

# Visual AI

CPSC 532R/533R – 2019/2020 Term 2

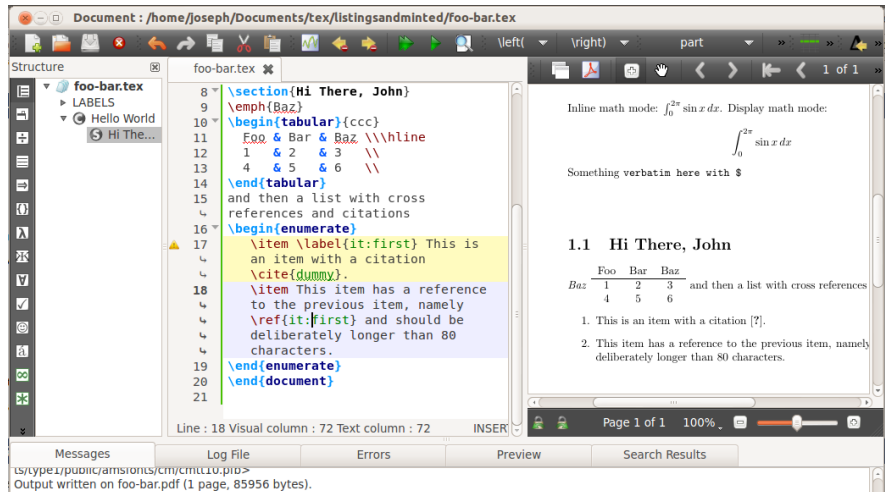
**Lecture 23.** How to write a paper/report

Helge Rhodin



# Text editor and file format

- Use latex!
  - good support for equations
  - floating figures and tables
  - nice line breaks and overall formatting
  - automatic references
  - automatic bibliography
  - many templates
- we use:
  - <http://cvpr2019.thecvf.com/files/cvpr2019AuthorKit.zip>



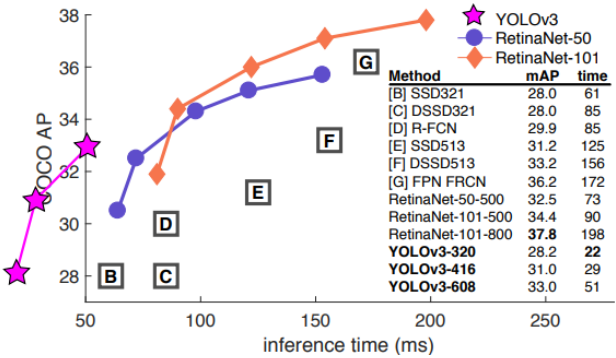
## Many latex editors available

- TEXstudio (offline)
- Overleaf (online, collaborative)
- ...

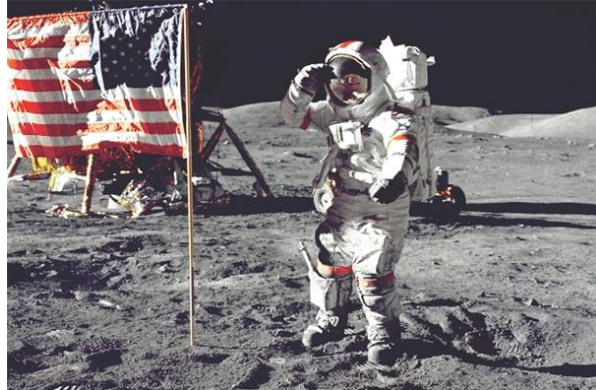
# Intention

- ~~get the best grade?~~
- ~~entertaining?~~
- tell a story?
- ~~build tension?~~
- inform?
- teach?
- document?
- archive?
- get most citations?
- trigger future research?
- plant a flag?
- discuss?
- clarify?
- make people remember?

a bit of everything!



[YOLOv3 paper]



# Who is reading your research paper?

- Sometimes: Researchers in related fields
  - ECE, Math, ...
  
- Often: Researchers in your general field
  - computer science
  
- Mostly: Researchers in your specialization
  - computer vision, computer graphics, machine learning
  - newcomers, veterans,  
BSc, MSc, PhD, ...



