

# Learned Single-Shot Image-Based Camera Localization

Master Thesis

Moyuan Zhou

September 15, 2019

Advisors: Dr. Bugra Tekin, Dr. Johannes L. Schönberger Supervisor: Prof. Marc Pollefeys

Department of Computer Science, ETH Zürich

#### Abstract

This example thesis briefly shows the main features of our thesis style, and how to use it for your purposes.

## Contents

C	onten	ts	iii
1	Intr	oduction	1
	1.1	Features	1
		1.1.1 Extra package includes	1
		1.1.2 Layout setup	2
		1.1.3 Theorem setup	2
		1.1.4 Macro setup	3
2	Wri	ting scientific texts in English	5
	2.1	Basic writing rules	5
	2.2	Being nice to the reader	5
	2.3	A few important grammar rules	6
	2.4	Things you (usually) don't say in English	9
3	Тур	ography	11
	3.1	Punctuation	11
	3.2	Spacing	12
	3.3	Choice of 'fonts'	13
	3.4	Displayed equations	13
	3.5	Floats	14
4	Intr	oduction	17
	4.1	Motivation	18
	4.2	Background	18
	4.3	Focus of this work	18
		4.3.1 Example Subsection	18
5	Exa	mple Chapter	19
	5.1		19
		5.1.1 Example Subsection	19

#### Contents

6	Dat	aset		21
	6.1	Datas	et types	21
		6.1.1	Synthetic data	21
		6.1.2	Realistic data	21
A	Dur	nmy A	ppendix	23
Bi	bliog	raphy		25

## Introduction

This is version v1.4 of the template.

We assume that you found this template on our institute's website, so we do not repeat everything stated there. Consult the website again for pointers to further reading about LATEX. This chapter only gives a brief overview of the files you are looking at.

#### 1.1 Features

The rest of this document shows off a few features of the template files. Look at the source code to see which macros we used!

The template is divided into T<sub>E</sub>X files as follows:

- 1. thesis.tex is the main file.
- 2. extrapackages.tex holds extra package includes.
- 3. layoutsetup.tex defines the style used in this document.
- 4. theoremsetup.tex declares the theorem-like environments.
- 5. macrosetup.tex defines extra macros that you may find useful.
- 6. introduction.tex contains this text.
- 7. sections.tex is a quick demo of each sectioning level available.
- 8. refs.bib is an example bibliography file. You can use BibTEX to quote references. For example, read [?] if you can get a hold of it[4].

#### 1.1.1 Extra package includes

The file extrapackages.tex lists some packages that usually come in handy. Simply have a look at the source code. We have added the following comments based on our experiences:

**REC** This package is recommended.

**OPT** This package is optional. It usually solves a specific problem in a clever way.

**ADV** This package is for the advanced user, but solves a problem frequent enough that we mention it. Consult the package's documentation.

As a small example, here is a reference to the Section *Features* typeset with the recommended *varioref* package:

See Section 1.1 on the preceding page.

#### 1.1.2 Layout setup

This defines the overall look of the document – for example, it changes the chapter and section heading appearance. We consider this a 'do not touch' area. Take a look at the excellent *Memoir* documentation before changing it.

In fact, take a look at the excellent Memoir documentation, full stop.

#### 1.1.3 Theorem setup

This file defines a bunch of theorem-like environments.

**Theorem 1.1** An example theorem.

**Proof** Proof text goes here.

Note that the q.e.d. symbol moves to the correct place automatically if you end the proof with an enumerate or displaymath. You do not need to use \qedhere as with amsthm.

**Theorem 1.2 (Some Famous Guy)** Another example theorem.

**Proof** This proof

1. ends in an enumerate.

**Proposition 1.3** *Note that all theorem-like environments are by default numbered on the same counter.* 

**Proof** This proof ends in a display like so:

$$f(x) = x^2$$
.

