HP AccuPage: A Toolkit for High-Quality Document Scanning

Working with commercially available OCR programs, the image processing transforms used in HP AccuPage 2.0 improve the accuracy of converting scanned images from a variety of documents to editable text and pictures at the same time.

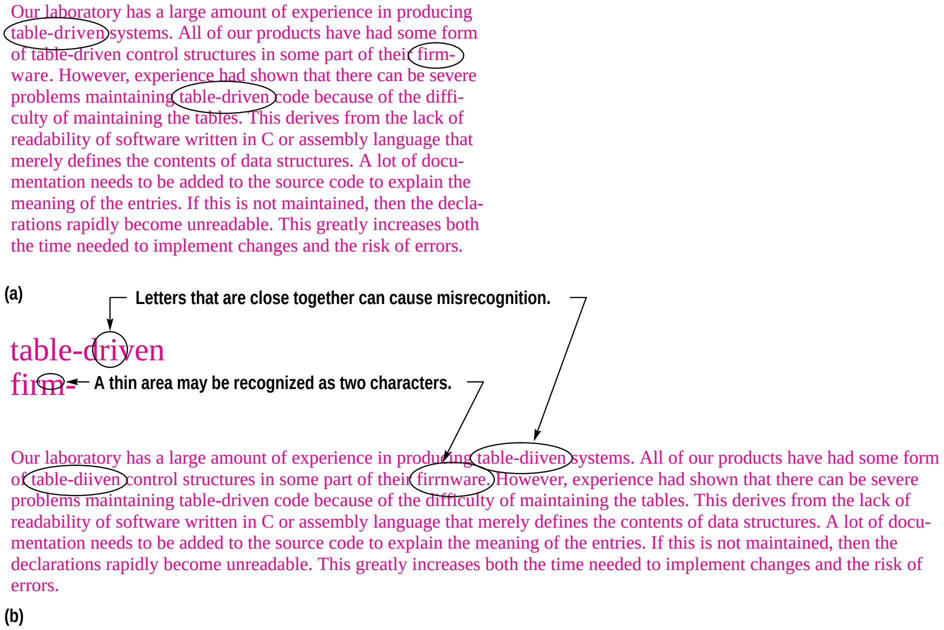
by Steven L. Webb, Steven G. Henry, Kevin S. Burke, and George Prokop

Desktop scanners are becoming more prevalent in both commercial and home office environments. Many people recognize that a scanner provides several benefits. First, desktop scanners continue to be a useful tool for improving the quality of published documents. Scanners easily allow the document author to add editable graphics to a publication, including color line drawings, diagrams, and photographs. Secondly, desktop scanners are increasingly being used to scan hardcopy documents and convert them into editable text using optical character recognition (OCR) programs. Instead of retyping text, scanner users can quickly convert documents into computer editable form, often at rates of up to 1000 words per minute.

HP continues to be a leading provider of desktop scanners for use with personal computers. HP produces color (HP ScanJet 3C) and black and white (HP ScanJet 3p) scanners that are designed to work with IBM-compatible and Macintosh personal computers. In addition to the hardware, HP also contributes software that, in partnership with software provided by independent software vendors (ISVs), provides a total solution.

HP Accupage 2.0

Two key contributions come from HP scanner software. One is to enable the scanner user to scan documents easily into a personal computer. The other is to do some sophisticated image processing on the scanned image. These two contributions help to improve the quality of the scanned image and enable much more accurate optical character recognition of text-based pages. Accurate OCR capability is provided by the image processing transforms contained in HP AccuPage 2.0.



Our laboratory has a large amount of experience in producing
table-driven systems. All of our products have had some form
of table-driven control structures in some part of their firm-
ware. However, experience had shown that there can be severe
problems maintaining table-driven code because of the diffi-
culty of maintaining the tables. This derives from the lack of
readability of software written in C or assembly language that
merely defines the contents of data structures. A lot of docu-
mentation needs to be added to the source code to explain the
meaning of the entries. If this is not maintained, then the decla-
rations rapidly become unreadable. This greatly increases both
the time needed to implement changes and the risk of errors.

