

# Recogniform Reader

## Desktop & Workgroup

BEYOND OPTICAL READING



# Recogniform Reader

## BEYOND OPTICAL READING

Recogniform Reader is the innovative solution for the automatic data capture from paper forms, suitable for little users as well big companies and services suppliers. Using Recogniform Reader you will immediately and simultaneously get:

- ▶ Costs DOWN
- ▶ Speed UP
- ▶ Quality 100% guaranteed

Compared to manual data-entry, Recogniform Reader allows you to reduce the number of human resources, always operating at the maximum hardware speed, without performance decreasing and errors caused by personnel's tiredness and boring.

## RECOGNITION ENGINES: FAST, UNLIMITED AND RELIABLE

Recogniform Reader includes several recognition engines, in order to read every kind of data:

- ▶ ICR  
*HAND-WRITTEN **CONSTRAINED** AND NOT*

- ▶ OCR  
Printed text / Typed text

- ▶ OCR-A, OCR-B, MICR E13B, MICR CMC7  
Standard codelines with numbers & symbols

- ▶ BCR  
Any tipe of barcode 

- ▶ OMR  
Check boxes 

- ▶ CHR  
*Curive Hand-written text*

Moreover:

- ▶ Cut and extrapolate whole pages or part of them, creating digital format images that can be easily filed: **IDE**
- ▶ Enable the assisted manual data-entry from image: **AMK**



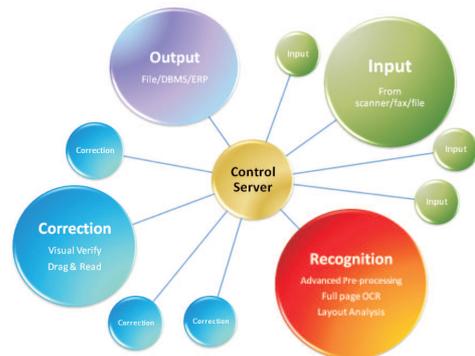
## CLIENT-SERVER ARCHITECTURE OR SINGLE STATION

While with the Desktop version all features are available on one single station, with the Workgroup solution all steps (input, recognition, validation, correction and output) can be distributed on several dedicated stations.

The Workgroup solution is made up of the following applications:

- ▶ Recogniform Control Server
- ▶ Recogniform Input Station
- ▶ Recogniform Recognition Station
- ▶ Recogniform Correction Station
- ▶ Recogniform Output Station

Workgroup's configuration can change easily, according to your specific requirements. Thanks to our innovative licensing system, you can use any number you wish of input, correction and output stations without paying any additional charges. For example, if the correction phase requires more resources, it's possible to arrange multiple correction stations; at the same time, if you have more scanners available, it would be convenient to arrange more input stations.



The communication between stations and server is TCP/IP based. In this way the Recogniform Workgroup Reader is suitable and ready for internet use, allowing you to place the stations in different locations.

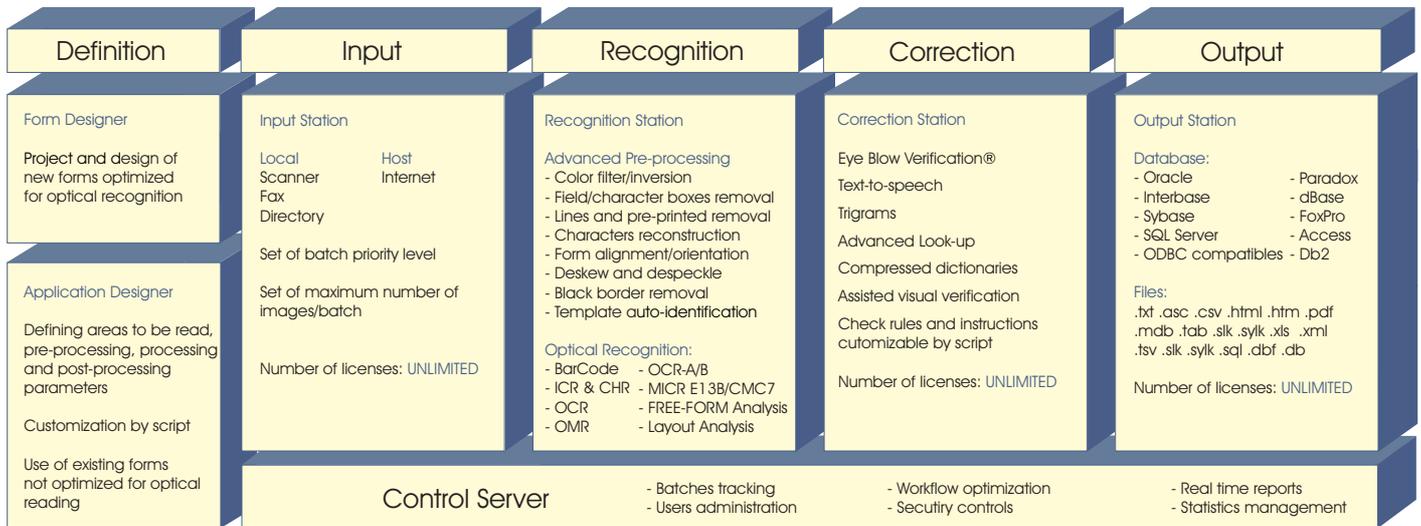
**UNLIMITED:** the number of input, correction and output stations. No extra charge for additional stations!

