# Improving Monocular Depth Estimation with Global Depth Histogram Matching using a Single SPAD Transient

Anonymous ECCV submission

Paper ID 3668

**Abstract.** Existing monocular depth estimation algorithms successfully predict the relative depth order of objects in a scene. However, because of the fundamental scale ambiguity associated with monocular images, these algorithms fail at correctly predicting an object's true metric depth. In this work, we demonstrate how a depth histogram of the scene, which can be readily captured using a single-pixel diffused single-photon avalanche diode (SPAD), can be fused with the output of existing monocular depth estimation algorithms to resolve the depth ambiguity problem. We validate this novel sensor fusion technique experimentally and in extensive simulation. We show that it dramatically improves the performance of several state-of-the-art monocular depth estimation algorithms. . . .

**Keywords:** monocular depth estimation, single-photon avalanche diode, histogram matching, LiDAR

## 1 Introduction

Estimating dense 3D geometry from 2D images is an important open problem with applications to robotics, autonomous driving, and medical imaging. Depth maps are a common representation of scene geometry and are useful precursors to higher-level scene understanding tasks such as pose estimation and object detection. Additionally, many computer vision tasks rely on depth sensing, including navigation [?], semantic segmentation [?,?,?], 3D object detection [?,?,?,?,?], and 3D object classification [?,?,?].

Traditional depth sensing techniques include those based on stereo or multiview, active illumination, camera motion, or focus cues [?]. However, each of these techniques has aspects that may make their deployment challenging. For example, stereo or multiview techniques require multiple cameras, active illumination techniques may have limited resolution or require time-consuming scanning procedures, and other techniques require camera motion or multiple exposures at different focus distances.

One of the most promising approaches to overcoming these challenges is monocular depth estimation (MDE), which requires only a single RGB image from a conventional camera to recover a dense depth map [?,?,?,?,?]. Recent