(a)
Letters that are close together can cause misrecognition.
table-driven
firm-
A thin area may be recognized as two characters.
(b)

Our laboratory has a large amount of experience in producing table-diiven systems. All of our products have had some form
of table-diiven control structures in some part of their firrnware. However, experience had shown that there can be severe
problems maintaining table-driven code because of the difficulty of maintaining the tables. This derives from the lack of
readability of software written in C or assembly language that merely defines the contents of data structures. A lot of docu-
mentation needs to be added to the source code to explain the meaning of the entries. If this is not maintained, then the
declarations rapidly become unreadable. This greatly increases both the time needed to implement changes and the risk of
errors.

Fig. 1. A portion of a document after being read in by a typical OCR application. (a) Portion of the original document. (b) Results after scanning and passing through an OCR application.

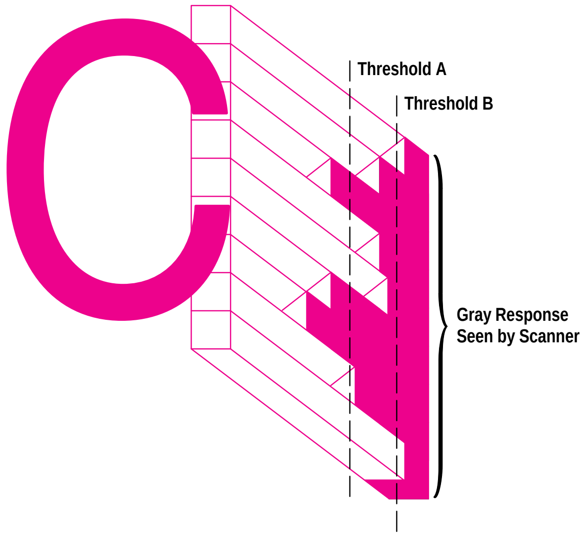


Fig. 2. A representation of the responses of the scan head as it passes over the letter c.

When HP AccuPage 2.0 is used with a third-party OCR utility, the accuracy of converting a scanned image to editable text is improved. Specifically, documents that are printed on colored backgrounds, that contain small point size text (5 to 6 point), and have complex layouts are converted with improved accuracy as a result of the HP AccuPage 2.0 image processing transforms. HP AccuPage 2.0 also allows both text and pictures to be captured from a document at the same time. This capability is becoming more important as both text and pictures increasingly make up the content of most office documents today.

In addition to providing better scanning capabilities to our customers, HP AccuPage 2.0 has also helped us to work more effectively with our ISVs to let our customers know about the value of HP AccuPage 2.0 and HP scanners. OCR independent software vendors help promote an overall document solution based around an HP scanner, HP AccuPage 2.0 technology, and the software that they provide.

This article will describe the limitations of OCR and the features provided in HP Accupage 2.0 that help to improve overall OCR accuracy. It will also describe some of the image processing techniques used to boost OCR accuracy.

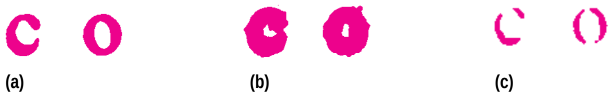


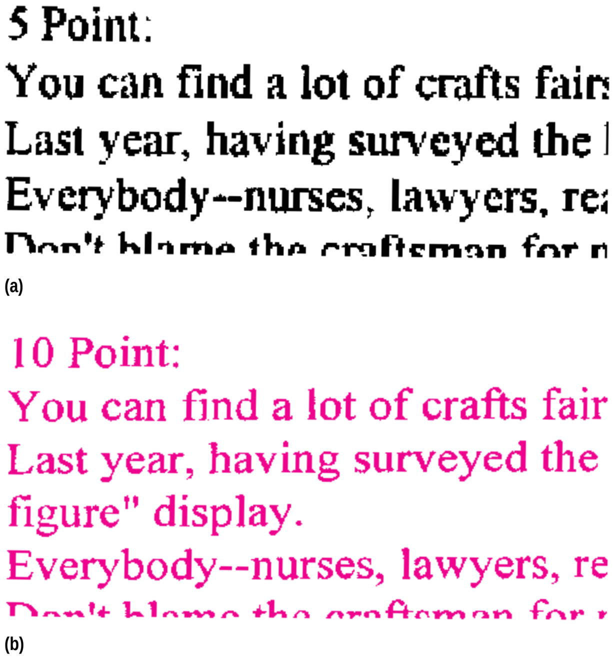
Fig. 3. The results of different threshold settings. (a) Normal setting. (b) Threshold set too low. (c) Threshold set too high.

