Univerza v *Ljubljani*
Fakulteta *za elektrotehniko*

ASSIGNMENT 2024/2025

ADVANCED COMPUTER VISION

ASSOC. PROF. JANEZ PERŠ

UL FE

DECEMBER 17, 2024

OVERVIEW

- Assignment: written and oral part
- Written part: Structure of your paper
- Experiments in computer vision
- Datasets in computer vision
- Plagiarism

ASSIGNMENT

- Assignment has two parts
  - Written paper (equivalent to written exam)
  - Oral presentation (equivalent to oral exam)
- Written paper:
  - **Maximum of 6 pages** in the predefined template
  - Fixed deadline we agree on
- Oral presentation
  - Presentation with slides, **10 min maximum**.
  - Discussion (questions)

WHAT IS "A PAPER"?

- It is a technical report
  - For your boss
  - For the archival in your company
  - For the people you supervise!
- In academia
  - Represents the result of your research

YOUR PAPER STRUCTURE

- Differs between fields
- General rules are taught in other classes
- Sections typically include:
  - Abstract
  - Introduction
  - Related work
  - Methods
  - Experiments & Results
  - Conclusion / Discussion
  - References
- Specifically for computer vision:
  - Description of the dataset
  - Analysis of results

ABSTRACT

- General rule:
  - Fellow engineers MUST be able to decide whether to read your report or not!
- If you fail with the abstract:
  - People will not read what you write, even if they should.
  - People will read your paper, and be disappointed.
- Good practice in computer vision (CV):
  - Include key contribution/result and the data you used at the end of the abstract.
  - For example, “Our tests on the ACME dataset show a 5.6% improvement in recognition performance, compared to state of the art.”

- Make your case why your work was needed
  - Your reader is past the abstract
  - You are addresing a person who thinks that your work is relevant
  - The reader knows something about the field
  - Don’t go into unnecessary explanations for the layman!

RELATED WORK

- Demonstrate that you are aware of state of the art!
  - You want to use methods proven to be successful
  - You want to avoid repeating other people's mistakes
- This is important regardless of whether you write:
  - Work report
  - Scientific paper
  - Patent application

METHODS

- Usually most important part of the paper.
- Contains
  - Technical description of your input and output data
  - Description of assumptions that you take
  - Technical description of algorithms (or other methods)
- In “Methods” (other title may be used) you explain your idea – that is the purpose of the section.

EXPERIMENTS & RESULTS

- One or two sections
  - If your experimental setup itself is novel and nonstandard, better make it a separate section.
  - But… your readers may forget what the experiments were by the time you get to results.
- No real rule, depends on final readability.
- Specify your data and dataset!
- Specify evaluation methods!
  - Avoid nonstandard evaluation, UNLESS that is the purpose of your paper.

CONCLUSION/DISCUSSION

- Longer papers, surprising results
  - Hypothesize why the discrepancy
  - Discuss if anything still unclear
  - Explain the results
  - State if more investigation is needed
- Expected, self explanatory results
  - Summarize results
  - Put results in context of your original task
- Sometimes people include both.

PAPER FORMAT & SUBMISSION

- Should follow ICCV 2023 format
  - Double column, **5-6 pages**
  - You may use LaTeX or Word
  - Full instructions: `https://iccv2023.thecvf.com/submission.guidelines-361600-2-20-16.php`
- Typesetting:
  - LaTeX: Use `www.overleaf.com` as LaTeX tool, use ICCV2023 template
  - Word: you are on your own
- Submission to e-classroom `e.fe.uni-lj.si`

PAPER FORMAT EXAMPLE

Deep geometry-aware camera self-calibration from video

Annika Hagemann[1,2], Moritz Knorr[1], Christoph Stiller[2]
[1]Bosch Research, Germany [2]Karlsruhe Institute of Technology
{annika.hagemann, moritzmichael.knorr}@de.bosch.com, stiller@kit.edu

Abstract

*Accurate intrinsic calibration is essential for camera-based 3D perception, yet, it typically requires targets of well-known geometry. Here, we propose a camera self-calibration approach that infers camera intrinsics during application, from monocular videos in the wild. We propose to explicitly model projection functions and multi-view geometry, while leveraging the capabilities of deep neural networks for feature extraction and matching. To achieve this, we build upon recent research on integrating bundle adjustment into deep learning models, and introduce a self-calibrating bundle adjustment layer. The self-calibrating bundle adjustment layer optimizes camera intrinsics through classical Gauß-Newton steps and can be adapted to different camera models without re-training. As a specific realization, we implemented this layer within the deep visual SLAM system DROID-SLAM, and show that the resulting model, **DroidCalib**, yields state-of-the-art calibration accuracy across multiple public datasets. Our results suggest that the model generalizes to unseen environments and different camera models, including significant lens distortion. Thereby, the approach enables performing 3D perception tasks without prior knowledge about the camera. Code is available at https://github.com/boschresearch/droidcalib.*

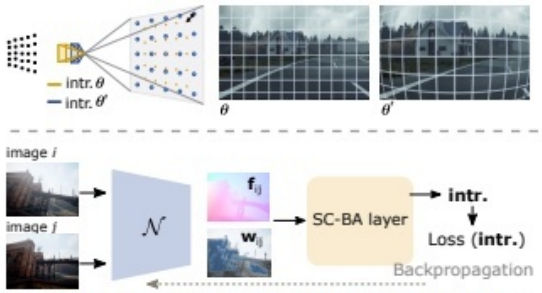

Figure 1. The proposed self-calibration infers the intrinsic camera parameters $\boldsymbol{\theta}$ that define a camera's projection function (top) without relying on calibration targets. A deep neural network $\mathcal{N}$ predicts weighted correspondences (flow $\mathbf{f}_{ij}$, confidence weights $\mathbf{w}_{ij}$), while the self-calibrating bundle adjustment (SC-BA) layer estimates the intrinsics through differentiable Gauß Newton steps (bottom).

