



OCR Software for Windows

User's Guide

ENGLISH - DEUTSCH - FRANÇAIS

ITALIANO - ESPAÑOL

Copyright

Copyright © 1994 by Xerox Imaging Systems, Inc. All rights reserved. No part of this publication may be transmitted, transcribed, reproduced, stored in any retrieval system or translated into any language or computer language in any form or by any means, mechanical, electronic, magnetic, optical, chemical, manual, or otherwise, without the prior written consent of Xerox Imaging Systems, Inc., 9 Centennial Drive, Peabody, Massachusetts 01960. Printed in the United States of America.

The software described in this book is furnished under license and may be used or copied only in accordance with the terms of such license.

Disclaimer

Xerox Imaging Systems, Inc. (XIS) provides this publication “as is” without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability or fitness for a particular purpose. Some states or jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you. XIS reserves the right to revise this publication and to make changes from time to time in the content hereof without obligation of XIS to notify any person of such revision or changes.

Credits

TextBridge is a trademark of Xerox Imaging Systems, Inc.

Ami Pro and *Lotus 1-2-3* are trademarks of Lotus Development Corp.

DisplayWrite is a trademark of IBM Corp.

Excel, *Word*, and *Windows* are trademarks of Microsoft Corp.

FrameMaker is a registered trademark of Frame Technology, Inc.

Interleaf and *Technical Publishing System* are trademarks of Interleaf Corp.

Multimate Advantage is a trademark of Multimate Corp.

OneScanner is a trademark of Apple Computer, Inc.

WordPerfect is a registered trademark of WordPerfect Corp.

WordStar is a trademark of MicroPro International Corp.

Other terms used in this manual are the trademarks of their respective holders.

Portions of this product copyright © 1990, 1991, 1992, 1993, 1994 Pixel Translations, Inc.

Portions of this product copyright © 1994 Mastersoft Corp.

Writer: Jim Cahill

© **Xerox Imaging Systems, Inc.**

9 Centennial Drive

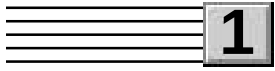
Peabody, Massachusetts 01960

(508)–977–2000 **Customer Support: (508) 977–0764**

TextBridge User's Guide

Part Number 00–07763–00

February 1994



OVERVIEW

Welcome to Xerox Imaging Systems' **TextBridge™ OCR for Windows**. OCR stands for **optical character recognition**, the capability to read an image and recognize and output text from it. Images can come from scanners, fax modems, or other sources.

TextBridge is a suite of applications that combines Xerox Imaging Systems' industry-leading OCR with simple Microsoft Windows interfaces.

TextBridge applications can convert recognized text to a number of word processing, spreadsheet, database, and other text formats.

WHAT IS TEXTBRIDGE OCR?

TextBridge OCR is software that turns image data into usable text files on your PC.

With TextBridge OCR, you can access the valuable data locked inside paper documents, as well as on-line faxes and page images from other sources.

TextBridge supports a number of popular desktop scanners and converts recognized text to many popular **text formats**.

Use your scanner to input hard copy documents to TextBridge, which takes the scanned images, performs OCR, converts the recognized text to the format of your choice, and stores it as a PC file.

TextBridge provides an easy-to-use interface and a powerful set of built-in capabilities. For example, the **preview** tool lets you view and zone pages before OCR. The **verifier** lets you interact with the OCR software to achieve the highest possible accuracy.

TextBridge incorporates Xerox Imaging Systems' industry-leading document recognition software, and includes a number of technologies developed by Xerox Palo Alto Research Center (PARC), where modern computer interfaces originated.

Consequently, TextBridge provides the most accurate OCR and format retention results on the widest range of documents:

- documents with point sizes ranging from 6-point to 72-point type in practically any typeface
 - + TextBridge can reliably recognize type smaller than 8-point only from images scanned at 400 dots per inch resolution.
- documents printed on typewriters, phototype-setters, and impact, ink-jet, and laser printers
- documents with single- or multiple-column layouts
- documents containing halftones
- on-line single- or multiple-page TIFF images from fax modems and other sources
- hard-copy faxes

SYSTEM REQUIREMENTS

To install and run TextBridge, your Windows-compatible PC must be equipped with the following:

- an Intel (or compatible) 80386, 80486, or Pentium™ microprocessor
- Microsoft Windows™, version 3.1 or later running in **enhanced mode** only
- four megabytes (4Mb) of random access memory (RAM); eight megabytes (8Mb) is recommended
- 8Mb to 16Mb of **permanent virtual memory**; 16Mb is recommended
 - + For information about configuring your system with permanent virtual memory, refer to the *Microsoft® Windows™ User's Guide* published by Microsoft Corporation.
- a hard disk with a minimum of 4Mb of free space in which to install TextBridge

ON-LINE HELP

TextBridge is designed to be easy to learn and use. However, if you need assistance, TextBridge provides a complete **hypertext**-based on-line Help system.

In any of TextBridge's screens, you can select Help and display a Help window about that particular screen.

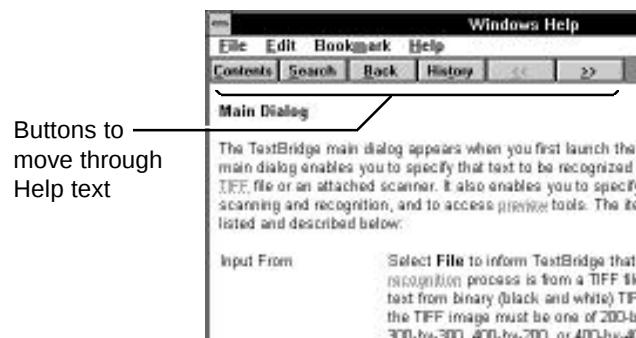


Figure 1. TextBridge on-line Help

By selecting the Contents button near the top of the window, you can display the top-level index for TextBridge Help. You can move through the TextBridge Help system in a number of ways:

- by selecting a topic from the Contents list
- by choosing a jump
- by searching for a topic
- by browsing forward or backward
- by backtracking

In any Help window, you can pull down the Help menu and select the “How to Use Help” topic for complete information about using Help tools.

