

## **4 TIPS UND TECHNIKEN**

Die ersten drei Kapitel dieses Benutzerhandbuchs sollen Ihnen helfen, TextBridge zu verstehen, zu installieren und seine Hauptfunktionen anzuwenden.

Dieses Kapitel beschäftigt sich mit der Optimierung von OCR-Ergebnissen. Die Themen lauten im einzelnen:

- Optimale Texterkennung
- Steigerung der Effizienz bei der Vorlagenverarbeitung
- Seite als Bild speichern
- Starten der TextBridge-OCR aus anderen Anwendungen heraus

### **OPTIMALE TEXTERKENNUNG**

Die TextBridge OCR-Software erreicht eine gleichbleibend hohe Qualität der Zeichenerkennung bei verschiedensten Vorlagen. Sie können jedoch auch selbst dazu beitragen, daß die Zeichenerkennung bei einem bestimmten Dokument wirklich optimal verläuft.

Dieser Abschnitt geht hierzu auf folgende Aspekte ein:

- Ordnungsgemäße Verwendung und Pflege Ihres Scanners
- Einstellung der Helligkeit
- Einstellung von Farben
- Verwendung des Faxfilters
- Separate Verarbeitung mehrerer Vorlagen
- Verwendung der Prüffunktion
- Verwendung des Befehls "Invers" bei der Seitenvorschau



## Ordnungsgemäße Verwendung und Pflege Ihres Scanners

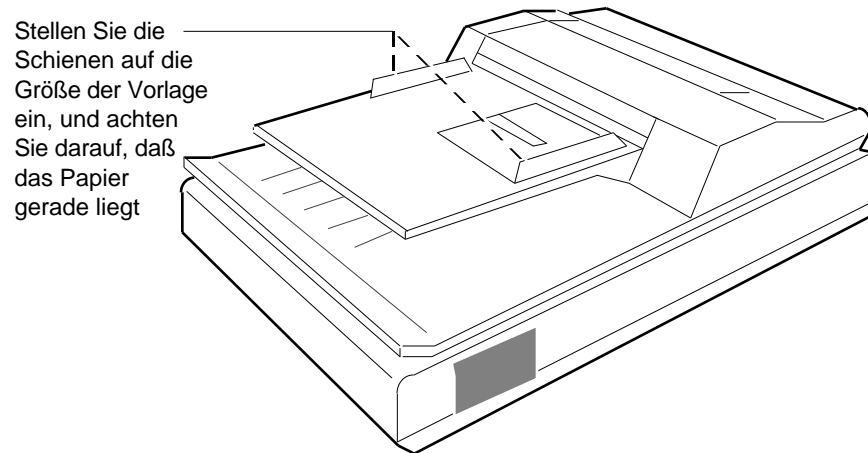
Die Verwendung und Wartung Ihres Scanners hat großen Einfluß auf die Qualität des Scannens. Sie sollten daher die folgenden Ratschläge beherzigen:

- **Lernen Sie Ihren Scanner genau kennen.**  
Lesen und verstehen Sie die gesamte zu Ihrem Scanner gehörige Dokumentation.
- **Pflegen Sie Ihren Scanner.** Halten Sie ihn möglichst staubfrei. Säubern Sie das Flachbett von Schmutz und von Rückständen, die beim Scannen entstanden sind.
- **Laden Sie den Scanner ordnungsgemäß.**  
Sorgen Sie dafür, daß die Vorlage nicht schief gescannt und somit die Zeichenerkennung erschwert wird.

Wenn Sie die automatische Vorlagenzuführung verwenden, sorgen Sie dafür, daß der Papiereinzug für die zu scannenden Seiten ordnungsgemäß ausgerichtet ist (Abbildung 4–1).

Sorgen Sie bei der Benutzung des Flachbetts dafür, daß das Seitenbild jeweils glatt und gerade auf der Auflagefläche liegt. Manchmal liegt die eigentliche Vorlage schief im Vergleich zu dem Papier, auf das sie gedruckt werden soll. Korrigieren Sie dies so weit wie möglich vor dem Scannen.





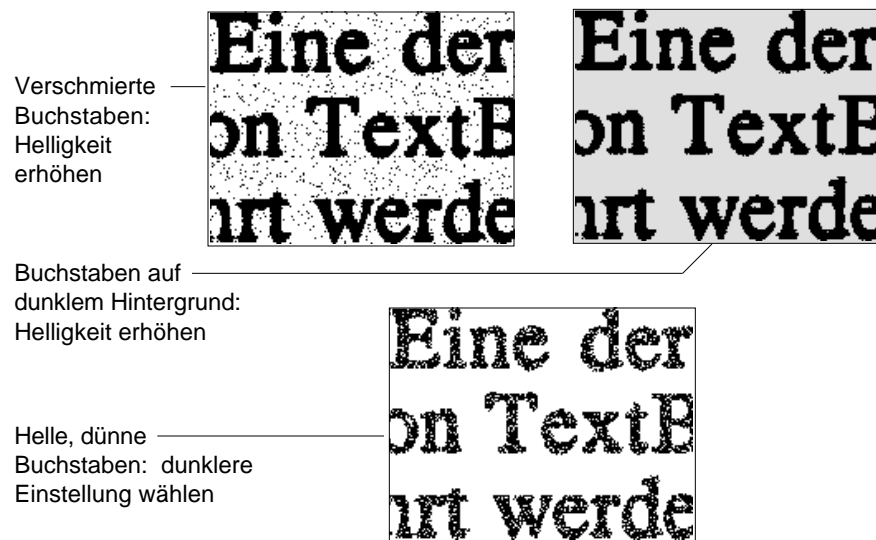
**Abbildung 4–1.** Seitenplatzierung auf dem Scanner

## Einstellung der Helligkeit

Eine der wichtigsten Einstellungen beim Scannen ist die **Helligkeit**. Wie in der Abbildung 4–2 dargestellt, können die zu scannenden Vorlagen sehr unterschiedlich sein.

Die Text- und Hintergrundfärbung, die Menge der **Störfaktoren** (Schmutz, verschmierte Tinte, Fingerabdrücke, Handschrift und anderes) auf der Seite können die Erkennungsgenauigkeit stark beeinflussen.



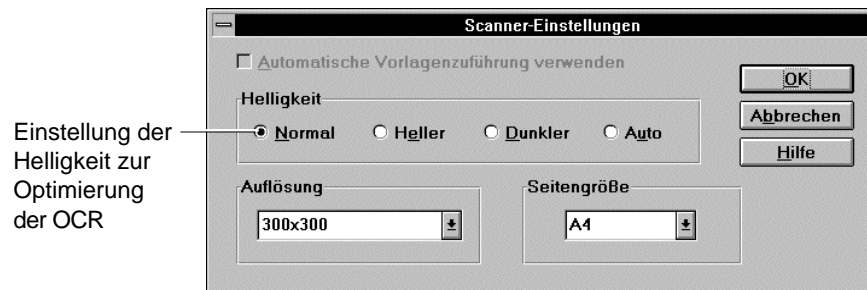


**Abbildung 4–2.** Vorlagenoriginale und Helligkeit

Im Dialogfeld “Scanner-Einstellungen” (Abbildung 4–3) können Sie durch Einstellen der Helligkeit eine schlechtere Druckqualität und einen dunklen Vorlagenhintergrund kompensieren. Auf die Scanner-Voreinstellungen können Sie vom Dialogfeld “Voreinstellungen” aus zugreifen.

