

PRIMEVIEW™
PRIMEVERIFY™

**User's Guide
Version 4.40**

PRIME RECOGNITION
PHONE: 425-895-0550
FAX : 425-484-6415
EMAIL: support@primerecognition.com
WEB: www.primerecognition.com
FTP: ftp.primerecognition.com

Copyright © 1994-2007 Prime Recognition. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the prior written permission of the publishers.

Printed in the United States of America.

Statement of Limited Warranty

Prime Recognition warrants to the original licensee of this program that it conforms to Prime Recognition's specifications. Should this program, in Prime Recognition's option, malfunction due to non-conformity with Prime Recognition's specifications, Prime Recognition will, at its option, repair, replace, or update the program at no charge, provided that the program has not been subjected to misuse, abuse, accident, disaster or non-Prime Recognition authorized alteration, modification, and/or repairs. In no event shall Prime Recognition be liable for incidental or consequential damages in connection with or arising out of the furnishing, performance, or use of any of these programs. Prime Recognition reserves the right to modify or revise all or part of this document without notice and shall not be responsible for any loss, cost, or damage, including consequential damage caused by reliance on these materials.

US Government Restricted Rights

The documentation and software are provided with restricted rights. Use, Duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (C)(1) (ii) of the Rights in Technical Data and Computer Software clause at DFARS 52.227-7013.

Trademarks

PrimeView, PrimeVerify, and PrimeOCR are trademarks of Prime Recognition. Other trademarks appearing in this manual are trademarks of their respective companies.

Table of Contents

| | |
|---|-------------------------------------|
| OTHER PUBLICATIONS..... | III |
| Chapter 1 Overview | |
| PRIME RECOGNITION APPLICATIONS | 1-1 |
| PRIMEVIEW | 1-1 |
| PRIMEVERIFY | 1-1 |
| Chapter 2 System Requirements | |
| HARDWARE..... | 2-1 |
| SOFTWARE | 2-1 |
| Chapter 3 Installing PrimeView & PrimeVerify | |
| INSTALL PRIMEOCR FIRST..... | 3-1 |
| INSTALLING PRIMEVIEW/PRIMEVERIFY | 3-1 |
| Chapter 4 Licensing PrimeView & PrimeVerify | |
| HARDWARE KEY VS. SOFTWARE KEY | 4-1 |
| HARDWARE KEY INSTALLATION | 4-1 |
| <i>LPTI based Key</i> | <i>Error! Bookmark not defined.</i> |
| <i>USB based Key</i> | 4-1 |
| SOFTWARE KEY INSTALLATION..... | 4-1 |
| SOFTWARE KEY HARDWARE/OS DEPENDENCIES..... | 4-2 |
| Chapter 5 PrimeView | |
| RUNNING PRIMEVIEW | 5-1 |
| OPENING AN IMAGE FILE | 5-1 |
| IMAGE DISPLAY ADJUSTMENTS | 5-2 |
| CONFIGURE IMAGE ATTRIBUTES | 5-3 |
| CREATING ZONES ON A PAGE..... | 5-4 |
| REMOVING ZONES FROM A PAGE..... | 5-4 |
| SPECIFYING ZONE SETTINGS..... | 5-5 |
| WORKING WITH TEMPLATES | 5-7 |
| CONFIGURE IMAGE PREPROCESSING FOR OCR..... | 5-8 |
| SELECT ZONING OPTIONS | 5-8 |
| SPECIFYING OUTPUT FILE FORMAT | 5-9 |
| SUBMITTING AN IMAGE FOR OCR | 5-11 |
| OTHER OPTIONS | 5-11 |
| EXITING PRIMEVIEW | 5-12 |
| Chapter 6 PrimeVerify | |
| OPENING AN OUTPUT FILE | 6-2 |
| OPENING AN OUTPUT FILE | 6-2 |
| IMAGE DISPLAY ADJUSTMENTS | 6-3 |
| TEXT DISPLAY ADJUSTMENTS | 6-4 |
| WORKING WITH CHARACTER CONFIDENCE LEVELS | 6-5 |
| TEXT AND IMAGE NAVIGATION | 6-6 |
| PAGE NAVIGATION | 6-7 |
| SEARCHING FOR TEXT..... | 6-7 |
| EDITING TEXT | 6-8 |
| SAVING THE OUTPUT FILE | 6-10 |
| EXPORTING TEXT | 6-10 |
| EXPORT OPTIONS | 6-11 |
| MULTIPLE EXPORT | 6-13 |

| | |
|---|------|
| CLOSING THE OUTPUT FILE | 6-13 |
| AUTO PROCESS | 6-13 |
| PRIMEVERIFY OPTIONS | 6-15 |
| EXITING PRIMEVERIFY | 6-15 |
| Chapter 7 Troubleshooting | |
| FILE INPUT/OUTPUT | 7-1 |
| INITIALIZATION PROBLEMS..... | 7-1 |
| LICENSING ERRORS | 7-1 |
| IMAGE DISPLAY PROBLEMS..... | 7-1 |
| Chapter 8 Programming Considerations | |
| COMMAND LINE START UP | 8-1 |
| PROCESSING A LIST OF FILES | 8-1 |
| Chapter 9 InputAccel Version | |
| OVERVIEW | 9-1 |
| SYSTEM REQUIREMENTS | 9-1 |
| CONFIGURING PRVRFIA | 9-1 |
| USING PRVRFIA | 9-1 |
| PRVRFIA SHORTCUTS..... | 9-2 |
| REMOVED COMMANDS FROM PRIMEVERIFY | 9-2 |

Contents of User's Guide

- **Chapter 1, Overview**
This chapter gives a brief overview of PrimeView and PrimeVerify. It also introduces other related Prime Recognition applications.
- **Chapter 2, System Requirements**
This chapter provides detailed descriptions of the hardware and software required to support the PrimeView/PrimeVerify applications.
- **Chapter 3, Installing PrimeView/PrimeVerify**
This chapter provides instructions on how to install PrimeView/PrimeVerify.
- **Chapter 4, Licensing PrimeView/PrimeVerify**
PrimeView/PrimeVerify require a license to be used. This chapter describes the licensing method used for PrimeView/PrimeVerify.
- **Chapter 5, PrimeView**
This chapter describes how to use the PrimeView application to submit images to PrimeOCR for Optical Character Recognition (OCR).
- **Chapter 6, PrimeVerify**
This chapter describes how to use the PrimeVerify application to review and edit the text results generated by PrimeOCR.
- **Chapter 7, Troubleshooting**
This chapter provides suggestions and highlights points that might be helpful in troubleshooting.
- **Chapter 8, Programming Considerations**
PrimeView and PrimeVerify are end user applications, however, this chapter describes some ways to access its functionality programmatically.
- **Chapter 9, InputAccel Version**
A version of PrimeVerify has been adapted to the InputAccel imaging system. This chapter describes the functionality of this version.

Other Publications

Refer to the *PrimeOCR Access Guide*, for information on how to install PrimeOCR for use with the PrimeView and PrimeVerify applications.

The "PTM" (template) and "JOB" (job) formats used by PrimeView and the "PRO" format used by PrimeVerify are defined in the *PrimeOCR Job Server* chapter of the *PrimeOCR Access Guide*.

Other publications that may be of interest include:

- *PrimeView Data Sheet*
- *PrimeVerify Data Sheet*
- *PrimeOCR Data Sheet*
- *Prime Recognition High Accuracy OCR Cost Justification*
- *Prime Recognition High Accuracy OCR "Cleaner Data" Justification*

Chapter 1 Overview

Prime Recognition Applications

PrimeOCR, PrimeView, and PrimeVerify are a set of Microsoft Windows-based products designed to assist the user in performing and verifying Optical Character Recognition (OCR). OCR is performed by PrimeOCR, a highly accurate software recognition engine that utilizes “Voting” technology to produce significantly fewer errors than the best conventional OCR products. PrimeView and PrimeVerify are end-user applications for image viewing, page zoning, character recognition and verification.

The scalable architecture of these products allows them to function seamlessly in either a stand-alone or networked imaging environment. One user can run all applications on the same PC, or each product can be loaded onto a separate PC, sharing data over the network. For example, one PC, dedicated to image preparation, can run PrimeView. A second, designated as an OCR server, can run PrimeOCR. And a third, used for OCR verification, can run PrimeVerify.

This modular design also allows multiple copies of PrimeView/PrimeVerify to access the same PrimeOCR server, maximizing your OCR engine utilization and investment.

PrimeView



PrimeView allows users to prepare image jobs for processing by PrimeOCR. OCR parameters are set through PrimeView, as well as the region(s) of the page (if not the full page) to be processed. Regions are created graphically, by drawing a rectangle around the area of the page to be recognized.

Users can save all of the OCR parameters and regions drawn into an image “Template” for subsequent use. Templates are a powerful feature, allowing the information used in preparing one image to be used by another. This is a particularly desirable feature for forms-based images and other types of images that follow the same page layout.