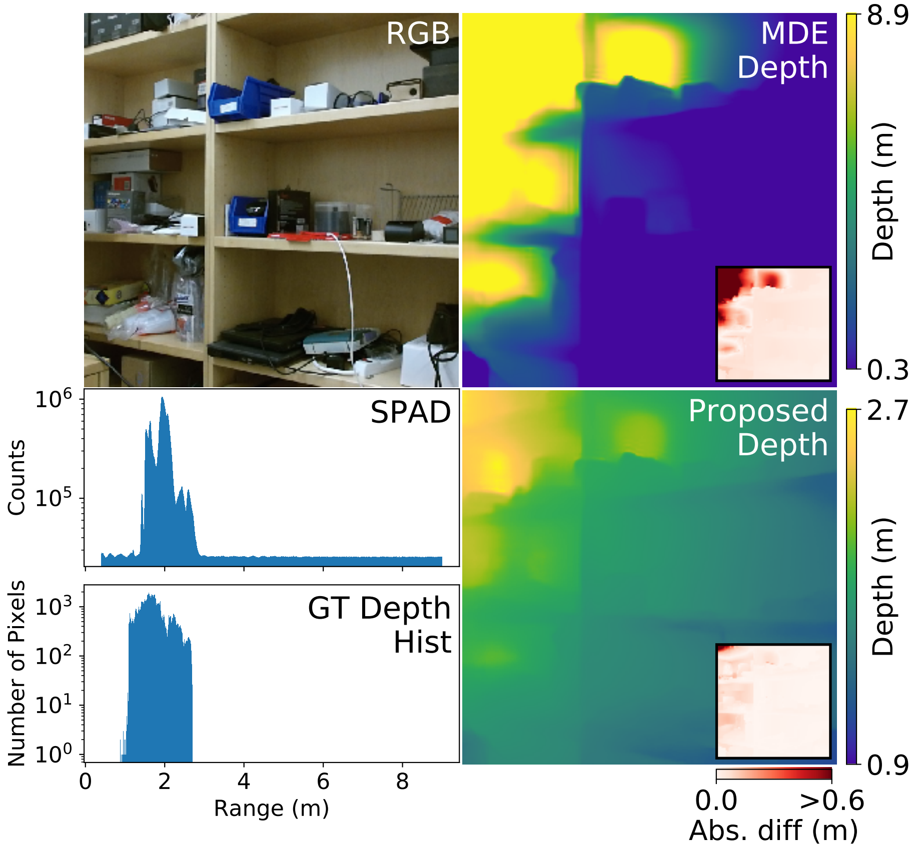


Fig. 1: Monocular depth estimation predicts a depth map (upper right) from a single RGB image (upper left). The ill-posedness of the problem prevents reliable absolute depth estimation, resulting in large errors (inset image). The proposed method uses a measurement from a diffused SPAD (lower left), which resembles the shape of a histogram of the ground truth depth map, to correct the output of the depth estimation and optimize the quality of the estimated absolute depth (lower right).

approaches to MDE employ neural network models that learn to predict depth by exploiting monocular image cues such as perspective, occlusion, shading, and relative object size. While such models have significantly improved over recent years, MDE approaches to date are incapable of reliably estimating absolute distances in a scene due to the inherent scale ambiguities of monocular image cues. Instead, these models excel in predicting ordinal depth, or the relative ordering of objects in a scene [**?**,**?**]. Interestingly, Alhashim and Wonka [**?**] recently showed that if the median ground truth depth of the scene is known, the initial output of a MDE network can be corrected to produce accurate absolute depth.

Although access to the median ground truth depth is impossible in a realistic scenario, low-cost sensors capable of capturing aggregated depth information from a scene are readily available. For example, the proximity sensor on recent generation Apple iPhones uses a low-power pulsed light source and a single-pixel, single-photon avalanche diode (SPAD) to sense distance to an object directly in front of the phone. SPADs have also been used to recover full 3D geometry, forming the backbone of emerging single-photon LiDAR systems [**?**,**?**,**?**]. However, SPADs have not been used for 3D imaging on consumer electronics because the requirement for ultra-fast timing electronics makes it difficult to produce high-resolution arrays at low cost, and because the scanning requirement for single-pixel systems introduces a point of mechanical failure and complicates high-resolution, high-framerate imaging.

Here, we propose to use a single-pixel SPAD and pulsed light source in an unconventional way: rather than optically focusing them to record the distance to a single scene point, we diffuse the emitted light and the detector over the entire scene. The resulting measurement resembles a histogram of the depth of the scene and we demonstrate that this can be used to achieve accurate absolute depth when combined with a monocular depth estimate (cf. Figure 1).

To this end, we develop a sensor fusion strategy that processes the ordinal depth computed by a monocular depth estimator to be consistent with the measurements captured by the diffused SPAD. We demonstrate in extensive simulations that our approach achieves substantial improvements in the quality of the estimated depth maps, regardless of which specific depth estimator is used. Moreover, we build an RGB-SPAD camera prototype and demonstrate in practice that our sensor fusion strategy achieves significantly higher-quality depth maps for a variety of captured scenes compared to MDE alone. With this work, we present a practical way to disambiguate depth estimation with RGB images using minimal additional sensing hardware. Specifically, we make the following contributions:

- We propose augmenting an RGB camera with a global depth histogram captured by a diffused SPAD to address scale ambiguity error in monocular depth estimators.
- We analyze this approach on indoor scenes using the NYU Depth v2 dataset and demonstrate that our approach is able to resolve scale ambiguity while being fast and easy to implement.

– We build a prototype RGB-SPAD camera and evaluate its efficacy on captured data, assessing both the quality and the ability of our method to help generalization of monocular depth estimators across scene types.

