEEE Transactions on Robotics (T RO), 2017.

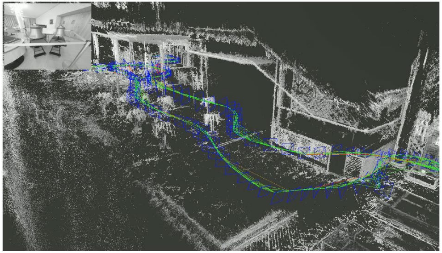
Direct SLAM Methods

➢ LSD-SLAM

✓ Supports both monocular and stereo cameras

✓ Direct (photometric error) + **Semi Dense** formulation

- 3D structure represented as semi dense depth map
- Minimizes photometric error
- Separately optimizes poses & structure



Engel, Schoeps, Cremers, LSD SLAM: Large scale Semi Dense SLAM , European Conference on Computer Vision (ECCV), 2014.

Direct SLAM Methods

➢ LSD-SLAM

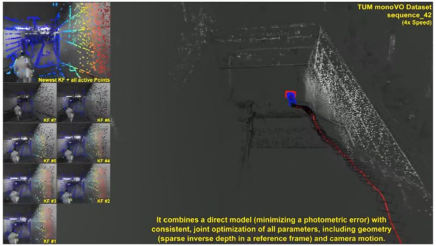
✓ Same workflow as PTAM (keyframe based, alternation of localization and mapping as independent threads)

✓ Includes:

- Loop closing
- Relocalization
- Final optimization

✓ Real-time (30Hz), however global optimization is not done in real time but asynchronously every once in a while.

Direct SLAM Methods



➢ DSO

✓ Supports both monocular and stereo cameras

✓ Direct (photometric error) + **Sparse** formulation

- 3D structure represented as sparse large gradients' depth map
- Minimizes photometric error
- Jointly optimizes poses & structure (sliding window)
- Incorporates photometric correction to compensate exposure time change $\boxed{(\Delta t_{k-1}, \Delta t_k)}$

$$P^i, R, K = \arg\min_{P^i, R, K} \sum_{i=1}^{N} \rho\left(I_{k-1}(p_{k-1}^i) - \boxed{\frac{\Delta t_{k-1}}{\Delta t_k}} I_k\left(\pi(P^i, K, R, T)\right)\right)$$

Engel, Koltun, Cremers, DSO: Direct Sparse Odometry , IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), 2017.

Direct SLAM Methods

- DSO

- ✓ Same workflow as PTAM (keyframe based, alternation of localization and mapping as independent threads)

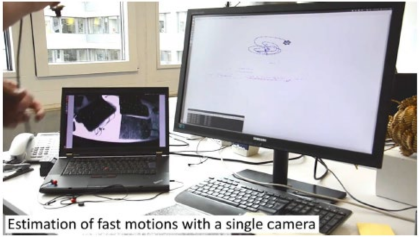
- ✓ Real time (30Hz), however global optimization is not done in real time but asynchronously every once in a while

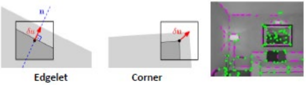
Direct SLAM Methods

➢ SVO

✓ Supports both monocular, stereo, multi camera systems as well as omnidirectional models (fisheye and catadioptric)

✓ Combines indirect + direct methods
- Direct methods for frame to frame motion estimation
- Indirect methods for frame to keyframe pose refinement





Direct SLAM Methods

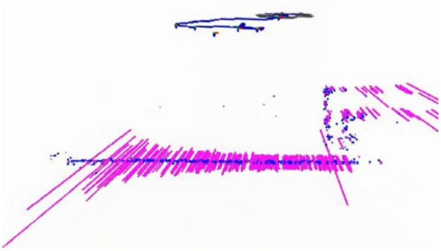
➢ SVO

✓ Mapping
- Probabilistic depth estimation (based on Gaussian distribution)

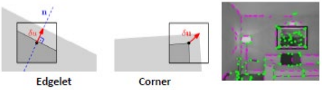
✓ Other Modules
- Loop closing,
- Relocalization
- Final optimization

✓ Same workflow as PTAM (keyframe based, alternation of localization and mapping as independent threads)

✓ Faster than real-time: up to 400 fps on i7 laptops and 100 fps on smartphone.