2 **INSTALLATION**

This chapter describes the TextBridge installation process, which requires as many as three stages:

1. Follow the scanner manufacturer's instructions to install and test the scanner and **TWAIN** source driver.
2. Run the TextBridge scanner setup program and test the scanner interface.

This chapter organizes information in the installation sequence listed above.

However, before describing the installation steps in more detail, this chapter discusses issues relating to "System Configuration and Performance."

Please read that section, as it describes ways to avoid problems associated with memory limitations and inefficient use of your PC's resources with TextBridge.

SYSTEM CONFIGURATION AND PERFORMANCE

TextBridge operates under Windows 3.1 (or later) in **enhanced mode** only. Thus, your PC must have at least an Intel 80386 microprocessor.

Also, to run TextBridge, your PC must have at least **four megabytes** (4Mb) of memory (RAM), and must be configured with at least **eight megabytes** (8Mb) of **permanent virtual memory** (swap space).

If you regularly intend to scan multiple-column or landscape pages of text, or pages with complex layouts, you should configure your PC with eight megabytes (8Mb) or more of system memory (RAM) and 16Mb of virtual memory.

With only 4Mb of memory, the minimum requirement, TextBridge will more often use **virtual memory** (swap space on your hard disk).

If you try to run TextBridge with less than 4Mb of RAM, the application informs you that it:

Cannot initialize server.

In general, the more RAM that is available when you use TextBridge, the less swapping to disk will be required during operation.

As a rule of thumb, you should configure your system with twice as much virtual memory as RAM. Thus, if you have 4Mb of memory, configure your system with at least 8Mb of virtual memory. If you have 8Mb of RAM, specify 16Mb of virtual memory.

Important

It is necessary to configure virtual memory as **perma-nent** virtual memory, especially for PCs with only 4Mb of RAM. This assures that TextBridge always has an adequate amount of contiguous swap space during OCR. Refer to your *Microsoft® Windows™ User's Guide* for information about configuring virtual memory.

Regardless of your RAM and virtual memory, a number of other system configuration choices can affect the availability of memory to TextBridge, and thus can affect performance.

Following is a list of items that can affect TextBridge system performance:

- **RAM disks**

If you have set up your system to use part of your extended memory as temporary file storage, called a RAM disk, this subtracts from available memory.

- **TSR (terminate-and-stay-resident) programs**

Some programs are designed to automatically load into memory when you start your system, or to stay in memory even after you exit them. These programs also affect the memory available to TextBridge.

- **Expanded Memory Drivers**

These are programs that use extended memory as expanded memory (memory used by the operating system), for example, a windows driver.

- **Other Drivers**

These are programs that provide some type of system control, for example, a network driver.

If you find that TextBridge's performance seems slow, check to see if your system is configured with any of these devices. If there are one or more of these devices that you can do without, remove them to improve TextBridge performance.

INSTALLING AND TESTING THE SCANNER

In addition, with its support of the **TWAIN** standard, TextBridge works with virtually any fully TWAIN-compliant device that provides a binary image in a supported size and resolution.

Scanners generally require a TWAIN source driver, which is provided by the scanner or interface card manufacturer. **Consult the scanner documentation for details about installing your scanner, interface card, and driver.**

INSTALLING TEXTBRIDGE SOFTWARE

Installing TextBridge software is a two-phase process. First, you install the software files. Second, you run a scanner setup program to link TextBridge to your scanner's system-level or TWAIN source driver.

Note If you are **not** using TextBridge with a supported scanner, you can ignore the second installation phase, running the scanner setup program. For example, you can use TextBridge to recognize TIFF files produced by fax modems

Run the scanner setup program

To use TextBridge with a supported scanner, run the **Scanner Setup** program. Scanner Setup links TextBridge to your scanner's system-level or TWAIN source driver. As Figure 2–2 shows, the Scanner Setup program icon is available in the TextBridge OCR group window.

This section provides two procedures, one to install and test a TWAIN source driver; the other, to install and test an ISIS driver.

- + If you have problems getting your scanner to work with TextBridge, consult Appendix A of this guide.

Install and test a TWAIN source driver

To install and test a TWAIN source driver, use the following procedure:

1. In the TextBridge OCR program group, double-click the Scanner Setup icon.

The Scanner Setup main window appears (Figure 2–3). Note that the default scanner type is ISIS.

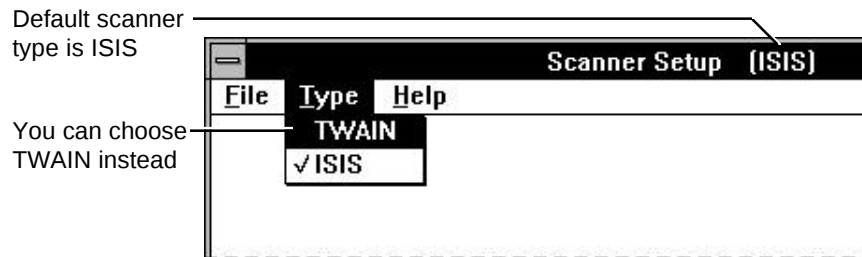


Figure 2–3. Scanner Setup main window

2. Specify the scanner type.

- Pull down the Type menu, and choose TWAIN as the scanner type.
 - + Select TWAIN only if your scanner is provided with a TWAIN source driver.

If your scanner is provided with a DOS system-level driver (.sys file) that must be referenced in a DEVICE statement in the config.sys file, select ISIS instead, and refer to the next subsection, “Install and test an ISIS driver.”

Check your scanner documentation for information about the provided driver(s).

- Go on to specify the TWAIN source driver to be used by TextBridge.

3. Specify the TWAIN source driver.

- Pull down the File menu, and choose the Select Source command. The TWAIN Select Source dialog is displayed
- All properly loaded TWAIN source drivers are shown. Highlight the one you want TextBridge to use.
- Click the Select button to complete the selection.

4. Test the scanner interface.

- Pull down the File menu and select the Acquire command. The TWAIN source’s **native user interface** should now appear.
- Use the TWAIN source’s native UI to specify and acquire an image. Refer to the documentation that came with the device.

5. Exit the Scanner Setup program.

Pull down the File menu and select Exit. You are now ready to use TextBridge with your TWAIN device. Refer to Chapters 3 and 4 for information.

3 USING TEXTBRIDGE

After you have installed TextBridge as described in Chapter 2, you are ready to use the application to turn paper documents and on-line page images into usable text files.