## 2 Related Work

*Monocular Depth Estimation* Estimating a depth map from a single RGB image has been approached using Markov Random Fields [?], geometric approaches [?], and non-parametric, SIFT-based methods [?]. More recently, deep neural networks have been applied to this problem. For example, Eigen et al. [?] use a multi-scale neural network to predict depth maps, Godard et al. [?] use an unsupervised approach that trains a network using stereo pairs, and Fu et al. [?] combine a logarithmic depth discretization scheme with an ordinal regression loss function. Various experiments using different types of encoder networks (ResNet, DenseNet) [?,?] have also been employed with some success, as have approaches mixing deep learning with conditional random fields [?], and attention-based approaches [?,?]. Recently, Lasinger et al. [?] improved the robustness of monocular depth estimation using cross-dataset transfer.

Despite achieving remarkable success on estimating ordinal depth from a single image, none of these methods are able to resolve inherent scale ambiguity in a principled manner. We introduce a new approach that leverages existing monocular depth estimation networks and disambiguates the output using depth histogram-like measurements obtained from a single, diffused SPAD. Other approaches to disambiguating monocular depth estimation use optimized freeform lenses [?,?] or dual-pixel sensors [?], but these approaches require custom lenses or sensors and specialized image reconstruction methods. In contrast, our approach is potentially compatible with existing camera systems (deployed on current cell phones) and our algorithms could be used in tandem with existing camera ISPs.

*Depth Imaging and Sensor Fusion with SPADs* Emerging single-photon LiDAR systems use single-photon avalanche diodes (SPADs) to record the time of flight of individual photons. SPAD detectors can be fabricated using standard CMOS processes, but the required picosecond-accurate time-stamping electronics are challenging to miniaturize and fabricate at low cost. For this reason, many single-photon 3D imaging approaches use a single SPAD combined with a raster scanning mechanism [?,?,?,?,?]. Unfortunately, this makes it challenging to scan dynamic scenes at high resolution and scanners can also be expensive, difficult to calibrate, and prone to mechanical failure. To reduce the scanning complexity to one dimension, 1D SPAD arrays have been developed [?,?,?], and 2D SPAD arrays are also an active area of research [?,?,?,?]. Yet, single-pixel SPADs remain the only viable option for low-cost consumer devices today.

The proposed method uses a single-pixel SPAD and pulsed light source that are diffused across the entire scene instead of aimed at a single point, as with proximity sensors. This unique configuration captures a measurement that closely resembles the depth histogram of the scene. Our sensor fusion algorithm achieves reliable absolute depth estimation by combining the SPAD measurement with the output of a monocular depth estimator using a histogram matching technique. While other recent work also explored RGB-SPAD sensor fusion [?], the RGB image was primarily used to guide the denoising and upsampling of measurements from a SPAD array.

*Histogram Matching and Global Hints* Histogram matching is a well-known image processing technique for adjusting an image so that its histogram matches some pre-specified histogram (often derived from another image) [?,?]. Nikolova et al. [?] use optimization to recover a strict ordering of the image pixels, yielding an exact histogram match. Morovic et al. [?] provide an efficient and precise method for fast histogram matching which supports weighted pixel values. In the image reconstruction space, Swoboda and Schnörr [?] use a histogram to form an image prior based on the Wasserstein distance for image denoising and inpainting. Rother et al. [?] use a histogram prior to create an energy function that penalizes foreground segmentations with dissimilar histograms. In a slightly different application area, Zhang et al. [?] train a neural network to produce realistically colorized images given only a black-and-white image and a histogram of global color information.

In our procedure, the diffused SPAD measurements closely resemble a histogram of the depth map where the histogram values are weighted by spatially varying scene reflectances and inverse-square falloff effects. We therefore adapt the algorithm in Morovic et al. [?] in order to accommodate general per-pixel weights during histogram matching.

# 3 Initial Submission

## 3.1 Language

All manuscripts must be in English.

## 3.2 Paper length

Papers submitted for review should be complete. The length should match that intended for final publication. Papers accepted for the conference will be allocated 14 pages (plus references) in the proceedings. Note that the allocated 14 pages do not include the references. The reason for this policy is that we do not want authors to omit references for sake of space limitations.