The Limitations of OCR

The process of converting a hardcopy page into editable text accurately is not an easy task. Many documents that we can easily read with our eyes cannot be accurately converted into editable text by an ordinary OCR program. For example, Fig. 1 shows a portion of a document before and after it is scanned and run through a typical OCR program. Several of the characters weren't converted accurately despite the fact that the text is easy to read for the human eye. Current OCR pattern-recognition algorithms still require the input characters to be well-formed, smooth, and large enough so that the individual character elements are very distinct.

The reason OCR cannot recognize characters as well as people is that we have the remarkable ability to look subconsciously at different facets of what we are reading and use multilevel thinking to figure out and recognize words. We not only take into account the individual shape of each letter, but our brain is also able to piece together patterns and make use of overall context in determining the meaning of each word. Modern text recognition algorithms are only now beginning to exploit the use of context in evaluating the recognition of characters as part of words.

Another limitation of OCR has more to do with our expectations than with the algorithms. Many of us have the expectation that if we can easily read a page, then a computer-based OCR program should be able to read the page even better (more accurately and faster). In fact, we often have the hope


5 Point:
You can find a lot of crafts fair
Last year, having surveyed the
Everybody--nurses, lawyers, re
Don't blame the craftsman for r

(a)

10 Point:
You can find a lot of crafts fair
Last year, having surveyed the
figure" display.
Everybody--nurses, lawyers, re
Don't blame the craftsman for r

(b)

Fig. 4. A comparison between five point and ten point text scanned at 300 dpi. The five point sample has been enlarged for comparison with the ten point sample. Both of these samples are larger than actual size. Note the loss of detail in the five point sample.

Glossary

Charge-coupled Device (CCD). A CCD is a miniature photometer that measures incident light and converts the measured value to an analog voltage. The CCDs in a scanner are arranged in an array.

Desktop Scanner. A desktop scanner is a device that uses a light source, a color-separation method, and a charge-coupled device (CCD) array to capture optical information about an object (e.g., photographs or documents) and transforms that information into a digital light-intensity map for computer processing (see Fig. 1). The digital data is a two-dimensional map of pixels in which each pixel holds an intensity measurement corresponding to the reflectance (for paper) or the transmittance (for transparencies) of the object at the location represented by that pixel.

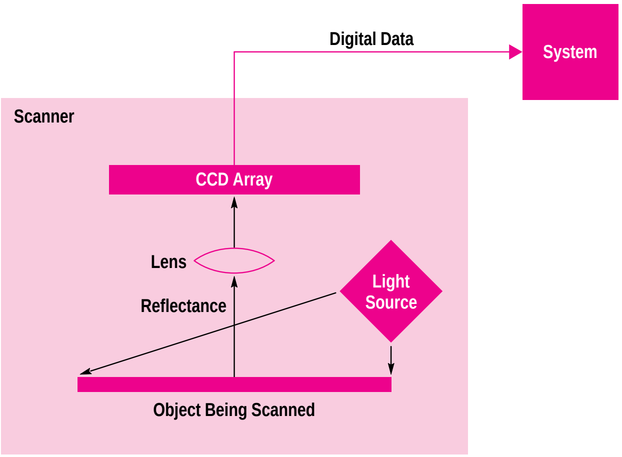


Fig. 1. The basic components of a desktop scanner.

dpi (dots per inch). The number of dots that can be printed per inch by a laser or inkjet printer.

Intensity. The amount of light reflected or transmitted by an object with black as the lowest intensity and white as the highest intensity.

Optical Sampling Rate. This is the number of samples, in pixels per inch (ppi), that are taken by a scanner per linear distance as determined by the CCD array, the optical system, and the motion of the carriage. The optical sampling rate for a scanner is specified as the pixels per inch in the x direction (across the page) and the y direction (down the page).