Image jobs submitted for OCR processing are stored in a directory or “job queue”.

PrimeVerify



Once an image is recognized by PrimeOCR, the text output can be viewed and edited through PrimeVerify. One of the best features in PrimeVerify is the ability to see both recognized text and the original image together for easy “heads-up” style editing. Characters flagged as “suspicious” by PrimeOCR are highlighted in red and a red box is also drawn around the area on the image where the suspicious character can be found.

Suspicious characters are highlighted based upon a user-defined confidence threshold (there are 9 to choose from) and users can jump between highlighted characters using the TAB key. The corresponding character on the original image is kept synchronized with the text being viewed. Error statistics for each page are also supplied. Users can make changes to any of the text and save the results in a variety of output formats.

Chapter 2 **System Requirements**

PrimeView and PrimeVerify run on Intel or compatible personal computers. The system requirements are listed below.

Hardware

- Intel PC or 100% compatible computer with Pentium, or 100% compatible CPU (e.g. AMD Athlon).
- A hard disk with 10 megabytes of available disk space for software installation.
- At least 64 MB of Random Access Memory (RAM).
- Active USB port (if using USB based hardware key) (LPT1 port if using an old style key). No ports are required if using software license.

Software

- Win2000, WinXP, or Windows2003 Server.

Chapter 3 Installing PrimeView & PrimeVerify

Install PrimeOCR First

PrimeOCR should be installed prior to installing PrimeView or PrimeVerify.

When deciding where to install PrimeOCR, keep in mind that it is designed to be an OCR server. OCR “jobs” can be submitted to the PrimeOCR engine from PrimeView running on the same PC or running on another PC connected to the PrimeOCR PC via a local area network (LAN). Multiple LAN-based PC’s, each running a copy of PrimeView, can submit jobs to the same PrimeOCR server.

It is recommended that PrimeOCR be installed on a PC dedicated to OCR processing. OCR processing speed will decrease if other applications are running on the same system.

Installing PrimeView/PrimeVerify



You must have administrator rights to install the software. You will also need administrator rights if you have a software license with a page limit.



Run SETUP.EXE from the main directory on installation CD and follow the directions presented. The installation program will handle all installation issues automatically, including decompressing all files into a new or existing directory selected by the user.

A program group entitled “Prime Recognition” will be created by the installation program to include icons for the applications - PrimeView and PrimeVerify, and an icon for the PrimeVerify InputAccel module.

Chapter 4 **Licensing PrimeView & PrimeVerify**

PrimeView and PrimeVerify require a license to run. This license may be either a hardware key or a software license.

Hardware Key vs. Software Key

PrimeView/Verify can utilize a hardware key or software key license. The advantage of a hardware key is that it can be easily moved to a different machine. Its disadvantages include the need for a working USB port, and keeping track of the hardware key itself.

A software license cannot be moved to another machine, it is tied to the specific machine and that machine's OS/hardware configuration (see details below). It also requires that the user have administrator rights if a page limit exists on the license (e.g. evaluation license). However, if the user does not have a working USB port, or does not want to work with hardware keys, then this is another choice.

Hardware Key Installation

USB based Key

After installing the software, install the key into an active USB port on the PC. The installation process should automatically load the hardware key driver, but if you must reload the driver for any reason you can go to the \Sent directory and run SETUP.EXE. The program will periodically read the hardware key so the key must be present during all engine processing.

If you are receiving hardware key errors:

- Make sure the key is securely attached to the USB port.
- Verify that USB port is active.
- Note that only the hardware key supplied with the product will pass the product's tests.

Software Key Installation

If a software key has never been installed on the target machine, and the user tries to run any other Prime Recognition product, a temporary software license will automatically be created.

Warning

The file name of the PRLICENSE.LOG file can be changed. If you are updating multiple software licenses at one time, make sure to name the log file something meaningful that will allow you to associate that file name to the specific PC it came from.

A permanent software license can be created (or a temporary software license extended) by running ...\\PRPROOF\\PRLICENSE.EXE. This will create the file ...\\PRPROOF\\PRLICENSE.LOG. Send this log file to your sales contact by email.

Your sales contact will return a program(s) or instructions on how to upgrade the license. The name of the program may be SoftwareKeyWrite.EXE or a name that corresponds to the name of the log file you sent (particularly if you sent in multiple log files). Place this program in your ...\\PRPROOF directory and run. You should get several dialog boxes indicating the progress and successful conclusion of the software license creation.

Software Key Hardware/OS Dependencies**Warning (for Software License Only)**

You can change either your CPU type or LAN adapter (but not both). You can change or reinstall your OS installation and you can change the hard drive on which Prime software was .

The software license is specific to a particular machine. It cannot be moved. The license is also tied to two key items: the operating system installation, and a combination of the CPU type, hard drives, and LAN adapter ID. This means that you can change anything else on the machine without affecting your license.

Chapter 5 PrimeView

Running PrimeView



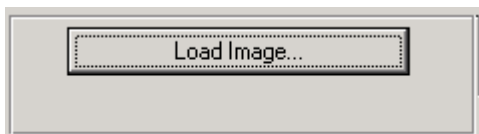
Open the Prime Recognition program group and double-click on the icon for PrimeView. You will be presented with a series of OCR controls on the left-hand side of the screen and an image display area on the right used to view the image you will be working with. To submit an image to the PrimeOCR engine, you will typically perform the steps listed below:

- **Open an Image File** - Choose and display the image you are submitting to PrimeOCR.
- **Make Image Display Adjustments** (optional) - Change how the image is displayed for easier viewing.
- **Configure Image Attributes** (optional) - Help PrimeOCR perform the best recognition possible by selecting the type of image you are working with.
- **Create Zones** - Choose one or more regions of the image to OCR and describe how PrimeOCR should recognize text within each.
- **Load or Create an Image Template** (optional) - Save all of the OCR and Image settings you have made to a Template file for use with other images.
- **Configure Image Preprocessing for OCR** (optional) - Help PrimeOCR perform the best recognition possible by preprocessing the image before OCR (for example, deskew).
- **Choose an Output File Format** - Select a format for PrimeOCR to store all recognized text in.
- **Submit the Image for OCR** (optional) - Create an entry for this image in the PrimeOCR Job Queue.
- **Select Zoning Options** (optional) - Settings that will make zoning faster or more convenient.

Many of the steps listed above are optional. If you do not need to change the Image and OCR between images, you can quickly submit new images to PrimeOCR using only a few keystrokes.

The remainder of this chapter describes each of the above steps in greater detail.

Opening an Image File



If no image file is loaded, the Path/File “Name” will say **(none)**. Press the Enter Key, click the “Load Image” button or choose “Load Image” from the VIEW menu to bring up the Open File dialog box. From here you can scroll through the directories on your system to find the image file to open.

PrimeView supports a variety of image file formats including TIFF, and PCX. TIFF files can be uncompressed, Group 3, Group 3 Fax,

or Group 4, multipage TIFF, including bitonal, color, and grayscale images. Image resolutions supported include 200, 240, 300, 400 and 600 DPI as well as standard Fax. To search using file types supported by PrimeView, use the "List Files of Type" drop-down list on the Open File dialog box.

Select the "OK" button to open the file you have chosen. The blank page on the right will now show the open image and the Image File Name and Path on the left will reflect the file you have opened.

Load Next\Previous Image - This menu command will automatically load the next\previous image file in the directory.

Move to ExceptDirX - This menu command will move the TIF file, along with any associated PTM, PRO, and TXT file to the **ExceptDirX** subdirectory (creating subdirectory if it does not already exist). Intended to be used in conjunction with "View/Load Next/Previous Image" commands. Useful to quickly QA images/templates and "reject" "bad" image/templates for later exception processing.

Multiple Page TIFF Files - If the image opened is a multiple page TIFF file then a "Page #" entry box is shown along a page scroll bar and "Next Page/Previous Page" menu commands. Any of these commands may be used to move between pages in a multipage tiff file. For example, if you type "6" into the "Page #" entry box PrimeView will display page 6 of the image file.

Image Display Adjustments

The initial display settings for an image may make viewing the image difficult.

Maximizing the PrimeView Window - Once an image is opened, it is shown in the image display area. To enlarge the image display area for better viewing, use the Maximize button on the upper right corner of the PrimeView window. This will expand the window to fit your entire screen.

Zooming To Full Screen - To have the image fill the entire width of the image display area, use the Zoom To Full Screen command from the VIEW menu. If it is still difficult to view individual characters or words in this mode, you will want to enlarge the image using Prime View's Zoom In feature.

Zooming In and Out on an Image - The Zoom In command on the VIEW menu will enlarge the image in predefined increments. Use the Zoom Out command on the VIEW menu to shrink the image by the same increments. At any time you can return to full width viewing by selecting Zoom to Full Screen from the VIEW menu.

