Computer Vision SS22 Assignment 1: Document Scanner

Felix Hamburger
Student ID: 35925
Computer Vision SS22
Computer Sience Master
Ravensburg Weingarten University
Email: felix.hamburger@rwu.de

Mario Amann
Student ID: 35926
Computer Vision SS22
Computer Sience Master
Ravensburg Weingarten University
Email: mario.amann@rwu.de

This paper presents a usecase for a DIN A4 document scanner. The explanations contained in this paper will be divided into three steps. First the corner points of the document will be detected utilizing the canny edge corner detection. By performing a perspective transofmation the document will then be provided in a top-view. With the help of a filter the document will then be changed into a binary format.

1 INTRODUCTION

In everyday life, there are often situations in which it is advantageous to scan a document that has been received and keep it as a copy. Unfortunately not everyone has a ready-to-use scanner with them to scan the document. Though most of the people have a device capable of doing just this - a cell phone with a camera. Using various methods, it is possible to scan documents with the help of the cell phone camera, similar to a scanner, and save them in a suitable foramt. This paper will give your step by step guide on how it is possible to implement these methodologies.

2 PREPROCESSING

The first step is to read the image. This is done with the imread function of openCV which outputs the image as numpy array with the shape (Height, Width, Channel). By default the decoded image is stored in the BGR-format. Which means, the first channel contains the blue color channel, the second one contains the green color channel and the last the red one. The original image is seperately stored, because for further steps the photo is needed in a grayscale format. This convertion is made with the cvtColor[1] function of openCV. The convertion of colorspace is made with the following formula[2]:

$$Y = B * 0.114 + G * 0.587 + R * 0.299$$

The next preprocessing step includes the smoothing/blurring of the grayscale image. This is done to remove noise from the image[3]. In our case the Gaussian Blurring[4] is used. The image is convolved with the Gaussian Kernel. As size of the kernel the size (5,5) is used. Because a document could be landscape or portrait format, it makes sense to use same values in x and y direction in the kernel size, as well as for the parameters sigmaX and sigmaY. These parameters are used to calculate the gaussian filter coefficients. In the function call sigmaX and sigmaY are set to zero, which means the Gaussian Kernel computes these variables from the given ksize[5].

$$\sigma = 0.3 * (0.5 * (ksize - 1) - 1) + 0.8$$

The gaussian filter coefficients are computated with the following formula[5]:

$$G_i = \alpha * e^{-\frac{(i-\frac{(ksize-1)}{2})^2}{2*\sigma^2}}$$

3 EDGE DETECTION

After we took our preprocessing step and chose a suitable noise reduction algorithm we start with the edge detection. For edge detection we use the Canny edge[6] detector which is a multi-staged algorithm:

1. Noise reduction.
2. Calculating the intensity gradient of the image.
3. Non-maximum supression.

4. Hysteresis thresholding.

The noise reduction step was allready described in the Preprocessing section of this paper.

The intensity gradient of the image is calculated using the **Sobel Filter**:

$$S(I(x,y)) := \sqrt{(S_x * I(x,y))^2 + S_y(*I(x,y))^2} \quad (1)$$

Generally we can define the gradient of the Image I as:

$$G(I(x,y)) := \sqrt{I_x^2 + I_y^2} \quad (2)$$

In addition to the gradient we calculate the the orientation given by:

$$\phi(x,y) = arctan(\frac{g_y}{g_x}) \quad (3)$$

After getting the gradient magnitude and orientation of each pixel, the edges have to be reduced to a thickness of one pixel which will be done with **non-maximum supression.**

To implement this methodologie in the document scanner, the function which implements the full Canny edge detector can be called.

$$Canny(blurred\_image, 75, 180,$$
$$L2gradient = True, apertureSize = 3)$$

The functions input is the noise reduced image described in Preprocessing. For Hysteresis Threshold 75 for the lower-bound and 180 for the upper bound are chosen. For several trials this settings could sufficiently generalize various documents. An apeerture size of 3 does also generate solid outputs. Instead of the predefined L1-Norm [7], the more precise L2-Norm[8] is used.

