

TUTORIALS

This chapter provides step-by-step tutorials designed to introduce you to some of the most important capabilities of TextBridge Pro 96.

To familiarize you with TextBridge Pro, the following topics are presented in this chapter:

- u TextBridge Pro interface
- u sample documents
- u automatic operation
- u capturing parts of a document (preview)
- u interactive training
- u Instant-Access OCR™ (running TextBridge Pro from within a text application)
- u document recomposition (this requires a word processor that supports multiple columns and embedded images, for example Word for Windows 6.0)
- u performing post-recognition proofing and editing from within your text application (this exercise requires Word for Windows 6.0 or later, or WordPerfect for Windows 6.0 or later)

TEXTBRIDGE PRO INTERFACE



This section provides a brief overview of the TextBridge Pro interface.

To start TextBridge Pro in Windows 95, double-click the Shortcut to Tb95 icon on the desktop. (Setting up this shortcut is explained in the Quick Start section at the beginning of this guide).

Otherwise, start TextBridge Pro in Windows 95 as follows:

1. **Click the Start button on the Windows taskbar.**
2. **Point to Programs, then point to the TextBridge Pro 96 folder.**
3. **Click the TextBridge Pro 96 icon.**

The TextBridge Pro main window is displayed (Figure 3–1).

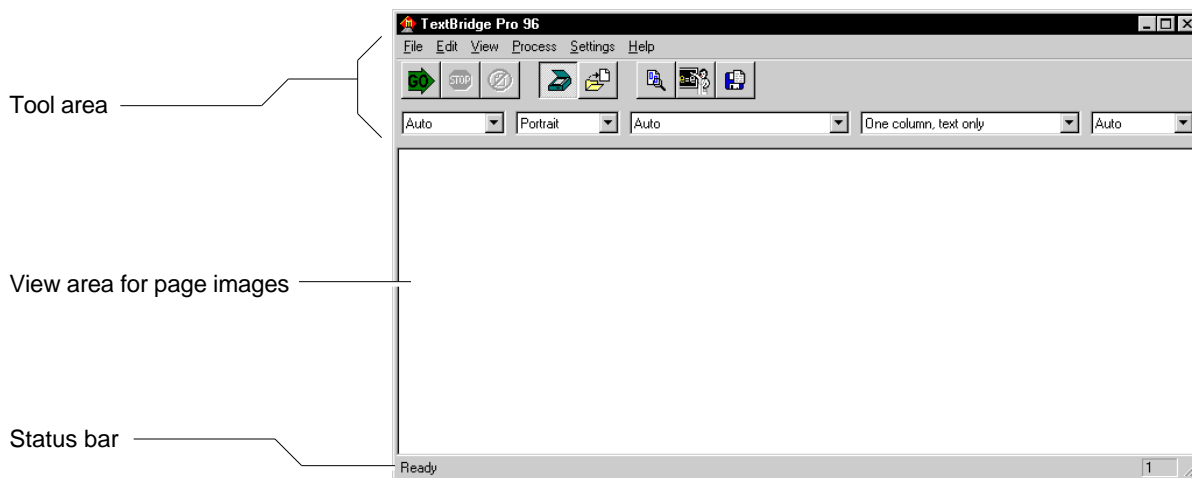


Figure 3–1. *TextBridge Pro main window*

The main window shown in Figure 3-1 follows the standards for Windows 95. The upper right corner provides standard Minimize, Maximize, and Close buttons.

Below the TextBridge Pro 96 title bar, the main window provides a **menu bar**, a **main toolbar**, and a **preferences panel**. These tools let you set up, start, and control the document recognition process. Note that when you place the cursor in the view area and click the right mouse button, a pop-up menu of TextBridge Pro commands appears.

At different stages of the process, additional toolbars—the **preview toolbar** and interactive **training toolbar**—are added automatically beneath the preferences panel (Figure 3-2).

Below the toolbar area, the largest area of the main window is the **view area**. Here, depending on the processing stage (preview, training, feedback), different views of the page image appear.



Figure 3-2. *Preview and training toolbars*

At the bottom of the main window, the **status bar** provides descriptions of toolbar tools when you place the cursor over them. During processing, it also displays messages that indicate the status of the job. Near the right edge of the status bar, an inset area contains the current **page number** being processed.