Papers with more than 14 pages (excluding references) will be rejected without review. This includes papers where the margins and formatting are deemed to have been significantly altered from those laid down by this style guide. The reason such papers will not be reviewed is that there is no provision for supervised revisions of manuscripts. The reviewing process cannot determine the suitability of the paper for presentation in 14 pages if it is reviewed in 16.

### 3.3 Paper ID

It is imperative that the paper ID is mentioned on each page of the manuscript. The paper ID is a number automatically assigned to your submission when registering your paper submission on the submission site.

All lines should be numbered in the initial submission, as in this example document. This makes reviewing more efficient, because reviewers can refer to a line on a page. Line numbering is removed in the camera-ready.

### 3.4 Mathematics

Please number all of your sections and displayed equations. Again, this makes reviewing more efficient, because reviewers can refer to a line on a page. Also, it is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the line numbering will not be present in the final copy, so is not an alternative to equation numbers). Some authors might benefit from reading Mermin's description of how to write mathematics: www.pamitc.org/documents/mermin.pdf.

## 4 Policies

### 4.1 Review Process

By submitting a paper to ECCV, the authors agree to the review process and understand that papers are processed by the Toronto system to match each manuscript to the best possible chairs and reviewers.

### 4.2 Confidentiality

The review process of ECCV is confidential. Reviewers are volunteers not part of the ECCV organisation and their efforts are greatly appreciated. The standard practice of keeping all information confidential during the review is part of the standard communication to all reviewers. Misuse of confidential information is a severe professional failure and appropriate measures will be taken when brought to the attention of ECCV organizers. It should be noted, however, that the organisation of ECCV is not and cannot be held responsible for the consequences when reviewers break confidentiality.

Accepted papers will be published by Springer (with appropriate copyrights) electronically up to three weeks prior to the main conference. Please make sure to discuss this issue with your legal advisors as it pertains to public disclosure of the contents of the papers submitted.

### 4.3 Dual and Double Submissions

By submitting a manuscript to ECCV 2020, authors acknowledge that it has not been previously published or accepted for publication in substantially similar form in any peer-reviewed venue including journal, conference, or workshop. Furthermore, no paper substantially similar in content has been or will be submitted to a journal, another conference or workshop during the review period (March 05, 2020 – July 3, 2020). The authors also attest that they did not submit substantially similar submissions to ECCV 2020. Violation of any of these conditions will lead to rejection and the violation will be reported to the other venue or journal, which will typically lead to rejection there as well.

The goals of the dual submission policy are (i) to have exciting new work be published for the first time at ECCV 2020, and (ii) to avoid duplicating the efforts of the reviewers. Therefore, all papers under review are checked for dual submissions and this is not allowed, independent of the page size of submissions.

For already published papers, our policy is based upon the following particular definition of "publication". A publication, for the purposes of the dual submission policy, is defined to be a written work longer than four pages that was submitted for review by peers for either acceptance or rejection, and, after review, was accepted. In particular, this definition of publication does not depend upon whether such an accepted written work appears in a formal proceedings or whether the organizers declare that such work "counts as a publication".

An arXiv.org paper does not count as a publication because it was not peer-reviewed for acceptance. The same is true for university technical reports. However, this definition of publication does include peer-reviewed workshop papers, even if they do not appear in a proceedings, if their length is more than 4 pages including citations. Given this definition, any submission to ECCV 2020 should not have substantial overlap with prior publications or other concurrent submissions. As a rule of thumb, the ECCV 2020 submission should contain no more than 20 percent of material from previous publications.

### 4.4 Requirements for publication

Publication of the paper in the ECCV 2020 proceedings of Springer requires that at least one of the authors registers for the conference and present the paper there. It also requires that a camera-ready version that satisfies all formatting requirements is submitted before the camera-ready deadline.