4 CORNER DETCTION
5 TRANSFORMATION
6 BINARY IMAGE
7 SUMMARY
8 TRANSFORMATION

Footnotes are referenced with superscript numerals and are numbered consecutively from 1 to the end of the paper[1]. Footnotes should appear at the bottom of the column in which they are referenced.

Beautiful Figure

Fig. 1. The caption of a single sentence does not have period at the end

9 MATHEMATICS

Equations should be numbered consecutively beginning with (1) to the end of the paper, including any appendices. The number should be enclosed in parentheses and set flush right in the column on the same line as the equation. An extra line of space should be left above and below a displayed equation or formula. LaTeX can automatically keep track of equation numbers in the paper and format almost any equation imaginable. An example is shown in Eqn. (4). The number of a referenced equation in the text should be preceded by Eqn. unless the reference starts a sentence in which case Eqn. should be expanded to Equation.

$$f(t) = \int_{0_+}^{t} F(t)dt + \frac{dg(t)}{dt} \quad (4)$$

10 FIGURES

All figures should be positioned at the top of the page where possible. All figures should be numbered consecutively and centered under the figure as shown in Fig. 1. All text within the figure should be no smaller than 7 pt. There should be a minimum two line spaces between figures and text. The number of a referenced figure or table in the text should be preceded by Fig. or Tab. respectively unless the reference starts a sentence in which case Fig. or Tab. should be expanded to Figure or Table.

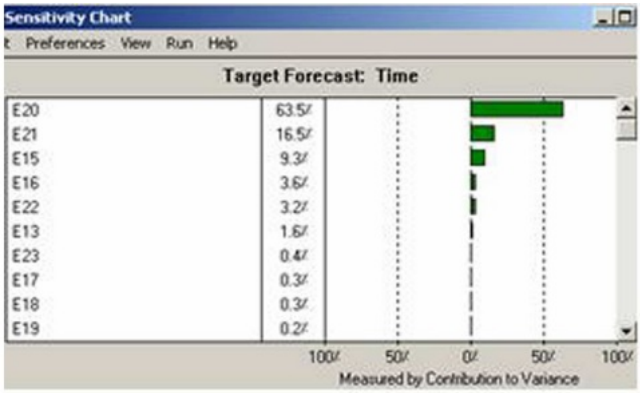
In the following subsections, I have inserted figures that have been provided by authors in order to demonstrate what to avoid. In each case the authors provided figures that are 3.25in wide and 600dpi in the .tif graphics format. The papers containing these figures have been held from production due to their poor quality.

10.1 The 1st Example of Bad Figure

In order to place the figure in this template using MSWord, select Insert Picture from File, and use wrap-

[0]Examine the input file, asme2ej.tex, to see how a footnote is given in a head.

[1]Avoid footnotes if at all possible.



(a) Time

(b) Cost

Fig. 2. Example taken from a paper that was held from production because the image quality is poor. ASME sets figures captions in 8pt, Helvetica Bold.

ping that is top and bottom. Make sure the figure is 3.25in wide.

Figure ‘2 was taken from a recent paper that was held from publication, because the text is fuzzy and unreadable. It was probably obtained by taking a screen shot of the computer output of the authors software. This means the original figure was 72dpi (dots per inch) on a computer screen. There is no way to improve the quality such a low resolution figure.