Except for an occasional **dialog box**, all program activities take place in the main window. It is your control center for TextBridge Pro's powerful document recognition tools.

SAMPLE DOCUMENTS



For use with the tutorial sessions provided in this chapter, TextBridge Pro provides five sample documents located in the installation subdirectory:

C:\Program Files\TextBridge Pro 96\tiffs

The sample documents are stored in TIFF format and are named:

3col	Markplan	Plexis
Resume	Zonepic	

The sample documents provide a cross-section of the document types that TextBridge Pro can process. They are designed to highlight the wide range of capabilities the program provides.

- + Refer to Appendix B of this document for descriptions and pictures of the sample documents.

In each of the six tutorial sessions provided in this chapter, you are asked to use a specific sample document.

To identify an on-line sample document for TextBridge Pro to use, follow these steps:

1. **From the TextBridge Pro main toolbar, push in the Read Image from File button:**



2. **Click the Go button.**



TextBridge Pro now displays the Open dialog (Figure 3-3).

Usually, the default directory (tiffs, that is, C:\Program Files\TextBridge Pro 96\tiffs) is open.

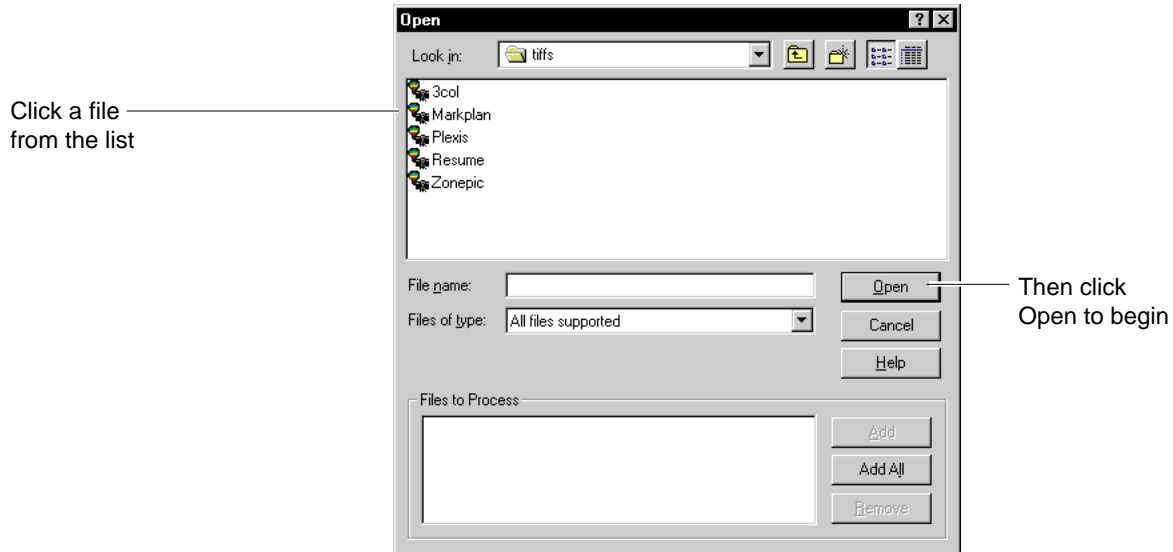


Figure 3-3. *Open dialog*

3. **If necessary, access the sample documents directory from the Open dialog:**

`C:\Program Files\TextBridge Pro 96\tiffs`

The five sample TIFF files should now be listed in the Open dialog.

4. **Double-click a file name to open it.**

After TextBridge Pro reads the on-line image and preprocesses it, the Save As dialog appears (Figure 3-4).

5. **Click Cancel to close the Save As dialog without saving the file.**

Please now proceed to the tutorial sessions to work with TextBridge Pro, and familiarize yourself with its capabilities.

SESSION 1: AUTOMATIC OPERATION



TextBridge Pro provides a range of features that no other OCR program can match.

However, TextBridge Pro is also designed to be very easy to use. For most documents, you can use default settings, and simply press the **Go** button to start the document recognition process.

- + For this example, use the on-line sample document named Markplan.
- 1. **On the main toolbar, select the Read Image from File button.**
- 2. **Click the Go button on the main toolbar.**
- 3. **In the Open dialog (refer to Figure 3–3), locate and select the sample document, Markplan, then click Open.**

TextBridge Pro reads the on-line image, and automatically performs OCR on it, as indicated by the feedback display in the view area of the main window. When it is finished, it displays the Save As dialog (Figure 3–4).