#### 1.1.4 Macro setup

For now the macro setup only shows how to define some basic macros, and how to use a neat feature of the *mathtools* package:

$$|a|$$
,  $\left|\frac{a}{b}\right|$ ,  $\left|\frac{a}{b}\right|$ .

## Writing scientific texts in English

This chapter was originally a separate document written by Reto Spöhel. It is reprinted here so that the template can serve as a quick guide to thesis writing, and to provide some more example material to give you a feeling for good typesetting.

#### 2.1 Basic writing rules

The following rules need little further explanation; they are best understood by looking at the example in the booklet by Knuth et al.,  $\S2-\S3$ .

Rule 2.1 Write texts, not chains of formulas.

More specifically, write full sentences that are logically interconnected by phrases like 'Therefore', 'However', 'On the other hand', etc. where appropriate.

**Rule 2.2** Displayed formulas should be embedded in your text and punctuated with it.

In other words, your writing should not be divided into 'text parts' and 'formula parts'; instead the formulas should be tied together by your prose such that there is a natural flow to your writing.

#### 2.2 Being nice to the reader

Try to write your text in such a way that a reader enjoys reading it. That's of course a lofty goal, but nevertheless one you should aspire to!

Rule 2.3 Be nice to the reader.

Give some intuition or easy example for definitions and theorems which might be hard to digest. Remind the reader of notations you introduced many pages ago – chances are he has forgotten them. Illustrate your writing with diagrams and pictures where this helps the reader. Etc.

#### Rule 2.4 Organize your writing.

Think carefully about how you subdivide your thesis into chapters, sections, and possibly subsections. Give overviews at the beginning of your thesis and of each chapter, so the reader knows what to expect. In proofs, outline the main ideas before going into technical details. Give the reader the opportunity to 'catch up with you' by summing up your findings periodically.

*Useful phrases:* 'So far we have shown that ...', 'It remains to show that ...', 'Recall that we want to prove inequality (7), as this will allow us to deduce that ...', 'Thus we can conclude that .... Next, we would like to find out whether ...', etc.

**Rule 2.5** Don't say the same thing twice without telling the reader that you are saying it twice.

Repetition of key ideas is important and helpful. However, if you present the same idea, definition or observation twice (in the same or different words) without telling the reader, he will be looking for something new where there is nothing new.

*Useful phrases:* 'Recall that [we have seen in Chapter 5 that] ...', 'As argued before / in the proof of Lemma 3, ...', 'As mentioned in the introduction, ...', 'In other words, ...', etc.

**Rule 2.6** Don't make statements that you will justify later without telling the reader that you will justify them later.

This rule also applies when the justification is coming right in the next sentence! The reasoning should be clear: if you violate it, the reader will lose valuable time trying to figure out on his own what you were going to explain to him anyway.

*Useful phrases:* 'Next we argue that ...', 'As we shall see, ...', 'We will see in the next section that ..., etc.

#### 2.3 A few important grammar rules

Rule 2.7 There is (almost) never a comma before 'that'.

It's really that simple. Examples:

We assume that ...
Wir nehmen an, dass ...

```
It follows that ... Daraus folgt, dass ...
```

'thrice' is a word that is seldom used.

'thrice' ist ein Wort, das selten verwendet wird.

Exceptions to this rule are rare and usually pretty obvious. For example, you may end up with a comma before 'that' because 'i.e.' is spelled out as 'that is':

```
For p(n) = \log n/n we have ... However, if we choose p a little bit higher, that is p(n) = (1 + \varepsilon) \log n/n for some \varepsilon > 0, we obtain that...
```

Or you may get a comma before 'that' because there is some additional information inserted in the middle of your sentence:

Thus we found a number, namely  $n_0$ , that satisfies equation (13).

If the additional information is left out, the sentence has no comma:

Thus we found a number that satisfies equation (13).

(For 'that' as a relative pronoun, see also Rules 2.9 and 2.10 below.)