In order to understand how poor the quality of this figure is, please zoom in slightly, say to 200%. Notice that while the font of the paper is clear at this size, the font in the figures is fuzzy and blurred. It is impossible to make out the small symbol beside the numbers along the abscissa of the graph. Now consider the labels Time and Cost. They are clearly in fonts larger that the text of the

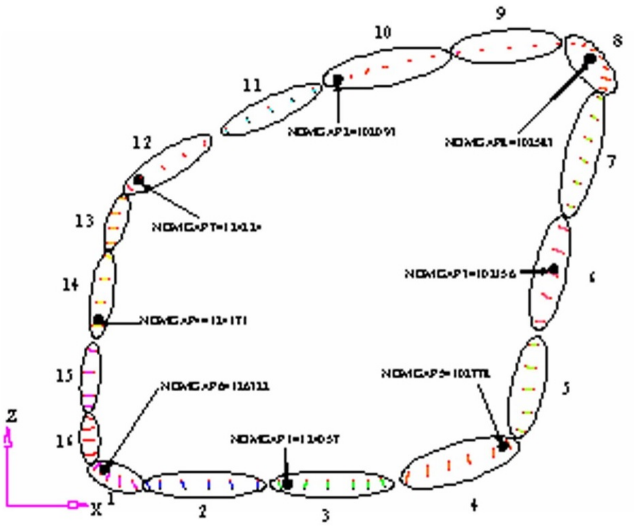


Fig. 3. While this figures is easily readable at a double column width of 6.5in, when it is shrunk to 3.25in column width the text is unreadable. This paper was held from production.

article, yet the pixilation or rasterization, associated with low resolution is obvious. This figure must be regenerated at higher resolution to ensure quality presentation.

The poor quality of this figure is immediately obvious on the printed page, and reduces the impact of the research contribution of the paper, and in fact detracts from the perceived quality of the journal itself.

10.2 The 2nd Example of Bad Figure

Figure 3 demonstrates a common problem that arises when a figure is scaled down fit a single column width of 3.25in. The original figure had labels that were readable at full size, but become unreadable when scaled to half size. This figure also suffers from poor resolution as is seen in the jagged lines the ovals that form the chain.

This problem can be addressed by increasing the size of the figure to a double column width of 6.5in, so the text is readable. But this will not improve the line pixilation, and a large low resolution figure is less desirable than a small one. This also significantly expands the length of the paper, and may cause it to exceed the JMD nine page limit. Additional pages require page charges of $200 per page. It is best to regenerate the figure at the resolution that ensures a quality presentation.

10.3 The 3rd Example of Bad Figure

An author provided the high resolution image in Fig. 4 that was sized to a single column width of 3.25in. Upon seeing the poor quality of the text, the publisher

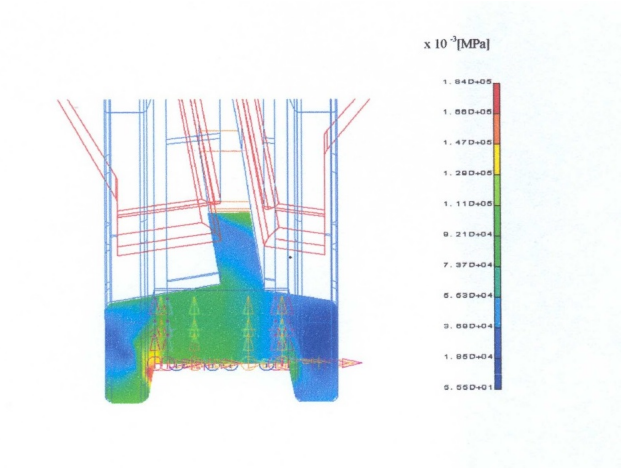


Fig. 4. Another example of a figure with unreadable text. Even when the paper was expanded to double column width the text as shown in Fig. 5 was of such low quality that the paper was held from production.

Table 1. Figure and table captions do not end with a period

Example	Time	Cost
1	12.5	$1,000
2	24	$2,000

scaled the image to double column width as shown in Fig. 5 at which point it took half of a page. The publisher went on to do this for all eight figures generating four pages of figures that the author did not expect. ASME stopped production of the paper even with the larger figures due to the pixilation of the font.

Clearly the text in this figure is unreadable, and it is doubtful that the author can print the output in a way that it is readable. This is a problem that the author must solve, not the publisher.