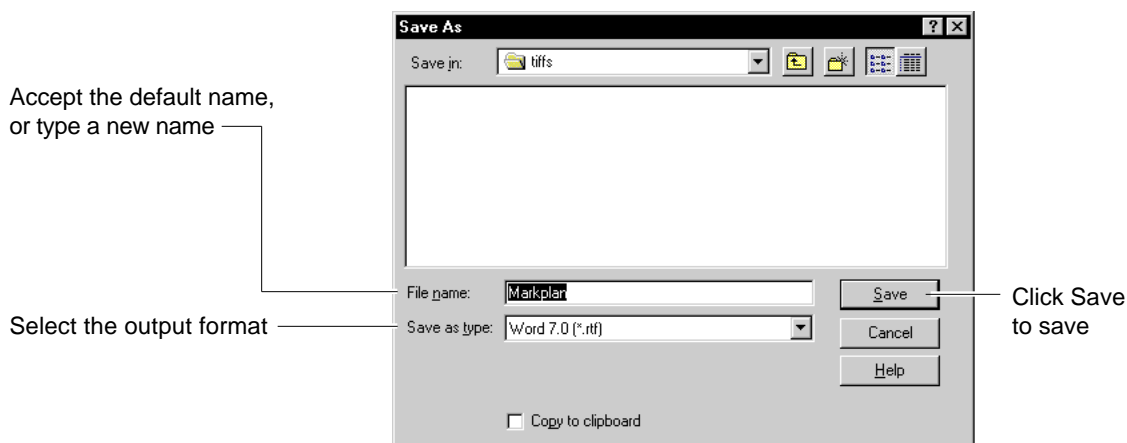


Figure 3-4. *Save As dialog*

4. In the Save As dialog, define the output file.

In the Save as type combo box, select the output format for your word processor or other text application.

In the File name box, type a file name, using the file extension provided (for example, if you selected Word, the default file extension is `.rtf`).

Click Save to convert and save the recognized data.

5. Open the file with your word processor or other text application.

Compare the recognized document in your word processor with the picture of the sample document, `markplan.tif`, in Appendix B of this document. With a word processor such as Word or WordPerfect, the recognized document should look virtually identical to the TIFF image. The difference is now you have formatted, fully editable text.

SESSION 2: CAPTURING PARTS OF A DOCUMENT



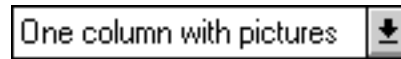
TextBridge Pro also enables you to capture selected parts (text and graphics) of a document. For this purpose, TextBridge Pro provides **preview** tools.

- + For this example, use the on-line sample document named `zonepic`. Note that you also need a text application that can display both text and graphics, such as WordPerfect.

1. On the main toolbar, push in the Read Image from File and Preview buttons.



2. In the Preferences Panel, fourth combo box (Output Document settings), select One column with pictures.



3. Click the Go button on the main toolbar.
4. In the Open dialog, locate and select the sample document, Zonepic, then click Open.

TextBridge Pro reads the on-line image, and in a few moments, displays it in the view area of the main window (Figure 3–5).

5. Zoom in on the page.

Select the Zoom In tool:



Position the mouse inside the view area at the upper left corner of the page image, and click once to magnify this area of the page.