At your direction, TextBridge can scan a document, or read on-line **TIFF** images, and perform **optical character recognition(OCR)**.

TextBridge can also display pages in a **Preview window**, where you can view, **zoom**, and **zone** the page before processing.

During OCR, you can **verify** recognized words as correct, or fix recognition errors. By verifying text, you teach TextBridge to improve recognition accuracy for the rest of the job

After TextBridge performs OCR, you can direct the application to convert the recognized text to a **text format** that you can use with your favorite word processing, desktop publishing, spreadsheet, database, or other application.

This chapter describes and provides step-by-step procedures for using the main TextBridge application.

Specifically, this chapter covers the following topics:

- About TextBridge preferences
- Scanning and converting a document
- Recognizing and converting on-line TIFF files
- Previewing pages before processing
- Verifying text during recognition

ABOUT TEXTBRIDGE PREFERENCES

TextBridge optical character recognition software has a built-in set of **preferences**, settings that control the OCR process. For most good-quality office documents, you can use TextBridge default preferences to provide excellent character recognition.

However, to fine-tune the OCR process for a variety of documents, TextBridge enables you to define preferences on a job-by-job basis. Before starting OCR, you can click the Preferences button in the Main dialog. This displays the Preferences dialog.

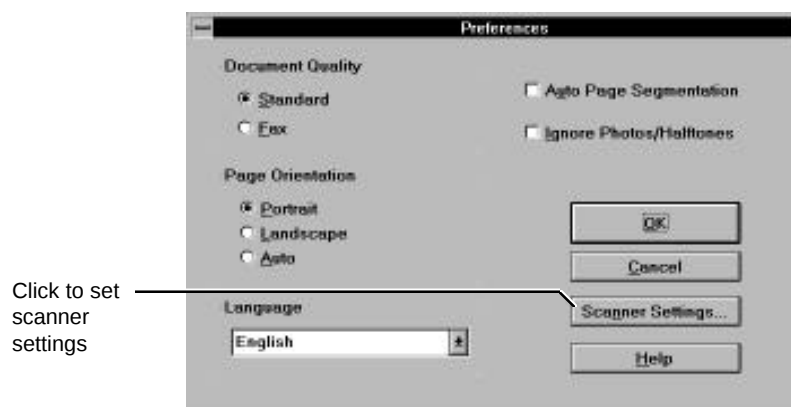


Figure 3–1. Preferences dialog

Specify one or more of the preferences to control how TextBridge will process your document(s). Table 3–1 describes the options available to you.

Note When you change a preference, it remains in place until you change it again, even when you exit Text-Bridge and run the program again later. This lets you lock in place the preferences that are appropriate for documents you process most often. Of course, you can always change one or more preferences at any time.

Table 3–1. Preferences

Preference	Function
Document Quality	- Choose Standard for most documents. - Choose Fax for on-line, fax-quality TIFF images (200x100 or 200x200 dots per inch), or if you are scanning hard copy faxes.
Page Orientation	- Click Portrait for most typical portrait-oriented office documents. - Click Landscape for wide documents that are scanned in sideways, and thus must be rotated in memory by 90-degrees before recognition begins. - Click Auto if your document contains a mixture of page orientations (for example, mostly portrait pages, with large rotated tables on landscape pages). Note that you may also want to click Auto if you are recognizing TIFF files and are not sure how the page image is oriented in the file. With Auto selected, TextBridge performs a pre-processing step to determine the orientation of the page. Thus, overall processing speed is slower with this option turned on.
Language	This category provides a pop-up menu of the Text-Bridge recognition languages that you have installed on your system. TextBridge can perform highly accurate optical character recognition on documents in German, French, Italian, and Spanish, as well as English. Select the primary language of the document to be recognized (for example, German).
Auto Page Segmentation	- Leave off for single-column documents. - Click the checkbox on for documents with two or more columns of text. With page segmentation on, TextBridge performs a pre-processing step to analyze the shape and location of text blocks on the page, so they are output in the right order. Note that, in the converted text file, TextBridge outputs text in galley (single-column) format.
Ignore Photos/Halftones	Click the Ignore Photos/Halftones checkbox on if the document to be recognized contains photographs or halftones (screened photographs). Otherwise, TextBridge tries to recognize the halftone dots as text, reducing recognition accuracy and speed.

SCANNING AND CONVERTING A DOCUMENT

One of the tasks that TextBridge enables you to accomplish is scanning a hard copy document into an on-line text file. The document can comprise one page or many pages.

During this process, TextBridge scans the first page, performs OCR on it, and saves the recognized text in a temporary file. It repeats the process for each subsequent page, appending the recognized text to the temporary file.

When you inform TextBridge that the entire document has been scanned, TextBridge displays the Save As dialog. In this dialog, you specify the output file name, disk and directory, and the text format to save it in.

Note The following procedure assumes that the scanner is properly connected to your PC and is powered on and ready. The procedure also assumes that TextBridge is installed and running.

To scan and convert a document, complete the following steps:

1. Load the page(s) into your scanner.

Depending on your scanner, you can load a stack of pages in the automatic document feeder (ADF), or a single page on the scanner's flatbed.

2. From the Main dialog, click Scanner in the Input From box (Figure 3–2, next page).

3. To save an image of each page to a TIFF file, click the Save Page Images check box.

+ Page images are stored as TIFF files with **CCITT** Group 3 compression.

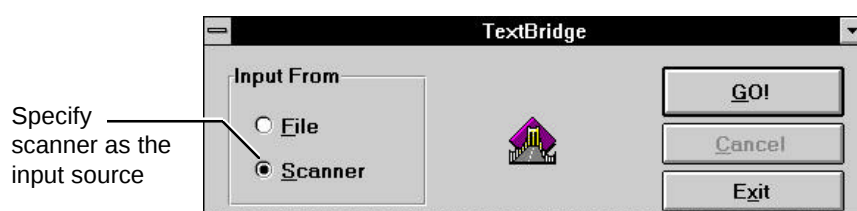


Figure 3–2. Scanner as input source

4. Optionally, click the Preferences button to further define the scanning and OCR process.

Use Table 3–1 as a guide. With Preferences, you can specify such things as page orientation, recognition language, and so on. When you are done specifying preferences, click OK to return to the Main dialog.