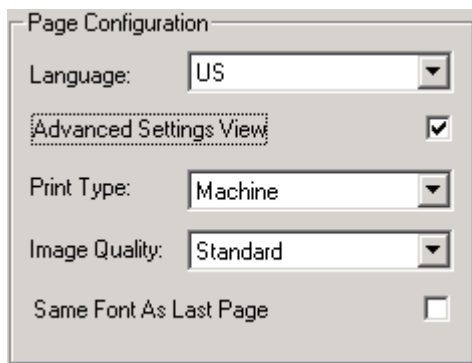
Rotate Image 90 degrees to Right – Rotates image in viewer. User should rotate image until text is upright before drawing zones.

Save Rotated Image as G4 TIFF - If selected then image, if it has been rotated, will be saved with new orientation when next image is loaded or application quits.

Use the other commands on the VIEW menu to improve the overall quality of the image display:

- **Invert** - This command will invert all pixels in the image, changing all whites to blacks, and vice versa. Any subsequent image opened will also be inverted as long as this command is checked.
- **Scale To Grey** - Scale-To-Grey is an image display enhancement feature that is extremely helpful in improving the contrast and readability of poor quality images. It is also important when you are applying high scaling factors, such as when you are displaying images on a low resolution monitor or are displaying thumbnail images. As with the Invert command above, Scale-To-Grey will apply to any subsequent image opened, as long as this command is checked.

Configure Page Attributes



To help PrimeOCR perform the best recognition possible, you will need to specify the correct Page settings for the open image.

OCR accuracy can vary greatly depending upon how these values are set, so you may want to experiment with different settings to see which is best for the type of image(s) you are working with.

Use the definitions below to determine your document's image quality:

- **Language** - This setting is only relevant when the optional "international language" version of PrimeOCR is installed. Choose the language of the image. Choices include English and 10 other western European languages.
- **Machine Print** - Tells PrimeOCR that the characters on the page were formed by a typewriter, laser printer, typeset, etc.
- **Dot Matrix** - Characters were formed by a dot matrix printer with spaces between the dots. If dots overlap, use the Machine Print setting.
- **Standard** - Choose this setting for the majority of images.
- **Poor** - Choose this setting to boost PrimeOCR speed and accuracy on very low quality images such as standard mode Fax.
- **Same Font As Last Page** - This setting tells PrimeOCR that the current image page is similar to the last page recognized in terms of character font used. Slight accuracy improvements can be achieved because PrimeOCR has the ability to "learn" between pages.

Creating Zones on a Page

You can select only a portion of the page to OCR by drawing a Zone around it, or you can OCR the whole page by defining the whole page as a Zone.

Zones Defined - A Zone is a frame or box drawn on an image and the section of the image the box encloses. Multiple zones can be drawn and each zone can have specific settings, as discussed in the section entitled Specifying Zone Settings.

Drawing Zones - To draw a zone around a region on the image to be OCR'd, position the mouse in the top left corner of the region, then press and hold the left mouse button. Drag the mouse down and to the right, until you have covered all of the text to be OCR'd. Release the mouse button. A red box outlines the region you have drawn and this new zone is assigned a number, which is displayed in the Current Zone control on the left.

Make sure that all of the text is completely inside of this box. If the box cuts across any of the text, OCR accuracy will be affected. Also try to avoid having the same text appear in more than one zone (overlapping zones), as this will affect OCR accuracy. Boxes can be cleared and redrawn using commands in the EDIT menu as discussed in the section entitled Removing Zones from a Page.

Repeat this procedure for all of the text regions on the page that you want to OCR. The Current Zone settings will change to reflect the values set for each zone. You can change how PrimeOCR recognizes text in the current zone by modifying these settings.

Removing Zones from a Page

The commands on the EDIT menu allow you to remove or redraw zones on the image page as follows:

- **Clear Zone** - This command will remove the current zone, which is highlighted with a red border. If no zone is current (i.e., all zones have a blue border), click anywhere inside of the desired zone with the mouse to highlight it and choose Clear Zone from the EDIT menu. This will remove the zone and renumber all zones above it.
- **Clear All Zones** - Use this command to remove all zones on the page. Selecting Clear All Zones from the EDIT menu will delete all zones drawn and make the full page the current zone.

Specifying Zone Settings

Current Zone

Number:

Content Restriction:

Lexical Check:

Accuracy Level:

Resize Increment: (1 - 100)

Zone Text:

Each time a new Zone is drawn, it receives a unique number and becomes the “Current Zone” as denoted by its red border. Zones that are not the current zone are outlined in blue. To make any one of these zones the current zone, simply click anywhere inside of the zone with the mouse. The border of the selected zone will change to red and its zone number and settings will be displayed in the Current Zone.

If there is no current zone, the setting under Current Zone are grayed and the words “Number: (none)” appear.

The settings specified in the Current Zone control affect the way PrimeOCR recognizes text in the current zone. Changes to the settings under this control will affect only the current zone. The settings available for each zone are:

Number: - You can change the order of zones by typing in a new zone number for the selected zone and pushing update button.

Keyboard Command

You can quickly select a content restriction via a keyboard command: ALT + #, where # is a number between 1-9.

Order of Listings

You can change the order of the Content Restriction listings through the creation of a PRZNCONT.INI file. Open the PRZNCONT.INI file in the program directory (e.g. \PRPROOF) with a text editor and see the instructions at the end of the file.

Content Restriction - This feature strongly biases recognition to a particular set of possible characters, which can significantly improve OCR accuracy. (For strict enforcement of content restriction see "Enable Lexical Checking" below). The options are:

- **None** - No restrictions.
- **Alphabetic** -Valid characters include: a-z A-Z , . and international characters.
- **Numeric** -Valid characters include: 0-9 and special characters: , . + - () \$ / @ £.
- **Uppercase** -"No restrictions" but with no lowercase characters.
- **Lowercase** -"No restrictions" but with no uppercase characters.
- **AlphaUppercase** -"Alphabetic" but with no lowercase characters.
- **AlphaLowercase** "Alphabetic" but with no uppercase characters.
- **OMR** -"Optical Mark Recognition", in other words define a zone in which marks occur. PrimeOCR will return the percentage of black pixels within this zone area.
- **Image** -PrimeOCR will save this zone as a separate Group 4 TIFF image in the image directory, with the file name "XXXX.YYY", where "XXXX" is the original file name, and "YYY" is the cumulative number of image zones in this document. Note that YYY must be between 000 and 999.
- **Anchor** -Prime Recognition offers certain special zoning tools, which require the definition of an anchor point.
- **Bar Code** -Can be used for 1D or 2D barcodes supported by PrimeOCR.

Lexical Check - This option tells PrimeOCR to perform a series of checks based upon zone content. If zone content is **None** or **Alphabetic**, and this setting is **Standard** then spelling and tri-gram analysis is performed on text in zone, and content restrictions, as noted above, are strictly enforced. Any characters recognized that are not in these groups are switched to "?" and confidence level "1". If this setting is increased to "Lexical Plus X", then an additional lexical step is performed. The definition of this extra processing is controlled by PRLEXICA.INI in the PrimeOCR engine. We encourage you to read *Chapter 4* of the *PrimeOCR Access Guide* for more information on Lexical Plus.

Accuracy Improvement - Through its voting technology, PrimeOCR can provide up to 6 levels of accuracy improvement. The level number in this dropdown list represents the number of OCR engines PrimeOCR will use to recognize text. Every additional engine increases accuracy but also slows down the recognition process.

For most zones, we recommend using the highest accuracy level available. Selected zones may not be as accuracy sensitive and can use a lower accuracy level to increase speed.

Approximate accuracy and speed rates for each level are:

- **Level 1** - provides conventional OCR accuracy and speed.
- **Level 2** - Provides 15% better OCR accuracy at roughly 2.1 times slower speed than conventional OCR.
- **Level 3** - Provides 58% better OCR accuracy at roughly 3.1 times slower speed than conventional OCR.
- **Level 4** - Provides 74% better OCR accuracy at roughly 5.8 times slower speed than conventional OCR.
- **Level 5** - Provides 76+% better OCR accuracy at roughly 7.1 times slower speed than conventional OCR.
- **Level 6** - Provides 80+% better OCR accuracy at roughly 8.0 times slower speed than conventional OCR.
- **Level 7** - Provides 84+% better OCR accuracy at roughly 11.1 times slower speed than conventional OCR.

When you submit an image for OCR, all of the zone information - the number, location and settings for each zone, is sent to PrimeOCR for processing. If you want to use this information on a subsequent image, see the Working with Templates section below.

Resize Zone - You can change the shape of the current zone by clicking on the small scroll bars around the "Resize" box. If the zone boundary moves out of the visible area the image will automatically scroll to keep this edge of the zone in view. If you want to move the zone boundary in larger steps then increase the "Resize Increment" setting.