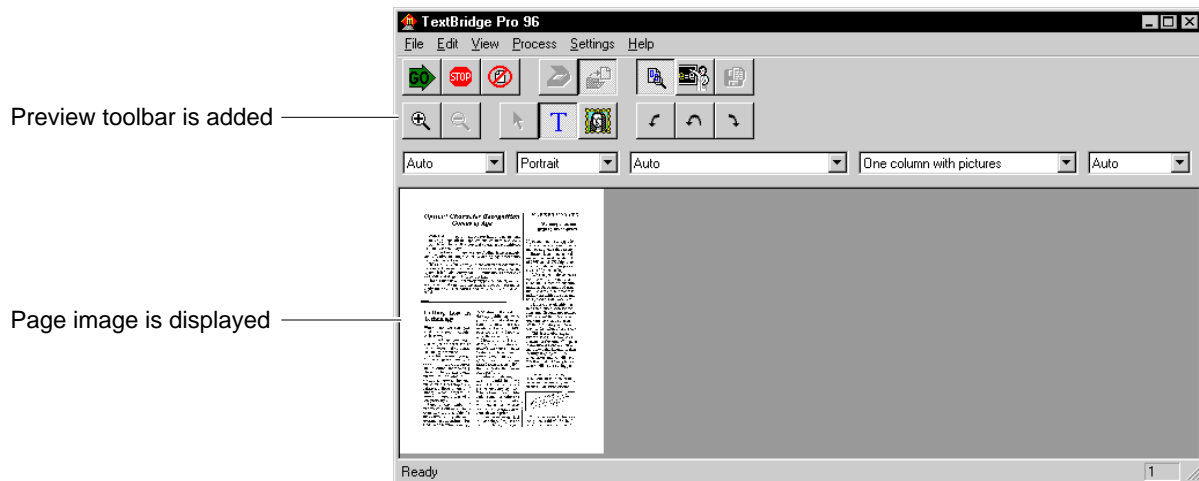


Figure 3–5. Main window in preview mode

6. Create a text zone.

Select the Text Zone tool:



Position the mouse inside the view area at the upper left corner of the page image.

Holding down the left mouse button, drag the mouse diagonally downward until the text zone rectangle outlines a block of text to be recognized (Figure 3–6).

Release the mouse button.

Text zone identifies area
to be recognized

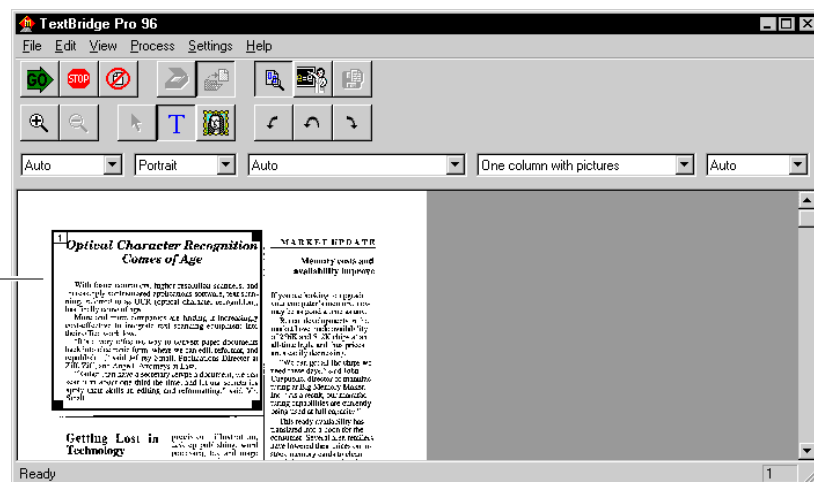


Figure 3–6. Text zone on previewed page

7. Use the vertical scroll bar to the right of the view area to display the bottom of the scanned page image.
8. Again click the Zoom In tool, and click once on the line art at the bottom right of the page image.
9. Create an image zone.

Click the Image Zone tool:



Position the mouse at the upper left of the line art on the previewed page.

Holding down the left mouse button, drag the mouse diagonally downward until the image zone rectangle outlines the line art (Figure 3-7). Release the mouse button.

Image zone identifies graphic to be captured

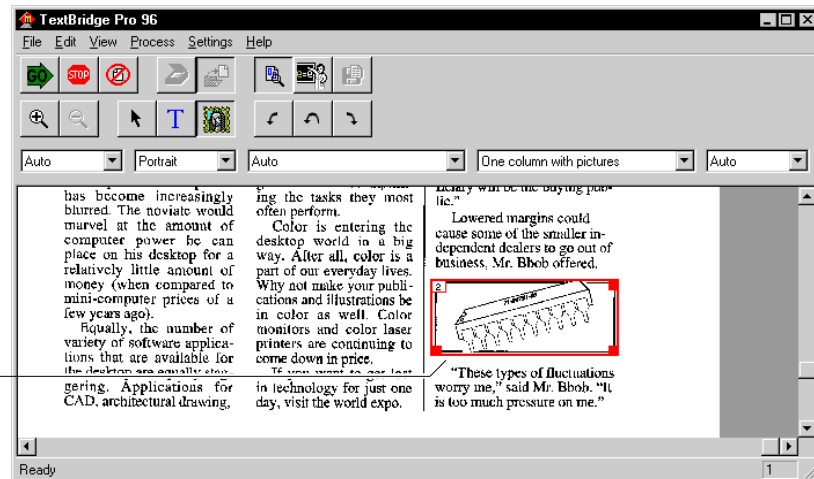


Figure 3-7. Image zone on the previewed page

10. **Click the Go button again to process the zoned text and image.**

When processing is complete, TextBridge Pro displays the Save As dialog (refer to Figure 3–4).