As you might expect, I have many more examples, but in the end the author is the best judge of what is needed in each figure. ASME simply requires that the image meet a minimum standard for font and line quality, specifically the font should be the appropriate size and not be blurred or pixilated, and that lines should be the appropriate weight and have minimal, preferably no, pixilation or rasterization.

11 TABLES

All tables should be numbered consecutively and centered above the table as shown in Table 1. The body of the table should be no smaller than 7 pt. There should be a minimum two line spaces between tables and text.

12 CITING REFERENCES

The ASME reference format is defined in the authors kit provided by the ASME. The format is:

> *Text Citation*. Within the text, references should be cited in numerical order according to their order of appearance. The numbered reference citation should be enclosed in brackets.

The references must appear in the paper in the order that they were cited. In addition, multiple citations (3 or more in the same brackets) must appear as a " [1-3]". A complete definition of the ASME reference format can be found in the ASME manual [9].

The bibliography style required by the ASME is unsorted with entries appearing in the order in which the citations appear. If that were the only specification, the standard BibTeX unsrt bibliography style could be used. Unfortunately, the bibliography style required by the ASME has additional requirements (last name followed by first name, periodical volume in boldface, periodical number inside parentheses, etc.) that are not part of the unsrt style. Therefore, to get ASME bibliography formatting, you must use the `asmems4.bst` bibliography style file with BibTeX. This file is not part of the standard BibTeX distribution so you'll need to place the file someplace where LaTeX can find it (one possibility is in the same location as the file being typeset).

With LaTeX/BibTeX, LaTeX uses the citation format set by the class file and writes the citation information into the .aux file associated with the LaTeX source. BibTeX reads the .aux file and matches the citations to the entries in the bibliographic data base file specified in the LaTeX source file by the `\bibliography` command. BibTeX then writes the bibliography in accordance with the rules in the bibliography .bst style file to a .bbl file which LaTeX merges with the source text. A good description of the use of BibTeX can be found in [10, 11] (see how two references are handled?). The following is an example of how three or more references [10, 9, 11] show up using the `asmems4.bst` bibliography style file in conjunction with the `asme2ej.cls` class file. Here are some more [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22] which can be used to describe almost any sort of reference.