In the x direction, the optical sampling rate depends on the CCD layout and the magnification of the optical system. For example, a CCD with 2,550 elements applied across an 8.5-in image width has an optical sampling rate of 2250/8.5 = 300 ppi in the x direction.

In the y direction, the optical sampling rate depends on the distance and speed at which the carriage moves relative to the exposure time of the of the CCD. For example, if the carriage moves 1/300 in during a CCD exposure time, the y-direction optical sampling rate is 300 ppi.

ppi (pixels per inch). Ppi is often used interchangeably with dpi, although a dot is a bilevel entity, either on or off, and a pixel can hold multiple levels of information. For example, for an eight-bit scanner, one pixel has 256 possible values.

Resolution. For a scanner, resolution is the degree to which the scanner can distinguish detail. Resolution is dependent on items such as optical sampling rate, lens quality, filter quality, and carriage motion.

Threshold. A value to which a signal is compared when transforming from a multilevel value to a binary value. In a binary scan, parts of the image below the threshold will be recorded as black and parts above the threshold will be recorded as white.

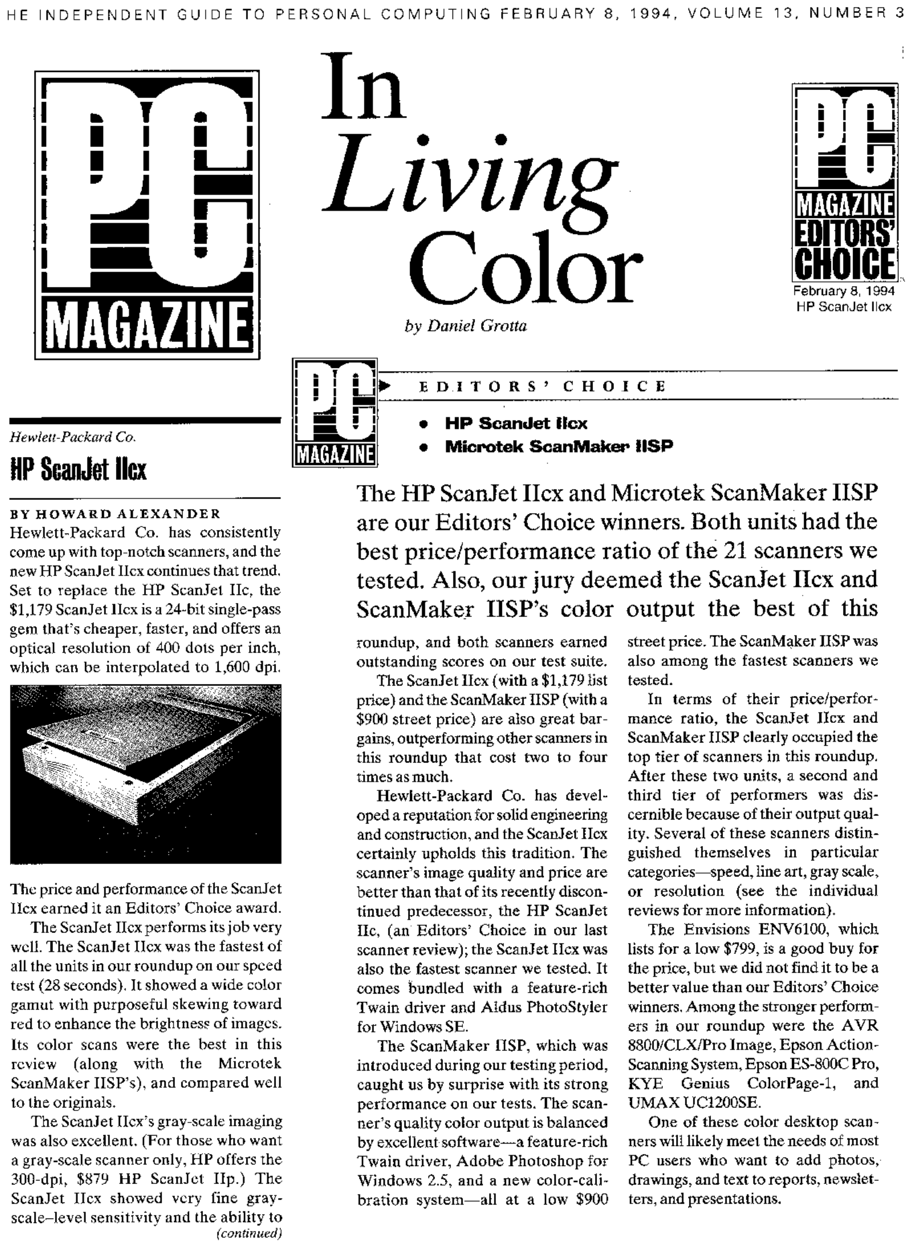
that if the page is somewhat illegible then the computer could convert it into legible format without any errors. Thus, many people (until their expectations are set correctly) are somewhat disappointed in using today's OCR technology. However, many customers quickly realize that for many high-quality pages, the number of errors is fairly small (typically 4 to 10 errors per 2000 characters), which is favorable compared to many professional typists and very favorable compared to the ability most of us have to type text accurately.