### 4.5 Double blind review

ECCV reviewing is double blind, in that authors do not know the names of the area chair/reviewers of their papers, and the area chairs/reviewers cannot, beyond reasonable doubt, infer the names of the authors from the submission and the additional material. Avoid providing links to websites that identify the authors. Violation of any of these guidelines may lead to rejection without review.

If you need to cite a different paper of yours that is being submitted concurrently to ECCV, the authors should (1) cite these papers, (2) argue in the body of your paper why your ECCV paper is non trivially different from these concurrent submissions, and (3) include anonymized versions of those papers in the supplemental material.

Many authors misunderstand the concept of anonymizing for blind review. Blind review does not mean that one must remove citations to one's own work. In fact it is often impossible to review a paper unless the previous citations are known and available.

Blind review means that you do not use the words "my" or "our" when citing previous work. That is all. (But see below for technical reports).

Saying "this builds on the work of Lucy Smith [1]" does not say that you are Lucy Smith, it says that you are building on her work. If you are Smith and Jones, do not say "as we show in [7]", say "as Smith and Jones show in [7]" and at the end of the paper, include reference 7 as you would any other cited work.

An example of a bad paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of our previous paper [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Removed for blind review

An example of an excellent paper:

> An analysis of the frobnicatable foo filter.
>
> In this paper we present a performance analysis of the paper of Smith [1], and show it to be inferior to all previously known methods. Why the previous paper was accepted without this analysis is beyond me.
>
> [1] Smith, L. and Jones, C. "The frobnicatable foo filter, a fundamental contribution to human knowledge". Nature 381(12), 1-213.

If you are making a submission to another conference at the same time, which covers similar or overlapping material, you may need to refer to that submission in order to explain the differences, just as you would if you had previously published related work. In such cases, include the anonymized parallel submission [**?**] as additional material and cite it as

> 1. Authors. "The frobnicatable foo filter", BMVC 2014 Submission ID 324, Supplied as additional material `bmvc14.pdf`.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [**?**]". Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the ECCV audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus.

You can handle this paper like any other. Don't write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

For sake of anonymity, it's recommended to omit acknowledgements in your review copy. They can be added later when you prepare the final copy.

## 5 Manuscript Preparation

This is an edited version of Springer LNCS instructions adapted for ECCV 2020 first paper submission. You are strongly encouraged to use $\LaTeX2_\varepsilon$ for the preparation of your camera-ready manuscript together with the corresponding Springer class file `llncs.cls`.

We would like to stress that the class/style files and the template should not be manipulated and that the guidelines regarding font sizes and format should be adhered to. This is to ensure that the end product is as homogeneous as possible.

### 5.1 Printing Area

The printing area is 122 mm $\times$ 193 mm. The text should be justified to occupy the full line width, so that the right margin is not ragged, with words hyphenated as appropriate. Please fill pages so that the length of the text is no less than 180 mm.

### 5.2 Layout, Typeface, Font Sizes, and Numbering

Use 10-point type for the name(s) of the author(s) and 9-point type for the address(es) and the abstract. For the main text, please use 10-point type and single-line spacing. We recommend using Computer Modern Roman (CM) fonts, Times, or one of the similar typefaces widely used in photo-typesetting. (In these typefaces the letters have serifs, i.e., short endstrokes at the head and the foot of letters.) Italic type may be used to emphasize words in running text. Bold type and underlining should be avoided. With these sizes, the interline distance should be set so that some 45 lines occur on a full-text page.

**Headings.** Headings should be capitalized (i.e., nouns, verbs, and all other words except articles, prepositions, and conjunctions should be set with an initial capital) and should, with the exception of the title, be aligned to the left. Words joined by a hyphen are subject to a special rule. If the first word can stand alone, the second word should be capitalized. The font sizes are given in Table 1.