11. **Define the output text file, then click Save in the Save As dialog.**

12. **Open the file with your word processor or other text application.**

The text contained in the text zone you created in Step 6 is followed by the line art in the image zone you created in Step 9.

SESSION 3: INTERACTIVE TRAINING



To assure the highest possible accuracy, TextBridge provides an **interactive training** capability. This feature enables you to participate in the OCR process, verifying correctly recognized words and correcting recognition errors.

Interactive training is especially effective for degraded documents, such as faxes and multi-generation photocopies.

As you correct and/or accept TextBridge Pro's recognition decisions for a page or two, you also train the program to improve its own accuracy rate for later pages of the document.

In addition, you can save and later reload **training data** to assure that other documents of the same type are recognized with the same high degree of accuracy.

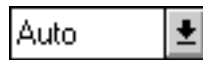
- + For this example, use the on-line sample document named Plexis.

1. On the main toolbar, push in the Read Image from File and Train OCR buttons:



+ Make sure the Preview button is not pushed in.

2. In the Preferences Panel, make sure the first combo box (Original document quality) is set to Auto or Fax.



3. Click the Go button on the main toolbar.
4. In the Open dialog, locate and select the sample document, Plexis, then click Open.

When TextBridge Pro finds the first **suspect word**, it adds the **training toolbar** to the main window and displays the suspect word in the Word edit box. The image of the word is highlighted immediately below in the view area (Figure 3–8).

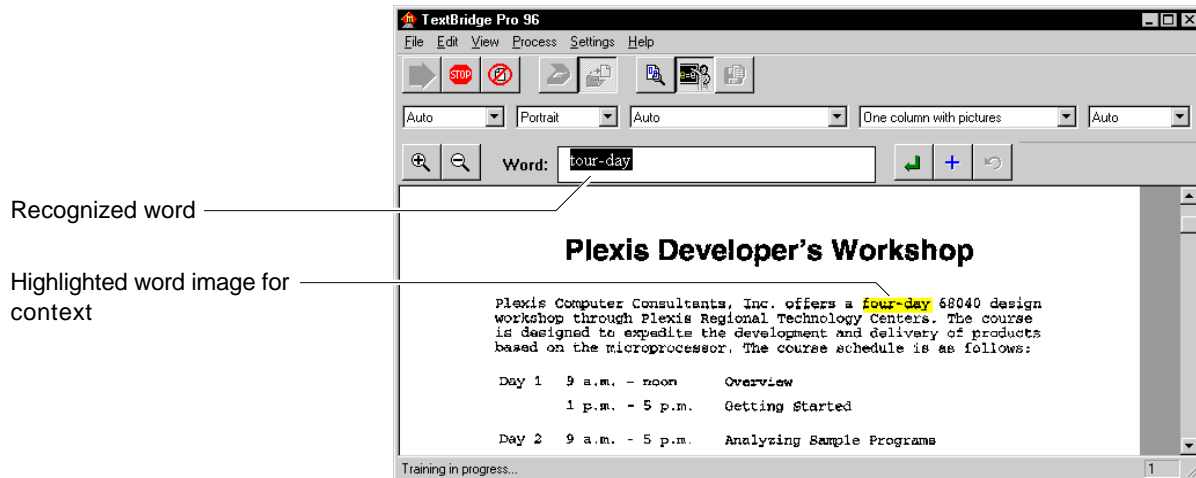


Figure 3–8. Main window in interactive training mode

5. Go on to train TextBridge Pro on the suspect words.

If the word is correct, simply press the Enter key on your keyboard, or click the Accept button on the Training toolbar:



If the word is incorrect, correct it in the Word edit box, then accept it.

Continue correcting and/or accepting words at your option. You can complete the entire page or only a portion. Generally, it is recommended that you complete at least one full page of a multi-page document to sufficiently train TextBridge Pro about the character shapes and sizes for that document.