# How will people read your research paper?

Papers aren't read front to back like a novel!

1. read title
  - is it relevant?
2. read abstract and teaser
  - is it relevant and insightful?
3. quick pass
  - get general idea about the paper
4. content pass
  - grasp paper contents, but skip details
5. details pass
  - understand the paper in depth

If you don't draw the attention, people might stop at 1. or 2. !

To appear in ACM TOG (SIGGRAPH Asia 2016).

### EgoCap: Egocentric Marker-less Motion Capture with Two Fisheye Cameras

Helge Rhodin<sup>1</sup>    Christian Richardt<sup>1, 2, 3</sup>    Dan Casas<sup>1</sup>    Eldar Insafutdinov<sup>1</sup>  
 Mohammad Shafiq<sup>1</sup>    Hans-Peter Seidel<sup>1</sup>    Bernd Schiele<sup>1</sup>    Christian Theobalt<sup>1</sup>

<sup>1</sup>Max Planck Institute for Informatics    <sup>2</sup>Intel Visual Computing Institute    <sup>3</sup>University of Bath

**Prototypes of EgoCap motion-capture rig**

1. Attached to bike helmet  
2. Attached to Oculus VR HMD

**Application scenario**

**EgoCap motion-capture applications**

Large-scale outdoor capture    Contrasted, crowded spaces    Embodied virtual reality

Input views    Reconstruction    motion track of skeletons    reconstructed skeletons    first person view in VR

**Figure 1:** We propose a marker-less optical motion-capture approach that only uses two head-mounted fisheye cameras (see rigs on the left). Our approach enables three new application scenarios: (1) capturing human motions in outdoor environments of virtually unlimited size, (2) capturing motions in space-constrained environments, e.g. during social interactions, and (3) rendering the reconstruction of one's real body in virtual reality for embodied immersion.

**Abstract**

Marker-based and marker-less optical skeletal motion-capture methods use an *outside-in* arrangement of cameras placed around a scene, with viewpoints converging on the center. They often create discomfort with marker suits, and their recording volume is severely restricted and often constrained to indoor scenes with controlled backgrounds. Alternative suit-based systems use several inertial measurement units or an exoskeleton to capture motion with an *inside-in* setup, i.e. without external sensors. This makes capture independent of a confined volume, but requires substantial, often constraining, and hard to set up body instrumentation. Therefore, we propose a new method for real-time, marker-less, and egocentric motion capture: estimating the full-body skeleton pose from a lightweight stereo pair of fisheye cameras attached to a helmet or virtual reality headset – an optical *inside-in* method, so to speak. This allows full-body motion capture in general indoor and outdoor scenes, including crowded scenes with many people nearby, which enables reconstruction in larger-scale activities. Our approach combines the strength of a new generative pose estimation framework for fisheye views with a ConvNet-based body-part detector trained on a large new dataset. It is particularly useful in virtual reality to freely roam and interact, while seeing the fully motion-captured virtual body.

**Keywords:** Motion capture, first-person vision, markerless, optical, inside-in, crowded scenes, large-scale

**Concepts:** •Computing methodologies → Motion capture;

**1 Introduction**

Traditional optical skeletal motion-capture methods – both marker-based and marker-less – use several cameras typically placed around a scene in an *outside-in* arrangement, with camera views approximately converging in the center of a confined recording volume. This greatly constrains the spatial extent of motions that can be recorded: simply enlarging the recording volume by using more cameras, for instance to capture an athlete, is not scalable. *Outside-in* arrangements also constrain the type of scene that can be recorded, even if it fits into a confined space. If a recording location is too small, cameras can often not be placed sufficiently far away. In other cases, a scene may be cluttered with objects or furniture, or other dynamic scene elements, such as people in close interaction, may obstruct a motion-captured person in the scene or create unwanted dynamics in the background. In such cases, even state-of-the-art *outside-in* marker-less optical methods that succeed with just a few cameras, and are designed for less controlled and outdoor scenes [Elhayek et al. 2015], quickly fail. Scenes with dense social interaction were previously captured with *outside-in* camera arrays of a few hundred sensors [Joo et al. 2013], a very complex and difficult to scale setup.

