

Thanks for your reply!

Firstly, I typed `ocrmypdf -l chi_sim --tesseract-psm 6 tt.pdf r.pdf`, and I got
`ocrmypdf: error: unrecognized arguments: --tesseract-psm r.pdf`.

I tried with `ocrmypdf --tesseract-pagesegmode 6 --sidecar a.txt -l chi_sim tt.pdf r.pdf`, the output txt doesn't have extra space, while the output pdf has. The test material mentioned above was in the attachment.

Secondly, for the recommendation to use verbose 2, the result is the following. I'm sorry for the typesetting as I don't know what's the better way to present it to you.

```
C:\Users\EYx\Desktop>ocrmypdf tt.pdf r.pdf -l chi_sim --verbose 2
```

```
[WinError 2] 系统找不到指定的文件。
```

```
DEBUG ocrmypdf - ocrmypdf 11.5.0
```

```
DEBUG ocrmypdf.subprocess - Running: ['Y:\\Program Files\\Tesseract-OCR\\tesseract - v5.0\\tesseract.EXE', '--list-langs']
```

```
DEBUG ocrmypdf.subprocess.tesseract - stdout/stderr = List of available languages (161):
```

```
afr  
amh  
ara  
asm  
aze  
aze_cyrl  
bel  
ben  
bod  
bos  
bre  
bul  
cat  
ceb  
ces  
chi_sim  
chi_sim_vert  
chi_tra  
chi_tra_vert  
chr  
cos  
cym  
dan  
deu
```

div
dzo
ell
eng
enm
epo
equ
est
eus
fao
fas
fil
fin
fra
frk
frm
fry
gla
gle
glg
grc
guj
hat
heb
hin
hrv
hun
hye
iku
ind
isl
ita
ita_old
jav
jpn
jpn_vert
kan
kat
kat_old
kaz
khm
kir
kmr
kor

lao
lat
lav
lit
ltz
mal
mar
mkd
mlt
mon
mri
msa
mya
nep
nld
nor
oci
ori
osd
pan
pol
por
pus
que
ron
rus
san
script/Arabic
script/Armenian
script/Bengali
script/Canadian_Aboriginal
script/Cherokee
script/Cyrillic
script/Devanagari
script/Ethiopic
script/Fraktur
script/Georgian
script/Greek
script/Gujarati
script/Gurmukhi
script/HanS
script/HanS_vert
script/HanT
script/HanT_vert

script/Hangul
script/Hangul_vert
script/Hebrew
script/Japanese
script/Japanese_vert
script/Kannada
script/Khmer
script/Lao
script/Latin
script/Malayalam
script/Myanmar
script/Oriya
script/Sinhala
script/Syriac
script/Tamil
script/Telugu
script/Thaana
script/Thai
script/Tibetan
script/Vietnamese
sin
slk
slv
snd
spa
spa_old
sqi
srp
srp_latn
sun
swa
swe
syr
tam
tat
tel
tgk
tha
tir
ton
tur
uig
ukr
urd

DEBUG ocrmypdf.subprocess - 1 Running: ['Y:\\Program Files\\Tesseract-OCR\\tesseract -v5.0\\tesseract.EXE', '-l', 'chi_sim', '-c', 'textonly_pdf=1', WindowsPath('C:/Users/EYx/AppData/Local/Temp/ocrmypdf.io.9m8h58e_/000001_ocr.png'), 'C:\\Users\\EYx\\AppData\\Local\\Temp\\ocrmypdf.io.9m8h58e_\\000001_ocr_tess', 'pdf', 'txt']

WARNING ocrmypdf._exec.tesseract - 1 [tesseract] lots of diacritics - possibly poor OCR

DEBUG ocrmypdf._graft - 1 Text rotation: (text, autorotate, content) -> text misalignment = (0, 0, 0) -> 0

DEBUG ocrmypdf._graft - 1 Grafting

DEBUG ocrmypdf._graft - 1 Page rotation: (content, auto) -> page = (0, 0) -> 0

OCR:

100%  | 1.0/1.0
[00:01<00:00, 1.68s/page]

INFO ocrmypdf._sync - Postprocessing...

DEBUG ocrmypdf.subprocess - Running: ['C:\\Program Files\\gs\\gs9.53.3\\bin\\gswin64c.EXE', '-dBATCHE', '-dNOPAUSE', '-dSAFER', '-dCompatibilityLevel=1.6', '-sDEVICE=pdfwrite', '-dAutoRotatePages=/None', '-sColorConversionStrategy=RGB', '-dAutoFilterColorImages=true', '-dAutoFilterGrayImages=true', '-dJPEGQ=95', '-dPDFA=2', '-dPDFACompatibilityPolicy=1', '-o', '-', '-sstdout=%stderr', 'C:\\Users\\EYx\\AppData\\Local\\Temp\\ocrmypdf.io.9m8h58e_\\fix_docinfo.pdf', 'C:\\Users\\EYx\\AppData\\Local\\Temp\\ocrmypdf.io.9m8h58e_\\pdfa.ps']

DEBUG ocrmypdf.subprocess.gswin64c - GPL Ghostscript 9.53.3 (2020-10-01)

DEBUG ocrmypdf.subprocess.gswin64c - Copyright (C) 2020 Artifex Software, Inc. All rights reserved.

DEBUG ocrmypdf.subprocess.gswin64c - This software is supplied under the GNU AGPLv3 and comes with NO WARRANTY:

DEBUG ocrmypdf.subprocess.gswin64c - see the file COPYING for details.

DEBUG ocrmypdf.subprocess.gswin64c - Processing pages 1 through 1.

DEBUG ocrmypdf.subprocess.gswin64c - Page 1

DEBUG ocrmypdf.subprocess.gswin64c - GPL Ghostscript 9.53.3: PDF/A doesn't allow images with Interpolate true.

DEBUG ocrmypdf.subprocess.gswin64c - GPL Ghostscript 9.53.3: PDF/A doesn't allow images with Interpolate true.

PDF/A conversion:
100%  | 1/1 [00:00<00:00,
4.87page/s]

DEBUG ocrmypdf.optimize - Treating 18 as an optimization candidate

DEBUG ocrmypdf.optimize - Skipping image 18 because it is an SMask

DEBUG ocrmypdf.optimize - Treating 19 as an optimization candidate

DEBUG ocrmypdf.optimize - XrefExt(xref=19, ext='.png')

DEBUG ocrmypdf.optimize - Optimizable images: JPEGs: 0 PNGs: 1

JPEG: 0image [00:00, ?image/s]
DEBUG ocrmypdf.optimize - Treating 18 as an optimization candidate
DEBUG ocrmypdf.optimize - Skipping image 18 because it is an SMask
DEBUG ocrmypdf.optimize - Treating 19 as an optimization candidate
DEBUG ocrmypdf.optimize - Optimizable images: JBIG2 groups: (0,)
JBIG2: 0item [00:00, ?item/s]
INFO ocrmypdf.optimize - Optimize ratio: 1.00 savings: 0.0%
DEBUG ocrmypdf_pipeline -
C:\Users\EYx\AppData\Local\Temp\ocrmypdf.io.9m8h58e_\optimize.pdf -> r.pdf
INFO ocrmypdf._sync - Output file is a PDF/A-2B (as expected)

Thanks for your time again!