Zone Text - Allows you to enter a string of up to 100 characters for each zone. This text will be written to the zone information line in an ACC style template file. This text may be used to add alternative

text for graphics in the production of PDF files that meet Section 508 Accessibility requirements. (This Zone Text field is only visible if menu item Options/Zone Text is selected).

Working with Templates

PrimeView allows you to save all of the zone information you supplied for an image into a file called a “Template”. PrimeOCR uses this Template file when processing an image, to determine what portion of the image to OCR and how. If you do not use the commands on the FILE menu to create your own template, one will be created for you when you choose any “Submit...” function from the OCR menu. The name of this file will combine the name of the open image file with a “.PTM” extension. If a Template file with this name already exists, you can overwrite it or choose another Template file name. If you cancel the Template File Save, no OCR will be performed. Otherwise, the image and its corresponding template will be submitted for OCR as described in the section entitled Performing OCR.

Save Template As - To manually save zoning information, choose Save Template As from the FILE menu. PrimeView will bring up the Save File As dialog box and provide a default name for the template file. You can change the name of the file as desired and, if no extension is supplied, PrimeView will automatically add the “.PTM” template file extension for you. After choosing OK to accept the Template file name, PrimeView will save all of the zone information, including the number of zones, their location and all settings for each zone, to this file. The PrimeView title bar will now show the name of the template just saved.

Save Template - If Save Template is chosen for a template that is already open, PrimeView will automatically save the latest zone changes to that file, otherwise it will bring up the Save Template As dialog box.

New Template - If you wish to remove the current Template or start with a blank Template, choose New Template from the FILE menu. If changes have been made to the current template, PrimeView will ask you to save the template before closing. Once closed, the Template name will be removed from the PrimeView title bar.

Open Template - To apply a saved Zone Template to another image, choose Open Template from the FILE menu. The Open File dialog box will appear, allowing you to scroll through the directories on your system to find the Template file to open. To search using the Template file type supported by PrimeView, use the “List Files of Type” drop-down list on the Open File dialog box. Select the “OK” button to open the file you have chosen. The PrimeView title bar will now show the name of the open Template and all zone information contained in the template file will be loaded.

Configure Image Preprocessing for OCR

Warning

Only select image processing steps that you know will help the image. Many image processing steps can harm the image if they are not relevant to the image, they are improperly configured, etc.

PrimeOCR offers optional functionality which preprocesses images before OCR, with powerful algorithms, for example, deskew, line removal, registration, and so on. PrimeOCR uses ScanFix from TMSSequoia to perform these steps along with its own algorithms. The functionality of each function is exactly that defined by ScanFix, therefore you may consult ScanFix documentation or review the PRIMAGE.INI documentation in *Chapter 3* of the *PrimeOCR Access Guide*.

Notes:

- You may select multiple processing steps at one time.
- Selecting a function in PrimeView merely turns "on" the function. The configuration of the preprocessing function is controlled by the PRIMAGE.INI file of PrimeOCR.
- For most functions you will want to experiment with the preprocessing settings to select the best combination of functions and respective configuration settings.

Strong AutoRotate – This is not a ScanFix feature. It was developed by Prime Recognition. It is slower but much more accurate than the ScanFix version. We strongly recommend you use this version of AutoRotate as well as the regular version of AutoRotate (both version compliment each other).

Save Processed Image – This will save a copy of the image used for recognition to the same directory and file name as the original image except with a “.fix” extension. However, the file will only be saved if deskew was turned on, and/or if any image processing was turned on, and/or Strong Auto Rotate was turned on and the best orientation was different from the original image. (If none of these conditions were true then we would just be saving the original image file).

Select Zoning Options

PrimeView offers several processing options that may speed your zoning effort. These options should typically be set at the beginning of the zoning session. The processing options are mutually exclusive, if you select one option it will automatically deselect any option previously set.

AutoZone – Text, Top to Bottom – Automatic zoning will be performed by the PrimeOCR engine. Only text zones will be identified (not image zones). The zones will be ordered from top of the page to the bottom. If no zones are identified a full page zone will be automatically created. If multiple zones are defined all zones except for the first will be ignored. On the first zone the coordinates will be ignored (auto zoning is always performed on the full page), however all the other zone settings will be used on all zones automatically found.

AutoZone – Text, Reading Order – Same as above except zones will be reported in the “reading order”. (This is hard to do on complex documents hence may be prone to error.)

AutoZone – Text + Image, Top to Bottom - Same as first AutoZone setting except images are now reported. (Image based data is only saved in PRO, RTF, and PDF Normal output. There is no advantage to selecting this with any other type of output format).

AutoZone – Text + Image, Reading - Same as above except zones are now reported in reading order. (This is hard to do on complex documents hence may be prone to error.)

Full Page Zone All Pages - Automatically creates a full page zone for all pages in the image using the existing settings. The user may not modify the zone size under this option, but may modify settings such as accuracy level, etc. Note that only the configuration settings active at the time of the template file "save" action will be used, intermediate changes, even when performed on different pages, are ignored, all pages receive the full page/current settings configuration. This option is most useful when you want to quickly set up a template which OCRs all pages within a file using the same full page settings.

Full Page Zone Default - Automatically creates a full page zone for all pages in the image which the user views (images not viewed by user are assumed to not require OCR). The net effect of this option is to automatically bring up a full page zone when a page is first loaded. In every other way PrimeView acts normally.

Autoload Template - If a template exists for the newly loaded image in the current template directory (which may not be the directory of the image) the template will automatically be loaded. This option is useful when using PrimeView to review/QA zones created by Prime Recognition's automatic zoning tools.

Retain Template between Images - Keeps existing template on screen and active even if image is changed. This option is useful to define and test whether a single template will work with multiple images.

Specifying Output File Format

Before submitting the image for OCR, you must decide what type of file format the recognized text should be stored in. Select one of the following formats from the OCR menu:

- **ASCII** - OCR results are stored as a straight, unformatted ASCII (American Standard Code for Information Interchange) text file. If zones were defined they are separated in the output by two blank lines.

- **FASCII** - Text output is padded with spaces to mirror the layout of the text on the original image. If zones were defined they are separated in the output by two blank lines.
- **PDA** - A file format that is defined in the WordScan 5.2 Development Kit. PrimeOCR's PDA format is Level 2 style. This format should only be of interest to developers who have already interfaced to the WordScan product and want to use their existing output parsing code for PrimeOCR output.
- **PrimeOCR** - This is an internal PrimeOCR format ("PRO"). In addition to recognized text, it stores line, word, and character coordinated as well as character confidence levels. Select this format if you are going to use PrimeVerify to edit OCR results.
- **ASCII No Zone Buffer** - "ASCII" as defined above except with no blank lines between zones.
- **RRI3** - Output format defined by Recognition Research Inc. Version 3.0.
- **FASCIITAG** - "FASCII" as defined above except with a tag line at bottom of page, which includes any string, added to the image file name line in job file.
- **ZyIndex** - "FASCII" as defined above except with a tag line at bottom of page, which includes image file name, time, and date of image file creation. (ZyIndex format version 3).
- **RTF** - Format defined by Microsoft, version 1.X. Used by many word processing programs. Includes links to image zones (if defined). Image zones may only be displayed in Word 7.0 or later and compatible word processors.
- **PDF** - Format defined by Adobe. Three versions of PDF are supported:
 - **Normal**. Includes formatted text plus image zones (if any) plus image "snippets" for low confidence words.
 - **Image Only**. Only image is included in output file, no text is included (if this format is of interest PrimeZone is a much direct way of producing Image Only output.)
 - **Image plus Hidden text**. Image is displayed with text available in the background for searching, cut & paste, etc.
- **Comma Delimited** - An ASCII text file. The contents of each zone are reported as plain ASCII with a comma between each zone's output. This format will typically be used by "forms" type applications, or reporting data from a table from a "Green bar" application.
- **HTML** - Format defined by Internet standards. Defines a format that can be read natively by browsers. Format and character attributes such as font type, size, bold/italic/underline, etc. are maintained. Useful, for example, if you wish to have OCR'd data published directly to a web site.

Submitting an Image for OCR

Once you have supplied all of the information required to process the open image - specified image attributes, selected text on the page to OCR, made any zone specific settings and determined an Output file format, you are ready to submit the image to the PrimeOCR engine for processing.

Job Queue Setup. The first step is to setup the directory to which you will be writing “job” files. PrimeOCR uses two job queues to manage image recognition. PrimeOCR looks for new images to process in these queues.

Select the Job Queue Setup menu item and double-click on the desired directory to select it. The full path to this directory will display under the word “Directories”. The directory that you choose here should be the same directory that PrimeOCR reads when looking for new jobs. See the PrimeOCR Guide for more information on setting the Job Queue Directory for PrimeOCR.

Submit Job. This menu item causes PrimeView to create a “job file” for your image and places that file in the specified job directory. The job file contains the location of both the image you have submitted and its corresponding Template. PrimeView creates the job file using the name of the open image with the addition of a “.JOB” extension. If a job file with this name already exists, you can overwrite it or choose another job file name.

