PRIME RECOGNITION™
High Accuracy Optical Character Recognition

Try our OCR on the Web: www.primerecognition.com

PrimeOCR Feature Description

PrimeOCR is award-winning OCR software designed for the Windows NT/2000 environment that uses "Voting" technology to generate 65-82% fewer errors than the best conventional OCR products.

A growing list of features makes it easy to customize PrimeOCR for your type of image, accuracy and performance requirements. PrimeOCR features address not only your character accuracy and throughput demands, but also a wide range of pre- and post-OCR requirements that may be part of your imaging conversion process.

PrimeOCR Access Methods

PrimeOCR can be accessed through a variety of means:

Job Server

The Job Server is designed to control the way PrimeOCR performs unattended, batch image processing. Its graphical interface provides on-screen progress reporting of image jobs and gives users numerous production OCR options such as:

- **Specify Input Job Directory/Subdirectory Locations and Polling Options** - The Job Server can span multiple directories while searching for image jobs to process.

  In addition, the Job Server can be set up to process all image jobs and terminate or continuously poll job input directories for new images. It even allows users to specify multiple job "queues" so jobs can be prioritized.

- **Modify PrimeOCR Initialization Settings** - A host of PrimeOCR configuration values can be set through this option including engine speed, image template overrides and OCR diagnostics.

- **Enter Custom PrimeOCR (String) Processing Options** - PrimeOCR includes many powerful pre- and post-OCR functions that can be invoked through String options in the Job Server.

  Examples include replacing or removing repetitive characters in the OCR output, generating a variety of statistics on characters or words for a given image zone or document, and adjusting RTF/PDF output file defaults.

- **Review/Reset Processing and Error Log Files** - The Job Server keeps a complete record of PrimeOCR image processing activities through log files. These files can be reviewed or reinitialized through the Job Server as well.

- **Set up PrimeOCR Error Recovery Alternatives** - PrimeOCR has a number of built-in fault tolerant features. The Job Server extends this capability by allowing users to select what to do when an error is encountered. For example, users can define specific error codes where the Job Server (and PrimeOCR) should restart.

- **Control OCR Results File and Input Image File I/O** - The Job Server automates a variety of file maintenance issues by allowing users to delete image files after processing, overwrite duplicate output files, change output directories and file naming conventions, and specify the location of document confidence reports.

PrimeView/PrimeVerify

These two graphical user interface applications were designed to allow end users to zone and send images to the Job Server and verify the PrimeOCR results. See the PrimeView/PrimeVerify data sheets for more information.

Software Developers Kit (SDK)

The SDK consists of 40 simple, orthogonal API calls accessible as a Dynamic Link Library by C /C++, Visual Basic or any language capable of accessing a DLL such as PowerBuilder.

The SDK also includes complete documentation and working source code examples in C and VB.

Recognition Data Types

PrimeOCR recognizes the following data types:

Characters

Omnifont machine print and dot matrix text in any of the following 11 languages:

- Danish
- Dutch
- English (US or UK)
- French
- German
- Italian
- Norwegian
- Portuguese
- Spanish
- Swedish

Optical Marks (OMR)

When an area on the image is "zoned" as OMR, PrimeOCR will return the percentage of black space contained within the zone. This percentage can be used to determine whether a user has marked a selection on the page.

Graphics

PrimeOCR normally ignores any graphics (e.g., pictures) found on an image. It can instead be instructed to save the graphic to a file. A path to the graphic is added to the text output for later page reconstruction.

Image Input

PrimeOCR will read images from either file or memory in the following formats:

- Tiff (single or multi-page) as uncompressed, G3 (FAX), G3, or G4 (black & white, color, or gray scale)
- PCX
- JPEG

Valid resolutions include 200, 240, 300, 400, and 600 DPI as well as Standard or Fine Mode FAX.

IMAGING MAGAZINE PRODUCT OF THE YEAR

"The voting approach implemented by Prime Recognition is the biggest leap in the technology since the omnifont recognition capability developed in the 1980's." **Dave Abbott, RTIS-Government Services**

Pre-Processing

PrimeOCR offers a variety of ways to enhance and define your image for optimal OCR:

Image Enhancement

- Improves image quality for better OCR using features such as Auto-Rotation, Crop, Deskew, Despeckle, Image Registration, Inverse Type, Line Removal, etc.

Image Zoning

- Manual Zoning
- Auto-Zoning
- Zone Content Restrictions include: None, Alphabetic, Alphabetic Upper/Lower Case, Numeric, Graphic, and OMR.

OCR Processing

PrimeOCR has several features that improve OCR accuracy, fault tolerance, and speed:

Configurable Accuracy

The base PrimeOCR configuration achieves 65% fewer errors than conventional OCR using a "3 engine" voting configuration. Even greater accuracy can be achieved through the following:

- **Lexical Plus** - Improves accuracy by up to 30% on documents that are high in language content (e.g., magazines, books, legal documents).
- **4, 5, or 6 Engines** - Add a 4th OCR engine to the base configuration for 75% fewer errors, a 5th engine for 80% fewer errors, or a 6th engine for 82% fewer errors.
- **Character Training** - PrimeOCR can be trained to recognize specific character sets or fonts.
- **Engine Customization** - Users may select which engines participate in the recognition process or even weigh engine results differently.

High Fault Tolerance

- **Automatic Engine Recovery** - A poor quality image can cause a conventional OCR product to "crash". To solve this problem, PrimeOCR can sense when an engine fails and automatically reinitialize it for the next image. This increases throughput by allowing PrimeOCR to run unattended, 24 hours a day!

Configurable Speed

- **Multi-Processor Support** - PrimeOCR can utilize up to six processors in a multi-CPU system for faster throughput.
- **Selective Voting** - While "Voting" takes longer than conventional OCR, you can speed up the processing on high quality images through Selective Voting. The result: faster OCR speeds on high quality documents and more processing power on lower quality documents.

Output

- **ASCII** - Text only output, left justified.
- **Formatted ASCII** - Spaces are added to text to mimic the original imaging layout.
- **"PRO"** - Prime Recognition's Output format. Provides recognized characters and information on each character, including recognition confidence, font style, location on image, point size, and bold/underline/italic attributes. Image location data is also reported for lines, zones, and words.
- **PDF** - Converts scanned images into PDF "Normal", "Image + Text" or "Image Only" formatted files.
- **RTF** - Retains original character attributes and page layout using frames and paragraph conventions. Imports easily into word processing programs.
- **HTML** - Format often used to place data on a web site.
- **Comma Delimited** - Used most often in "forms" applications to export data to a database or a spreadsheet.
- **RRI3** - RRI's FormWorks compatible format.
- **ZYINDEX** - ZyLab's ZyIndex compatible format.

System Requirements

Software:

- Microsoft Windows NT 4.0 Workstation/Server
- Microsoft Windows 2000 Professional/ Server
- Microsoft Win95/98 (demo only)

Hardware:

- Intel Pentium compatible computer
- One CD drive to install software
- A hard disk with 20 MB of space for installation
- At least 64 MB of RAM (128+MB recommended)

PRIME RECOGNITION ❖ PHONE: (425) 895-0550 FAX: (425) 895-9580
WEB: WWW.PRIMERECOGNITION.COM EMAIL: SALES@PRIMERECOGNITION.COM