**UNLIMITED:** the number of forms you can recognize. Process as many forms you wish, with no quantity restrictions!

**UNLIMITED:** the speed of our recognition engines. Speed depends only from the hardware you use, and can easily reach several hundreds of characters recognized per second!



# Recogniform Technologies



Recogniform Workgroup Reader's architecture

## FREE FORM AND LAYOUT ANALYSIS

Recogniform Free Form and Layout Analysis Engine is the optional module for Recogniform Reader allowing to recognize not-structured forms or semi-structured forms. This means that we can recognize a field even if we don't know exactly its position. Fields position is not fixed, but it's variable. This module is the solution for invoice processing: through the free-form recognition engine we can identify a field according to several parameters: label, format, layout. At the basis of the Free-Form processing engine there's the use of our powerful OCR Full text recognition engine, together with a complex layout analysis algorithm. Thanks to these two instruments, we can identify horizontal and vertical lines, text blocks, text elements and verify the consistency of the conditions and rules that have been set.

## ADVANCED PRE-PROCESSING AND QUALITY CONTROL

An adequate preventive images elaboration increases the quality of the recognition process and reduces the files sizes. Recogniform Reader includes a lot of functions covering every specific requirement.

**Deskew** Correction of the slope that sometimes is caused by the scanner using ADF.

**Black Border Removal** Removal of the possible black edges due to the excessive sizes of the scanning area.

**Form Alignment** Alignment of the form to compensate the horizontal and vertical shift.

**Form Removal** Removal of the pre-printed elements when the form is not printed with blind-ink.

**Lines Removal** Removal of horizontal and vertical lines rebuilding the characters crossed by them.

**Despeckle** Removal of black isolated points and spots.

**Field Box Removal** Removal of the box delimiting a single field, if it is not printed with blind-ink.

**Character Box Removal** Removal of the box delimiting the single characters of constrained fields, if it is not printed with blind-ink.

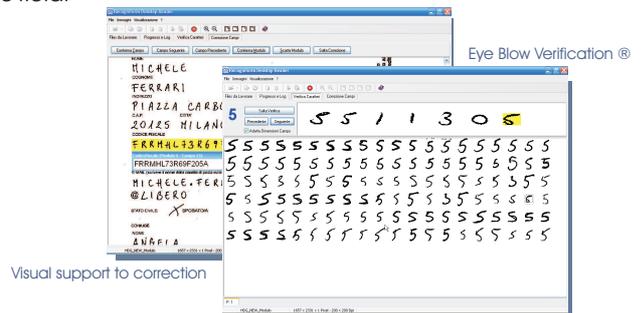
**Color Inversion** Inversion of colours in case of white characters on dark background.

These cleaning functions can be used on the whole image or locally, defining the specific areas to be processed.

## EYE BLOW VERIFICATION®

Thanks to the Recogniform Technologies exclusive system called **Eye Blow Verification®**, the operator can easily verify thousands of characters a minute with minimum effort and maximum accuracy. This system is based on the principle that, inside a group of similar elements, the eye is automatically attracted from the different ones. For example, if in a group of people, everybody suits a blue t-shirt except a person who has got a red t-shirt, the last one would be fast noticed without a detailed examination of the whole group!