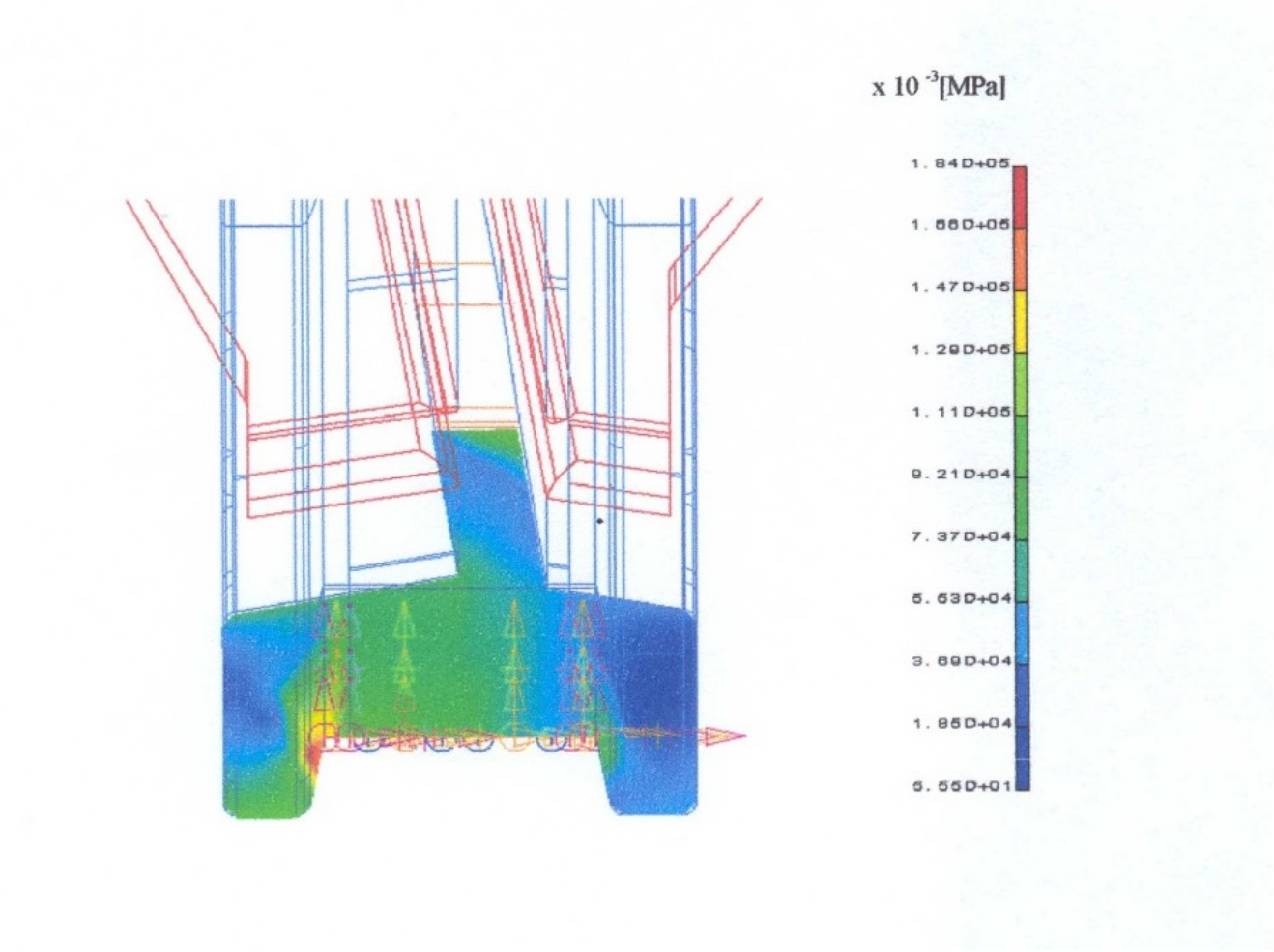


Fig. 5. A figure expanded to double column width the text from Figure 4

13 CONCLUSIONS

The only way to ensure that your figures are presented in the ASME Journal of Mechanical Design in the way you feel is appropriate and meets the requirement for quality presentation is for you to prepare a double column version of the paper in a form similar to that used by the Journal.

This gives you the opportunity to ensure that the figures are sized appropriately, in particular that the labels are readable and match the size of the text in the journal, and that the line weights and resolutions have no pixilation or rasterization. Poor quality figures are immediately obvious on the printed page, and this detracts from the perceived quality of the journal.

I am pleased to provide advice on how to improve any figure, but this effort must start with a two-column version of the manuscript. Thank you in advance for your patience with this effort, it will ensure quality presentation of your research contributions.

14 DISCUSSIONS

This template is not yet ASME journal paper format compliant at this point. More specifically, the following features are not ASME format compliant.

1. The format for the title, author, and abstract in the cover page.
2. The font for title should be 24 pt Helvetica bold.

If you can help to fix these problems, please send us an updated template. If you know there is any other non-compliant item, please let us know. We will add it to the above list. With your help, we shall make this template compliant to the ASME journal paper format.

ACKNOWLEDGEMENTS

ASME Technical Publications provided the format specifications for the Journal of Mechanical Design, though they are not easy to reproduce. It is their commitment to ensuring quality figures in every issue of JMD that motivates this effort to have authors review the presentation of their figures.

Thanks go to D. E. Knuth and L. Lamport for developing the wonderful word processing software packages TEX and LATEX. We would like to thank Ken Sprott, Kirk van Katwyk, and Matt Campbell for fixing bugs in the ASME style file `asme2ej.cls`, and Geoff Shiflett for creating ASME bibliography stype file `asmems4.bst`.