Table 1: Font sizes of headings. Table captions should always be positioned *above* the tables. The final sentence of a table caption should end without a full stop

| Heading level | Example | Font size and style |
|---|---|---|
| Title (centered) | **Lecture Notes . . .** | 14 point, bold |
| 1st-level heading | **1 Introduction** | 12 point, bold |
| 2nd-level heading | **2.1 Printing Area** | 10 point, bold |
| 3rd-level heading | **Headings.** Text follows . . . | 10 point, bold |
| 4th-level heading | *Remark.* Text follows . . . | 10 point, italic |

Here are some examples of headings: "Criteria to Disprove Context-Freeness of Collage Languages", "On Correcting the Intrusion of Tracing Non-deterministic Programs by Software", "A User-Friendly and Extendable Data Distribution System", "Multi-flip Networks: Parallelizing GenSAT", "Self-determinations of Man".

**Lemmas, Propositions, and Theorems.** The numbers accorded to lemmas, propositions, and theorems etc. should appear in consecutive order, starting with the number 1, and not, for example, with the number 11.

### 5.3 Figures and Photographs

Please produce your figures electronically and integrate them into your text file. For LaTeX users we recommend using package `graphicx` or the style files `psfig` or `epsf`.

Check that in line drawings, lines are not interrupted and have constant width. Grids and details within the figures must be clearly readable and may not be written one on top of the other. Line drawings should have a resolution of at least 800 dpi (preferably 1200 dpi). For digital halftones 300 dpi is usually sufficient. The lettering in figures should have a height of 2 mm (10-point type). Figures should be scaled up or down accordingly. Please do not use any absolute coordinates in figures.

Figures should be numbered and should have a caption which should always be positioned *under* the figures, in contrast to the caption belonging to a table, which should always appear *above* the table. Please center the captions between the margins and set them in 9-point type (Fig. 2 shows an example). The distance between text and figure should be about 8 mm, the distance between figure and caption about 5 mm.

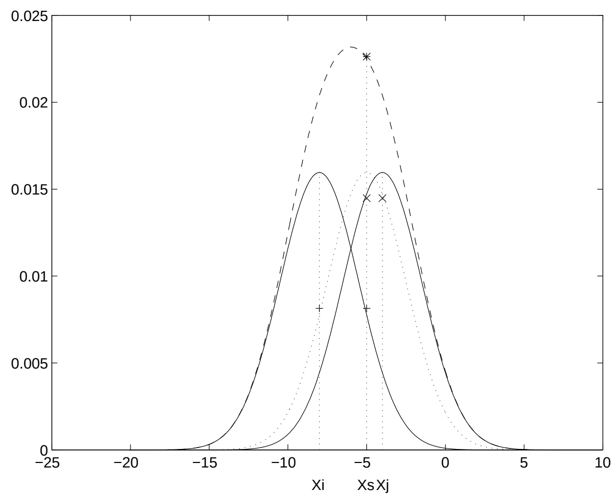


Fig. 2: One kernel at $x_s$ (*dotted kernel*) or two kernels at $x_i$ and $x_j$ (*left and right*) lead to the same summed estimate at $x_s$. This shows a figure consisting of different types of lines. Elements of the figure described in the caption should be set in italics, in parentheses, as shown in this sample caption. The last sentence of a figure caption should generally end without a full stop

If possible (e.g. if you use LaTeX) please define figures as floating objects. LaTeX users, please avoid using the location parameter "h" for "here". If you have to insert a pagebreak before a figure, please ensure that the previous page is completely filled.

### 5.4 Formulas

Displayed equations or formulas are centered and set on a separate line (with an extra line or halfline space above and below). Displayed expressions should be numbered for reference. The numbers should be consecutive within the contribution, with numbers enclosed in parentheses and set on the right margin. For example,

$$\psi(u) = \int_0^T \left[ \frac{1}{2} \left( \Lambda_0^{-1} u, u \right) + N^*(-u) \right] dt \tag{1}$$
$$= 0? \tag{2}$$

Please punctuate a displayed equation in the same way as ordinary text but with a small space before the end punctuation.

