

Veel pagina's scannen, één pdf

Abstract

Voor een koor of een orkest is het soms nodig uit een muziekboek een stuk te kopiëren, met het resultaat een aantal gescande afbeeldingen – meestal van het type jpeg of pdf. Hieronder beschrijf ik een methode om met een paar LaTeX-macro's de kantlijnen van alle pagina's recht en symmetrisch te krijgen, de scans op de gehele pagina weer te geven en het resultaat als één pdf bruikbaar te maken voor afdruk. Bijvoorbeeld de trapeziumvorm van een scan corrigeren, dat gaat alleen met meer specifieke software zoals DigiKam.

Gebruik van een editor met kolom-editmogelijkheden kan handig zijn.

Inleiding

De hier beschreven methode is bedoeld om met behulp van LaTeX meerdere gescande pagina's tot één A4-document samen te stellen.

Iedere scan apart bewerken met een programma als Digikam of Gimp is een alternatief, zeker als het nodig is om perspectivisch te vervormen, de pixeldichtheid aan te passen, de jpeg-kwaliteit te variëren, of vlekken in een scan wegpoetsen.

De benadering waarvoor ik hier gekozen heb is over algemeen voldoende, werkt louter met LaTeX, meer heb je niet nodig.

We gaan uit van een aantal scans van bladmuziek

Die scans worden, als het goed is, zo netjes mogelijk gemaakt, maar vaak zitten er zwarte repen of vlekken aan de rand van zo'n scan omdat het origineel niet in het midden en/of niet precies recht heeft gelegen of je wilt dat het resultaat vergroot, pagina-vullend wordt afgebeeld.

De hier gebruikte werkwijze is niet zo bijzonder maar wel effectief: lelijke randen kunnen worden weggehaald, de scan kan uitgespreid worden over de hele pagina en zo nodig (om)gedraaid worden. Ook een voordeel is dat je meerdere pagina's in hetzelfde document tegelijk kunt aanpassen en dat na compileren het eindresultaat van al die wijzigingen meteen zichtbaar is in het einddocument.

De opzet van het systeem, het werkbestand

De basis van het systeem is het LaTeX-werkdocument, waarvan hieronder een voorbeeld wordt weergegeven.

Let op de zeven à acht parameters, met name \Raster, \Witrans en \Bron – waarmee de omschakeling met % van de belangrijkste features beheerd wordt – en verder: \Naam, \Title, \Subtitle, \Author, de info voor titels op een eventuele voorpagina.

Bij het werkdocument hoort, behalve scancorr.cls, een hulpbestand in Postscript dat wordt gebruikt om een raster af te beelden over alle pagina's zodat het gemakkelijker is de afstanden *te schatten* waarmee de randen van scans moeten worden getrimd.

```

%!lualatex
\documentclass{scancorr}
%\pageav % optioneel de output op A5 zetten
\begin{document}
  \Bron{.} % bronmap bijv. {/c/EPSON_scan}
  \Naam{Ernst van der Storm}
  \Title{Ah! mio cor}
  \Subtitle{Giorgio Federico Haendel}
  \Author{Aria}
  \Info % meta info for pdf
  \Raster % hulpraster
  \Witrand % witte rand

%\TitlePage % frontpagina
%\EmptyPage % lege pagina

% #1: bron waaruit pagina gehaald wordt
% #2 #3 #4 en #5 zijn trimwaarden in mm
% #6: draaihoek in graden (+=antiklok)
% #7: paginanummer in bron
%
% #1 #2 #3 #4 #5 #6 #7
% bron lft rgt top bot ang pag
\jpgpdftrim{example0.pdf} {000}{000}{000}{000}{0.00}{001}
\jpgpdftrim{example0.pdf} {000}{000}{000}{000}{0.00}{002}
\jpgpdftrim{example0.pdf} {000}{000}{000}{000}{0.00}{003}
\jpgpdftrim{example0.pdf} {000}{000}{000}{000}{0.00}{004}
\jpgpdftrim{example0.pdf} {000}{000}{000}{000}{0.00}{005}
\end{document}

```

Figuur 1. Broncode van een werkbestand

Ook kan men later met een witte rand van 1 cm breed eventuele *ongerechtigheden* van het scannen *aan de rand van de pagina's* maskeren.

In het weergegeven werkbestand staan het raster en de witte rand beide aan, men kan ze deactiveren door het %-teken ervoor te zetten. In het voorbeeld is na %\TitlePage een optie voor een (nagenoeg) lege pagina \EmptyPage toegevoegd en je kunt dat commando naar behoefte gebruiken, bijvoorbeeld om ervoor te zorgen dat de oneven pagina's rechts komen.