Wenn die Buchstaben auf Ihrer Seite zu fett erscheinen, sich gegenseitig berühren oder zu geringe Abstände aufweisen (wie in manchen Zeitschriften), versuchen Sie es mit einer helleren Einstellung. Das Erkennen von Vorlagen mit Störfaktoren, Rastern oder Farbe im Hintergrund wird durch Erhöhen der Helligkeit beträchtlich verbessert.





**Abbildung 4–3.** Helligkeitseinstellungen des Scanners

Eine dunklere Einstellung ist dann empfehlenswert, wenn die Buchstaben auf Ihrer Seite schwach, brüchig oder dünn erscheinen.

Falls die Einstellung “Automatisch” von Ihrem Scanner unterstützt wird, erreichen Sie damit die besten Resultate.

Eine andere Art zur Feststellung der richtigen Helligkeitseinstellung ist die maximale Vergrößerung der Seite im Vorschaufenster. Sie zeigen damit genau das an, was auch der Scanner “sieht” (Abbildung 4–4). Wenn die angezeigte Seite nicht die richtige Helligkeitseinstellung aufweist, korrigieren Sie die Einstellung und scannen Sie die Vorlage im Anschluß daran.





Abbildung 4–4. Vergrößerte Vorschauanzeige

## Einstellung von Farben

Alle Scanner können eine oder mehrere Farben nicht lesen. Diese Farben werden **Blindfarben** genannt. Finden Sie anhand Ihres Scanner-Handbuches die Blindfarben Ihres Scanners heraus.

- + Wenn Ihre Scanner-Dokumentation nicht auf Blindfarben eingeht, untersuchen Sie die Farbe des Scanner-Lichtes, während sich der Scanner über das Flachbett bewegt. Diese Farbe ist dann die Blindfarbe. Viele Geräte haben ein hellgrünes Scanner-Licht – die entsprechende Blindfarbe ist also hellgrün.

Weiterhin gibt es Farben, die Ihr Scanner nur mit Schwierigkeit erkennt. Wenn Ihr Text/Ihr Bild farbig ist oder sich auf einem farbigen Hintergrund befindet, müssen Sie die Helligkeit entsprechend regulieren.

Wenn das nicht funktioniert, fotokopieren Sie die Seite, und scannen Sie die Schwarzweißkopie.



## Verwendung des Faxfilters

Eine Verwendungsmöglichkeit von TextBridge ist das Erkennen von Text in Faxabbildern. Faxabbilder haben im Normalfall eine niedrige Auflösung (100 mal 200, 200 mal 100 oder 200 mal 200 dpi). Auch die Faxe mit einer sogenannten hohen Auflösung von 200 mal 200 dpi sind oft nur schwierig zu lesen (Abbildung 4–5).

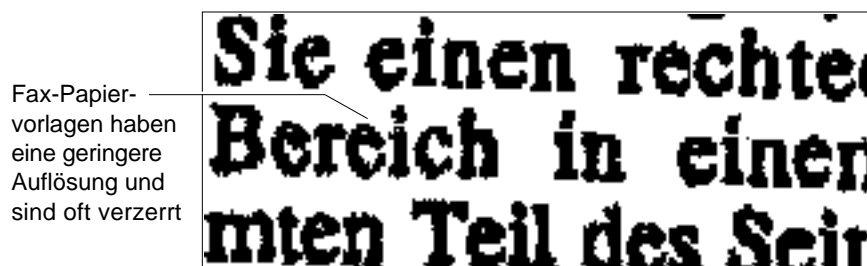


Abbildung 4–5. Faxabbild

Zur Erkennung von Faxabbildern verfügt TextBridge über eine **Faxeinstellung** im Dialogfeld “Voreinstellungen”. Dieser Filter leitet eine Vorverarbeitung ein, die das Faxabbild noch vor der OCR verbessert.

Die Faxoption funktioniert bei Faxabbildern, die in TIFF-Dateien gespeichert sind **und** bei Fax-Papiervorlagen, die bei höherer Auflösung gescannt werden (z.B. 300 dpi).



---

**Hinweis** Verwenden Sie den Faxfilter **nicht** für andere Vorlagen (d.h. Vorlagen, die keine Faxvorlagen sind), egal ob sich um zu scannende oder Online-Vorlagen handelt. Andernfalls verschlechtert sich eventuell die OCR-Genauigkeit. Wenn Sie weiterhin feststellen, daß die OCR-Qualität bei elektronisch erstellten Faxabbildern (z.B. eine von einem Faxmodem gesendete Textverarbeitungsvorlage), zu wünschen übrig läßt, schalten Sie den Faxfilter aus, und versuchen Sie, einen Teil des Textes während der OCR zu überprüfen.

---

## Separate Verarbeitung mehrerer Vorlagen

Die TextBridge-OCR-Software verwendet eine Reihe von Texterkennungstechniken.

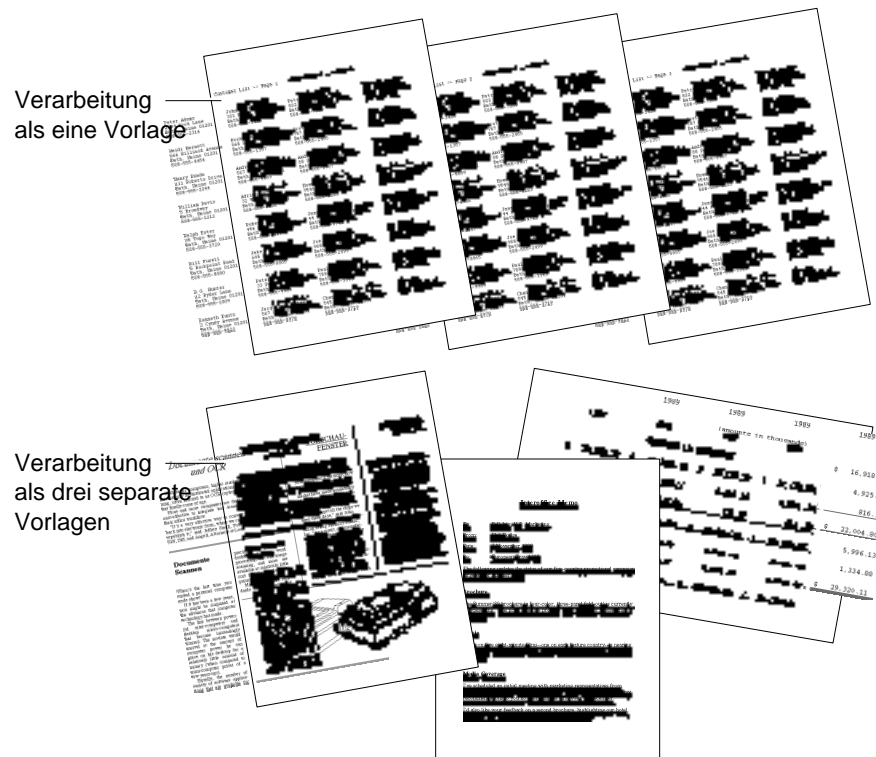
Mit Hilfe dieser Techniken bringt TextBridge sich Teile dieses Erkennungsvorgangs sogar selbst bei.

