

Smart Text Recognition

A new version of Autolmage Pro for AutoCAD 2000/i and AutoCAD LT 2000/i also offers the improved engine for text recognition.

You can never expect 100 percent accuracy from text vectorization - especially if it is from a hand done drawing - but you can significantly improve the results and save time by selecting the appropriate algorithm and tuning parameters.

Text Areas

The processing of raster text is divided into two stages – searching for text strings and applying additional parameters to retrieved raster text.

Firstly, open the *Conversion Options* dialog and turn on the *Text Area* option in the *Recognition* tab. This action you select the algorithm for text recognition.

You may go further by defining additional parameters for text recognition, such as *None*, *Text Areas*, *Polylines*, *Outlines* and *OCR* (figure one).

The first four options described below are recommended for manual drawings with text that can hardly be recognized by the software.

Select *None*, if you don't want to vectorize the text. The areas, containing raster text will be ignored.

Text Areas - creates a vector rectangle boundary around raster text strings. Areas, containing found raster text are not vectorised (figure two). Vector texts can be entered manually with the help of the review and correction procedure.

Polylines – approximates the central lines of the raster text strings with polylines.

Outlines - approximates boundaries of the raster text strings by outline polylines.

OCR - recognizes raster text strings and creates the AutoCAD text objects.

The algorithm uses the value of the *Text Height* of the *Options* tab as the maximum height of upper case raster text characters.

Setting up OCR module.

Setting up the Autolmage Pro OCR module is performed under the *Text* tab (figure three).

The standard package contains a file of OCR character templates, which the program uses to recognize the characters of English alphabet, digits punctuation marks and special characters. If OCR can not recognize the character, it is substituted by the tilde sign.

In *Orientation* mode you can define accepted raster text orientation:

Horizontal - searches for horizontal text lines only.

Horizontal and vertical - searches for horizontal and vertical

Arbitrary - searches for all text lines.

If you switch on the option *Overlapped by Graphic*, Autolmage Pro will search for the raster texts crossing other objects. but remember that this option can reduce the speed for text area search.

With the *Standalone Letters* options on, the software will search for standalone text characters. If this option is off, Autolmage Pro will not find single text characters, but also will not identify graphic objects such as markers, dash, etc.

In the *Height Table* you can specify possible text heights. If the check box is on, then during the generation of recognized texts, the OCR module will create text objects with heights from the list, rounding the recognized height to the nearest value specified in the list.

Using the *Place to Layer* option you can set the name of an AutoCAD layer, on which vectorized text will be placed.

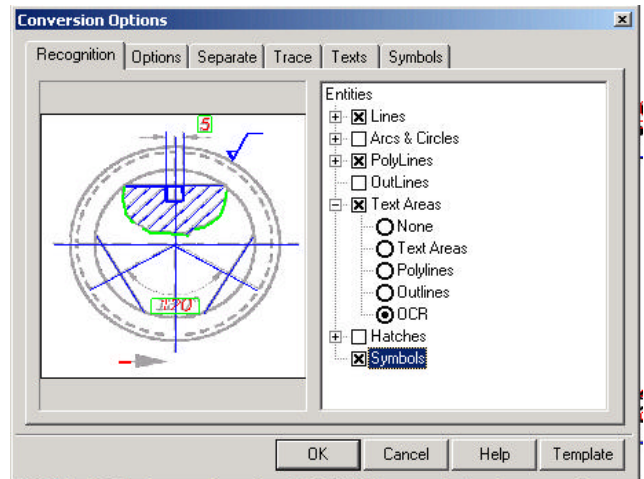


Figure one: selecting algorithm for text recognition

Editing recognized texts

Autolmage OCR checkmarks all created vector strings to enable you to review and edit the texts using the special dialog box (figure four).

To modify a text line, enter a new value in the box. To move a text string, press the *Move* button and specify a new position for text insertion point.

To modify the text height, press the *Modify Height* button and specify on the screen or enter a value of text height in the command line.

If you accept the edited text, press the *Accept* button. The program will exclude this text string from the recognised text set and will automatically proceed to the next recognised text.

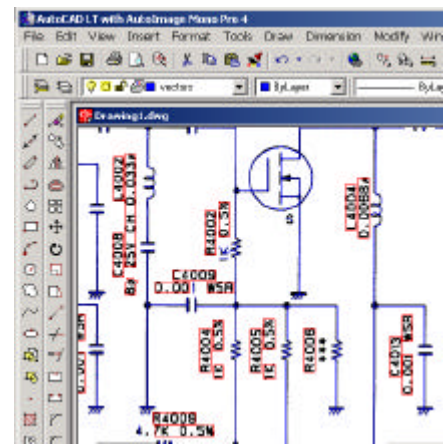


Figure two: example of text areas recognition

For more information, please visit the Rastertech web site:

www.rastertech.com.au

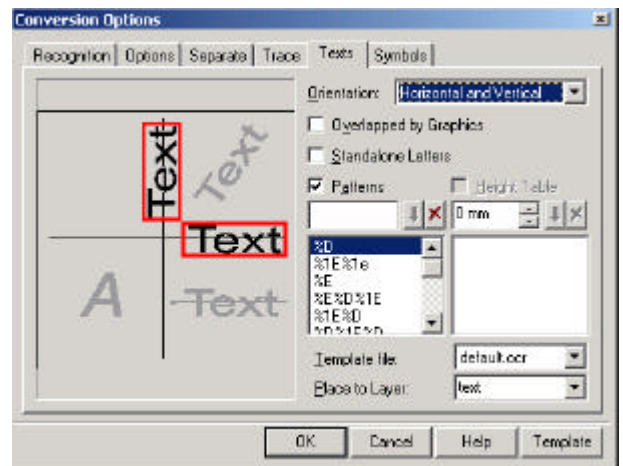


Figure three: setting up OCR parameters

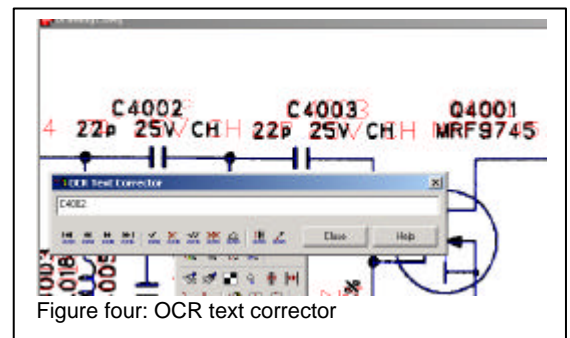


Figure four: OCR text corrector