**Rule 2.8** There is usually no comma before 'if'.

Example:

A graph is not 3-colorable if it contains a 4-clique. *Ein Graph ist nicht 3-färbbar, wenn er eine 4-Clique enthält.* 

However, if the 'if' clause comes first, it is usually separated from the main clause by a comma:

```
If a graph contains a 4-clique, it is not 3-colorable . Wenn ein Graph eine 4-Clique enthält, ist er nicht 3-färbbar.
```

There are more exceptions to these rules than to Rule 2.7, which is why we are not discussing them here. Just keep in mind: don't put a comma before 'if' without good reason.

Rule 2.9 Non-defining relative clauses have commas.

Rule 2.10 Defining relative clauses have no commas.

In English, it is very important to distinguish between two types of relative clauses: defining and non-defining ones. This is a distinction you absolutely need to understand to write scientific texts, because mistakes in this area actually distort the meaning of your text!

It's probably easier to explain first what a *non-defining* relative clause is. A non-defining relative clauses simply gives additional information *that could* also be left out (or given in a separate sentence). For example, the sentence

The Weird Sort algorithm, which was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

would be fully understandable if the relative clause were left out completely. It could also be rephrased as two separate sentences:

The Weirdsort algorithm is theoretically best possible but difficult to implement in practice. [By the way,] Weirdsort was found by the famous mathematician John Doe.

This is what a non-defining relative clause is. *Non-defining relative clauses are always written with commas.* As a corollary we obtain that you cannot use 'that' in non-defining relative clauses (see Rule 2.7!). It would be wrong to write

The WeirdSort algorithm, that was found by the famous mathematician John Doe, is theoretically best possible but difficult to implement in practice.

A special case that warrants its own example is when 'which' is referring to the entire preceding sentence:

Thus inequality (7) is true, which implies that the Riemann hypothesis holds.

As before, this is a non-defining relative sentence (it could be left out) and therefore needs a comma.

So let's discuss *defining* relative clauses next. A defining relative clause tells the reader *which specific item the main clause is talking about*. Leaving it out either changes the meaning of the sentence or renders it incomprehensible altogether. Consider the following example:

The WeirdSort algorithm is difficult to implement in practice. In contrast, the algorithm that we suggest is very simple.

Here the relative clause 'that we suggest' cannot be left out – the remaining sentence would make no sense since the reader would not know which algorithm it is talking about. This is what a defining relative clause is. *Defining relative clauses are never written with commas*. Usually, you can use both 'that' and 'which' in defining relative clauses, although in many cases 'that' sounds better.

As a final example, consider the following sentence:

For the elements in  $\mathcal{B}$  which satisfy property (A), we know that equation (37) holds.

Table 21: Things you (usually) don't say

	, -,	<u>-                                      </u>
It holds (that)	We have	Es gilt
('Equation (5) holds.' is fine,	though.)	
$x$ fulfills property $\mathcal{P}$ .	$x$ satisfies property $\mathcal{P}$ .	$x$ erfüllt Eigenschaft ${\cal P}.$
<del>in average</del>	on average	im Durchschnitt
estimation	estimate	Abschätzung
composed number	composite number	zusammengesetzte Zahl
with the help of	using	mit Hilfe von
surely	clearly	sicher, bestimmt
monotonously increasing	monotonically incr.	monoton steigend
(Actually, in most cases 'incre	easing' is just fine.)	

This sentence does not make a statement about all elements in  $\mathcal{B}$ , only about those satisfying property (A). The relative clause is *defining*. (Thus we could also use 'that' in place of 'which'.)

In contrast, if we add a comma the sentence reads

For the elements in  $\mathcal{B}$ , which satisfy property (A), we know that equation (37) holds.

Now the relative clause is *non-defining* – it just mentions in passing that all elements in  $\mathcal{B}$  satisfy property (A). The main clause states that equation (37) holds for *all* elements in  $\mathcal{B}$ . See the difference?