TextBridge kann daher die Erkennungsgenauigkeit und -geschwindigkeit während des Scannens und Erkennens aufeinanderfolgender Seiten steigern.

Diese Lernfähigkeit setzen Sie jedoch aufs Spiel, wenn Sie die Seiten verschiedener Vorlagen an die gleiche Ausgabedatei weiterleiten.

TextBridge nimmt an, daß die zweite und die folgenden Seiten einer Vorlage die gleiche Schriftart enthalten, wie die erste erkannte Seite (Abbildung 4–6).





**Abbildung 4-6.** Verarbeitung mehrerer Vorlagen

Wenn die zweite Seite zu einer völlig anderen Vorlage mit anderen Schriftarten und -größen gehört, wird das Wissen, das TextBridge über die ersten Seiten erworben hatte, nutzlos.

TextBridge muß dann für die zweite und die folgenden Seiten mit dem Lernprozeß von vorn beginnen.

Wenn Sie mehrere Vorlagen mit besten Erkennungsergebnissen scannen wollen, scannen Sie jede Vorlage einzeln.



## Verwendung der Prüffunktion

Wenn Sie mit den Erkennungsergebnissen bei einem bestimmten Dokument nicht zufrieden sind, können Sie die Erkennungsgenauigkeit mit Hilfe der Prüffunktion verbessern.

Durch Ihre Interaktion mit dem OCR-Prozeß im Prüfungsfenster, (Abbildung 4–7) geben Sie TextBridge Informationen über die Buchstaben und Wörter in der Vorlage. Die Erkennungsgenauigkeit kann dadurch bedeutend gesteigert werden.

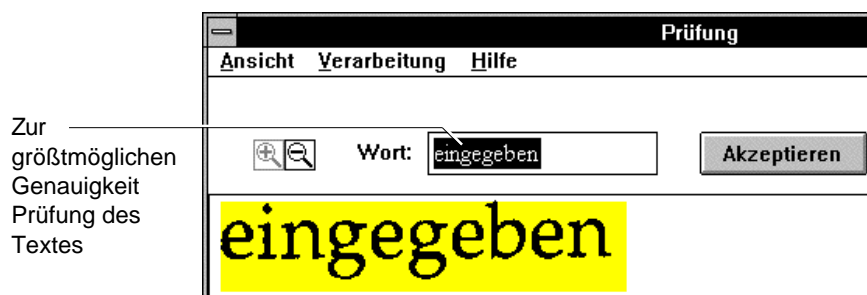


Abbildung 4–7. Das Prüfungsfenster

Jedes Wort, das TextBridge nicht (oder nur mit Schwierigkeiten) erkennen kann, erscheint im Wortbearbeitungsfeld im oberen Bereich des Prüfungsfensters. Das Bild des Wortes wird darunter in seinem Kontext hervorgehoben.

Mit Hilfe der Prüfung können Sie den erkannten Text durchlaufen und die Entscheidungen von TextBridge übernehmen oder korrigieren. Ihre Eingaben helfen TextBridge dann bei der Erkennung des restlichen Teils der Vorlage.

Beenden Sie also die Überprüfung einer mehrseitigen Vorlage nach ein oder zwei Seiten.

Prüfen Sie hingegen kleine (ein- oder zweiseitige) Vorlagen vollständig. Weitere Informationen über die Prüfung und ihre Optionen finden Sie in Kapitel 3.



## Verwendung des Befehls “Invers” bei der Seitenvorschau

TextBridge ist in der Lage, Online-TIFF-Dateien zu erkennen, die von Faxmodems oder anderen Quellen stammen.

Hin und wieder werden Bilddaten so gespeichert, daß die Bildelemente (**Pixel**) in der daraus resultierenden Datei umgekehrt werden: Der weiße Seitenhintergrund ist schwarz und der Text weiß. Dies kommt z.B. bei Intel FAXability-Dateien häufiger vor.

TextBridge **kann** solche Dateien **nicht** erkennen. Eine Online-TIFF-Datei kann nur dann erkannt werden, wenn schwarze Schrift auf weißem Hintergrund vorliegt.

TextBridge ermöglicht die Erkennung solcher Dateien jedoch unter Verwendung des Befehls “Invers” im “Ansicht”-Menü des Vorschaufensters (Abbildung 4–8).

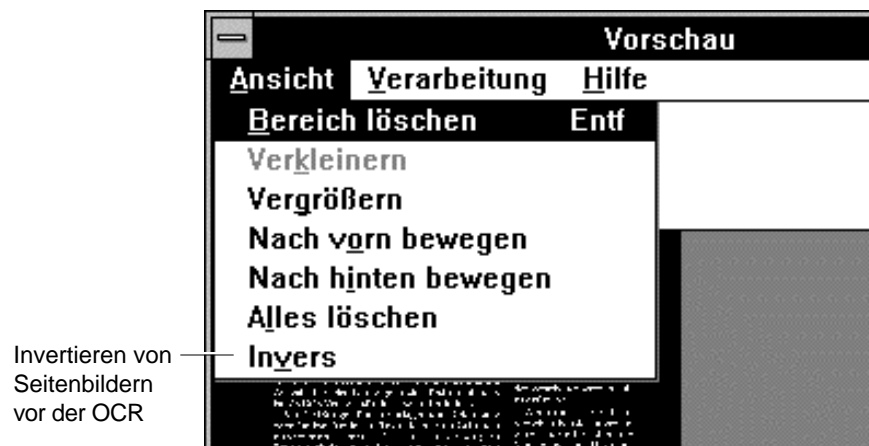


Abbildung 4–8. Das Invertieren einer Vorlage



Wenn Sie sich nicht sicher sind, ob eine Online-Datei umgekehrt ist, zeigen Sie sie vor der OCR im Vorschauenfenster an. Wenn weißer Text auf schwarzem Hintergrund angezeigt wird, öffnen Sie das Menü “Ansicht”, und klicken Sie auf den Befehl “Invers”. TextBridge kehrt das Seitenbild dann entsprechend um.

Anschließend können Sie mit dem OCR-Vorgang beginnen. Eine Invertierung muß bei jeder auf diese Weise gespeicherten Datei manuell korrigiert werden.

Weitere Informationen zur Seitenvorschau finden Sie in Kapitel 3.

## **TIPS FÜR EINE EFFIZIENTE VERARBEITUNG**

Wenn Sie TextBridge zum ersten Mal verwenden, ist es am einfachsten, zunächst einmal mit den Voreinstellungen zu arbeiten. Hiermit können schon gute Ergebnisse erzielt werden.

Um diese jedoch zu optimieren, können Sie vor der OCR folgendes tun:

- Verwendung des Bereichswerkzeuges bei der Seitenvorschau
- Verwendung der Einstellung “Fotos/Halbtonraster ignorieren”
- Verwendung der automatischen Seitenausrichtung, falls nötig
- Verwendung der automatischen Seitenaufteilung bei mehrspaltigen Vorlagen

Diese Funktionen gewährleisten, daß das System nur die wesentlichen Teile einer Seite verarbeitet und dies richtig.

Die hier genannten Funktionen können sich bei richtiger Anwendung auf eine oder mehrere Vorlagen äußerst zeitsparend auswirken.