The open image is closed after submission for OCR and the page once again becomes blank.

If you have chosen the “Retain Template between Images” Option then the active Template remains open for use with another image. Otherwise the template is erased.

Submit Job As. This option lets you select a new directory and/or change the file name of the job file.

Submit Job and Load Next. After job is submitted it automatically does a “F3” type action, loading the next image in the directory.

Other Options

PrimeView offers an option that can further speed your processing:

Template Put In Image Path (Typical). If selected, this will put the template file in the same directory as the image. If not selected then the template will be placed in the path that was created by the user in a “Save As” action.

Auto F3 if F12 on last page of document. If you are processing multiple page documents and you reach the last page of the document this will automatically load the next document.

Show Zone Text. Displays Zone Text area in Current Zone part of User Interface.

Accessible Style Templates (.acc) (Typically used to produce Section 508 compatible/Accessible PDF Files.) Templates will be produced with acc extension. Content Restriction choices will be Text, Figure, Table Row, and Table Column. These Content Restriction labels will be added to the PTM on each zones' description, along with any zone text entered by the user. (This extra data is used by PrimeOCR in producing the Section 508 compatible PDF files).

Exiting PrimeView

You can terminate the PrimeView application by choosing Exit from the FILE menu. If you have created one or more zones on the image page, PrimeView will allow you to save the zone information to a Template file. Any changes you have made to the default Image Attributes, Current Zone or Output File control settings, as well as changes made to any "checkable" menu items, will be saved automatically upon exit and loaded as defaults the next time you run PrimeView.

Chapter 6

PrimeVerify



Open the Prime Recognition program group and double-click on the icon for PrimeVerify. You will be presented with an application window that has a split screen view. The top section, called the image display area, will show the original image used by PrimeOCR to generate the Output file. The bottom section, called the edit display area, will contain the recognized text in editable form.

The dual screen view allows users to quickly compare recognized text with the text as it appears on the original image. This, combined with the PrimeVerify proofreading features described below, makes for efficient and comprehensive manual error correction.

To proofread PrimeOCR results, you will typically perform the following steps:

- **Open an Output File** - Choose and display both the image and OCR results from a PrimeOCR job.
- **Make Image Display Adjustments** (optional) - Change how the image is displayed for easier viewing.
- **Make Text Display Adjustments** (optional) - Change the font size of text in the edit display for easier proofreading.
- **Highlight Low Confidence Characters** (optional) - Low Confidence characters, marked by PrimeOCR, are highlighted in red using a user-defined confidence threshold.
- **Text and Image Navigation** (optional) - Move through the text using keystrokes or mouse clicks to locate the characters you wish to edit.
- **Edit Text** (optional) - Select and replace characters, words, or whole lines using keyboard commands.
- **Save the Output File** (optional) - Store intermediate editing results for later completion.
- **Export Text** (optional) - When you are done proofreading the OCR output, you can export the recognized text to variety of output formats such as ASCII or RTF.
- **Configure Options** (optional) - PrimeVerify includes several processing options that can speed verification.

The remainder of this chapter describes the proofreading process in greater detail.

Opening an Output File

PrimeVerify requires an Output file that is in the PRO format. Most commonly these files have a ".PRO" extension but that is not required. The PRO file type is generated by PrimeOCR and contains all of the recognized text from the image submitted by PrimeView, as well as character confidence values and location information. PrimeOCR will only generate a PRO output file if the Output File Type within PrimeView for this image was set to "PrimeOCR Output". Choosing another Output File Type assumes that the user wants to bypass proofreading and output the recognized text directly to one of the supported export formats.

Locating an Output File - PrimeOCR deposits all Output files into the directory specified by the PrimeOCR Job Server Setup screen. Refer to the *PrimeOCR Access Guide* for more information on running PrimeOCR Job Server, or check with your systems administrator for the location of all PrimeOCR output.

Open File dialog box - Choose Open PRO File from the FILE menu to bring up the Open File dialog box. From here you can scroll through the directories on your system to find the Output file to open. To search using the file type supported by PrimeOCR (e.g., ".PRO"), use the "List Files of Type" drop-down list on the Open File dialog box.

Select the "OK" button to open the file you have chosen. The image display area on the top portion of your screen will now show the image associated with this Output and the edit display area on the bottom portion of the screen will show the recognized text.

If the image associated with this Output file cannot be found, you will be asked to locate the Image file.

Locating an Image File - The location of the image file is set when the image is first opened in PrimeView. The path to the image is added to the Job file PrimeOCR processes and also included in the Output file read by PrimeVerify. If the image file has been moved to another location, PrimeVerify will not be able to find the image based upon the original path.

Image Location

PrimeVerify will always look for the original image using the path specified in the Output file. Only after failing to find the image using this path will it use any Default Image Directory that you specify.

There are two opportunities within PrimeVerify to specify the new image file location. The first comes when you initially try to open an Output file where the associated image path is incorrect. A warning message to that affect appears and then an Open File dialog box is presented, allowing you to specify the new image file location. After specifying the new location, PrimeVerify will ask if you want to use the directory for this image as the Default Image Directory. If you accept, PrimeVerify will look for images associated with subsequent Output files in this directory, if it cannot find an image using the Output file's image path. The second way to assign a Default Image Directory is to use the Directory Setup command from the IMAGE menu.

Image Directory Setup - Choosing Directory Setup from the IMAGE menu brings up the Select Directory dialog box. Double-click on the desired directory to select it. The full path to this directory will display under the word "Directories". PrimeVerify will look for images associated with subsequent Output files in this directory if it cannot find an image using the Output file's image path.

Now that you have opened an Output file and its associated image, you are ready to make image and text display adjustments in preparation for the proofreading process.

Image Display Adjustments

The initial display settings for an image may make viewing the image difficult.

Maximizing the PrimeVerify Window - Once an image is opened, it is shown in the image display area. To enlarge the image display area for better viewing, use the Maximize button on the upper right corner of the PrimeVerify window. This will expand the window to fit your entire screen.

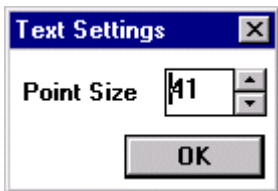
Zoom To Full Screen - To have the image fill the entire width of the image display area, use the Zoom To Full Screen command from the IMAGE menu. If it is still difficult to view individual characters or words in this mode, you will want to enlarge the image using PrimeVerify's Zoom In feature.

Zooming In and Out on an Image - The Zoom In command on the IMAGE menu will enlarge the image in predefined increments. Use the Zoom Out command on the IMAGE menu to shrink the image by the same increments. At any time you can return to full width viewing by selecting Zoom to Full Screen.

Other Image Display Adjustments - there are two other commands on the IMAGE menu that can improve the overall quality of the image display:

- **Invert** - This command will invert all pixels in the image, changing all whites to blacks, and vice versa. Any subsequent image opened will also be inverted as long as this command is checked.
- **Scale To Grey** - Scale-To-Grey is an image display enhancement feature that is extremely helpful in improving the contrast and readability of poor quality images. It is also important when you are applying high scaling factors, such as when you are displaying images on a low resolution monitor or are displaying thumbnail images. As with the Invert command above, Scale-To-Grey will apply to any subsequent image opened, as long as this command is checked.

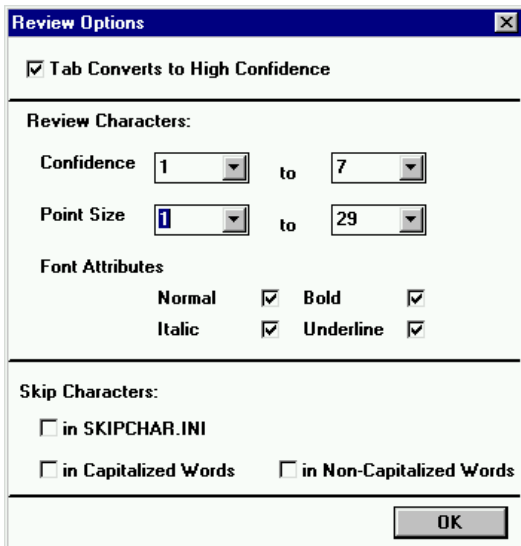
Text Display Adjustments



Black Boxes - Suspicious characters are surrounded by a black box instead of a red box. This option is helpful on monochrome screens.

Size - Similar to Zoom In and Zoom Out for the image display area, the size of text shown in the edit display area can be changed using the Text Settings command from the Main Menu. Selecting this command brings up the Text Settings dialog box where you can change the font size of displayed text for easier proofreading.

Review Options:



Tab converts to high confidence - If the user presses the TAB key then the last character confidence is set at “9” or high confidence, indicating that the character was reviewed and found to be correct.