5. Click GO! in the Main dialog to begin the scanning and recognition process.

If you are using a TWAIN scanner, the **native user interface** of the source driver appears. Here you can control the scanning process.

If you are using an ISIS scanner, the page(s) in the scanner are scanned and recognized automatically.

When you dismiss the native UI (TWAIN), or the page(s) are all scanned, TextBridge displays the Add More Pages dialog (Figure 3–3).

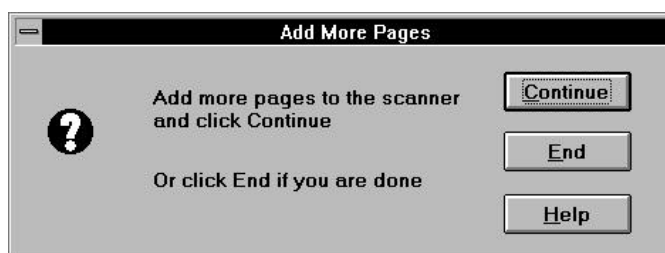


Figure 3–3. Add More Pages dialog

**6. To scan more pages, go to step 7.
To end scanning and OCR, go
directly to step 8.**

**7. Prepare your scanner, then click
Continue.**

If you are using a TWAIN scanner, the native UI reappears, and you can resume scanning and recognition. When you are finished, TextBridge again displays the Add More Pages dialog (refer to Figure 3–3). Proceed from step 6.

8. Click End in the Add More Pages dialog.

TextBridge now displays the Save As dialog.

9. Specify the output file information.

- Specify the file name, destination disk and directory.
- Also specify the output text format in the Save File as Type pop-up menu.
- Click OK.

TextBridge converts the recognized text to the specified format, saves the converted text to the specified file, and closes the Save As dialog. The Main dialog remains, ready for the next job.

RECOGNIZING AND CONVERTING TIFF FILES

If you have one or more TIFF files that contain page images, you can use TextBridge to produce a text file from them. To TextBridge, an on-line TIFF image file is similar to one or more scanned page images.

Some TIFF files contain more than one page image. These multiple-page TIFF files are most commonly generated from **fax modems**. TextBridge is equally capable of processing both single- and multiple-page TIFF files.

TextBridge can process TIFF (*.TIF) files in the resolutions and formats described in Table 3–2:

Table 3–2. Supported TIFF Resolutions and Formats

Resolutions*	TIFF Formats
100-by-200	Uncompressed (Intel header)
200-by-100	CCITT-3 (Intel header)
200-by-200	CCITT-4 (Intel header)
300-by-300	Uncompressed (Motorola header)
400-by-200	CCITT-3 (Motorola header)
400-by-400	CCITT-4 (Motorola header)
	Intel FAXability

* All TIFF files processed to one document must be of the same resolution. If TextBridge has processed several files that are 200-by-200 dpi, for example, and the next file that it encounters is 300-by-300 dpi, the program generates an error message, then displays the Save As dialog box so you can save the recognition results up to that point to an output document.

At your direction, TextBridge opens the TIFF file(s), reads the image data into computer memory, performs OCR on the image(s), and then saves the recognized text to a formatted text file.

To recognize and convert one or more on-line TIFF files to a text file, use the following procedure:

1. From the Main dialog, click File in the Input From box.

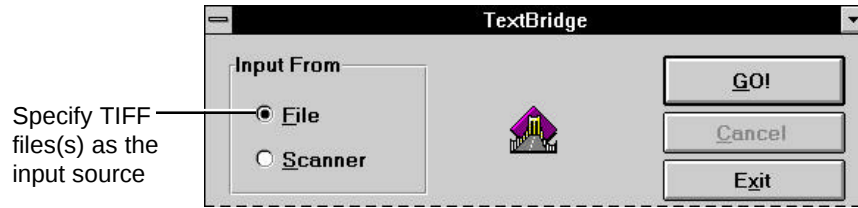


Figure 3-5. Input From File

2. Optionally, specify Preferences to control the recognition process.

- Click the Preferences button to display the Preferences dialog (refer to Figure 3-1).
- Change one or more of the preferences in place. Refer to Table 3-1 for information.
- When you are done specifying preferences, click OK to return to the Main dialog.

3. Click GO! in the Main dialog.

TextBridge displays the Open dialog (Figure 3-6).

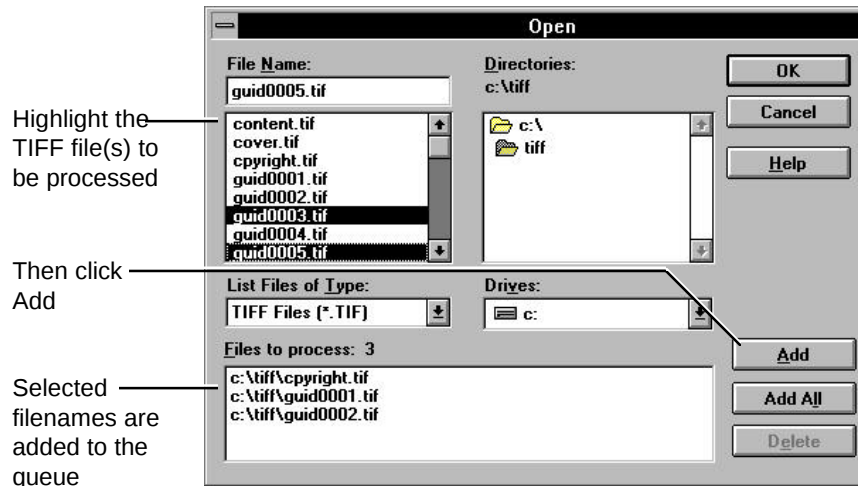


Figure 3-6. Open dialog

4. Specify the TIFF file(s), then click OK.

- + To select a single file, click on it in the File list box, then click Add to add it to the **Files to process** list. Or, simply double-click the file.

To select a range of files, point to the first file in the range, click, hold, and drag the mouse downward in the File list box. Or, click the first file in a range, then Shift-click (hold down the Shift key and click the mouse) on the last file in the range. With the file range selected, click Add to add the files to the process list.

To select TIFF files that are not in sequence, click the first file, then Control-click (hold down the Control key and click the mouse on) subsequent files in the file list.