## Eingrenzung eines zu bearbeitenden Bereichs

Manche Vorlagen enthalten Logos, Grafiken, lebende Kolumnentitel (oben oder unten) oder andere Bereiche, die Sie aus Zeitgründen nicht erfassen möchten.

Mit Hilfe des Bereichswerkzeuges im Vorschaufenster können Sie genau den Teil der Seite(n) definieren, den Sie erfassen möchten (Abbildung 4–9).

In Kapitel 3 finden Sie weitere Informationen zu den Vorschau-Tools.

Eingrenzen  
eines zu  
verarbeitenden  
Bereichs

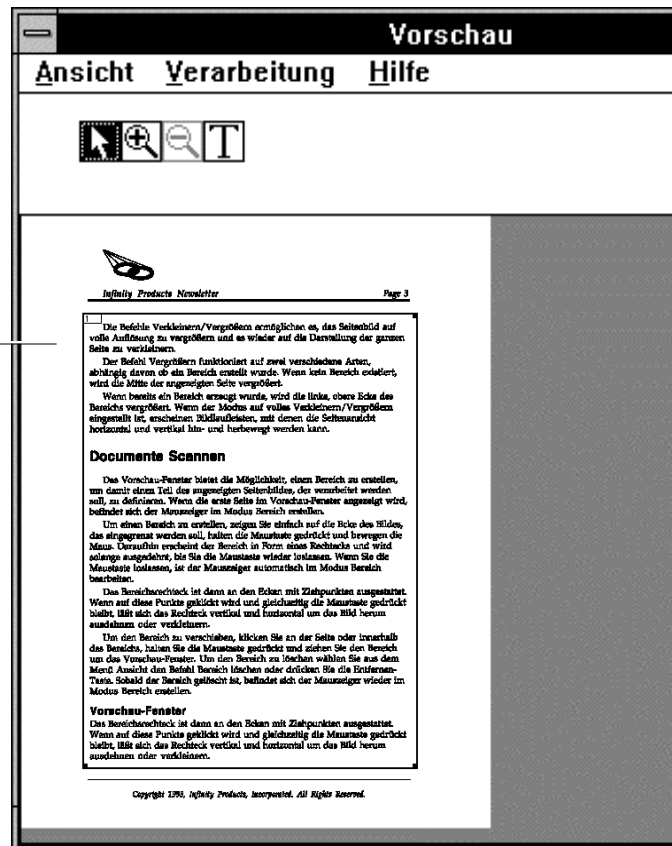


Abbildung 4–9. Vorschau eines Bereichs



## Verwendung der Option “Fotos/Halbttonraster ignorieren”

In einer gedruckten Vorlage besteht ein Halbttonraster aus verschiedenen großen schwarzen Punkten. Im Normalfall würde TextBridge versuchen, die Rasterpunkte als Text zu erkennen.

Nach einer Weile würde TextBridge herausfinden, daß es sich bei (einem Teil) der Vorlage um ein Halbttonraster handelt und würde das Raster ignorieren. Um Zeit zu sparen, können Sie jedoch von vornherein die Einstellung **Fotos/Halbttonraster ignorieren** auswählen.

In diesem Fall scannt TextBridge das Seitenbild und filtert Halbttonraster heraus, **bevor** die OCR beginnt (Abbildung 4–10). Die eigentliche Zeichenerkennung verläuft dann schneller und effizienter.

TextBridge versucht, Halbttonraster als Text zu erkennen

Vor der OCR werden mit Hilfe eines Filters Halbttonraster entfernt

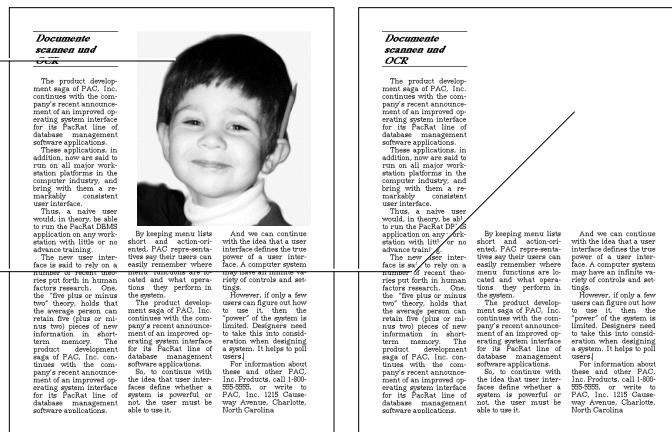


Abbildung 4–10. Der Filter “Fotos/Halbttonraster ignorieren”

Wählen Sie zur Verwendung der Option “Fotos/Halbttonraster ignorieren” die Schaltfläche “Voreinstellungen” aus dem Hauptdialogfeld. Klicken Sie dann im Dialogfeld “Voreinstellungen” auf das Kontrollkästchen “Fotos/Halbttonraster ignorieren”.

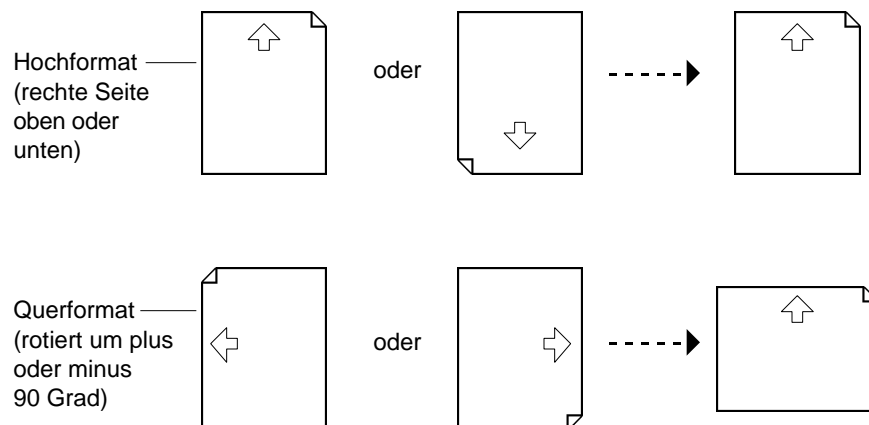


- + Obwohl das Herausfiltern von Halbtönern nur wenig Zeit in Anspruch nimmt, sollten Sie diese Option nur dann auswählen, wenn Ihre Vorlage auch wirklich Halbtöner enthält.

## Verwendung der automatischen Seitenausrichtung

TextBridge verfügt auch über eine Funktion, die automatisch die Ausrichtung einer Seite festlegt, diese im Arbeitsspeicher rotiert und dann mit der OCR beginnt (Abbildung 4–11).

Wählen Sie im Dialogfeld “Voreinstellungen”, auf das Sie vom Hauptdialogfeld zugreifen können, **Automatische Seitenausrichtung**. Diese Funktion ist in bestimmten Fällen äußerst nützlich, z.B.:



**Abbildung 4–11.** Funktionsweise der automatischen Seitenausrichtung

- bei der Verarbeitung von Vorlagen mit unterschiedlicher Seitenausrichtung
- bei der Verarbeitung einer TIFF-Datei, deren Seitenausrichtung unklar ist



Im ersten Beispiel könnten Sie eine Vorlage verarbeiten, die hauptsächlich aus Hochformaten (gemischt mit einigen Querformaten) besteht.