Review Characters – Select criteria to include characters that will be reviewed. Note that all criteria will impact characters to be displayed, so that if, for example, no Font Attributes are selected, then no characters will be displayed as “suspicious”, even if meets all other suspicious parameters (like low confidence).

Confidence Level Marking - The display will highlight, in red, all characters that are within the specified confidence range (so long as they meet all other criteria as well). Since confidence level 9 is the highest ranking, selecting this level will highlight all characters in the recognized text. Most users will select a range of 1 to 6, or 1 to 7.

Point Size – Most users should select a range of 1 to 39 (all characters), but if you wish to review only a select range, you may do so with this control.

Font Attributes – Most users should select all choices here, but if you wish to only review italic characters, for example, then deselect all attributes except Italic. Note that for the purposes of this display, “Normal” is defined as a character that is not Bold, or Italic, or Underline.

Skip Characters – Once characters have been chosen to be reviewed, you can also then skip some of these chosen characters using the following parameters:

Skip characters in SKIPCHAR.INI – Select check box, and add the target characters, one per line, to the SKIPCHAR.INI file in your installation directory.

Capital/Non Capital Words – Any character in a word that is capitalized or not capitalized, may be skipped. This lets you, for example, focus your verification on proper names, or the opposite, ignore proper names.

Working with Character Confidence Levels

PrimeOCR uses a character ranking system to help identify text that may not have been interpreted correctly. Characters with a low ranking were more difficult for the OCR engine to identify and are therefore more likely to be incorrect. Similarly, characters that were easy to identify have a high ranking and a higher probability of matching the characters on the original image.

The ranking system used by PrimeOCR is divided into 9 “Confidence Levels”, with 9 being the highest ranking and 1 being the lowest. You can highlight characters in the edit display area based upon their confidence level ranking through the Text\Review Options\Confidence menu command.

Image Statistics - Select the Image Statistics command from the IMAGE menu to see confidence level ranking for all characters in the current file. The meaning of each field is described below:

| Image Statistics | | |
|---------------------------|-------------------|--------------|
| Confidence Level | Characters Marked | Cumulative % |
| 1 | 1 | 0 |
| 2 | 0 | 0 |
| 3 | 0 | 0 |
| 4 | 0 | 0 |
| 5 | 1 | 0 |
| 6 | 6 | 1 |
| 7 | 0 | 1 |
| 8 | 8 | 1 |
| 9 | 1522 | 100 |
| Total | 1538 | 100 |
| Average Confidence | | 8.97 |

- Confidence Level - Shows one of the 9 ranks that a character can have. The ranking is assigned by PrimeOCR.
- Characters Marked - Contains the number of characters in the current Output file that have been assigned the confidence level indicated at the left. Using the image statistics chart shown, if you were to set the Maximum confidence level marked field to “1”, then only 1 character in the entire document would be highlighted in red. If the maximum level was “6”, then a total of eight characters would appear in red (one from level 1, one from level 5 and six from level 6).
- Cumulative % - Tells you the percentage of characters marked up to the confidence level indicated at the left, out of the total number of characters in the Output file. This number is especially useful for imaging projects that only want to verify OCR output if the number of low confidence characters (say level 6 or below) exceeds a certain percentage of the total # of characters.
- Average Confidence - This value gives you an idea of how well the entire document OCR'd.

Highlighting characters not only helps to identify text for error correction, but also aids in the text navigation process. The cursor moves between highlighted character in the edit display area when the Tab key is pressed, as detailed in the next section

Text and Image Navigation

PrimeVerify has been designed for keyboard intensive proofreading and error correction. Special keys and key combinations have been set up to allow for fast location and replacement of incorrect characters, words or even full lines of text. Image and text navigation operations are described below.

Image Synchronization - One of the best features in PrimeVerify is the ability to see both recognized text and the original image together for easy “heads-up” style editing. To reduce the amount of searching required to verify text, whenever text is selected in the edit display, a corresponding box is drawn around the area on the image where the original text can be found. For easy comparison, the image display also moves to center the original text above the recognized text in the edit display area.

Regardless of the type of text navigation used, the image display will always try to synchronize the area on the image that corresponds with the current cursor position or selected text in the edit display.

Tabbing to the Next Highlighted Character - You can quickly locate and edit low confidence characters that have been highlighted by pressing the Tab key. The Tab key will only move the cursor between highlighted characters, jumping over characters that are not red. When it reaches the next low confidence character, the cursor changes into a box that is drawn around the entire character. The box means that the character is selected and edit operations (such as typing the correct character if the highlighted character is wrong) will automatically remove the character before performing the operation.

Once you have tabbed to the last highlighted character, pressing the tab key again will only remove the box drawn around the last character, indicating that you have reached the end of the proofreading process. As an alternative to moving between and selecting highlighted characters, you can move between and highlight entire words that contain low confidence characters by using Tab in combination with the Shift and Ctrl keys.

Tabbing Between Words That Contain Highlighted Characters - When you hold down the Shift key and press Tab, PrimeVerify will locate the next highlighted character and draw a box around the entire word that contains it. The entire word is now selected and the word will automatically be removed when any text editing operation is performed. This option is useful when it is faster to type over the entire word than to replace individual characters in the word.

You can also move to the previous word containing one or more highlighted characters by holding down the Ctrl key and pressing Tab.

Navigation Keys

Need a quick reminder on the PrimeVerify navigation keys? Just choose KEY TABLE from the Main menu for a description of all Arrow, Return, and Tab key actions.

Text Navigation/Selection using other Options

Return Key - The Return key can be used to quickly select the current word in the edit display. Hold down the Shift key and press Return to select the next word. To select the previous word, press Return while holding down the Ctrl key.

Arrow Keys - Use the arrow keys to move the current cursor position left, right, up, or down. Holding down the Shift key while pressing the left or right arrow will draw a box around the previous or next character, respectively. Selecting the same arrow key again while the Shift key is depressed will enlarge the box to include the next character in that direction. If you hold down the Ctrl key while selecting the left arrow key, the current cursor position will jump to the beginning of the previous word. Similarly, a Ctrl+Right arrow key combination moves the cursor to the beginning of the next word.

Mouse - Clicking anywhere in the edit display area with the mouse will reposition the cursor at that location. The mouse can also be used in combination with the vertical scroll bars to move either display area up or down.

You can also select text using the mouse by drawing a box around it. Click and hold down the left mouse button. As you drag the mouse left or right, a box will be drawn over the selected text. Release the left mouse button to complete the selection. Edit operations can then be performed on the text as described below. To have PrimeVerify select an entire word, simply double-click with the mouse on any character within the word.

Page Navigation

PrimeVerify will read multiple page PRO files. Typically these files have been created by OCR'ing a multiple page TIFF file. The Page menu item offers several commands to move between pages in a multiple page image file:

Moving between Pages To move between pages select "Next Page", "Previous Page", or "Go to Page" from the PAGE menu. Note that any "global" operations such as Find or Image Statistics apply only to the current page, not to all pages in the full multipage document.

Auto Next Page - If the user presses the TAB key (to go to the next suspicious character) after the last suspicious character then the next page in a multiple page PRO file is automatically opened.

Searching for Text

PrimeVerify supplies two menu commands for quickly locating text in the edit display area.

Find - Choose Find from the EDIT menu to display the Find dialog box.



Enter the text you want to locate in the “Find What” field and whether the search should be Case Sensitive by selecting the “Match Case” checkbox. If the checkbox is checked, PrimeVerify will only display a match if it has the exact upper/lower case attributes of the text you entered in the Find field. If unchecked, PrimeVerify will ignore case restrictions when looking for a match. Match Case does not apply to numerical text.

Find Next - Once a match has been found, subsequent searches for the same text can be performed using the Find Next command on the EDIT menu. For rapid searching, use F3 - the Find Next Shortcut.

Both Find and Find Next will start at the current cursor position and continue until either a match is found or the search loops back to the current cursor position. Matches in the edit display are shown with boxes drawn around them to make editing easier. If no match is found, a message saying that the search is complete will appear once all of the text has been scanned.

Editing Text

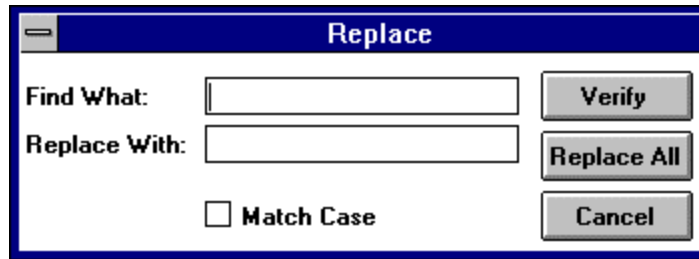
Text in the Edit display area can be modified similar to the way you would edit text in a word processor. The Edit operations that you can perform will remove, replace, or insert text.