For the purposes of this example, correct and accept about 10 words, then turn training off by clicking the Train OCR button on the main toolbar again.

TextBridge Pro finishes recognition of the page, then provides a prompt that asks:

Do you want to save training data?

6. Click No to discard training data.

TextBridge Pro now displays the Save As dialog (refer to Figure 3-4).

7. Define the output text file, then click Save in the Save As dialog.

TextBridge Pro converts the document to the selected format.

8. Open the file with your word processor.

Notice that, even though the input document was a low-quality fax image, TextBridge Pro recognized it with a high degree of character recognition and formatting accuracy.

SESSION 4: INSTANT-ACCESS OCR

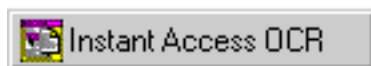


TextBridge Pro is the only high-end document recognition program that can be launched from virtually any Windows text application, and can automatically paste recognized data directly into that text application's open document.

This capability is referred to as **Instant-Access OCR™**.

When you install TextBridge Pro (refer to Chapter 2), the setup program automatically places a copy of the Instant-Access OCR program in your Windows StartUp folder.

So, whenever you start Windows, the program is automatically launched and minimized on your Windows desktop:



This assures that TextBridge Pro is always immediately available from your text application.

- + For this example, use the on-line sample document named Resume.
- 1. **From Windows, launch your text application, and open a new document.**
- 2. **Verify that the Instant Access OCR program is active on your Windows desktop.**

Go directly to Step 3 if the Instant Access OCR program is already active on your Windows desktop. (It's a minimized button on the taskbar.)

If the program is not already running, click the Start button on the Windows taskbar. Point to Programs, then point to the TextBridge Pro 96 folder. Click the TextBridge Pro 96 Instant Access OCR icon.

The program starts up minimized on your Windows desktop.

3. Register your text application.

Click the Instant Access OCR button on the taskbar to display the Registration dialog (Figure 3–9). In the Registration dialog, your text application should appear in the top list.

Click your text application in the top list, and click the Add button to add it to the bottom list of registered applications. Skip this step if your text application already appears in the bottom list.

Click the OK button in the Registration dialog. The Instant Access OCR program is again minimized on your Windows desktop.

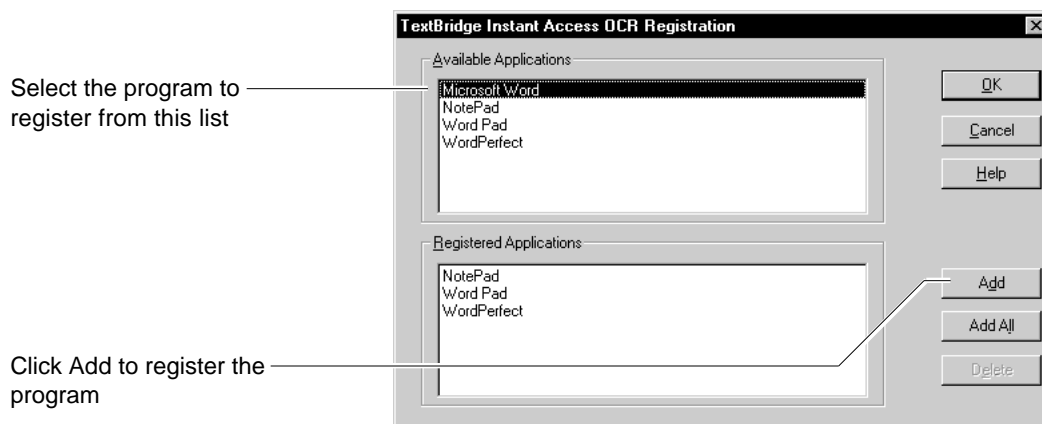


Figure 3–9. *Instant Access OCR Registration dialog*

4. Access TextBridge Pro from your text application.

Switch back to your text application.

Pull down the File menu of your text application, and select the now-present **TextBridge OCR** command. The TextBridge Pro main window now appears (refer to Figure 3–1).

Otherwise, switch to TextBridge Pro, if it's already running, by clicking its button on the taskbar.

- + Make sure that Save Proofing Data in the Process menu is not checked.
- 5. **On the main toolbar, select the Read Image from File button.**
- 6. **Click the Go button on the main toolbar.**
- 7. **In the Open dialog, locate and select the sample document, Resume, then click Open.**

TextBridge Pro reads the on-line image and automatically performs OCR on it.