## 2.4 Things you (usually) don't say in English – and what to say instead

Table 21 lists some common mistakes and alternatives. The entries should not be taken as gospel – they don't necessarily mean that a given word or formulation is wrong under all circumstances (obviously, this depends a lot on the context). However, in nine out of ten instances the suggested alternative is the better word to use.

## **Typography**

#### 3.1 Punctuation

Rule 3.1 Use opening (') and closing (') quotation marks correctly.

In LaTeX, the closing quotation mark is typed like a normal apostrophe, while the opening quotation mark is typed using the French *accent grave* on your keyboard (the *accent grave* is the one going down, as in *frère*).

Note that any punctuation that *semantically* follows quoted speech goes inside the quotes in American English, but outside in Britain. Also, Americans use double quotes first. Oppose

"Using 'lasers,' we punch a hole in ... the Ozone Layer," Dr. Evil said.

to

'Using "lasers", we punch a hole in . . . the Ozone Layer', Dr. Evil said.

Rule 3.2 Use hyphens (-), en-dashes (-) and em-dashes (--) correctly.

A hyphen is only used in words like 'well-known', '3-colorable' etc., or to separate words that continue in the next line (which is known as hyphenation). It is entered as a single ASCII hyphen character (-).

To denote ranges of numbers, chapters, etc., use an en-dash (entered as two ASCII hyphens --) with no spaces on either side. For example, using Equations (1)–(3), we see...

As the equivalent of the German *Gedankenstrich*, use an en-dash with spaces on both sides – in the title of Section 2.4, it would be wrong to use a hyphen instead of the dash. (Some English authors use the even longer emdash (—)

instead, which is typed as three subsequent hyphens in LATEX. This emdash is used without spaces around it—like so.)

#### 3.2 Spacing

#### Rule 3.3 Do not add spacing manually.

Rule 3.4 Judiciously insert spacing in maths where it helps.

This directly contradicts Rule 3.3, but in some cases TEX fails to correctly decide how much spacing is required. For example, consider

$$f(a,b) = f(a+b,a-b).$$

In such cases, inserting a thin math space \, greatly increases readability:

$$f(a,b) = f(a+b, a-b).$$

Along similar lines, there are variations of some symbols with different spacing. For example, Lagrange's Theorem states that |G| = [G:H]|H|, but the proof uses a bijection  $f: aH \to bH$ . (Note how the first colon is symmetrically spaced, but the second is not.)

**Rule 3.5** Learn when to use  $\setminus$  and  $\setminus$ 0.

Unless you use 'french spacing', the space at the end of a sentence is slightly larger than the normal interword space.

The rule used by TEX is that any space following a period, exclamation mark or question mark is sentence-ending, except for periods preceded by an upper-case letter. Inserting \ before a space turns it into an interword space, and inserting \ 0 before a period makes it sentence-ending. This means you should write

Prof.\ Dr.\ A. Steger is a member of CADMO\@. If you want to write a thesis with her, you should use this template.

which turns into

Prof. Dr. A. Steger is a member of CADMO. If you want to write a thesis with her, you should use this template.

The effect becomes more dramatic in lines that are stretched slightly during justification:

Prof. Dr. A. Steger is a member of CADMO. If you

Rule 3.6 Place a non-breaking space (~) right before references.

This is actually a slight simplification of the real rule, which should invoke common sense. Place non-breaking spaces where a line break would look 'funny' because it occurs right in the middle of a construction, especially between a reference type (Chapter) and its number.

#### 3.3 Choice of 'fonts'

Professional typography distinguishes many font attributes, such as family, size, shape, and weight. The choice for sectional divisions and layout elements has been made, but you will still occasionally want to switch to something else to get the reader's attention. The most important rule is very simple.

Rule 3.7 When emphasising a short bit of text, use \emph.

In particular, *never* use bold text (\textbf). Italics (or Roman type if used within italics) avoids distracting the eye with the huge blobs of ink in the middle of the text that bold text so quickly introduces.

Occasionally you will need more notation, for example, a consistent typeface used to identify algorithms.