- + To add all files in the current directory to the process list, click the Add All button.

To add files from another directory, change to the other directory, and add the files as noted here. Note that the full pathnames of the files are added to the process list.

To delete files from the process list, you can select one at a time, and click the Delete button. Or, you can select a sequence of files, or a non-sequential group of files in the same manner as described above, and click the Delete button.

Files are processed in the order in which they are added to the queueAs you add files, the number of files in the process list is tracked by a counter.

After you add the file(s) and click OK in the Open dialog, TextBridge performs OCR on the page image(s). When OCR is complete, TextBridge displays the Save As dialog (refer to Figure 3–4).

5. In the Save As dialog, specify the output file information.

- Specify the file name, destination disk and directory.
- Also specify the output text format in the Save File as Type pop-up menu.
- When done, click OK.

TextBridge converts the recognized text and saves the converted text to a file of the given name, disk and directory location. It then displays the Main dialog, ready for the next job.

PREVIEWING PAGES BEFORE PROCESSING

To view a page before processing, or to define areas of a page to process, TextBridge provides the Preview window (Figure 3–8).

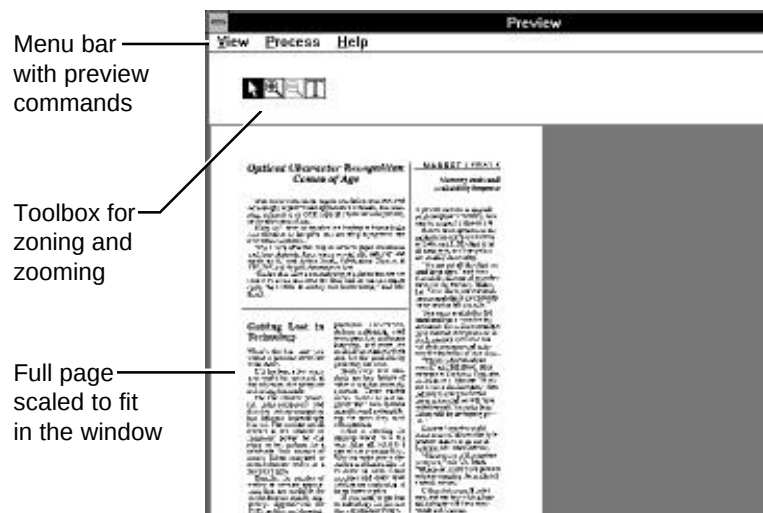


Figure 3–8. Preview Window

In the Preview window, you can use the Zoom tools in the Preview toolbox to zoom in and out on the page at several magnification levels:



Zoom In



Zoom Out

With the Zoom tools, you can magnify a page to **fullresolution**, zoom out to fit the page entirely in the window, or display the image somewhere in between.

To define a portion of the page to be processed, you can use the Text Zone tool to draw up to 127 rectangular **zones** around specific areas on the page:



Text Zone

Zones are numbered to show the order in which you created them and the order in which contained text will be output in the finished file. Overlapping zones are opaque; the topmost zone “owns” any common text it shares with the underlying zone(s).

The Select Zone tool lets you select any of the zones you have created:



Select Zone

A selected zone is identified by solid corner**handles**. After you select a zone, you can move it, resize it, delete it, or put it in front or in back of another zone.

For processing purposes, you can adjust the zones page-by-page or have the zones take effect for all pages of the document. When the job is complete, TextBridge automatically clears the zones.

Note The following step-by-step procedure assumes that, if you are using a scanner, it is properly connected to your PC, powered on and ready. The procedure also assumes that TextBridge is properly installed and running.

To preview a document before processing it, use the following procedure:

1. If you are scanning, load the page(s) into your scanner, then go to step 2. If you are processing a TIFF file, start at step 2.

2. In the Main dialog, define the input source.

- Select either File or Scanner in the Input From box (refer to Figures 3–2 and 3–5).
- Click the Preview box in the Main dialog.

3. Optionally, click the Preferences button to further define the scanning and OCR process.

Use Table 3–1 as a guide. When you are done with preferences, click OK to return to the Main dialog.

4. Click GO! in the Main dialog.

If you are scanning with a TWAIN scanner, the native UI for the scanner appears, and you can execute scanning from this interface. If you are scanning with an ISIS scanner, TextBridge automatically scans a page from the scanner.

If you are reading the page image(s) from one or more TIFF files, TextBridge first displays the Open dialog (refer to Figure 3–6). In the Open dialog, specify the name(s) of the TIFF file(s), and their directory and drive location, then click OK.

TextBridge opens the Preview window, and displays the first scanned or on-line page image (refer to Figure 3–8).

5. Optionally, zoom the page image.

The page first appears zoomed out to fit in the window. To quickly zoom to full resolution, you can pull down the View menu and choose the **Zoom Max** command. To zoom in on the page image by steps, click the Zoom In icon from the Preview toolbox, and click on the page display.

When the page is zoomed in, scroll bars appear to let you shift the display horizontally and vertically (Figure 3–9).

To zoom all the way out, pull down the View menu and choose the **Zoom Min** command. To zoom out in steps, click the Zoom Out icon, and click on the page display.

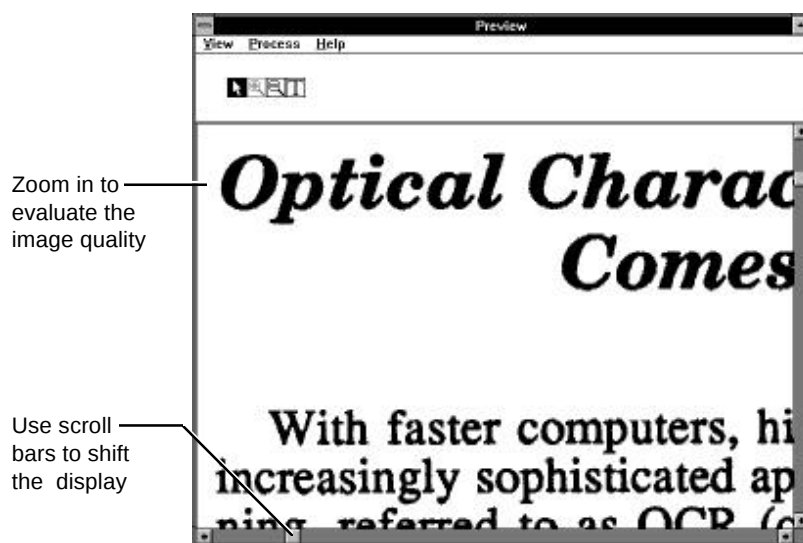


Figure 3–9. Previewed page zoomed in

6. Optionally, create one or more zones to define the page area(s) to be processed.

Up to 127 zones can be created, as follows:

- Click the Text Zone tool, point to a corner of the area to be zoned, click and hold the left mouse button, and drag the mouse. A zone rectangle appears as you are moving the mouse.
- When the zone is fully sized as you intend, release the mouse.
- + The zone number appears in the upper left corner. Handles appear on the zone rectangle for resizing purposes (Figure 3–10).