Removing Text - If there is a box around one or more characters in the edit display, that text is considered “selected”. Pressing either the Delete or Backspace key will remove the selected text and the cursor will now appear as a vertical bar (“|”) at the position where the text was removed. Selected text can also be removed and placed onto the Windows Clipboard using the Cut command from the EDIT menu.

If no text is selected (i.e., the cursor is a vertical bar), pressing the Delete key will remove the character to the right of the cursor. Pressing the Backspace key will remove the character to the left of the cursor. The Cut command does not apply to unselected text.

Replacing Text - Replace is actually a combination of removing selected text and inserting new text. You can perform a manual replace by first selecting text and then typing new text. PrimeVerify will delete the selected text and insert the new text as you type it.

To replace text automatically, choose the Replace command from the EDIT menu. The Replace dialog box will appear.



Enter the text to search for in the “Find What” field. Enter the text to be inserted in its place in the “Replace With” field. You can limit the replace operation to exact text matches using the Match Case field (see Find above). Choose either the “Verify” or “Replace All” button to start the replace.

Verifying Replacements - Select the Verify button if you want to manually confirm replacements. When a match is found, PrimeVerify will present the Verify dialog box. Choose “Yes” to perform the replace, “No” to skip this replace or “Cancel” to end the replacement process. Unless Cancel is selected, the replacement process will continue to look for subsequent matches, presenting the Verify dialog box again if it finds one.



Replace All - Choosing Replace All will tell PrimeVerify to perform replacements without asking for verification. If a match is found, the text will be deleted and replaced with the text you specified.

When all text has been scanned, a message will appear saying that the search is complete.

Special Symbols

Special characters such as the copyright symbol, ©, or international characters can be inserted by holding the ALT key down while typing the ANSI decimal value of the symbol (e.g., 169 for ©).

Inserting Text - Text can be inserted on any line simply by typing text at the current cursor position. Existing text to the right of the insertion point will shift to accommodate the new text. Note: Since PrimeVerify is tailored for finding and replacing incorrect text, using the Return/Enter key to insert new lines is not supported.

Changing Text to High Confidence – You can change the confidence of a character/word to “9” or high confidence indicating you have reviewed and found the character to be correct. The key sequence F4/F5 or the TAB key (assuming the proper option has been set) will make this change.

Convert Selection to UPPERCASE/lowercase - Selected text is converted to respective case.

Saving the Output File

PrimeVerify will allow you to save modification made to the recognized text in the “.PRO” Output file by choosing Save Output from the FILE menu. If you need to close the open Output file before proofreading is complete, the Save command will store your modifications to-date and allow you to continue where you left off the next time the file is opened. Since a Save operation will overwrite the existing Output file as generated by PrimeOCR, you may want to use Save Output As from the FILE menu to save your changes to another file instead.

Choosing Save Output As will display the Save File As dialog box, where you can enter a new file name to store intermediate results. If you do not supply a file extension, the “.PRO” extension will be added to the file name for you. Choose OK to save changes to the new file.

Exporting Text

You may want to export the verified text into a format other than “.PRO” once the proofreading process is complete.

Choose Export from the FILE menu to display the Save File As dialog box. Under File Name, enter the name of the file you want to save the exported text to. Under File Type, choose one of the supported export formats listed below:

- **ASCII** - Text from the edit display area is saved in a condensed, standard ASCII format. Page layout attributes such as leading spaces and blank lines are removed from the text. Use this export format to save space when the text layout information from the original image is not required.
- **Formatted ASCII** - Text from the edit display area is saved in the same format as the original image. Spaces are inserted into the ASCII output to retain the original page layout of the text.
- **PDA2** -A file format that is defined in the WordScan 5.2 Development Kit. PrimeOCR’s PDA format is Level 2 style. This format should only be of interest to developers who have already interfaced to the WordScan product and want to use their existing output parsing code for PrimeOCR output.
- **ZyIndex** – ZyIndex format (version 3.0), a format defined by ZyLab, for use within their full text searching programs.
- **RTF** - Rich Text Format (version 1.x), a format defined by Microsoft, and widely used by word processing programs. If zones have been defined the zones are defined as frames and text or images are inserted into the frames. Image zones can only be viewed correctly in Word 7.0 or later.
- **PDF** – Image Only, Image+Text, and Normal are supported.
- **Comma Delimited** – Content of each zone will be followed by a “,” character in output. Usually only used in “forms” type applications.

- **HTML** – Generate pages to display output on the web. The output does not include image or OMR zones. If the whole page was defined as one zone then the output will look like the original page. If zones were defined on the page then each zone is reported linearly down the page, separated from the next zone by two inserted blank lines. With this version of HTML output, there is no ability to define columns, or zones that appear next to each other in the output so the zones must be reported linearly down the page.

Export Options

These options may be accessed under File\Export Options menu command.

Point Size: You may increase or decrease the size of all font sizes reported in the output file. Many users find that a -15 or -20 (%) setting (decrease reported font sizes by 20%) works well here. (This setting is only relevant for output formats which include point size reporting, such as RTF, PDF, or PRO.)

ASCII:

Wrapped Paragraphs: If checked it removes the hard carriage return from the end of each line of text, except where the end of a paragraph is sensed. (This option will apply to both ASCII and FASCII formats).

HTML:

Wrapped Paragraphs: If checked it removes the hard carriage return from the end of each line of text, except where the end of a paragraph is sensed.

Default Font Type: If checked the font reported by PrimeOCR will not be inserted into the HTML file, forcing the HTML viewer, when displaying the file, to use its default font type.

Insert Page Breaks: If checked will insert page break characters in the HTML file between pages in a multi page document. (Many applications that display HTML do not need these characters.)

PDF:

Suspect Level: (Only relevant for PDF Normal format). A word with one or more characters with this confidence (or less) will display in the PDF file as an image snippet. (The text will also be in the file behind the snippet.) A setting of "0" turns off this feature.

Desample DPI: Any image data in the file will be changed in resolution to the indicated resolution. Usually you will downsample to 200 dpi to save file space. Valid values are from 1-600. A value of 0 or blank will turn off this feature.

Optimize/byte serve: The PDF file may be optimized for downloading across the internet. This requires a copy of Acrobat 4.0 or later to be installed on the same machine.

(Bookmarks may be automatically added to the PDF file. See BOOKMARK.INI for more information.)

Make Accessible: Identifies content within the PDF file to improve accessibility for users with disabilities. (Sometime referred to as Section 508, a government regulation.)
Modifications can include: setting up reading order of text, identifying paragraphs, providing alternative text descriptions of graphics, identifying reading order of tables. Some of these features require extra data to be supplied. Please refer to Chapter 5 of PrimeOCR, section "PDF_DEFAULTS" for more details.

Thumbnails: Inserts thumbnails into PDF file. Acrobat Reader 5.0 and above will automatically generate thumbnails so this setting is generally not necessary, except for:

- PDF files will be viewed within a browser and Acrobat Reader version is less than 6.0.

- PDF file has a large number of pages, and user does not want to wait for Acrobat Reader to create thumbnails in real time.

Zip Compression: Color and grayscale images are compressed using Zip compression. Zip compression is lossless. Some customers, including some Federal Government agencies, require the use of lossless compression techniques. (Lossless compression results in larger file sizes than lossy compression.)

Compression: Only valid for color or gray scale images (which use default JPEG compression). 0 leaves the compression as is, -100 is low levels of compression, +100 is high level of compression (which reduces file size and quality of image).

RTF:

Wrapped Paragraphs: If checked it removes the hard carriage return from the end of each line of text, except where the end of a paragraph is sensed.

Page Width/Height: Forces the page size to this dimension.

Margins: Indent text from the relevant margin by this amount.

Notes on Page and Margin variables:

- All coordinates are expressed in BMUs (1/1200 of an inch) and are relative to the top left hand corner of the page.
- Any coordinate can be listed as "-1". In this case the real value is reported.

- Any coordinate can be “-99”, in which case ALL coordinates will NOT be reported in the RTF output.
- If a coordinate is listed this fixed coordinate will be reported for all pages of OCR output in the RTF output.

Proportional/Fixed Fonts: You may select the fonts that will be reported in the RTF file for proportional, proportional sans serif, and fixed pitch fonts. Make sure you enter font names that are spelled exactly as they appear in your target word processor (or other application).

Multiple Export

This command, under the "File" menu, allows you to export many output files at one time, without opening and exporting each file individually. You may use this feature, for example, by generating all OCR output as ".pro" output. Based on the confidence level report generated by the Job Server you might segment a large number of files as exceeding your minimum confidence threshold. These files you export, using the Multiple Export command, without verification. The remaining files are verified using PrimeVerify.

- First, select one output file. This act targets the whole directory in which this file resides for export (not just the target file).
- Second, specify the type of export format you wish.
- Third, mark the "Include 1st Level Subdirectories" box if you wish to also export all similar files in the direct subdirectories of the target directory. (Subdirectories two or more levels removed from the target directory are not processed.)
- Fourth, click OK to begin processing.
- A message box will appear when processing is complete to indicate the completion of processing.