Probabilistic Depth Estimation



Direct SLAM Methods

➢ Comparisons Between Various Methods

✓ Efficiency

- Processing times

	Mean	CPU@20 fps
SVO Mono	2.53	55 ±10%
ORB Mono SLAM (No loop closure)	29.81	187 ±32%
LSD Mono SLAM (No loop closure)	23.23	236 ±37%
DSO	20.12	181 ±27%

Processing time in milliseconds

CPU load (100% = 1 core)

Forster, Zhang, Gassner, Werlberger, Scaramuzza, SVO: Semi Direct Visual Odometry for Monocular and Multi Camera Systems, IEEE Transactions on Robotics (T-RO), 2017.

Photometric Calibration

- Motivation

✓ Recall that photometric error relies on the brightness consistency assumption.
✓ However, in practice, this assumption may be affected by different exposure times, vignetting and other factors.

Ideal case: Consistent brightness

Photometric Calibration

➢ Motivation

✓ We try to reduce the various effects to meet the brightness consistency assumption, which is called "photometric calibration".

✓ In our class, I will only give you an overview. For detail, please refer to [1].

Before photometric calibration

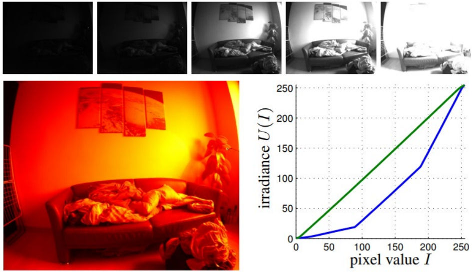
After photometric calibration

[1] P. Bergmann, R. Wang and D. Cremers, Online Photometric Calibration of Auto Exposure Video for Real-time Visual Odometry and SLAM, In IEEE Robotics and Automation Letters (RA-L), volume 3, 2018.

Photometric Calibration

- Response function

✓ Camera receives the light energy. We call the energy per unit time “irradiance”.
✓ In essence, we leverage the **irradiance consistency** between two images.
✓ Reponses function is to map this energy to digital signal (intensity or brightness of pixel).
✓ This function is a non-linear function. Accordingly, using brightness is less scientific than using irradiance. To use irradiance, we should calibrate this response function.



Photometric Calibration

- Exposure time

- ✓ Intuitively, a longer exposure time will lead to a brighter image.
- ✓ In practice, a pair of images may have different exposure times. For example, our cell phone may automatically adjust the exposure time.
- ✓ Given that we consider the consistency of irradiance (energy per in unit time), we have to calibrate the exposure time.

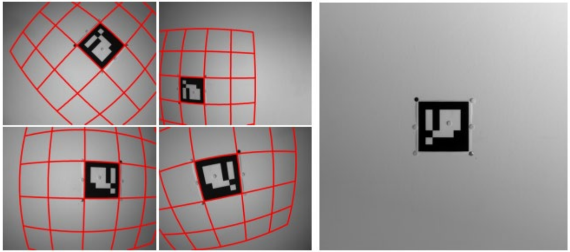
Images obtained by different exposure times

Photometric Calibration

- Vignetting

- ✓ Vignetting is a reduction of an image's brightness toward the periphery compared to the image center. It is mainly caused by the manufacturing flaw of camera.
- ✓ To apply photometric loss, we should remove this effect.

Representative illustrations



Vignetting calibration

Summary

- Overview and Motivation
- Photometric Error
- Direct SLAM Methods
- Photometric Calibration

Thank you for your listening!
If you have any questions, please come to me :-)