The zone number indicates the order in which recognized text blocks will be output in the finished text file.

To **resize** a zone, click and hold on a corner handle and drag the mouse.

To **move** a zone, click and hold inside the zone, and drag the mouse.

To **position** a zone relative to another zone, make sure the zone is selected (handles are black), then pull down the View menu and select the **Move to Front** or **Move to Back** command.

Overlapping zones are opaque, meaning that any image area shared by more than one zone is output as part of the topmost zone.

To **delete** a zone, pull down the View menu and choose the Clear Zone command (or simply press the Delete key).

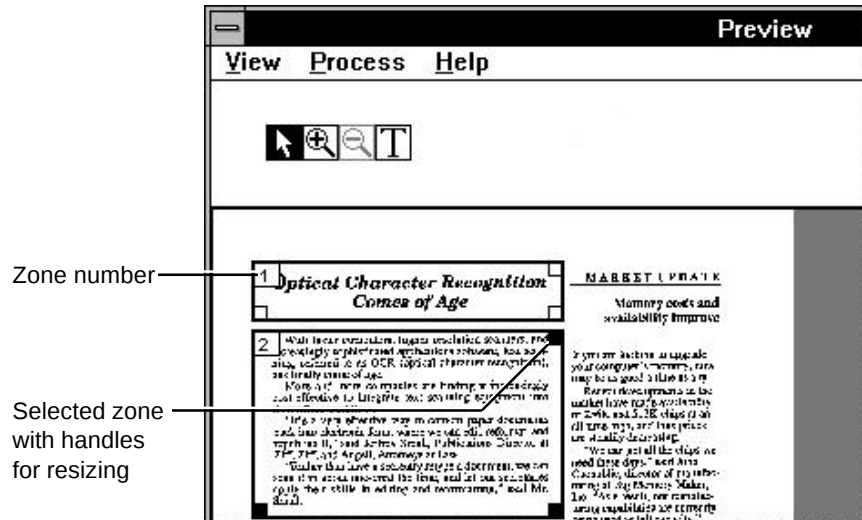


Figure 3-10. Zones in Preview

7. When you are done previewing the page, start the OCR process.

Pull down the Process menu and select This Page or All Pages to start/resume OCR.

- + Select All Pages to close the Preview window and process all pages of the job to the zones in place.

Select This Page to preview and process every page individually. TextBridge will process the first page, scan, or read, and display the next page in the Preview window. TextBridge then goes into idle mode. Proceed from step 5 to preview the now-displayed page.

When you have previewed the last page of the document, select the All Pages command to close the Preview window and have TextBridge finish OCR of the document.

When OCR is completed, TextBridge displays the Save As dialog.

8. Specify the output file information.

- Specify the file name, destination disk and directory.
- Also specify the output text format in the Save File as Type pop-up menu.
- When done, click OK.

TextBridge converts the recognized text and saves the converted text to a file of the given name, disk and directory location. The Main dialog remains, ready for the next job.

VERIFYING TEXT DURING RECOGNITION

To achieve the highest possible OCR accuracy, even on difficult documents, TextBridge provides a **word verifier** (Figure 3-11).

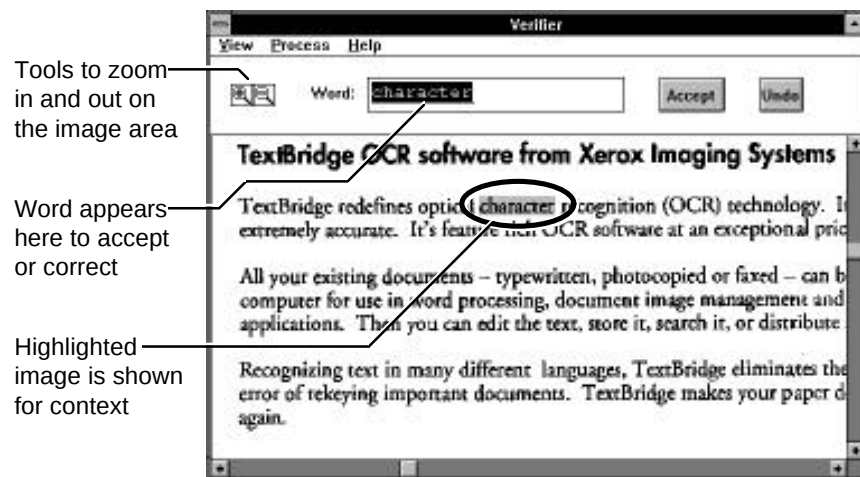


Figure 3–11. Word Verifier window

The word verifier displays **questionable words** for you to correct and/or verify during recognition.

Questionable words are those that fall below a **confidence threshold** built into TextBridge. During OCR, TextBridge assigns a confidence value to each word. If the value falls below the confidence threshold, and you are using the Verifier, TextBridge displays the word as questionable.

By correcting errors and verifying correct words, you help TextBridge improve its recognition accuracy for the rest of the document. TextBridge uses your input to improve recognition decisions as the job progresses.

The Verifier window, which is shown in Figure 3–11, is similar to the Preview window. In part of the Verifier window, TextBridge shows a zoomed image of the page with the questionable word highlighted for context.

Above the image area, zoom tools enable you to zoom further in or out on the page image:



Zoom In



Zoom Out

Next to the toolbox is the **Wordedit box** containing the recognized word, which is highlighted for correction or verification. To the right of the edit box are the Accept and Undo buttons.