TextBridge scannt jede Seite und bestimmt, ob sie im Hochformat, Querformat (90 Grad oder 270 Grad) oder verkehrt herum vorliegt, und rotiert sie vor der OCR in die Hochformatstellung (0 Grad).

Wenn das TIFF-Seitenbild seitlich verdreht oder völlig verkehrt herum liegt (zweites Beispiel), wird es von TextBridge automatisch rotiert und dann gescannt.

- + Die automatische Ausrichtung geschieht vor der OCR. Um also eine OCR möglichst schnell durchzuführen, sollte dieser Schritt **nur** dann durchgeführt werden, wenn es wirklich nötig ist.

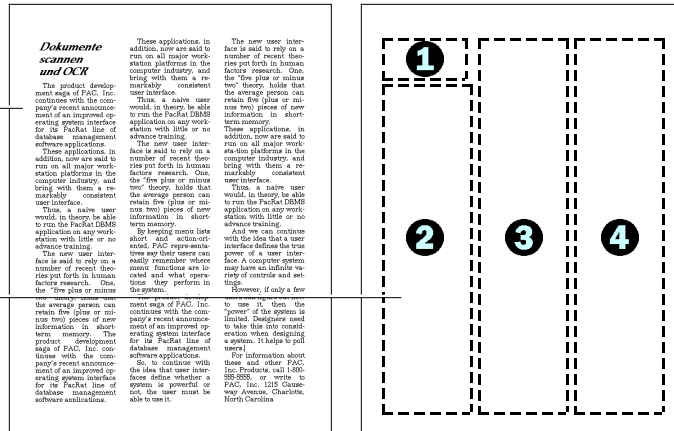
## Verwendung der automatischen Seitenaufteilung

TextBridge verfügt weiterhin über eine Funktion, die automatisch **Textbereiche** auf der Seite lokalisiert, ihre Anordnung definiert und anschließend mit der OCR beginnt. Diese Funktion zur **automatischen Seitenaufteilung** ist für solche Seiten von Bedeutung, die mehr als eine Spalte und/oder ein ungewöhnliches Layout haben (Abbildung 4–12).



Automatische  
Seitenaufteilung  
stellt Layout fest

Dann werden  
die einzelnen  
Bereiche  
identifiziert  
und geordnet



**Abbildung 4-12.** Seitenaufteilung und Bereichsanordnung

## Wichtig

Sie **müssen** die automatische Seitenaufteilung **einschalten**, wenn Sie Seiten mit mehr als einer Spalte Text verarbeiten. Andernfalls gibt TextBridge Bereiche mit erkanntem Text eventuell in der falschen Reihenfolge aus. Verwenden Sie die Funktion jedoch nicht für einspaltige Vorlagen.

Die automatische Seitenaufteilung wird vor der OCR durchgeführt.

Sie können einen Bereich in der “Vorschau” erstellen und trotzdem die automatische Seitenaufteilung verwenden, jedoch nur innerhalb des Bereichs. Wenn z.B. zwei von drei Spalten zu einem Bereich gehören sollen, erkennt und ordnet die Seitenaufteilungsfunktion die zwei Spalten zur Vorbereitung des OCR-Vorgangs.



## SEITE ALS BILD SPEICHERN

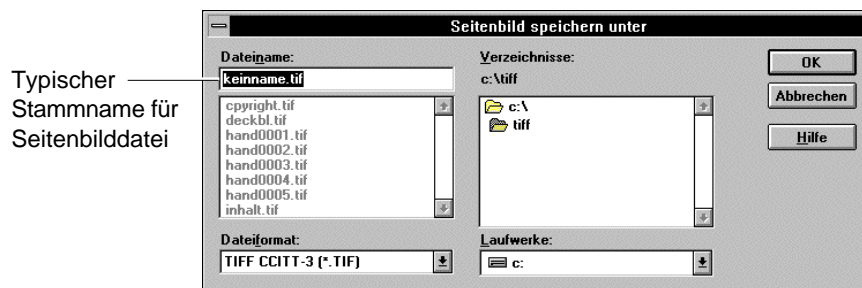
Eines der Kontrollkästchen im Hauptdialogfeld heißt **Seite als Bild speichern**. Diese Option ist verfügbar, wenn Sie “Eingabe aus Scanner” wählen. Sie ermöglicht das Speichern eines binären (schwarzweißen) Bildes von jeder Seite, die während einer TextBridge-OCR-Sitzung gescannt wird.

---

**Hinweis** TextBridge speichert Seitenbilder als TIFF-Dateien mit CCITT-Komprimierung (Gruppe 3). Gruppe 3 ist ein Komprimierungsstandard, der vom CCITT (Consultative Committee of International Telephone and Telegraph), dem Beratungsausschuß für internationale Standards im Bereich der Telekommunikation, festgelegt worden ist.

---

Nach dem Klicken auf “Starten!” und dem Scannen der ersten Seite, zeigt TextBridge das Dialogfeld “Seitenbild speichern unter” an (Abbildung 4–13).



**Abbildung 4–13.** Das Dialogfeld “Seitenbild speichern unter”

Das Dialogfeld “Seitenbild speichern unter” ähnelt einem Standard-Windows-Dialogfeld (“Speichern unter”), insofern, als Sie hier einen Dateinamen, Dateiformat, ein Ausgabeverzeichnis und ein Laufwerk angeben können.



Der Dateiname ist der **Stammname**, auf dem die Seitenbilddateinamen basieren. Er ist gleichzeitig der **Dokumentname**, der standardmäßig im Dialogfeld “Speichern unter” von TextBridge angezeigt wird, wenn die OCR beendet ist. Der Standarddateiname ist “keinname” mit der Erweiterung “.tif”.

Angenommen, Sie geben den Namen “handbuch” im Dialogfeld “Seitenbild speichern unter” an. Seitenbilddateien würden dann mit folgendem Namensformat gespeichert:

handnnnn.TIF

nnnn entspricht dabei der Seitenzahl der Vorlage mit davorstehenden Nullen (z.B. 0001, 0002, usw.).

Seitenbilddateien werden innerhalb des Verzeichnisses der Reihe nach benannt. Wenn eine Datei mit demselben Namen (z.B. “hand0001.tif”) schon existiert, verwendet TextBridge die nächsthöhere Zahl.

Nach der Verarbeitung erscheint der Vorlagename (z.B. “handbuch”) automatisch im Dateinamenfeld im Dialogfeld “Speichern unter”.

Im Dialogfeld “Seitenbild speichern unter” entspricht das erste Arbeitsverzeichnis dem Verzeichnis, von dem aus TextBridge gestartet wird:

C:\TXBRIDGE\BIN

Sie können zum Speichern der Seitenbilder jedoch auch jedes andere Laufwerk und Verzeichnis angeben. Dieses wird dann das neue **Arbeitsverzeichnis** für die entsprechende Vorlage und erscheint ebenso wie der Dokumentname nach der OCR im Dialogfeld “Speichern unter”.



Für das Seitenbilddateiformat bietet das Menü "Dateiformat" nur eine Auswahl, nämlich TIFF CCITT-3 Intel. Seitenbilder werden so gespeichert, wie sie gescannt werden, also im binären (schwarzweißen) Format.