De voorpagina wordt gegenereerd uit de teksten die zijn opgegeven in \Title, \Subtitle en \Author.

De voorbereiding

Ik beschrijf hier de basisstappen van de werkwijze met pdf's. Werken met andere afbeeldingsbestanden gaat net zo, alleen is het geven van een paginanummer bij jpg's of png's in parameter#7 niet nodig, maar soms misschien handig als geheugensteun bij een groot aantal pagina's. Het is ook mogelijk van willekeurige pagina's uit een ander pdf-bestand een 'nieuw origineel' te maken, dat werkt in principe op dezelfde manier.

De voorbereiding voor gebruik van de software is als volgt:

- sla de tekst van scancorr.cls en van het hulpbestand hulprastercm.eps op in de map waar je werkbestanden staan (later kunnen die, als alles werkt, beter naar de gebruikelijke plaatsen verhuisd worden)
- converteer het hulp-eps-bestand bij voorkeur meteen naar hulprastercm.pdf bijv. met epstopdf
- verzamel alle gescande paginabestanden op één plaats: in de broncode staat het pad er naartoe in de definitie van \Bron
- verander indien nodig in het werkbestand de definitie van \Bron{.}, de map dus waar de scans in staan



origineel

met raster

na correctie

De steeds opnieuw genereren-fase

De productie is een steeds opnieuw compileren van het onderhanden bestand, want de getallen die voor het 'trimmen' ingevuld moeten worden kunnen alleen goed worden na het opnieuw beoordelen van het resultaat, feitelijk dus met *trial and error*.

N.B. het raster is vooral bedoeld als richtlijn voor de marges van het resultaat, je kunt er niet echt aan afmeten *hoeveel* je moet corrigeren

Het best is te beginnen met het hulpraster \Raster actief en de witte rand van 1 cm geblokkeerd: %\Witrand.

Ter illustratie is het werkbestand ingevuld met een voorbeeldbestand – de scans – example0.pdf.

- bepaal hoeveel pagina's het geheel moet krijgen (evenveel als er zijn gescand, maar een originele pdf kan – zoals hier – ook meerdere bronpagina's bevatten)
- voer de benodigde regels in, voor iedere pagina van het te genereren stuk één regel, zoals:

```
\jpgpdftrim{example0.pdf} {024}{017}{020}{029}{0.00}{001}
\jpgpdftrim{example0.pdf} {025}{019}{020}{028}{-0.3}{002}
\jpgpdftrim{example0.pdf} {026}{020}{020}{029}{0.50}{003}
\jpgpdftrim{example0.pdf} {027}{019}{019}{028}{0.12}{004}
\jpgpdftrim{example0.pdf} {024}{019}{019}{028}{0.00}{005}
```

Daarin zijn de argumenten:

1. de naam van de onderhavige bronpagina
2. linker trimwaarde
3. rechter trimwaarde
4. top-trimwaarde
5. onder-trimwaarde
6. hoek (in graden)
7. paginanummer (nodig bij pdf's).

- de vier getallen van de trimwaardes moeten zo goed als zeker aangepast worden want de eerste vier getallen zijn allemaal nul (of zijn blijven staan van ‘de vorige keer’)
- negatieve waardes zorgen voor *extra* witmarge, bijvoorbeeld als de originele pagina *te dicht langs de rand* is gescand
- hieraan de nodige tijd besteden, steeds opnieuw compileren...

Pagina's kantelen of omdraaien

Parameter vijf, die de pagina kantelt, is meestal 0.00 (niet kantelen) en hoeft, als het om kleine correcties gaat, doorgaans pas in een later stadium aangepast te worden; bijvoorbeeld 1.0° is al veel. Een positieve hoek is kantelen naar links, een negatief getal kantelt naar rechts.

Het is hier ook mogelijk om een pagina *op z'n kop* te zetten door 180 in te vullen, of eventueel een kwartslag, 90° of 270°. Dat moet in de voorkomende gevallen wél eerst ingesteld worden want het omdraaien heeft gevolgen voor de randen die getrimd worden. De volgorde van de trimparameters blijft steeds links, rechts, boven, onder.