These strong constraints on recording volume and scene density prevent the use of optical motion capture in the majority of real-world scenes. This problem can partly be bypassed with *inside-in* motion-capture methods that use body-worn sensors exclusively [Menache 2010], such as the Xsens MVN inertial measurement unit suit. However, the special suit and cabling are obstructive and require tedious calibration. Shiratori et al. [2011] propose to wear 16 cameras placed on body parts facing *inside-out*, and capture the skeletal motion through structure-from-motion relative to the environment. This clever solution requires instrumentation, calibration and a static background, but allows free roaming. This design was inspirational for our egocentric approach.

arXiv:submit/1673729 [cs.CV] 23 Sep 2016

# Structure

- title
- abstract
- introduction
  - teaser figure
- related work
- method
  - overview figure
- results/evaluation
  - discussion/limitations/future work
  - conclusion
  - references
  - appendices/supplemental document
    - additional images
    - implementation details
  - supplemental video
  - website/code

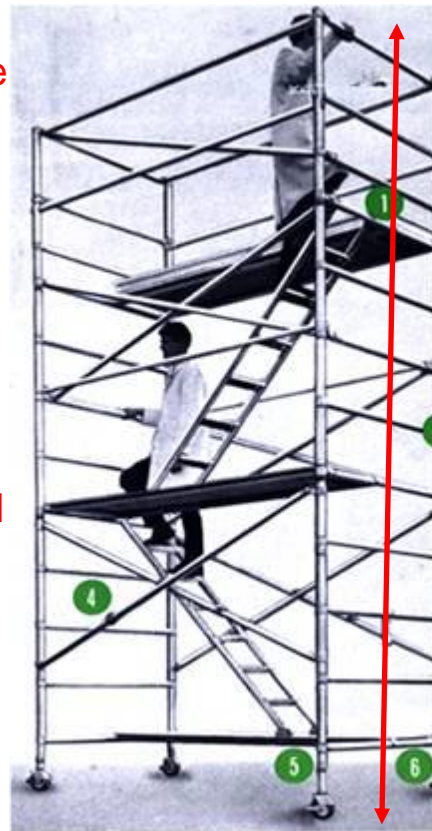


# Length

8 pages, double column format of CVPR:  
<http://cvpr2019.thecvf.com/files/cvpr2019AuthorKit.zip>

- title **1-2 lines**
- abstract **< ¼ of a column**
- introduction **1 page**
  - teaser figure
- related work **1 page**
- method **2 pages**
  - overview figure
- results/evaluation **3 pages**
  - **a** table (quantitative evaluation)
  - **a** graph (analyzing trends)
  - **many** pictures (qualitative results)
    - **many** good results
    - **a few** characteristic limitation cases
- discussion/limitations/future work **½ page**
- conclusion **< ¼ of a column**
- references **does not count for page limit**
- appendices/supplemental document
  - additional images
  - implementation details

**1-4 pages, not required**
- supplemental video **< 5 minutes, optional**



# How to start writing?

Writing is an iterative process

## 1. Decide on a structure

- sometimes the order of sections can be changed
  - e.g., related work after the method section
- sometimes sections can be merged
  - e.g., limitations and future works together

## 2. Add subsections

- different components in method section
- different experiments in evaluation

## 3. Outline story and content with bullet points

- motivation
- core method components
- major results





# How to structure the abstract & intro?

by Jean-luc Doumont

## % Sentence 1: CONTEXT - why now?

It is now possible to capture the 3D motion of the human body on consumer hardware and to puppet in real time skeleton-based virtual characters.

## % Sentence 2: NEED - why does the reader care?

However, many characters do not have humanoid skeletons. Characters such as spiders and caterpillars do not have boned skeletons at all, and these characters have very different shapes and motions. In general, character control under arbitrary shape and motion transformations is unsolved - how might these motions be mapped?

## % Sentence 3: TASK - what do we do?

## % Sentence 4: OBJECT - what does this document do?

We control characters with a method which avoids the rigging-skinning pipeline --- source and target characters do not have skeletons or rigs. We use interactively-defined sparse pose correspondences to learn a mapping between arbitrary 3D point source sequences and mesh target sequences. Then, we puppet the target character in real time.