According to these considerations, all the characters recognized and classified in the same manner will be showed in the same context, allowing the operator to immediately look and notice the presence of intruders generated by wrong classification. In this way it's possible to find and correct all the possible errors of recognition introduced by not eliminable dirtiness, defects of scanning, ambiguity, etc. Also in case of ambiguous cases, Recogniform Correction Station instantly allows you to solve the problem, automatically recalling on the video the image of the whole field.



## EXCLUSIVE INTEGRATED FEATURES

### Text-to-speech

To save time during the correction phase, the operator can listen the read data from the computer, using the text-to-speech technology: in this way, instead of reading data two times, the operator can read only the data on the image, listening from the computer the read data.

### Trigrams

This system uses statistical information relative to all the possible combinations of three consecutive characters in the target language, analyzing thousands of words: these combinations are called **trigrams**. In this way all the trigrams combinations, the allowed one and the forbidden one, as well as their using frequency, are fully and clearly known. For example, if a recognition engine reads the word "SMITK", the system will correct this word changing it into "SMITH", without operator's intervention and with no risk to give a wrong result. In fact, using a new technology called CREP® (Common Recognition Errors Proofing) and with the help of an Artificial Intelligence system, Recogniform Reader is able to investigate all proofing combinations, selecting the most probable one. All this process takes place in full autonomy.

### Advanced look-up and compressed dictionaries

To validate the read data, you can switch your databases into special compressed dictionaries. Look-up speed is very high: hundreds of thousands of comparisons/second. Compression is around millions of words/Mb. Moreover, thanks to **auto-enriching** feature, the dictionaries can grow as well as the correction goes on, getting your job faster, easier and more accurate.

Many ready-to-use dictionaries are available for every need.

## CUSTOMIZATION AND SCRIPT

A specific software procedure, **Recogniform Application Designer**, allows you to set all the necessary parameters to extrapolate data from the forms, using a visual interface: it's also possible to use a **flexible scripting language** to get new customized functions. With this language you can associate a specific procedure written by user to each scheduled event, modifying the standard behaviour of the program or integrating new features. This feature gives you the possibility to implement new personal functions like controls of squaring between fields, multiple readings of the same field with different pre-processing options, or output of personalized layouts. The advantages are as simple as great: lower costs and better reliability and flexibility.

**RECOGNIFORM TECHNOLOGIES** is a private Italian company, leader in high technology image processing and image recognition software.

Constantly providing solutions to many Industry and Services private Companies, Government Agencies and Educational Institutions all over the world, Recogniform Technologies is the only Italian company to develop all the optical recognition engines used in its products.

All Recogniform Reader features are also available as SDK.



## OUTPUT FOR EVERY NEED

Before being stored, the read data can be formatted in several ways, allowing the output to adapt to any necessity: conversion to uppercase or lowercase, empty spaces deletion, characters substitution, addition of suffix or prefix, etc.

Supported Databases:

Oracle, Interbase, Sybase, Microsoft SQL Server, DB2, Paradox, dBase, FoxPro, ODBC/ADO/dbExpress, Access.

Supported file formats:

ASCII text (.txt), Fixed spaced text (.asc), Comma Separated Value (.csv), Tab Separated Value (.tsv/.tab), Hypertext Markup Language (.htm/.html), Extensible Markup Language (.xml), Microsoft Excel (.xls), Symbolic Link Interchange Format (.slk/.sylk), SQL scripts (.sql), dBase (.dbf), Paradox (.db), Access (.mdb), Portable Document Format (.pdf), Tagged Interchange File Format (.tif), Bitmap (.bmp), Joint of Photography Expert Group (.jpg).

In addition to the listed formats, it's now available a mixed format "data + images" that allows to get in output a single searchable PDF file; in this file, recognized data are put on elaborated: the first ones are not visible, but it's possible to search, cut and paste. This feature is useful to create an autonomous PDF file from homogeneously filled documents, keeping the original scanned images and the recognized data or data inserted by keying for indexing the content.

If you wish, you can also overlay a pre-printed form to the PDF output, useful to re-build the original form if the scanning has been made removing blind-ink.