Besides the problems that can occur with degraded characters, text recognition can also be hampered by the scanning process. For instance, like a photocopy machine, a scanner can be set to produce images that can be lightened or darkened. If the lighten or darken setting (i.e., scanner intensity) is not set correctly the OCR might produce broken or incorrect character readings. Also, when scanning text on colored background with a black and white scanner, the colored background behind the text can actually obscure the text and render it unreadable by OCR algorithms.

To understand why these problems occur it is helpful to understand something about the OCR process. Most OCR programs use a binary (only black and white) image during the process of converting a bitmap image into text. Many of the problems of character recognition can be attributed to the intensity (lightness or darkness setting of the scanner) used during the scan when the binary bitmap image is created. The small space between strokes of a character will be seen by a scanner as a shade of gray (see Fig. 2). If the intensity (or threshold) is set low then more of the gray areas of the page will be set to black. As a result of this, if the shade of gray in the small space between strokes is mapped to black, it can cause the small gaps to close in the final binary image (see Fig. 3b). This can make a lowercase c look like a lower-case o. Setting the intensity too low is also a problem if the text is printed on a colored background. If the background shade is above the threshold then the image presented to the OCR engine is solid black.

Problems also occur if the threshold is set too high, because instead of characters joining they end up with breaks (Fig. 3c).

Another problem that can occur with some scanners is that the scan won't be completely uniform across the page. This can be caused by a nonuniform light profile across the page



HE INDEPENDENT GUIDE TO PERSONAL COMPUTING FEBRUARY 8, 1994, VOLUME 13, NUMBER 3

PC MAGAZINE

In *Living* Color

by Daniel Grotta

PC MAGAZINE EDITORS' CHOICE
February 8, 1994
HP ScanJet IIcx

EDITORS' CHOICE

- **HP ScanJet IIcx**
- **Microtek ScanMaker IISP**

The HP ScanJet IIcx and Microtek ScanMaker IISP are our Editors' Choice winners. Both units had the best price/performance ratio of the 21 scanners we tested. Also, our jury deemed the ScanJet IIcx and ScanMaker IISP's color output the best of this roundup, and both scanners earned outstanding scores on our test suite.

The ScanJet IIcx (with a $1,179 list price) and the ScanMaker IISP (with a $900 street price) are also great bargains, outperforming other scanners in this roundup that cost two to four times as much.

Hewlett-Packard Co. has developed a reputation for solid engineering and construction, and the ScanJet IIcx certainly upholds this tradition. The scanner's image quality and price are better than that of its recently discontinued predecessor, the HP ScanJet IIc, (an Editors' Choice in our last scanner review); the ScanJet IIcx was also the fastest scanner we tested. It comes bundled with a feature-rich Twain driver and Aldus PhotoStyler for Windows SE.

The ScanMaker IISP, which was introduced during our testing period, caught us by surprise with its strong performance on our tests. The scanner's quality color output is balanced by excellent software—a feature-rich Twain driver, Adobe Photoshop for Windows 2.5, and a new color-calibration system—all at a low $900 street price. The ScanMaker IISP was also among the fastest scanners we tested.