Wenn Sie im Dialogfeld "Seitenbild speichern unter" auf "Abbrechen" klicken, wird das Dialogfeld geschlossen und die Bearbeitung der Vorlage beendet. Das Hauptdialogfeld bleibt geöffnet und ist bereit für den nächsten Auftrag. (Wenn Sie z.B. keine Seitenbilder speichern wollen, können Sie das Kontrollkästchen "Seitenbild speichern unter" deaktivieren und mit der Bearbeitung von vorn beginnen.)

## VERWENDUNG VON TEXTBRIDGE AUS ANDEREN ANWENDUNGEN

TextBridge OCR für Windows besteht eigentlich aus mehreren Anwendungen, die es Ihnen ermöglichen, eine OCR aus praktisch jeder anderen Windows-Anwendung zu starten.

Zusätzlich zum Hauptdienstprogramm, welches als selbständiges Programm verwendet werden kann und welches über die größte Funktionalität verfügt, gibt es TextBridge OCR in zwei anderen Anwendungsweisen:

- **TextBridge Anwendungs-Server**, ein Programm, das einer Menüoption innerhalb einer beliebigen Windows-Textanwendung entspricht (Textverarbeitung, Desktop-Publishing-Programm, Tabellenkalkulation, Datenbank-anwendung, etc).
- **TextBridge OCR Printer**, eine Komponente, mit der Sie ein Seitenbild in beliebigem Format an eine Version von TextBridge OCR senden können und die wie ein konventioneller Druckertreiber funktioniert.



Der folgende Abschnitt informiert Sie über diese beiden Anwendungsweisen von TextBridge OCR.

---

<b>Hinweis</b>	TextBridge unterstützt auch eine DDA-Schnittstelle. Systementwickler können sich bei Bedarf beim Xerox Imaging Systems-Kundendienst nach Details erkundigen.
----------------	--

---

## Verwendung des TextBridge-Anwendungs-Servers

Der TextBridge-Anwendungs-Server (TAS) ist ein Windows-Programm, das mit anderen Windows-Textanwendungen verbunden und von dort aus gestartet werden kann.

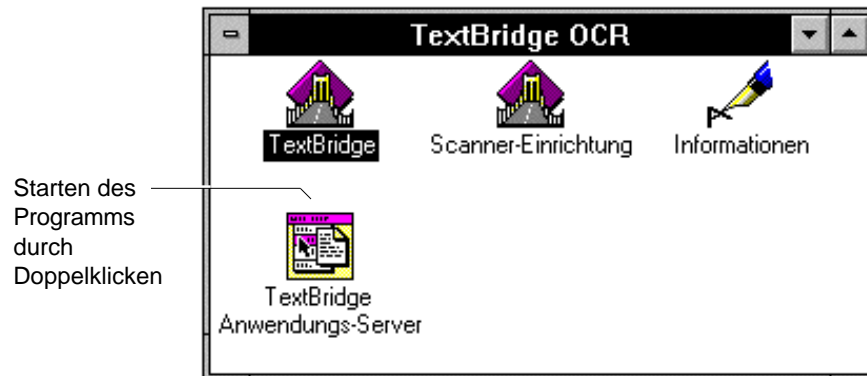
In diesem Fall erscheint dann TAS im Dateimenü der Host-Anwendung als Befehl **TextBridge OCR**. Wenn Sie diesen Befehl wählen, wird das Hauptdialogfeld von TextBridge angezeigt, als handelte es sich um ein Dialogfeld der Host-Anwendung. Von hier aus können Sie eine OCR genauso initiieren, wie aus dem Standard-TextBridge-Programm.

### Starten von TAS und Registrieren von Anwendungen

TAS wird zusammen mit TextBridge installiert (siehe Kapitel 2 dieses Handbuchs).

Nach Beendigung der Installation erstellt das TextBridge-Setup-Programm eine TextBridge-OCR-Programmgruppe, welche die TextBridge-Anwendungs-Server-Programmooption enthält (Abbildung 4–14).





**Abbildung 4–14. TAS-Programmoption**

Bevor Sie TAS aus irgendeiner Anwendung heraus verwenden können, müssen Sie das Programm starten und eventuell die Anwendung registrieren:

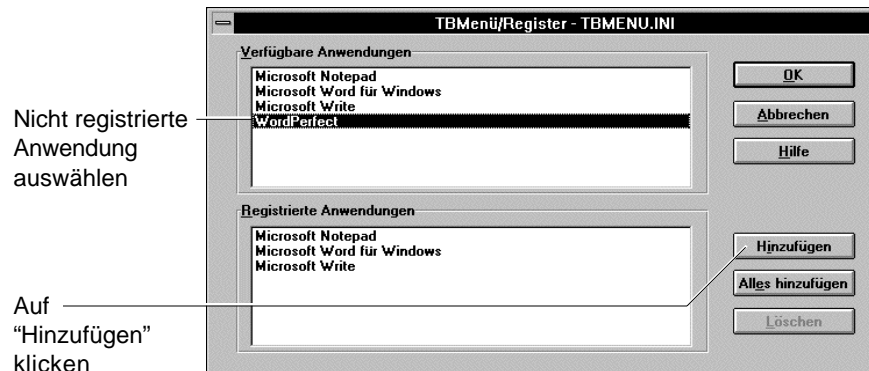
- 1. Klicken Sie doppelt auf die TAS-Programmoption in der TextBridge-OCR-Programmgruppe.**

Das Programm wird gestartet und erscheint als verkleinertes Symbol auf Ihrem Windows-Bildschirm.

- 2. Klicken Sie doppelt auf das entsprechende Symbol auf Ihrem Windows-Bildschirm.**

Das TAS-Registrierungsfeld wird angezeigt (Abbildung 4–15).





**Abbildung 4-15.** TAS-Registrierungsdialogfeld

### 3. Registrieren Sie, falls erforderlich, die Anwendung.

- Heben Sie im oberen Bereich des Registrierungsdialogfeldes die entsprechende Anwendung hervor.
- Klicken Sie auf “Hinzufügen”, um die aktuelle Anwendung der Liste der registrierten Anwendungen hinzuzufügen.
- Klicken Sie zum Schluß auf “OK”.

Sie können TextBridge OCR nun aus Ihrer registrierten Anwendung heraus verwenden.

### Starten von TAS aus Ihrer Anwendung

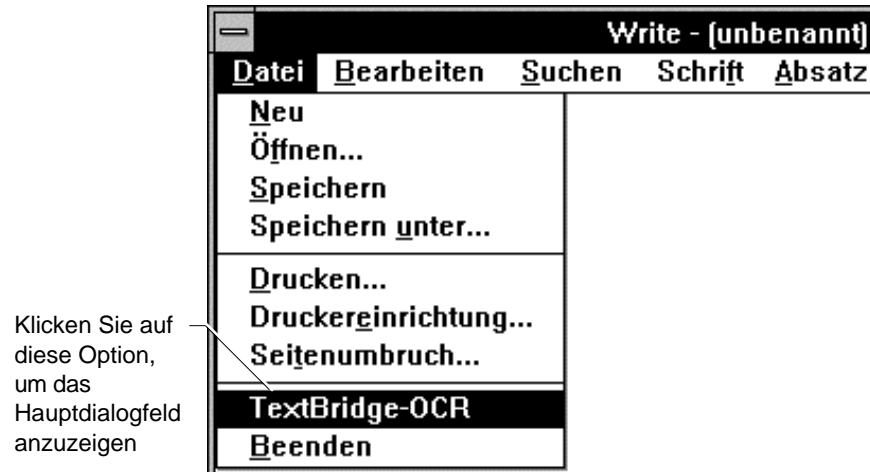
Im “Datei”-Menü jeder aktiven registrierten Anwendung wird der Befehl “TextBridge OCR” als letztes, direkt vor dem Befehl “Beenden”, angezeigt.