To verify words during OCR, use the following procedure.

Note The procedure assumes, if you are scanning, that the scanner is properly connected to your PC and is powered on and ready. It also assumes that TextBridge is properly installed and running.

1. If you are scanning, load the page(s) into your scanner, then go to step 2. Otherwise, start at step 2.

2. In the Main dialog, define the input source.

- Select either File or Scanner in the Input From box (refer to Figures 3–2 and 3–5).
- Click the Verifier checkbox in the Main dialog.

3. Optionally, click the Preferences button to further define the scanning and recognition process.

Use Table 3–1 as a guide. When you are done specifying preferences, click OK to return to the Main dialog.

4. Click GO! in the Main dialog.

If you are scanning with a TWAIN scanner, the native UI for the scanner appears, and you can execute scanning from this interface. If you are scanning with an ISIS scanner, TextBridge automatically scans a page from the scanner.

If you are reading the page image(s) from one or more TIFF files, TextBridge first displays the Open dialog (refer to Figure 3–6). In the Open dialog, specify the name(s) of the TIFF file(s), and their directory and drive location, then click OK.

TextBridge begins the OCR process. When it encounters the first questionable word, it opens the Verifier window, and displays the question-able word highlighted in a Word edit box (refer to Figure 3–11). Below the edit box, the Verifier window displays the image of the word highlighted for context.

5. Verify and/or correct the questionable word in the Word edit box, then click the Accept button.

- + If you make a mistake during verification, simply click the **Undo** button. The edit box restores the last word you edited, and you can then correct the mistake.

You can control the frequency of questionable words to verify with the **Verify** command, which has five possible settings. Pull down the Process menu, select the Verify command, and from its walking menu, select a setting from among Most, More, Normal, Fewer, or Fewest. More and Most will cause more words to be displayed for verification. Less or Least will show fewer words. Normal is the default.

Some documents have words with characters that are not available on your keyboard (accents, symbols, and so on). To verify such characters, pull down the View menu and select **Show Special Characters**. This displays the special character keypad. To enter special characters in the Word edit box, select them from the keypad by clicking on them with the mouse (Figure 3–12).

- + Occasionally, TextBridge mistakes **noise** (marks on the page) or a horizontal line as text. In the Verifier window, the Word edit box contains characters, while the image area shows the non-word highlighted. In such cases, delete all the text in the Word edit box, then click Accept. TextBridge ignores the noise, and proceeds to the next questionable word.

Also, if the image of a particular word is poor in comparison with the rest of the document, you should correct the text without training TextBridge on the image. Simply correct the word in the Word edit box, then hold down the Control key and press the Enter key (or click Accept). The corrected text is output without TextBridge being trained on the image.

The image area in the Verifier window is zoomed in to approximately the middle of the zoom range. If you want to get more of an idea of the page location of the word being verified, click the Zoom Out icon, then click inside the Verifier image area. The full page image appears in the image area of the Verifier window. Conversely, if you want to further magnify the display, use the Zoom In icon.

6. Repeat step 5 until you verify enough words.

- + Verify at least one page of a multiple-page document to teach the system about the entire document.

7. Close the Verifier.

Pull down the Process menu and select the **End Verification** command.

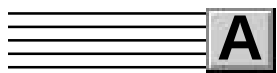
- + The Close command in the Verifier window's control menu is equivalent to the End Verification command.

The Verifier window closes and OCR continues automatically. When you are finished processing all pages, TextBridge displays the Save As dialog (refer to Figure 3–4).

8. Specify the output file information.

- Specify the file name, destination disk and directory.
- Also specify the output text format in the Save File as Type pop-up menu.
- When done, click OK.

TextBridge converts the recognized text and saves the converted text to a file of the given name, disk and directory location. The Main dialog remains, ready for the next job.



TROUBLESHOOTING AND ERROR CORRECTION

TextBridge is designed to be easy to install and use, and, under typical circumstances, you should rarely experience problems.

However, should you encounter a problem during installation and/or use of TextBridge, first consult this appendix to try to resolve the problem yourself.

TextBridge error messages appear in a standard Windows error dialog box

For information to resolve an error condition, refer to the appropriate section in this appendix. This appendix is organized in two sections:

- Troubleshooting common problems
- Correcting error conditions

TROUBLESHOOTING COMMON PROBLEMS

This section describes typical problems with TextBridge, and provides suggestions to resolve them. It also discusses other issues. Specifically, this section is organized into three topics:

- TWAIN scanner problems
- Virtual memory problems
- Other problems

TWAIN scanner problems

TWAIN is an emerging industry standard for the development of scanner and other image-capture device drivers.

TextBridge supports **fully-TWAIN compliant** hand-held scanners and other devices.

For your TWAIN scanner to work with TextBridge, it must have the following software:

- **TWAIN source manager** (TWAIN.DLL)—This software manages the communication between your scanner's TWAIN source driver and TextBridge. It is provided by your scanner manufacturer and must be loaded into the Windows directory, typically C:\WINDOWS.
- **TWAIN source driver**—This is the actual scanner driver. It is provided by your scanner manufacturer and typically is loaded in C:\WINDOWS\TWAIN, or a subdirectory of this directory path.

This section describes some of the problems that you can encounter with a TWAIN scanner while using it with TextBridge, specifically:

- Problems with scanner setup
- Problems with buffered memory

- TWAIN errors

Problems with Scanner Setup

The Scanner Setup program links your scanner's TWAIN source driver with TextBridge.

If your scanner's TWAIN software is not properly installed, you can encounter one of the following problems:

Under the Type menu, TWAIN is dimmed

TextBridge cannot find the TWAIN source manager (TWAIN.DLL) in the Windows directory, or there is no TWAIN subdirectory in the Windows directory.

Check to see that a file named TWAIN.DLL resides in the C:\WINDOWS directory. Check also to see if there is a C:\WINDOWS\TWAIN subdirectory.

If either of these conditions is untrue, repeat all TWAIN installation steps described in your scanner documentation. Verify the existence of the TWAIN source manager and the TWAIN source subdirectory, as above. **Restart your PC.** Then try running Scanner Setup again.

If the TWAIN type is still dimmed in the Scanner Setup Type menu, call XIS Customer Support.