Rule 3.8 Vary one attribute at a time.

For example, for WEIRDSORT we only changed the shape to small caps. Changing two attributes, say, to bold small caps would be excessive (LATEX does not even have this particular variation). The same holds for mathematical notation: the reader can easily distinguish  $g_n$ , G(x), G and G.

**Rule 3.9** Never underline or uppercase.

No exceptions to this one, unless you are writing your thesis on a typewriter. Manually. Uphill both ways. In a blizzard.

#### 3.4 Displayed equations

**Rule 3.10** Insert paragraph breaks *after* displays only where they belong. Never insert paragraph breaks *before* displays.

LATEX translates sequences of more than one linebreak (i.e., what looks like an empty line in the source code) into a paragraph break in almost all contexts. This also happens before and after displays, where extra spacing is inserted to give a visual indication of the structure. Adding a blank line in these places may look nice in the sources, but compare the resulting display

$$a = b$$

to the following:

$$a = b$$

The first display is surrounded by blank lines, but the second is not. It is bad style to start a paragraph with a display (you should always tell the reader what the display means first), so the rule follows.

#### Rule 3.11 Never use eqnarray.

It is at the root of most ill-spaced multiline displays. The *amsmath* package provides better alternatives, such as the align family

$$f(x) = \sin x,$$
  
$$g(x) = \cos x,$$

and multline which copes with excessively long equations:

$$P[X_{t_0} \in (z_0, z_0 + dz_0], \dots, X_{t_n} \in (z_n, z_n + dz_n]]$$

$$= \nu(dz_0)K_{t_1}(z_0, dz_1)K_{t_2-t_1}(z_1, dz_2) \cdots K_{t_n-t_{n-1}}(z_{n-1}, dz_n).$$

#### 3.5 Floats

By default this style provides floating environments for tables and figures. The general structure should be as follows:

```
\begin{figure}
  \centering
  % content goes here
  \caption{A short caption}
  \label{some-short-label}
\end{figure}
```

Note that the label must follow the caption, otherwise the label will refer to the surrounding section instead. Also note that figures should be captioned at the bottom, and tables at the top.

The whole point of floats is that they, well, *float* to a place where they fit without interrupting the text body. This is a frequent source of confusion and changes; please leave it as is.

Rule 3.12 Do not restrict float movement to only 'here' (h).

If you are still tempted, you should avoid the float altogether and just show the figure or table inline, similar to a displayed equation.

## Introduction

In recent years, image-based camera localization has been a key task in the areas of augmented reality, virtual reality, robotics, autonomous driving, etc. In this thesis, we focus on camera localization from RGB images and aim for real-time, efficient, and robust camera localization.

Traditionally, the PnP algorithm [2] is used to solve the camera localization problem by computing the 6D camera pose given the 2D and 3D corresponding coordinates of multiple points. The fundamental problems of traditional approaches relying on local image features are textureless environments and robustness against strong changes in illumination/occlusion/viewpoint/etc. between the localized image and the given 3D model. To address these limitations, recently, deep learning has been applied to predict the 2D and 3D coordinates of these control points [1] and has achieved superior results. However, these approaches are typically very time-consuming to train and evaluate. In this thesis, we want to combine the benefits of both worlds and develop a learned approach that is efficient and robust.

The problem of 6D object pose estimation from RGB images is related to the problem of camera localization. Recently, an efficient, single-shot approach for simultaneously detecting an object in an RGB image and predicting its 6D pose without requiring multiple stages [3] has been proposed. This approach resulted in an improvement over the state-of-the-art in terms of accuracy and efficiency, and addressed the challenges on keypoint occlusion and multiple object pose estimation. In this project, we aim to adapt this approach for efficient image-based camera localization. To this end, we will define keypoints on the 3D room layout for indoor environments and on 3D building models for outdoor environments and predict the projections of these 3D keypoints. The 6D pose will then be computed using PnP [2] based on the correspondences between 2D predictions and 3D reference keypoints. The major challenges of the project include limited data for training the localization task, occlusion, and motion. To address these issues, one

initial idea is to increase the number of keypoints without slowing down the method, which is a direction to go for higher accuracy and robustness. [1] provided a trainable RANSAC approach for larger set of control points and can be integrated with the current model. Cutting edge methods could also be learned and utilized to achieve a higher performance.