Extra voorpagina genereren en de afwerking

Om de paginanummering van het gescande origineel correct te houden – even pagina's links, oneven rechts – kun je met \TitlePage eenvoudig een voorpagina toevoegen. Nodig is dan wel om de bij de titelpagina behorende informatie in te voeren: \Title, \Subtitle en \Author. Met \EmptyPage kan zo nodig een lege pagina worden toegevoegd.

Aan het eind moet het raster weggehaald worden door die regel met %\Raster uit te zetten. Zo nodig kun je de witte rand activeren met \Wi trand om ongerechtigheden aan de paginarand te maskeren.

Uiteraard kunnen er tussen het gescande materiaal ook vlekken of aantekeningen zitten die je niet in het eindresultaat wil zien. Om dat goed te krijgen moet je de betreffende scan vooraf met een ander programma opschonen, dat is met deze software niet mogelijk. Liedteksten die slecht leesbaar zijn verbeteren, idem.

Als gescand is met een voldoende hoge resolutie levert het vergroten door *include-graphics* – ook al wordt een blad gekanteld – toch een goed resultaat op.

Voor een (kopie van) een partituur is het logisch om de paginagrootte op A4 te houden, maar als de bronbestand afkomstig is van bijvoorbeeld een klein boekje dat oorspronkelijk als pagina-afmeting 20x21 cm2 heeft, kan het handig zijn het resultaat kleiner te printen. Daarvoor is in het werkbestand optie uit het package *memoir* in de class file \pageav te activeren; daarmee komt alle output op A5.

Broncode van de *scancorr* class-file

```
\NeedsTeXFormat{LaTeX2e}[1999/12/01]
\ProvidesClass{scancorr}
    [2018/12/06 v1.03 manually trim scans]
\LoadClass[a4paper,fleqn]{memoir}

\RequirePackage{graphicx}
\RequirePackage{xcolor}
\RequirePackage[a4paper,margin=0mm]{geometry}
\RequirePackage{polyglossia}
\setdefaultlanguage{dutch}
\RequirePackage{fontspec}
\RequirePackage{alltt}
\def\Wi{\paperwidth}
\def\Hi{\paperheight}
\usepackage[absolute,overlay]{textpos}\TPGrid{\Wi}{\Hi}
\pagestyle{empty}
```

```

\parindent0pt

\newcommand{\Rasterpagina}{\par%
\begin{textblock*}{\Wi}{0pt,0pt}%
\begin{minipage}[c][\Hi][b]{\Wi}%
\includegraphics[width=1\Wi,totalheight=1\Hi]{hulprastercm}%
\end{minipage}%
\end{textblock*}}
\newcommand{\raster}{\ifx\RA\undefined\else\Rasterpagina\fi}

% \begin{textblock*}{hsize}[ho,ve](hpos,vpos) %\rule{breedte}{hoogte}
\newcommand{\Randpagina}{\par%
\begin{textblock*}{\Wi}{0pt,0pt}\color{white}\rule{\Wi}{.033\Hi}\end{textblock*}
\begin{textblock*}{\Wi}{0pt,0pt}\color{white}\rule{.047\Wi}{\Hi}\end{textblock*}
\begin{textblock*}{\Wi}{.951\Wi,0pt}\color{white}\rule{.049\Wi}{\Hi}\end{textblock*}
\begin{textblock*}{\Wi}{0pt,.9657\Hi}\color{white}\rule{\Wi}{.0343\Hi}\end{textblock*}
}
\newcommand{\wirand}{\ifx\WR\undefined\else\Randpagina\fi}

% bb= llx lly urx ury
\newlength{\wisselLRTB}
\newcommand{\jpgpdftrim}[7]{
\enlargethispage*{100pt}%
\wiselLRTB=#6pt % 180.0pt komt hierin of 0.1pt enz.
\message{==trim=#2mm #3mm #4mm #5mm,angle=#6,page=#7}
\ifdim\wiselLRTB>260pt% kanteel 3/4%
\includegraphics[trim=#4mm #2mm #5mm #3mm,angle=#6,page=#7,width=\paperwidth,totalheight=\paperheight,clip]{\bronMAP/#1}
\else
\ifdim\wiselLRTB>175pt% kanteel op zijn kop
\includegraphics[trim=#3mm #4mm #2mm #5mm,angle=#6,page=#7,width=\paperwidth,totalheight=\paperheight,clip]{\bronMAP/#1}
\else
\ifdim\wiselLRTB>85pt% kanteel een kwart
\includegraphics[trim=#5mm #3mm #4mm #2mm,angle=#6,page=#7,width=\paperwidth,totalheight=\paperheight,clip]{\bronMAP/#1}
\else%
\includegraphics[trim=#2mm #5mm #3mm #4mm,angle=#6,page=#7,width=\paperwidth,totalheight=\paperheight,clip]{\bronMAP/#1}
\fi
\fi
\fi
\raster\wirand\newpage}

% titelpagina genereren:
\def\TitlePage{\maketitle\newpage}
% lege pagina toevoegen:
\def\EmptyPage{%
\mbox{}
\vskip30mm%
\centerline{\jobname}%
\centerline{\tiny Deze pagina is met opzet leeg gelaten}%
\newpage}

\def\Bron#1{\def\bronMAP{#1}}
\def\Naam#1{\def\eigenaam{#1}}
\def\Title#1{\def\Titel{#1}\title{#1}}
\def\Subtitle#1{\def\stitle{#1}}\def\stitle{\relax}
\def\Author#1{\def\Auteur{#1}\author{#1}}
\def\Raster{\def\RA{}}
\def\Witrans{\def\WR{}}
\def\Info{
\pdfextension info{
/Title (\jobname)
/Creator (LaTeX)
/Producer (\eigenaam)
/Author (\Auteur)
/Subject (\Titel, \stitle)
}
}

\pretitle{\vskip30mm\begin{center}\fontsize{43}{43}\selectfont}
\posttitle{\vskip10mm\fontsize{21}{21}\selectfont\mbox{}\\stitle\end{center}}
\preauthor{\vskip30mm\begin{center}\fontsize{25}{25}\selectfont}
\postauthor{\end{center}}
\date{}
\endinput