No TWAIN sources installed

Under the Type menu, you have selected TWAIN. When you run the Select Source command under the File menu, this message is displayed. This means that TextBridge cannot locate any TWAIN drivers in the C:\WINDOWS\TWAIN subdirectory.

Repeat all TWAIN scanner installation steps as instructed by your scanner documentation. **Restart your PC.** Then try running Scanner Setup again.

If you continue to get this message, call XIS Customer Support.

No TWAIN source selected

When you run the Acquire command under the File menu, this error message is displayed.

Run Select Source first, select a TWAIN source driver, then run the Acquire command again.

Problems with buffered memory mode

To support TWAIN on 4Mb systems, TextBridge requests the TWAIN source driver to use **buffered memory** mode.

If the TWAIN source driver correctly supports buffered memory mode, it uses no more than 64 kilobytes (Kb) of memory at a time, passing the scanned image to TextBridge in segments.

TextBridge then copies these segments into the memory it has set aside to store the page image it is about to recognize.

Ideally, buffered memory mode reduces the total amount of memory the TWAIN source driver and TextBridge use to manage the scanned image.

Some TWAIN source drivers do not properly support buffered memory mode, and have problems delivering a clean image to TextBridge.

In such cases, the image tends to be severely slanted or otherwise garbled, and TextBridge cannot perform legible OCR on it.

If you encounter this problem, you can direct TextBridge to request **native memory mode**

In native memory mode, the TWAIN source driver allocates enough memory to store the entire page image before it passes it to TextBridge.

To specify native memory mode:

1. Edit the TextBridge initialization file:

c:\windows\txbridge.ini

2. At the bottom of the file, insert the following line exactly as shown.

memory=native

3. Save the changes to the file.

4. Quit TextBridge and start it again, then try scanning with your TWAIN scanner.

Note Native memory mode requires 1Mb of memory. If your PC is a 4Mb machine, you can encounter memory problems running in native memory mode. In this case, you should upgrade your PC to 8Mb of RAM.

Virtual memory problems

Some problems in using TextBridge are related to not allocating enough virtual memory in Windows.

Because TextBridge must process large image files during OCR, the program requires a minimum of eight megabytes (8Mb), and preferably 12 to 16Mb, of **permanent** virtual memory, especially on PCs that only have 4Mb of RAM (random access memory).

Permanent virtual memory is a contiguous block of swap space on your hard drive. It cannot be located on a compressed drive.

Note On systems with 8Mb of RAM, it is possible, although not recommended, to run TextBridge using **temporary** (non-contiguous) virtual memory.

Following are a few examples of problems that can be related to virtual memory:

- error message “Invalid index for language”
 - + This error can also be caused for other reasons; see the section, “Other problems.”
- error message “General protection fault”
- TextBridge hangs while acquiring the image
- the scanner stops during a scan

If you have installed TextBridge and experience problems while using it, check and, if necessary, change your virtual memory setting in Windows. Use the following procedure:

1. From the Program Manager, open the Main program group, and select Control Panel.

2. Double-click the 386 Enhanced icon.

3. Click the Virtual Memory button to display the current swap file size and type (permanent or temporary).

If virtual memory is less than 8Mb (preferably 16Mb) in size, or is temporary, you may need to change it, as described in the next step.

4. Change virtual memory.

- Click the Change button to display Maximum Size, Recommended Size (not in DOS 6.0), and New Size values.
- If Type is not Permanent, specify Permanent. This is **required** on 4Mb systems, recommended on 8Mb systems.
- Using the now-displayed Recommended Size value (not on DOS 6.0), enter this value as the New Size, or simply accept the New Size amount.
 - + If there is not enough contiguous space to create a large enough permanent swap file (8–16Mb), you will need to defragment your disk. Use a utility such as Norton Utilities' SpeedDisk to perform this operation.

With the appropriate virtual memory set up on your Windows-based PC, you should be able to use TextBridge successfully.

If you still encounter any of the problems listed above, or problems that you cannot otherwise resolve, contact TextBridge Customer Support.

Other problems during TextBridge operation

Following are some other common problems that TextBridge users have encountered:

Invalid index for language

If your PC has 4Mb of physical RAM (random access memory), and you get this message while using TextBridge, you are probably running DOS Version 6 with Memmaker, or have a number of devices being loaded into high memory.

If you are using Memmaker, you will have to undo it, as TextBridge requires the space in upper memory that Memmaker allocates to files. Close out of Windows, back up your autoexec.bat and config.sys files, then, at the DOS prompt, type:

```
Memmaker /undo ø
```

If you are not using Memmaker, and you get this error, it may be because you are using too many LH (load high) statements in your autoexec.bat file, or DH (device high) statements in your config.sys file.

Back up these files, then try reducing the number of LH and/or DH statements in these files. Restart your system and try using TextBridge again.

If you still get the error message while running TextBridge, it could be a problem with virtual memory. Refer to the previous section, “Virtual memory problems,” for information.

CORRECTING ERROR CONDITIONS

Occasionally, during TextBridge operation, you may receive an error message. TextBridge error messages are designed to be self-explanatory. Usually, you can simply correct the situation and proceed.

However, if you require more detail about how to correct an error condition, consult this section. Each error message is listed here, along with a description of the cause and a recommended course of action.

Canceling. Page too complex for selected mode.

You are trying to process a complex page with halftones, or which is improperly oriented, without specifying the correct preferences for TextBridge to operate.

In Preferences, specify Auto Page Segmentation and/or Auto Page Orientation, and try again.

Must select a different language

In Preferences, you have selected a recognition language that is not loaded on your system.

Run the TextBridge SETUP program to re-install TextBridge, making sure to include all the recognition languages you intend to use. Refer to Chapter 2 of this manual for information.

If you did install the language pack, and experience this problem, the language pack file in the TextBridge BIN directory has been deleted inadvertently, renamed, or corrupted.

Re-install TextBridge from the original installation diskettes. Refer to Chapter 2 of this manual for information.

Parameter combination not supported

This error is usually the result of trying to have TextBridge process a TIFF file that is inappropriate for OCR. For example, the TIFF file could be extremely low resolution (lower than 100-by-200 dpi), could be color or grayscale, or some combination of the two.

TextBridge can only process binary (black and white) TIFF images with resolutions greater than 100-by 200 dpi.