#### 4.1 Motivation

Image-based camera localization is to compute camera poses under some world coordinate system from images or video captured by the cameras. [5] In recent years, image-based camera localization has great and wide applications in multiple areas. Autonomous localization and navigation is needed for a moving robot Augment reality on images: camera pose or localization View virtual environment: corresponding viewing angle should be computed Ubiquitous Relocalization Real time

#### 4.2 Background

There are many localization tools nowadays, among which GPS is very commonly used outdoors, but it cannot be used indoors. Lidar, UWB, WiFi AP et al are effective indoor localization tools. However, they require special devices or data collection in advance. There are also many methods replying on RGB-D cameras which are quite robust and accurate Compared to other localization tools, camera localization is the most flexible and low cost one.

Currently, there are many effective and robust methods relying on depth information acquired by RGB-D cameras. However, the RGB-D cameras are power consuming and not ubiquitous as 2D cameras. Thus, in this work we focus on RGB images that could be acquired by 2D cameras, and do not rely on depth information.

#### 4.3 Focus of this work

#### 4.3.1 Example Subsection

Dummy text.

**Example Subsubsection** 

Dummy text.

**Example Paragraph** Dummy text.

Example Subparagraph Dummy text.

## **Example Chapter**

Dummy text.

## 5.1 Example Section

Dummy text.

#### 5.1.1 Example Subsection

Dummy text.

**Example Subsubsection** 

Dummy text.

**Example Paragraph** Dummy text.

Example Subparagraph Dummy text.

## **Dataset**

Dummy text.

## 6.1 Dataset types

Dummy text.

#### 6.1.1 Synthetic data

Dummy text.

#### 6.1.2 Realistic data

Dummy text.

Scannet

**Example Paragraph** Dummy text.

Example Subparagraph Dummy text.

## Appendix A

## **Dummy Appendix**

You can defer lengthy calculations that would otherwise only interrupt the flow of your thesis to an appendix.

## **Bibliography**

- [1] Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. Dsac-differentiable ransac for camera localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6684–6692, 2017.
- [2] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Epnp: An accurate o (n) solution to the pnp problem. *International journal of computer vision*, 81(2):155, 2009.
- [3] Bugra Tekin, Sudipta N Sinha, and Pascal Fua. Real-time seamless single shot 6d object pose prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 292–301, 2018.
- [4] Jian Wu, Liwei Ma, and Xiaolin Hu. Delving deeper into convolutional neural networks for camera relocalization. In 2017 IEEE International Conference on Robotics and Automation (ICRA), pages 5644–5651. IEEE, 2017.
- [5] Yihong Wu, Fulin Tang, and Heping Li. Image-based camera localization: an overview. *Visual Computing for Industry, Biomedicine, and Art*, 1(1):1–13, 2018.



Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

## **Declaration of originality**

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

respective electronic versions.				
ecturers may also require a declaration of originality for other written papers compiled for their courses.				
I hereby confirm that I am the sole author of the in my own words. Parts excepted are correction	e written work here enclosed and that I have compiled it ons of form and content by the supervisor.			
Title of work (in block letters):				
<b>Authored by</b> (in block letters): For papers written by groups the names of all authors are	required.			
Name(s):	First name(s):			
	agiarism described in the 'Citation etiquette' information			
<ul><li>sheet.</li><li>I have documented all methods, data and</li><li>I have not manipulated any data.</li></ul>	processes truthfully.			
- I have mentioned all persons who were si	ignificant facilitators of the work.			
I am aware that the work may be screened ele	ectronically for plagiarism.			
Place, date	Signature(s)			

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.