### 5.5 Footnotes

The superscript numeral used to refer to a footnote appears in the text either directly after the word to be discussed or, in relation to a phrase or a sentence, following the punctuation sign (comma, semicolon, or full stop). Footnotes should appear at the bottom of the normal text area, with a line of about 2 cm in TEX and about 5 cm in Word set immediately above them.[1]

### 5.6 Program Code

Program listings or program commands in the text are normally set in typewriter font, e.g., CMTT10 or Courier.
*Example of a Computer Program*

```
program Inflation (Output)
  {Assuming annual inflation rates of 7%, 8%, and 10%,...
   years};
   const
     MaxYears = 10;
   var
     Year: 0..MaxYears;
     Factor1, Factor2, Factor3: Real;
   begin
     Year := 0;
     Factor1 := 1.0; Factor2 := 1.0; Factor3 := 1.0;
     WriteLn('Year  7% 8% 10%'); WriteLn;
     repeat
       Year := Year + 1;
       Factor1 := Factor1 * 1.07;
       Factor2 := Factor2 * 1.08;
       Factor3 := Factor3 * 1.10;
```

[1] The footnote numeral is set flush left and the text follows with the usual word spacing. Second and subsequent lines are indented. Footnotes should end with a full stop.

```
    WriteLn(Year:5,Factor1:7:3,Factor2:7:3,Factor3:7:3)
  until Year = MaxYears
end.
```

(Example from Jensen K., Wirth N. (1991) Pascal user manual and report. Springer, New York)

### 5.7 Citations

The list of references is headed “References” and is not assigned a number in the decimal system of headings. The list should be set in small print and placed at the end of your contribution, in front of the appendix, if one exists. Please do not insert a pagebreak before the list of references if the page is not completely filled. An example is given at the end of this information sheet. For citations in the text please use square brackets and consecutive numbers: [**?**], [**?**], [**?**] ...

## 6 Submitting a Camera-Ready for an Accepted Paper

### 6.1 Converting Initial Submission to Camera-Ready

To convert a submission file into a camera-ready for an accepted paper:

1. First comment out

   ```
   \usepackage{ruler}
   ```

   and the line that follows it.
2. The anonymous title part should be removed or commented out, and a proper author block should be inserted, for which a skeleton is provided in a commented-out version. These are marked in the source file as

   ```
   % INITIAL SUBMISSION
   ```

   and

   ```
   % CAMERA READY SUBMISSION
   ```

3. Please write out author names in full in the paper, i.e. full given and family names. If any authors have names that can be parsed into FirstName LastName in multiple ways, please include the correct parsing in a comment to the editors, below the

   ```
   \author{}
   ```

   field.
4. Make sure you have inserted the proper Acknowledgments.

### 6.2 Preparing the Submission Package

We need all the source files (LaTeX files, style files, special fonts, figures, bib-files) that are required to compile papers, as well as the camera ready PDF. For each paper, one ZIP-file called XXXX.ZIP (where XXXX is the zero-padded, four-digit paper ID) has to be prepared and submitted via the ECCV 2020 Submission Website, using the password you received with your initial registration on that site. The size of the ZIP-file may not exceed the limit of 60 MByte. The ZIP-file has to contain the following:

1. All source files, e.g. LaTeX2e files for the text, PS/EPS or PDF/JPG files for all figures.
2. PDF file named "XXXX.pdf" that has been produced by the submitted source, where XXXX is the four-digit paper ID (zero-padded if necessary). For example, if your paper ID is 24, the filename must be 0024.pdf. This PDF will be used as a reference and has to exactly match the output of the compilation.
3. PDF file named "XXXX-copyright.PDF": a scanned version of the signed copyright form (see ECCV 2020 Website, Camera Ready Guidelines for the correct form to use).
4. If you wish to provide supplementary material, the file name must be in the form XXXX-supp.pdf or XXXX-supp.zip, where XXXX is the zero-padded, four-digit paper ID as used in the previous step. Upload your supplemental file on the "File Upload" page as a single PDF or ZIP file of 100 MB in size or less. Only PDF and ZIP files are allowed for supplementary material. You can put anything in this file – movies, code, additional results, accompanying technical reports–anything that may make your paper more useful to readers. If your supplementary material includes video or image data, you are advised to use common codecs and file formats. This will make the material viewable by the largest number of readers (a desirable outcome). ECCV encourages authors to submit videos using an MP4 codec such as DivX contained in an AVI. Also, please submit a README text file with each video specifying the exact codec used and a URL where the codec can be downloaded. Authors should refer to the contents of the supplementary material appropriately in the paper.