#### Hinweis

Damit TAS funktioniert, muß die Host-Anwendung über ein “Datei”-Menü und den dazugehörigen Befehl “Beenden” verfügen. Bei der Mehrheit der Windows-Anwendungen ist dies standardmäßig der Fall.



Die Abbildung 4–16 zeigt ein Beispiel des Befehls “TextBridge OCR” im MS Write®-“Datei”-Menü.



**Abbildung 4–16.** Der Befehl “TextBridge-OCR”

Zur Verwendung von TAS und zum Importieren von erkanntem Text direkt in ein offenes Dokument Ihrer Host-Anwendung gehen Sie folgendermaßen vor:

**1. Starten Sie den TextBridge-Anwendungs-Server.**

Klicken Sie doppelt auf die betreffende Programmoption in der TextBridge-OCR-Programmgruppe (siehe Abbildung 4–14).

- + Damit TAS automatisch mit Windows gestartet wird, platzieren Sie es in der StartUp-Programmgruppe.

**2. Vergewissern Sie sich, daß die Host-Anwendung registriert ist.**

Richten Sie sich dabei nach dem vorhergehenden Abschnitt, “Starten von TAS und Registrieren von Anwendungen”.



### 3. Starten Sie die Host-Anwendung.

Öffnen Sie in der Host-Anwendung ein neues oder existierendes Dokument, in das Sie den erkannten Text importieren wollen.

### 4. Öffnen Sie das “Datei”-Menü der Host-Anwendung, und wählen Sie den Befehl “TextBridge-OCR”.

Folgende Statusmeldungen werden angezeigt:

Verbinden mit den TextBridge Services...

Verbindung mit den TextBridge Services hergestellt.

Kurz danach wird das TextBridge-Hauptdialogfeld angezeigt.

### 5. Beginnen Sie mit der Einrichtung und Initiierung der OCR vom Hauptdialogfeld aus.

Wählen Sie entweder “Datei” oder “Scanner” als Eingabequelle aus dem Hauptdialogfeld, bestimmen Sie die Voreinstellungen, und gehen Sie anschließend so vor, als ob Sie TextBridge als selbständige Anwendung verwenden würden.

- + Die Funktionen “Vorschau”, “Prüfung” und “Seite als Bild speichern” sind in der TAS-Version von TextBridge **nicht** enthalten. Wenn Sie sie benötigen, starten Sie TextBridge als selbständige Anwendung, und speichern Sie erkannten Text in Ihrem Textverarbeitungs- oder einem anderen Textformat.

Die Verwendung von TextBridge wird in Kapitel 3 Schritt für Schritt erklärt; Tips und Techniken zur Optimierung der Ergebnisse finden Sie am Anfang dieses Kapitels.



Nach der OCR wird TAS geschlossen; der erkannte Text erscheint an der Cursorposition im offenen Dokument Ihrer Anwendung und kann dort bearbeitet werden.

- + TAS verwendet die Windows-Zwischenablage zum Ausschneiden und Einfügen von Text in Ihr Dokument, und zwar entweder als formatierten **RTF** (Rich Text Format) oder als einfachen **ASCII**-Text. Wenn Ihre Anwendung das Einfügen von RTF-Text vom Zwischenspeicher unterstützt, wird RTF verwendet. Ansonsten wird erkannter Text ohne Formatierung (fett, kursiv, etc.) eingefügt.

## Verwendung des TextBridge OCR Printers

In seinen anderen Verwendungsweisen kann TextBridge OCR für Windows optische Informationen nur dann erkennen, wenn sie im TIFF-Format gespeichert sind. Manche Anwendungen, insbesondere einige Faxprogramme, speichern Faxseitenbilder nur im PCX-, DCX- oder einem anderen anwendungseigenen Format.

Um eine OCR bei anderen, Nicht-TIFF-Seitenbildern durchzuführen, verwenden Sie den **TextBridge OCR Printer**. Der OCR Printer erscheint in Windows-Anwendungen einfach als ein anderer Zieldrucker. Sie können damit ein Seitenbild von einer Windows-Anwendung "drucken" und eine erkannte und formatierte Textdatei erstellen.

Eine typische Verwendung des OCR Printers ist das Erkennen eines Seitenbildes, das direkt aus einer Fax- oder Grafikanwendung stammt. Der OCR Printer ähnelt dem Modell vieler Faxprogramme, die eine ähnliche Funktion zur Versendung von Faxen verwenden. Das Faxseitenbild wird nämlich an das Faxmodem weitergeleitet und dann an ein anderes Faxmodem oder -gerät verschickt.



Der OCR Printer bietet den zusätzlichen Vorteil, daß er praktisch jedes Textformat erkennt (DCX, PCX, Corel, TIFF, etc.). Fast jedes Windows-Programm, das über eine graphische Komponente verfügt, kann den OCR Printer verwenden.

Wie die Verwendung des OCR Printers vorbereitet werden muß, erfahren Sie im folgenden Abschnitt "Hinzufügen des OCR Printers". Lesen Sie anschließend den Abschnitt "Verwendung des OCR Printers in einer Grafikanwendung"; hier wird das Herunterladen von optischen Informationen und die Erstellung einer Textdatei beschrieben.

### Hinzufügen des OCR Printers

Die Programmdateien des OCR Printers werden mit der Haupt-TextBridge-Anwendung installiert.

Wie bei anderen Druckertreibern muß auch der OCR Printer der Liste von Druckern, die den Windows-Anwendungen auf Ihrem PC zur Verfügung stehen, hinzugefügt werden.

Im folgenden wird davon ausgegangen, daß TextBridge, wie in Kapitel 2 beschrieben, auf Ihrem PC installiert ist.

- 1. Öffnen Sie im Windows-Programm-Manager die Hauptgruppe, und klicken Sie doppelt auf das Symbol "Systemsteuerung":**



Dadurch wird das Systemsteuerungsfenster mit Symbolen zu verschiedenen Systemfunktionen geöffnet.



2. **Klicken Sie doppelt auf das Druckersymbol im Systemsteuerungsfenster.**



Damit öffnen Sie das Dialogfeld mit den Windows-Druckern (Abbildung 4–17).

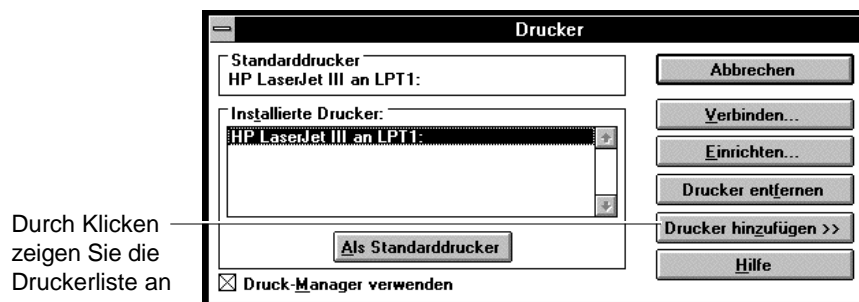


Abbildung 4–17. “Drucker”-Dialogfeld

3. **Klicken Sie auf die Schaltfläche “Drucker hinzufügen” im “Drucker”-Dialogfeld.**

Das Dialogfeld wird erweitert und zeigt die Liste der hinzufügbaren Drucker an.