```

Broncode hulpbestandje hulprastercm.eps

```

%%!PS-Adobe-3.0 %% hulprastercm.eps
%%BoundingBox: 0 0 596 842
1 0 1 setrgbcolor %% lichtpaars
/hg 842 def %% hoogte=lengte verticale lijn
/br 595 def %% lengte horizontale lijn
/m { 28.35 } def %% maat van het vakje, komt overeen met 1 cm
/vl { hg dup 0 exch rlineto neg 0 exch rmoveto } bind def %% verticaal
/hl { br dup 0 rlineto neg 0 rmoveto } bind def %% lijn horiz.
/shr { m 0 rmoveto } bind def %% stap naar rechts
/shl { m neg 0 rmoveto } bind def %% stap naar links
/svu { 0 m rmoveto } bind def /svd { 0 m neg rmoveto } bind def
/cirkel { 0 0 8 0 360 arc stroke } def
/raster { gsave
%% eerst verticale: 2 links en 2 rechts
0 1 moveto 0 1 2 { vl shr } for
br 1 moveto 0 1 2 { vl shl } for
%% dan horiz. 4 onder en 4 boven
0 1 moveto 0 1 4 { hl svu } for
0 hg moveto 0 1 4 { hl svd } for
0.002 setlinewidth stroke %% dunne lijnen
grestore } bind def
raster %% op 0 1
gsave m 1.05 mul 2.85 m mul translate cirkel grestore
gsave m 1.05 mul 10.85 m mul translate cirkel grestore
gsave m 1.05 mul 18.85 m mul translate cirkel grestore
gsave m 1.05 mul 26.85 m mul translate cirkel grestore
gsave m 19.95 mul 2.85 m mul translate cirkel grestore
gsave m 19.95 mul 10.85 m mul translate cirkel grestore
gsave m 19.95 mul 18.85 m mul translate cirkel grestore
gsave m 19.95 mul 26.85 m mul translate cirkel grestore
10 819 moveto %% infotekst plaatsen:
/Courier findfont 25 scalefont setfont
%%(afstand der lijnen is grofweg 1 cm) show
%%10 790 moveto %% infotekst plaatsen:
(positieve waarden halen links,) show
10 791 moveto %% infotekst plaatsen:
(rechts, onder en boven randen weg van) show
10 763 moveto %% infotekst plaatsen:
(het document (dus beeld wordt groter)) show
gsave
573 777 moveto %% infotekst plaatsen:
270 rotate
(& negatieve - laten MEER rand zien van het beeld) show
grestore
11 749 moveto %% infotekst plaatsen:
270 rotate
/Courier findfont 15 scalefont setfont
(de buitenste strook wordt rondom met wit afgedekt met \\def\\WR{ja}) show
%%90 rotate
%%10 734 moveto %% infotekst plaatsen:
%%(van het origineel) show
showpage %%EOF

```

Ernst van der Storm
evdstorm@xs4all.nl