In terms of their price/performance ratio, the ScanJet IIcx and ScanMaker IISP clearly occupied the top tier of scanners in this roundup. After these two units, a second and third tier of performers was discernible because of their output quality. Several of these scanners distinguished themselves in particular categories—speed, line art, gray scale, or resolution (see the individual reviews for more information).

The Envisions ENV6100, which lists for a low $799, is a good buy for the price, but we did not find it to be a better value than our Editors' Choice winners. Among the stronger performers in our roundup were the AVR 8800/CLX/Pro Image, Epson Action-Scanning System, Epson ES-800C Pro, KYE Genius ColorPage-1, and UMAX UC1200SE.

One of these color desktop scanners will likely meet the needs of most PC users who want to add photos, drawings, and text to reports, newsletters, and presentations.

Hewlett-Packard Co.

HP ScanJet IIcx

BY HOWARD ALEXANDER

Hewlett-Packard Co. has consistently come up with top-notch scanners, and the new HP ScanJet IIcx continues that trend. Set to replace the HP ScanJet IIc, the $1,179 ScanJet IIcx is a 24-bit single-pass gem that's cheaper, faster, and offers an optical resolution of 400 dots per inch, which can be interpolated to 1,600 dpi.

The price and performance of the ScanJet IIcx earned it an Editors' Choice award.

The ScanJet IIcx performs its job very well. The ScanJet IIcx was the fastest of all the units in our roundup on our speed test (28 seconds). It showed a wide color gamut with purposeful skewing toward red to enhance the brightness of images. Its color scans were the best in this review (along with the Microtek ScanMaker IISP's), and compared well to the originals.

The ScanJet IIcx's gray-scale imaging was also excellent. (For those who want a gray-scale scanner only, HP offers the 300-dpi, $879 HP ScanJet IIp.) The ScanJet IIcx showed very fine gray-scale–level sensitivity and the ability to

(continued)

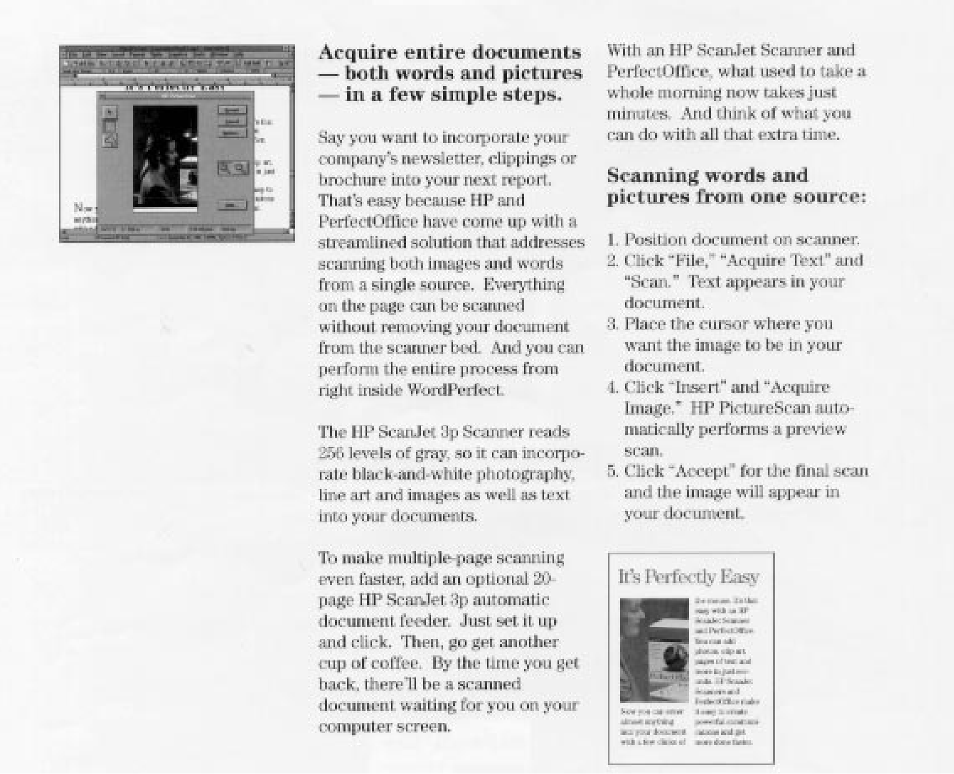
Fig. 5. Complex page with multiple columns, colors, graphics, titles, and so on. (Reprinted from *PC MAGAZINE*, February 8, 1994 Copyright© 1994 Ziff-Davis Publishing Company L.P.)