Camera self-calibration aims at inferring camera intrinsics based on arbitrary images or image sequences, without the need for a calibration target [14]. Yet, achieving accuracy and robustness comparable to target-based calibration remains challenging. Single-image self-calibration approaches (*e.g.* [52, 51, 22, 19]) have proven effective for image undistortion and for cases in which only a single image is available, however, they have to rely on known or

EXPERIMENTS IN COMPUTER VISION

EXPERIMENTS IN COMPUTER VISION

- Image = a digital sample of real world
  - Computer vision: obtain information about the world from images.
  - Any realistic experiment must contain real-world images.
  - Synthetic images only for training.
- 3rd party or your own data set?
  - Your own data: to illustrate your problem.
  - To compare with the others, use public datasets.
  - Middle ground: publish your own data (novel problems, etc.)

DATASET SPLIT FOR DEEP LEARNING

- Never train and test on the same images!
- Proper split of the dataset:
  - Train set: to train the model, use as you please.
  - Validation set 1: to monitor progress of (deep) learning.
  - Validation set 2: to adjust the methods between the trials.
  - Test set: to test your solution. Ideally, **you use it only once!**
- Any synthetic images are included in the **train set**!
- Train set can be multiplied by augmentation!
  - Use only the augmentations that make sense for your problem!
  - Rotation, scaling
  - Noise
  - Color manipulation, etc.

NETWORK INITIALIZATION FOR DEEP LEARNING

- Never start with weights w_i set to zero!
- Use one of the following initializations:
  - All parameters set to random values
  - Pre-trained network on ImageNet
  - Pre-trained network on data closer to your problem (best)

RESULTS IN COMPUTER VISION

- Always provide quantitative results!
  - Qualitative: “As it can be seen from Figure 3, our proposed detector detects many more matching interest points”
  - **Quantitative**: “The proposed detector detects 33% more matching interest points”
  - Even better: “On the dataset A, available for download from B, consisting of 1000 image pairs, the new detector detects on average 33%±5% more matching interest points.”
- Use tables and graphs, if needed.
- Illustrate perfectly working **and** failure cases.
  - This is considered **qualitative** evaluation and is welcomed.

PLAGIARISM

UNDERSTANDING PLAGIARISM

What is plagiarism?

Plagiarism is the act of using someone else's work or ideas without giving proper credit. It can be an intentional act of copying entire works, paraphrasing without acknowledgment, or accidental, often due to negligence or misunderstanding of citation rules. In academic and professional settings, plagiarism is considered unethical and can lead to serious consequences.

TYPES OF PLAGIARISM

- Direct Plagiarism:
  - Copying text word-for-word from a source without attribution.
- Self-Plagiarism:
  - Reusing one's own previously published work without citing it.
- Mosaic Plagiarism:
  - Piecing together texts from various sources to create a new text, without proper citation.
- Accidental Plagiarism:
  - Unintentionally failing to cite sources, often due to lack of knowledge or attention to detail.

WHEN IS PLAGIARISM ACCEPTABLE?

Never!

Plagiarism is **never** acceptable. It may result in significant sanctions (even expulsion from academic institution). It may lead to loss of repuation, or in commercial setting, to **lawsuits** and monetary damages. Different types of plagiarism influence only the degree of sanctions – all types are unacceptable.

YOUR TASK 2024/2025

TASK: VISION FOR USVS

- USV = unmanned surface vehicle
- Semantic segmentation: which pixels belong to
  - Water surface
  - Obstacles (things you don't want to hit)
  - Sky (no accuracy needed there)
- Object detection: find bounding boxes of
  - Boats
  - People in boats, SUP, etc.

SEMANTIC SEGMENTATION

OBJECT DETECTION

THE ASSIGNMENT

- You will be provided with:
  - Dataset (images of water environments)
  - Ground truth
- Your (segmentation) task will be:
  - Train basic segmentation algorithm on ground truth.
  - Do the evaluation
- Same for detection.
- as. Jon Muhovič will provide further details
  - After the labs are finished

ANY QUESTIONS?

ASSIGNMENT DETAILS

PART 1: MAXIMUM GRADE 7

- You will will perform **segmentation** of water vs everything else on a USV dataset.
- Ground truth will be provided as separate images.
- Ground truth will be **pixel-accurate**.
- You may use U-net.
- You will train your network!

PART 2: MAXIMUM GRADE 8

- You will will perform **detection** of obstacles on the same dataset.
- Ground truth will be provided as bounding boxes.
- You may use Yolo (any version, we recommend v8)
- You will train your network!

PART 3: MAXIMUM GRADE 9

- You will will perform **segmentation** of water vs everything else on a **thermal images**.
- Ground truth will be provided as separate images.
- Ground truth will be **pixel-accurate**.
- You may use U-net.
- You will train your network!

PART 4: MAXIMUM GRADE 10

- You will will perform **detection** of obstacles on a **thermal images**.
- Ground truth will be provided as bounding boxes.
- You may use Yolo (any version, we recommend v8)
- You will train your network!