## % Sentence 5: FINDINGS - what did we discover?

We demonstrate the versatility of our method through results on diverse virtual characters with different input motion controllers.

## % Sentence 6: CONCLUSIONS - so what?

## % Sentence 7: PERSPECTIVES - what now?

Our method provides a fast, flexible, and intuitive interface for arbitrary motion mapping which provides new ways to control characters for real-time animation.

Copy & paste this to your report!

# How to structure the related work section?

- Start with a sentence that explains the focus on a particular research field
- Divide your sections into about three paragraphs with different streams of related work
- cover the major related work but don't overload it with irrelevant papers
  - what are relevant papers?
    - most recent (still, don't ignore everything > 5 years)
    - published in major venues (if you know about similar work from smaller venues include it)
    - published by major groups (only an indicator)

## Personal advice:

- cite at least one paper from every group working in this area. These are likely to be your reviewers and you want to please them!
- praise how great the previous work is before you highlight one weakness that sets your new contribution apart. I had one case where I did not and the reviewer (seemingly the author of a prior work) got angry, up to the level of insults...

# How to structure the method section?

- Give an overview of your goals and methods (**not an index/overview of the paper structure!**)
- Introduce the input and output
- Introduce notation for the main quantities/variables
- Explain the variable dimensions (width, height, dimensionality, feature size)
  - does not need to be numbers, e.g., “in  $\mathbb{R}^N$  where N is the number of frames”
- Personal advice:
  - don't call your method a pipeline, it implies that the focus is on a complex system rather than a novel idea
  - don't put novel in every second sentence. If you do use 'novel', explain why it is novel in relation to what was there before
  - explain your method as simple as possible, don't overcomplicate
    - put equations only when necessary and minimize the number of variables

# Past or present tense?

Abstract, Intro, Methods

- mostly **present tense**
- past tense for earlier research results

Related work

- **present tense** for your opinion
- past tense for existing discoveries

Results

- **present tense** to refer to figures and tables
- past tense can be used to refer to completed experiments;

**I prefer present tense**

Discussion

- past tense possible, **I prefer present tense**

Conclusion and future work

- present perfect and future tense, **present tense possible**

## Active or passive voice?

- prefer active over passive for things you did
  - to highlight what your contributions are
- use we instead of I, even in your thesis (a matter of taste)

## Any other advice?

- copy from papers that you like
  - structure
  - introduction sentences
  - concluding sentences
  - common sentence structures
  - figure style
  - number and size of figures
- get feedback from others
  - experts
  - non-experts
  - proofreading software

# Common mistakes

- Make sure you refer to all figures in your text
  - “Figure 10 shows that ...”
- Make sure that you number all equations, even if you don’t refer to them yourself
  - such that future work can refer to it: “we extend Equation 4 from paper ...”
  - use `\begin{equation} ... \end{equation}` or `\begin{align} ... \end{align}`
    - **don’t use `$$ ... $$`**
    - use `$ ... $` within normal text flow, e.g., to define a variable
- Equations are part of the text, embed them with proper punctuation
  - if the sentence continues ,use a “,”
  - if the sentence ends, use a “.”
  - try to avoid “:”, rather link it with words “We write, `\begin{equation} ... \end{equation}`”

## Common mistakes (cont.)

- Figures and tables are not part of the text and should generally not be interleaved
  - align figures to the top with: `\begin{figure}[t] ...`
  - align tables to the bottom with: `\begin{figure}[b] ...`
  - exception: **small** inlets placed right to the reference in the text via text warping
- Capitalize when referring to a particular figure, table or equation:
  - Equation 1, Figure 2, Table 3 (Figure 2 is the title/name of that figure)
  - but not when referring to multiple equations 1 and 2 (here we refer to equations in general).
  - “Figure 10 shows that ...”
- Anything else?

# Have fun writing!

Additional resources:

[Write Good Papers](#) and [Notes on writing](#)

by Frédo Durand (more details on motivation)

[Modern myths – Shortcomings in scientific writing](#)

by Jean-luc Doumont (structured writing) **excellent!**

[How to write a good CVPR submission](#)

by Bill Freeman (more details on reviewing process)

[What's wrong with these equations?](#)

by David Mermin (details on equations)