or heating effects along the length of the page during the scan. This means that the correct setting for the threshold in the top middle of the page may not be the correct threshold setting for other areas on the same page.

Another challenge for OCR algorithms is the recognition of small point-size characters (in the range of five to seven points).* Many scanners scan at a resolution of 300 dots per inch (dpi), which makes it difficult to provide well-formed character shapes to the text recognition algorithms. (See Fig. 4 for an example of five point characters versus ten point characters scanned at 300 dpi.) Small-point-size characters are frequently encountered as captions or as the fine print on legal documents. When read with current OCR algorithms at a standard 300 dpi scanning resolution, many errors result. Again, these are character point sizes that we can easily read by eye.

So far we have focused the discussion on the limitations of OCR algorithms for character and word recognition. Another important task of an OCR algorithm is to recreate the overall look or layout of the scanned page. For many typical documents, this is straightforward for existing OCR algorithms. Techniques exist to identify the location of paragraphs and to keep them in order. However, as the layout of the page gets more complicated, OCR algorithms that recognize and retain the page format begin to fall apart. Complexity can result because of elements such as multiple columns, sidebars, and graphic elements such as figures, tables, and

* There are 72 points per inch.



Acquire entire documents — both words and pictures — in a few simple steps.

Say you want to incorporate your company's newsletter, clippings or brochure into your next report. That's easy because HP and PerfectOffice have come up with a streamlined solution that addresses scanning both images and words from a single source. Everything on the page can be scanned without removing your document from the scanner bed. And you can perform the entire process from right inside WordPerfect.

The HP ScanJet 3p Scanner reads 256 levels of gray, so it can incorporate black-and-white photography, line art and images as well as text into your documents.

To make multiple-page scanning even faster, add an optional 20-page HP ScanJet 3p automatic document feeder. Just set it up and click. Then, go get another cup of coffee. By the time you get back, there'll be a scanned document waiting for you on your computer screen.

With an HP ScanJet Scanner and PerfectOffice, what used to take a whole morning now takes just minutes. And think of what you can do with all that extra time.

Scanning words and pictures from one source:

1. Position document on scanner.
2. Click "File," "Acquire Text" and "Scan." Text appears in your document.
3. Place the cursor where you want the image to be in your document.
4. Click "Insert" and "Acquire Image." HP PictureScan automatically performs a preview scan.
5. Click "Accept" for the final scan and the image will appear in your document.

It's Perfectly Easy

[illegible]

Fig. 6. Some of the graphics items in a document contain text, making it difficult to identify the basic object type—is it a picture, text, or something else?

photographs. Fig. 5 shows an example of a complex page that might give a typical OCR program some difficulty. Although, in most cases the errors in formatting produced by the OCR algorithms are easily identified and can be fixed, formatting errors are a major area of concern for most customers who use today's OCR products.

Page layout recognition is the job of the page segmentation algorithms. Page segmentation involves separating the page into parts that contain images of text to be converted to ASCII text and photos and drawings that are to be kept as they are. Inaccurate page segmentation causes inappropriate bitmap images to be sent to the OCR engine. Line art (or any nontextual image) sent to an OCR engine causes long delays and results in bad data being inserted into the output file. Incorrect grouping of table data can result in loss of table integrity and a loss of formatting, and incorrect column identification can cause column merging.

