

NAME

cjb2 - Simple DjVuBitonal encoder.

SYNOPSIS

cjb2 [*options*] *inputfile* *outputdjvufile*

DESCRIPTION

This is a simple encoder for bitonal files. Argument *inputfile* is the name of a PBM or bitonal TIFF file containing a single document image. This program produces a DjVuBitonal file named *outputdjvufile*.

The default compression process is lossless: decoding the DjVuBitonal file at full resolution will produce an image exactly identical to the input file. Lossy compression is enabled by options **-losslevel**, **-lossy**, or **-clean**.

OPTIONS**-dpi** *n*

Specify the resolution information encoded into the output file expressed in dots per inch. The resolution information encoded in DjVu files determine how the decoder scales the image on a particular display. Meaningful resolutions range from 25 to 1200. The default resolution for TIFF files is the resolution is the resolution specified by the input file. The default resolution for PBM files is 300 dpi.

-lossless

Ensure that the encoded image is pixel-per-pixel equal to the initial image. This option is is equivalent to **-losslevel 0** and is the default.

-clean

Only remove flyspecks from the input image. This option enables a heuristic algorithm that removes very small marks. Such marks are often causes by noise and dust during the scanning process. The threshold mark size is chosen according to the resolution specified with option This option is is equivalent to **-losslevel 1**.

-lossy

Substitute patterns with small variations. In addition to the flyspeck removal heuristic, this option enables an algorithm that encodes certain characters by simply replicating the shape of a previously encoded character with a similar shape. This option is is equivalent to **-losslevel 100**.

-losslevel *x*

Specify the aggressiveness of the lossy compression. Its argument ranges from 0 to 200. Higher values generate smaller files with more potential distortions. Loss level 0 corresponds to lossless encoding. Loss level 1 performs image cleaning but does not perform character substitution at all. Loss level 100 is intended to provide a good compromise. Higher loss levels provide marginally better compression at the risk of unacceptable character substitutions.

-verbose

Display informational messages while running.

REMARKS

Lossless encoding is competitive with that of the Lizardtech commercial encoders.

Lossy encoding has made much progress thanks to Ilya Mezhirov from the minidjvu project. This also means that the lossy encoding performance can change from version to version. When lossy compression yields inadequate results, simply revert to only using option **-clean** or reduce the parameter of option **-losslevel**.

Two features are still missing:

- * Half-tone detection. Collecting small marks belonging to half-tone patterns would improve compression speed.
- * Multi-page compression. Matching characters on several pages would improve the compression ratios for multi-page documents.

CREDITS

This program was initially written by L on Bottou <leonb@users.sourceforge.net> and was improved by Bill Riemers <docbill@sourceforge.net> and many others. The pattern matching algorithm for lossy compression was contributed by Ilya Mezhirov <ilya@mezhirov.mccme.ru>. TIFF input routines are inspired by the ones contributed by R. Keith Dennis <dennis@rkd.math.cornell.edu> and Paul Young.

SEE ALSO

djvu(1), **pbm(5)**.