Roughly 2-4 files per second are converted from ".pro" format to the export format. The export files are placed in the same directory as the original output files, with the same file name but with a ".txt" or ".rtf" extension.

Closing the Output File

When you have completed the proofreading process for the current Output file, you can close it before moving on to the next file. Choose Close Output from the FILE menu. If changes have been made since the file was opened, PrimeVerify will ask if you want to save them before closing. After the file is closed, both the image and edit display areas will be blank and you can open another Output file or exit PrimeVerify.

Auto Process

The user can define several actions to take place automatically on initiation of the Auto Process. The automatic closing, saving,

exporting, and opening of the next file is a particularly powerful feature that can save significant amounts of operator time.

Auto Process Initiation

The start of the Auto Process can occur via two methods:

1. Auto Process menu item under File menu.
2. End of file reached and "When End of Current File is Reached Via Tab Key" is selected under Configure Auto Process menu item. If a file is a multipage file then Auto Process will only occur on the last page within the file.

Auto Process Configuration

Select Configure Auto Process menu item from File menu.

Invoke Auto Process:

Select "When End of Current File is Reached Via Tab Key" to cause automatic initiation of Auto Process as discussed above.

Output:

Select "Save To" if you wish to save PRO file during Auto Process. You can choose to save PRO file as is, or save it to another fixed directory.

Select "Export To" if you wish to export PRO file during Auto Process. You can choose to export PRO file to the same directory as the PRO file, or export it to another fixed directory. You can select the export format used for the process.

Note that you can both save and export PRO file. These options are not mutually exclusive.

Input:

You can select to delete current PRO file. It does not make much sense to "Save To/Same Directory as Input File" and delete current PRO file so check to see that both of these items are not checked.

You can automatically open the next PRO file in the original PRO directory at the end of the Auto Process. The "next" file is defined by sorting the files alphabetically and choosing the file right after the current file in the list.

Skip Files by Confidence

You can skip files that are below or above a user selected range of confidence. Check "Do Not Verify Files..." and enter the lowest average confidence level that you would like to review and the highest level. (Good numbers for each might

be "800" and "895"). Only files with average confidence levels between these settings will be brought up for review.

PrimeVerify Options

Special Characters

Tables are an easy way to enter special ANSI/ASCII characters not available on many keyboards such as © or international characters. Contact Prime Recognition if you need help creating these tables.

Show Insert Table X

Brings up a list of text strings in a table format. If the user double clicks on a text string it will be inserted at the current cursor position. If text in the display window has been selected this text will be cut before the insertion. Note that a file called "iTablex.ini" must exist in the same directory as PRVERIFY.EXE (where "x" is the number of the table) to implement this option. This ASCII format file has the following format:

```
Line 1: Identifier      "PrimeVerify OCR Verifier iTable2.ini File"
Line 2:Version         "Version X.X"
Line 3:Table X location "100"
Line 4:Table Y location "100"
Line 5:Table Width     "500"
Line 6:Table Height    "1000"
Line 7: Word 1         "TEST"
Line 8+: Word 8+      "TEST2"
```

Lines 3-6 reflect the location and size of the table when it is placed on the screen. These numbers are modified if the user changes the location of the tables.

Many lines in positions 7->? can be created to add strings to the table.

Log Opening of Files

If selected, a file called \PRPROOF\PRVERIFYREAD.LOG is created which lists the date and time each PRO file is opened. Note that this is not per page, it is per document (which might be a multiple page document).

Exiting PrimeVerify

Choose Exit from the FILE menu to terminate the PrimeVerify application. If you have an open Output file, PrimeVerify will ask if you want to save any changes before closing.

Chapter 7 Troubleshooting

Review the list of error conditions and corrections shown below if you experience problems while running PrimeView or PrimeVerify.

File Input/Output

If an error occurs while trying to open, close, save, export or update a file, make sure that the file is not in use by another application, and that the file's drive is accessible to your system and is not full.

Initialization Problems

PrimeView relies on the Initialization files PRVIEW.INI and PrimeVerify relies on PRVERIFY.INI respectively, to retain user defaults between application sessions. If, when running either application, you receive a message that says

“Only Use the Drop Down List Choices”

“The Initialization file has been corrupted”

then delete the initialization file (either PRVIEW.INI or PRVERIFY.INI) for the application you are running.

The next time you run the application, a new initialization file will be created automatically, after you receive the message:

“Could not open PRVERIFY.INI/PRVIEW.INI file. A new initialization file will be created.”

Starts Up OK but Can't See Window - Again the PRVIEW.INI and PRVERIFY.INI has been corrupted. Delete appropriate INI file start application again.

Licensing Errors

Licensing errors will occur if you have not properly installed or licensed the software. If you receive a message such as:

“Could not find a license for PrimeView on this hardware key”

review the steps listed in the “*Licensing PrimeView & PrimeVerify*” section of this manual. If you are still having problems, contact Prime Recognition.

Image Display Problems

If the image display area does not show an image after opening an image file in PrimeView or an Output file in PrimeVerify, you are probably experiencing a licensing problem. See “Licensing Errors” above.

If the image display area shows an image at extreme magnification, or you cannot move the image displayed, simply use the Zoom Out command from the IMAGE menu to re-adjust the display.

Chapter 8 **Programming Considerations**

Some users may wish to integrate PrimeView or PrimeVerify into their application programmatically. Each program offers several levels of programming integration.

Command Line Start Up

You may startup either PrimeView or PrimeVerify using a command line. The syntax is the executable name followed by a space followed by the full path to the target file. For example:

```
"C:\PRPROOF\PRVIEW.EXE C:\PRDEV\TEST1.TIF"
```

For PrimeView the appropriate target file is an image. For PrimeVerify the appropriate target is a "PRO" file.

Processing a List of Files

You can process an arbitrary list of files by creating an ASCII format text file that has a full path to each file, one per line. This could be a list of images for PrimeView, or a list of PRO files for PrimeVerify. When you go to load an image into PrimeView, or open a PRO file in PrimeVerify, choose this text file instead. Use F3 to proceed to the next image in PrimeView, use the Autoprocessing option to proceed to the next PRO file in PrimeVerify.

Chapter 9 InputAccel Version

A version of PrimeVerify is available for the InputAccel imaging system by Captiva Software. Prime Recognition is a development partner of Captiva. PrimeVerify (and PrimeOCR) has been certified by Captiva to work within InputAccel as a native application.

This version is basically the same as the PrimeVerify described previously in this manual with some differences and restrictions noted below. The InputAccel version of PrimeVerify ships with the standard version of PrimeVerify, its file name is PRVRFIA.EXE.

Overview

PRVRFIA receives PRO format files from the PrimeOCR engine, allows user to verify OCR output, and automatically returns verified output to InputAccel in a PRO format (for processing by PrimePostIA), PDA2 format, or other supported formats such as PDF. PrimeVerifyIA's output can be sent directly to standard IA export modules, such as the Text Export module, or the PDF Export module.

System Requirements

PRVRFIA is a standard InputAccel 32 bit application.

Operating System: Win2000, WinXP, Windows 2003 Server.

Memory: At least 64MB of RAM on top of the minimum configuration for the respective operating system.

The InputAccel version of PrimeVerify requires a software or hardware key license.

MDF, PCF, IPP, and IAP files for PRVRFIA ship with the PrimeOCR engine and are located in the \BIN directory (if InputAccel option was selected during PrimeOCR installation process). See *PrimeOCR Access Guide* for more details on these files.

Configuring PRVRFIA

The output format for PrimeVerifyIA is set in the IPP/PCF file.

Other items, typically user preference items, are setup within PrimeVerifyIA via menu items. (Setup through IA Supervisor is not required or supported.)

Using PRVRFIA

Run PRVRFIA from installation directory.

Log on to server as per a typical InputAccel client.

Selecting “All Batches” from “File” menu signals the server to send you a (or next available) page for verification from any available batch.

Selecting “Single Batch” from “File” menu and then a batch that has PrimeVerify tasks ready signals the server to send you a page for verification from that specific batch.

Selecting “Stop” from “File” menu sends the output back to the server (in format selected in PCF file).

Using one of the Reject commands will send the open page back to the server but set the MDF value of ”IsReject” to 1.

PRVRFIA Shortcuts

If a file is already open for verification and you select “All Batches or Single Batch” it will automatically close the open file and send the results to the server prior to opening next file.

You can use “Auto Process” to automatically send open page to server and get the next page on CTRL+U or reaching end of page via TAB key.

If you wish you may skip files for manual processing that fall outside a lower or upper threshold of confidence.

Removed Commands from PrimeVerify

The following menu items are removed from PRVRFIA, as they are not relevant in an InputAccel environment:

- Save commands
- Export commands and options
- Image Directory Setup
- Next/Previous Page and Auto Next Page