Check that the upload of your file (or files) was successful either by matching the file length to that on your computer, or by using the download options that will appear after you have uploaded. Please ensure that you upload the correct camera-ready PDF–renamed to XXXX.pdf as described in the previous step as your camera-ready submission. Every year there is at least one author who accidentally submits the wrong PDF as their camera-ready submission.

Further considerations for preparing the camera-ready package:

1. Make sure to include any further style files and fonts you may have used.
2. References are to be supplied as BBL files to avoid omission of data while conversion from BIB to BBL.

3. Please do not send any older versions of papers. There should be one set of source files and one XXXX.pdf file per paper. Our typesetters require the author-created pdfs in order to check the proper representation of symbols, figures, etc.
4. Please remove unnecessary files (such as eijkel2.pdf and eijkel2.eps) from the source folder.
5. You may use sub-directories.
6. Make sure to use relative paths for referencing files.
7. Make sure the source you submit compiles.

Springer is the first publisher to implement the ORCID identifier for proceedings, ultimately providing authors with a digital identifier that distinguishes them from every other researcher. ORCID (Open Researcher and Contributor ID) hosts a registry of unique researcher identifiers and a transparent method of linking research activities to these identifiers. This is achieved through embedding ORCID identifiers in key workflows, such as research profile maintenance, manuscript submissions, grant applications and patent applications.

### 6.3 Most Frequently Encountered Issues

Please kindly use the checklist below to deal with some of the most frequently encountered issues in ECCV submissions.

**FILES:**

- My submission package contains ONE compiled pdf file for the camera-ready version to go on Springerlink.
- I have ensured that the submission package has all the additional files necessary for compiling the pdf on a standard LaTeX distribution.
- I have used the correct copyright form (with editor names pre-printed), and a signed pdf is included in the zip file with the correct file name.

**CONTENT:**

- I have removed all `\vspace` and `\hspace` commands from my paper.
- I have not used `\thanks` or `\footnote` commands and symbols for corresponding authors in the title (which is processed with scripts) and (optionally) used an Acknowledgement section for all the acknowledgments, at the end of the paper.
- I have not used `\cite` command in the abstract.
- I have read the Springer author guidelines, and complied with them, including the point on providing full information on editors and publishers for each reference in the paper (Author Guidelines – Section 2.8).
- I have entered a correct `\titlerunning{}` command and selected a meaningful short name for the paper.
- I have entered `\index{Lastname,Firstname}` commands for names that are longer than two words.
- I have used the same name spelling in all my papers accepted to ECCV and ECCV Workshops.

- I have inserted the ORCID identifiers of the authors in the paper header (see http://bit.ly/2H5xBpN for more information).
- I have not decreased the font size of any part of the paper (except tables) to fit into 14 pages, I understand Springer editors will remove such commands.

**SUBMISSION:**

- All author names, titles, and contact author information are correctly entered in the submission site.
- The corresponding author e-mail is given.
- At least one author has registered by the camera ready deadline.

## 7 Conclusions

The paper ends with a conclusion.

Page 17 of the manuscript.

Page 18 of the manuscript.
This is the last page of the manuscript.

Now we have reached the maximum size of the ECCV 2020 submission (excluding references). References should start immediately after the main text, but can continue on p.15 if needed.