REFERENCES

[1] Documentary, O. cvtcolor https://docs.opencv.org/3.4/d8/d01/group__imgproc__color__conversions.html#ga397ae87e1288a81d2363b61574eb8cab Accessed: 2022-05-09.

[2] Documentary, O. Color convertion rgb to gray https://docs.opencv.org/3.4/de/d25/imgproc_color_conversions.html#color_convert_rgb_gray Accessed: 2022-05-09.

[3] Documentary, O. Smoothing images https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html Accessed: 2022-05-09.

[4] Documentary, O. Gaussianblur https://docs.opencv.org/4.x/d4/d86/group__imgproc__filter.html#gaabe8c836e97159a9193fb0b11ac52cf1 Accessed: 2022-05-09.

[5] Documentary, O. getgaussiankernel https://docs.opencv.org/4.x/d4/d86/group__imgproc__filter.html#gac05a120c1ae92a6060dd0db190a61afa Accessed: 2022-05-09.

[6] Canny, J., 1986, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **PAMI-8**(6), pp. 679–698.

[7] Weisstein, E. W. L1-norm https://mathworld.wolfram.com/L1-Norm.html Accessed: 2022-05-09.

[8] Weisstein, E. W. L2-norm https://mathworld.wolfram.com/L2-Norm.html Accessed: 2022-05-09.

[9] ASME, 2003, *ASME Manual MS-4, An ASME Paper*, latest ed. The American Society of Mechanical Engineers, New York See also URL `http://www.asme.org/pubs/MS4.html`.

[10] Lamport, L., 1986, *LATEX: a Document Preparation System* Addison-Wesley, Reading, MA.

[11] Goosens, M., Mittelbach, F., and Samarin, A., 1994, *The LATEX Companion* Addison-Wesley, Reading, MA.

[12] Author, A., Author, B., and Author, C., 1994, "Article title," *Journal Name,* **1**(5), May, pp. 1–3.

[13] Booklet, A., 1994, Booklet title On the WWW, at `http://www.abc.edu`, May PDF file.

[14] Inbook, A., ed., 1991, *Book title*, 1^{st} ed., Vol. 2 of *Series Title* Publisher Name, Publisher address, Chap. 1, pp. 1–3 See also URL `http://www.abc.edu`.

[15] Incollection, A., 1991, "Article title," In *Collection Title*, A. Editor, ed., 3^{rd} ed., Vol. 2 of *Series title*. Publisher Name, Publisher address, May, Chapter 1, pp. 1–3 See also URL `http://www.abc.edu`.

[16] Inproceedings, A., 1991, "Article title," In Proceedings Title, A. Editor and B. Editor, eds., Vol. **1** of *Series name*, Organization Name, Publisher Name, pp. 1–3 Paper number 1234.

[17] Mastersthesis, A., 2003, "Thesis Title," MS Thesis, University of Higher Education, Cambridge, MA, May See also URL `http://www.abc.edu`.

[18] Misc, A., 2003, Miscellaneous Title On the WWW, May URL `http://www.abc.edu`.

[19] Proceedings, A., ed., 1991, Volume Title, Vol. 1 of *Proceedings Series*, Organization Name, Publisher Name See also URL `http://www.abc.edu`.

[20] Phdthesis, A., 2003, "Thesis Title," PhD Thesis, University of Higher Education, Cambridge, MA, May See also URL `http://www.abc.edu`.

[21] Techreport, A., 2003, Techreport title Progress report 1, University of Higher Education, Cambridge, MA, May See also URL `http://www.abc.edu`.

[22] Unpublished, A., 2003, Unpublished document title See also URL `http://www.abc.edu`, May.

APPENDIX A: HEAD OF FIRST APPENDIX

Avoid Appendices if possible.

APPENDIX B: HEAD OF SECOND APPENDIX

Subsection head in appendix

The equation counter is not reset in an appendix and the numbers will follow one continual sequence from the beginning of the article to the very end as shown in the

following example.

$$a = b + c. \tag{5}$$