The program converts the recognized text to two formats (RTF and Windows ASCII), and copies it to the Windows clipboard. It then automatically pastes one of the formats into your text application's open document.

- + If your application can import RTF (Microsoft's Rich Text Format), then the RTF-formatted data on the clipboard is pasted. Otherwise, Windows-ASCII is pasted. Formatting characteristics, such as bold, italic, underlining, columns, and so on, are lost.

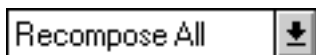
When the paste is complete, the TextBridge Pro main window closes.

- 8. **You can now go on to edit or otherwise use the recognized text in your text application.**

The TextBridge Instant Access OCR program remains active and minimized on the Windows desktop.

- 9. **To close the Instant Access OCR program, click its button on the taskbar with your right mouse button, then click Close.**

SESSION 5: DOCUMENT RECOMPOSITION



TextBridge Pro is the first OCR application capable of recomposing the layout of a document, including text and graphics, while maintaining full editability in the output file.

Consider, for example, a three-column newsletter that includes a picture. Assuming your text application supports these elements, TextBridge Pro not only can recognize the text, it can correctly recompose the column layout and output a copy of the picture in its original location.

- + For this session, use the sample document, 3col. Also, for full recomposition, TextBridge Pro supports output only to Word for Windows (2.x, 6.x, and higher than 6.x) and WordPerfect 6.1 (or higher).

1. **Launch TextBridge Pro. On the main toolbar, select the Read Image from File button.**

2. **Set up preferences appropriately.**

In the Preferences Panel, in the third combo box (Original document layout), select the Auto setting.

In the Preferences Panel, in the fourth combo box (Document recomposition), select Recompose all.

3. **Click the Go button on the main toolbar.**

4. **In the Open dialog, locate and select the sample document, 3col, then click Open.**

TextBridge Pro reads the on-line image, and automatically performs OCR on it. When it is finished, it displays the Save As dialog (refer to Figure 3-4).

5. **Define the output text file, then click Save in the Save As dialog.**
6. **Open the file with your word processor.**

Display the document so that its full layout is shown. For example, in Word for Windows 6.0, you must select the Page Layout command from the View menu.

Notice that the document is composed in three columns, and that the picture of the butterfly is output in the center of the second column.

Now click inside the first column, and start typing. Notice that the text flows appropriately across column boundaries. TextBridge Pro is the only OCR program that provides true editability in the recomposed output.

Notes About Document Recomposition

For some documents, you may want only the text in simple galley (one-column) form. Recomposition modes in TextBridge Pro are not appropriate here.

TextBridge Pro provides two recomposition modes. In one, Recompose Text, only the text is recomposed. This is useful, for example, when your document contains cell tables and you want them output as **cell tables**. (Word and WordPerfect support tabular data in gridded rows and columns; these are referred to as cell tables.)

In the second recomposition mode, Recompose All, text and automatically-detected (or manually-zoned) graphics are output in their original column layout and locations.

For best results in recomposing a document with graphics, you can use TextBridge Pro's **Smart Zones™** feature in preview mode. Before you process a page, display it in preview mode, and create image zones around **all** the graphics on the page (halftones and line art). Then click Go to process the page. The original column layout is maintained, and the pictures are output appropriately.

The reason why Smart Zones improve recomposition is that, quite frequently, pictures in a document include text, or some element that TextBridge Pro mistakes for text. Without Smart Zones, this can disrupt the recomposition process. With Smart Zones, TextBridge Pro only recognizes the areas not zoned.

It is important to note that, in reconstructing the layout of the original document, TextBridge Pro is limited by the composition capabilities of the target word processor.

You should **not** expect to scan a complex magazine page originally created with a desktop publishing program and get identical output in your word processor. Even the most powerful word processors (Word, WordPerfect) do not have some of the composition capabilities of desktop publishing software.

In addition, some free-form layouts defeat TextBridge Pro's recomposition capabilities. For these types of documents, it is often best to preview pages and manually zone text and image zones that you want to capture.

SESSION 6: IN-PLACE PROOFING

A rectangular button with a light gray background and a thin black border. The word "Replace" is written in a bold, black, sans-serif font, centered within the button.