4. **Heben Sie die folgende Option in der Liste hervor, und klicken Sie auf die Schaltfläche “Installieren”.**

Nicht aufgeführten oder aktualisierten  
Drucker installieren

Das Dialogfeld “Treiber installieren” wird angezeigt. Es enthält Anweisungen zur Angabe des Lauf-werks und des Verzeichnisses des Drucker-treibers.

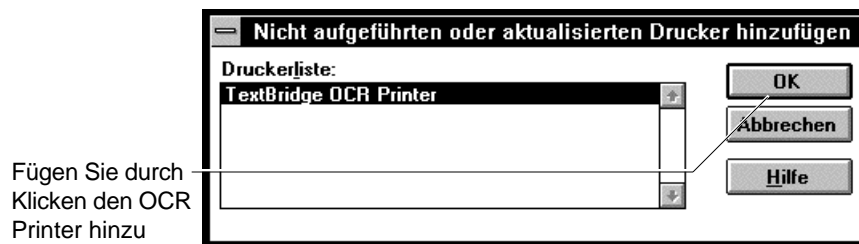


- 5. Im Dialogfeld “Treiber installieren” geben Sie den TextBridge BIN-Pfadnamen ein:**

c:\txbridge\bin

- 6. Klicken Sie auf “OK” (oder drücken Sie die Eingabetaste).**

Das Dialogfeld “Nicht aufgeführten oder aktualisierten Drucker hinzufügen” wird angezeigt (Abbildung 4–18).



**Abbildung 4–18.** Das Dialogfeld “Nicht aufgeführten oder aktualisierten Drucker hinzufügen”

- 7. Wählen Sie die Eingabe “OCR Printer” und klicken Sie auf “OK”.**

Das Dialogfeld “Nicht aufgeführten oder aktualisierten Drucker hinzufügen” wird geschlossen, das “Drucker”-Dialogfeld bleibt geöffnet.

- 8. Klicken Sie im “Drucker”-Dialogfeld auf “Schließen”, um das Hinzufügen zu beenden.**

Sie können den TextBridge OCR Printer nun wie im nächsten Abschnitt beschrieben verwenden.



## Verwendung des OCR Printers in Ihrer Grafik-Anwendung

Der OCR Printer kann mit jeder Windows-Anwendung verwendet werden, die eine Grafikdatei öffnen und anzeigen kann.

WinFax Pro (von Delrina Technology Inc.) verfügt z.B. über einen "Image Viewer" zum Anzeigen, Bearbeiten und Drucken von Faxseitenbildern. Bei Verwendung des Druckbefehls im "Datei"-Menü des WinFax Image Viewers können Sie den OCR Printer angeben und ihn zur Zeichenerkennung bei einer Faxvorlage angeben.

Vor der OCR können Sie mit "OCR Printer" auf die Voreinstellungen von TextBridge zugreifen. Nach der OCR können Sie den Dateinamen, den Auffindungsort und das Format des Ausgabetexts im Standarddialogfeld "Speichern unter" angeben.

Gehen Sie bei der Verwendung des OCR Printers in Ihrer Anwendung folgendermaßen vor:

### **1. Öffnen Sie die Grafikanwendung, und zeigen Sie das zu erkennende Seitenbild an.**

- + Das Seitenbild muß binär (d.h. schwarzweiß) sein und im Bereich der von TextBridge unterstützten Auflösungen liegen. TextBridge kann Bilder erkennen, deren Auflösung 100 mal 200, 200 mal 100, 200 mal 200, 300 mal 300 und 400 mal 400 dpi beträgt.



**2. Geben Sie in Ihrem Dialogfeld zur Druckereinrichtung den TextBridge OCR Printer als Zieldrucker an.**

Der OCR Printer sollte bereits dem Windows-Druckerprogramm hinzugefügt sein, wie im Abschnitt “Hinzufügen des OCR-Printers” beschrieben.

**3. Wahlweise: Legen Sie die Voreinstellungen fest.**

- Klicken Sie auf die Schaltfläche “Optionen” oder “Einrichten” im Druckereinstellungs-Dialogfeld Ihrer Grafikanwendung. Ein sekundäres Dialogfeld zur weiteren Einrichtung wird angezeigt.
  - Klicken Sie auf die Schaltfläche “Voreinstellungen” im Dialogfeld zur Druckereinrichtung, um das Dialogfeld “Voreinstellungen” anzuzeigen.
  - Legen Sie “Standard” oder “Faxvorlagenqualität”, “Seitenausrichtung”, “Automatische Seitenaufteilung”, etc. fest.
  - Klicken Sie im Dialogfeld “Voreinstellungen” auf “OK”, und verlassen Sie alle noch offenen Dialogfelder.
- + **Aktivieren** Sie im Druckereinstellungs-dialogfeld Ihrer Anwendung (falls vorhanden) die Option zur Druckerauflösung.



**4. Starten Sie den OCR-Vorgang für das angezeigte Seitenbild.**

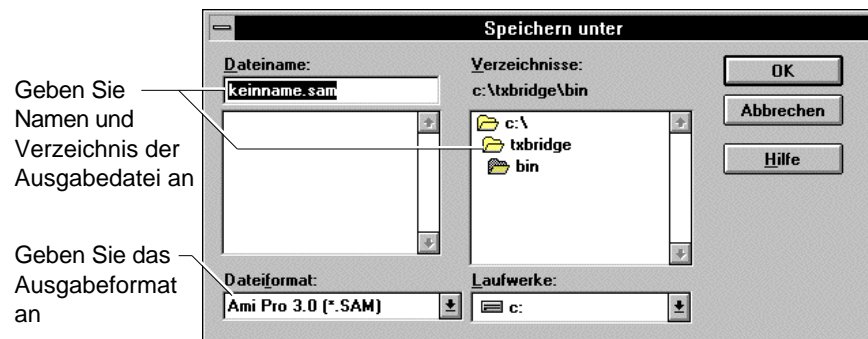
- Öffnen Sie das “Datei”-Menü Ihrer Anwendung und wählen Sie den Druckbefehl.
- Klicken Sie im Dialogfeld “Drucken” auf “OK”. Im Laufe der OCR werden dann Statusmeldungen angezeigt:

Wird verarbeitet...

Erfassen des Seitenbildes...

Erkennen von Text...

Nach der OCR wird das Dialogfeld “Speichern unter” angezeigt (Abbildung 4–19).



**Abbildung 4–19.** Das Dialogfeld “Speichern unter”

**5. Geben Sie Namen, Format, Ziellaufwerk und -verzeichnis der Ausgabedatei an, und klicken Sie auf “OK”.**

Die erkannte Textdatei wird nun in das angegebene Format konvertiert und auf Ihrer Festplatte gespeichert.