

Page Curl Correction - Training Material | WP3

Author: Sebastian Kirch (Fraunhofer IAIS)

Tool Description

For old books and newspapers it is often not possible to remove the binding before digitizing the individual pages. This often results in distorted pages images due to the warping implied by the books binding (Figure 1).

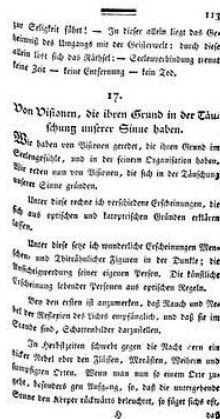


Figure 1: A scanned book page with page curl

Other reasons for distorted pages can be environmental conditions like humidity (which can cause page shrinking over time) or a wrong camera setup.

The Page Curl Correction is a command-line tool capable of detecting such distortions and correcting them automatically. This can significantly improve the results of a subsequent text recognition step since most OCR algorithms neglect this kind of distortion. More information about the tool can be found here: <http://www.digitisation.eu/tools/browse/image-enhancement/geometric-correction-page-curl>

System Requirements

Operating System	Windows
Hardware Dependencies	-
Software Dependencies	-
Non-technical Requirements	Page Curl Correction does not have a graphical user interface (GUI) and therefore requires some basic knowledge on how to use and execute the command-line programs.

Installation

Files

Filename	Description
Page_Curl_Correction_setup.exe	Installation wizard for installing the tool on a Windows system.

Installation Instructions

Page Curl Correction comes with an installer wizard that guides the user through the installation process (It is currently available for Microsoft Windows only). The wizard extracts the installation package to your local C:\ drive.

Quick Start Guide

To get started with the Page Curl Correction tool, just choose a distorted sample image and try one of the examples in the “Examples” section. A description of the parameters for the configuration of the tool can be found in the following section.

Documentation

Configuration and Customization

The Page Curl Correction tool can be configured using command-line parameters. The most important parameter is the one for controlling coarse or fine-grained rectification. Coarse-grained rectification is a computationally low cost transformation which addresses the projection of a curved surface (the original page) to a 2D rectangular area (the scanned image). Fine-grained rectification is a more advanced technique based on text line & word segmentation (See http://users.iit.demokritos.gr/~nstam/page_files/IEEE_TIP_2011_Stamatopoulos.pdf for more information).

The basic call pattern looks like this:

```
Page_Curl_Correction [0/1] [in] [out]
```

Parameter	Req.	Description
[0/1]	✓	If this parameter is set to 1 only coarse rectification is applied. Otherwise coarse and fine rectification are applied.
[in]	✓	Input file
[out]	✓	Output file

More information can be found in the following slides: <http://de.slideshare.net/impactproject/ncsr-page-curl-correction>.



Workflow Integration

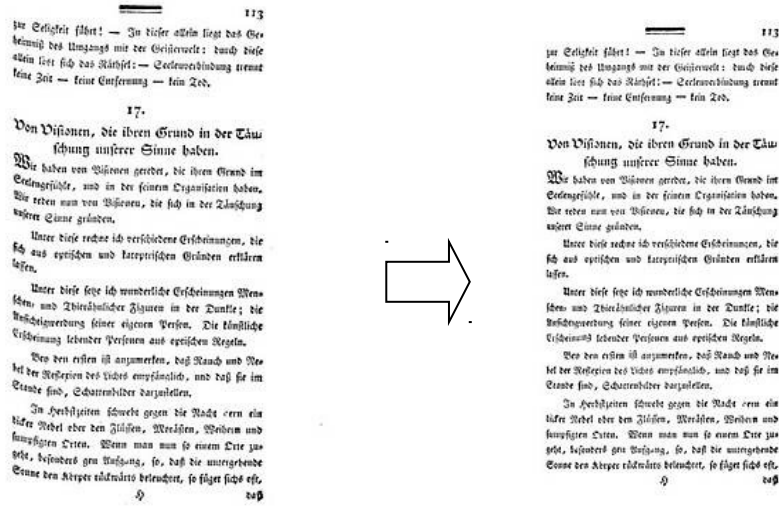
The Page Curl Correction tool can be integrated into any workflow or application that allows the execution of command-line tools. The configuration is rather straight forward since the tool has only three parameters: input, output and fine/coarse-grain adjustment.

Examples

Perform a page curl correction on a sample image, use coarse reification:

```
Page_Curl_Correction 1 sample.tif result.tif
```

Results:



Licensing

The tool is produced by the Computational Intelligence Laboratory, Institute of Informatics and Telecommunications, National Center for Scientific Research "Demokritos" (NCSR). For more information on terms and conditions to use the tool, please contact NCSR.