Replace

If you work with Word for Windows 6.0a (or later) or WordPerfect 6.1 (or later), TextBridge Pro provides a unique proofing capability to assure the highest productivity and OCR accuracy.

The **TextBridge Proofreader** capability is embedded directly in your word processor, providing powerful proofing and editing tools to verify or correct suspect words in the recognized output document.

- + For this example, use the on-line document, `Plexis`. Also, you must have either Word for Windows 6.0a (or later) or WordPerfect for Windows 6.1 (or later) to use the TextBridge Proofreader capability.

1. **Install the TextBridge Proofreader capability in your word processor.**

Please refer to Chapter 2 of this guide for instructions.

2. **Launch TextBridge Pro. From the Process menu, select:**

`Save Proofing Data`

3. **On the main toolbar, select the Read Image from File button.**

Make sure all state buttons on the main toolbar are off.

4. **Click the Go button on the main toolbar.**

5. **In the Open dialog, locate and select the sample document, `Plexis`, then click Open.**

TextBridge Pro reads the on-line image, and automatically performs OCR on it. When it is finished, it displays the Save As dialog (refer to Figure 3-4).

6. Define the output file, then click Save in the Save As dialog.

Save the file either in Word for Windows 6.0 (or later) or WordPerfect 6.1 (or later) format.

Click Yes if you are prompted to replace the existing file.

7. Open the file with your word processor.

Notice that some of the words in the document are shown in a color other than black. These are **suspect words** and their colors indicate the level of uncertainty that TextBridge has assigned to them, as follows:

Green	=	High confidence
Dark Yellow	=	Medium confidence
Red	=	Low confidence

8. Pull down the word processor's File menu, and select the TextBridge Proofreader command.

The TextBridge In-Place Proofreader dialog now appears in the middle of your screen (Figure 3–10). The first suspect word in the document is shown in the Suspect edit box.



Figure 3–10. *TextBridge Proofreader dialog*

- + You may find it easier to perform proofing by moving the dialog to a corner of your monitor, so you can see as much of the actual document as your screen size allows.

9. Go on to proof suspect words in your document.

Following are a few features of the TextBridge Proofreader that you should note:

To proof a suspect word, correct it in the Suspect edit box if needed, then click Replace. If the word is correct already, simply click Replace to move to the next word.

To start, you may want to correct all the Low confidence (red) words first. Set the Stop At category to Low Confidence only. Then as you proceed, the TextBridge Proofreader will move only through the low confidence words.

To speed up the process, you can set the Accept All range accordingly, and simply Accept All. For example, if the High confidence (green) and Medium confidence (dark yellow) words look mainly correct, you can click these items on in the Accept All range, then click Accept All.

If TextBridge Pro seems to have repeatedly made the same recognition error on all instances of a suspect word, you can correct it once, then click the Replace All button.

If a suspect word is garbled beyond recognition, you can click the Show Image button to display the original scanned image of that word (Figure 3–11).

Suspect word image is highlighted and shown in context

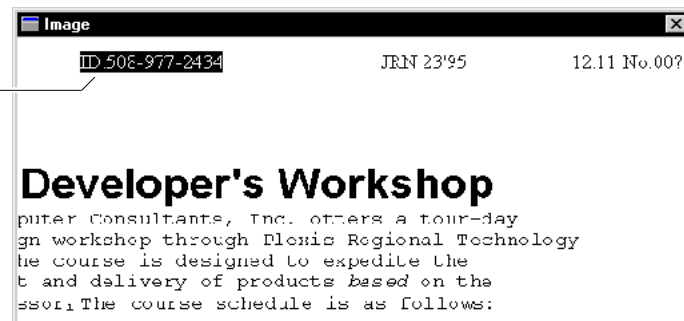


Figure 3–11. TextBridge Proofreader Image window

When you have completed post-recognition proofing, the TextBridge Proofreader dialog (and Image window, if showing) closes automatically.

If you do not want to proof all suspect words, you can click Close any time to end proofing. If you do, a dialog asks you if you want to save the remaining proofing data. Click Yes or No as desired.

After you are done with proofing, you can still run the word processor's native spell-checker to verify that all errors have been corrected.

WHERE TO GO FROM HERE



The tutorial sessions in this chapter were designed to give you a solid basis on which to use TextBridge Pro for your own documents.

For complete information about TextBridge Pro, please refer to the on-line Help built into the application.