Once the text information has been identified and formatted, most of the current OCR algorithms completely ignore and eliminate any pictures and graphical information on the page. Many scanner users have a difficult time understanding why the picture and graphics information isn't available in the final OCR results, especially since charts, diagrams, and pictures are becoming a standard part of business communications. One challenge has been in correctly identifying the text and image regions on the page, especially when the two are located next to each other. In addition, many picture and graphical areas on a page contain textual data, which adds to the challenge (see Fig. 6). Lastly, most scanner users also expect their scanned pictures to match the original image very closely. This requires a scan to be done with moderate resolution (150 dpi) with eight bits of gray information per pixel. Most OCR software solutions today rely on 300-dpi resolution with one bit per pixel of information.

HP AccuPage 2.0 Technology

Many of the issues mentioned above have been reasonably solved with HP AccuPage 2.0 technology working in combination with the latest OCR software available from several manufacturers. When a customer purchases an HP scanner solution, they receive the scanner, HP AccuPage 2.0 software, and an OCR software package that uses the HP AccuPage 2.0 image processing transforms. Fig. 7 shows a block diagram of the components that make up the HP AccuPage 2.0 document reading solution.

By using HP AccuPage 2.0 as part of the solution, customers can get better OCR character accuracy, improved page format retention, and the ability to capture both text and high-quality images from their scanned document. Better character accuracy results from the ability to optically read text on colored backgrounds and the ability to read smaller-point-size characters. Improved page format retention and the

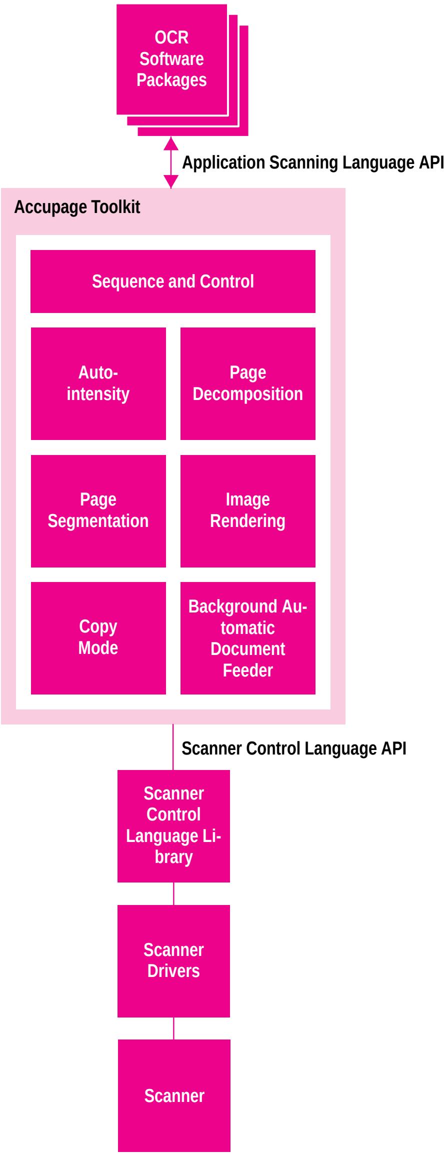


Fig. 7. The components that make up the HP AccuPage 2.0 document reading solution.

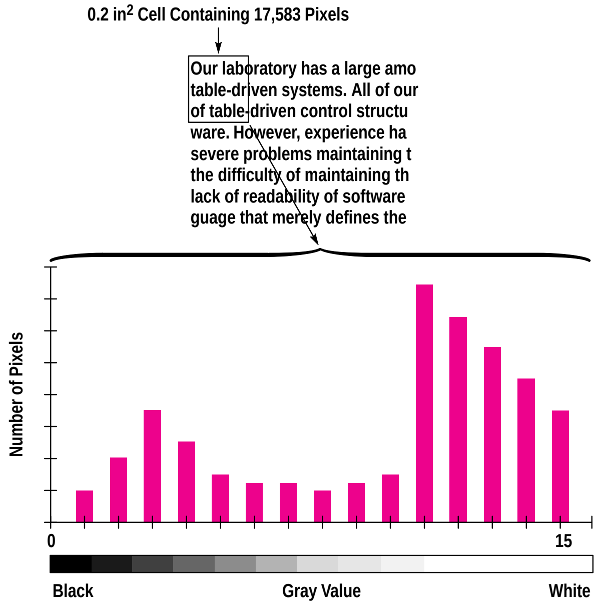
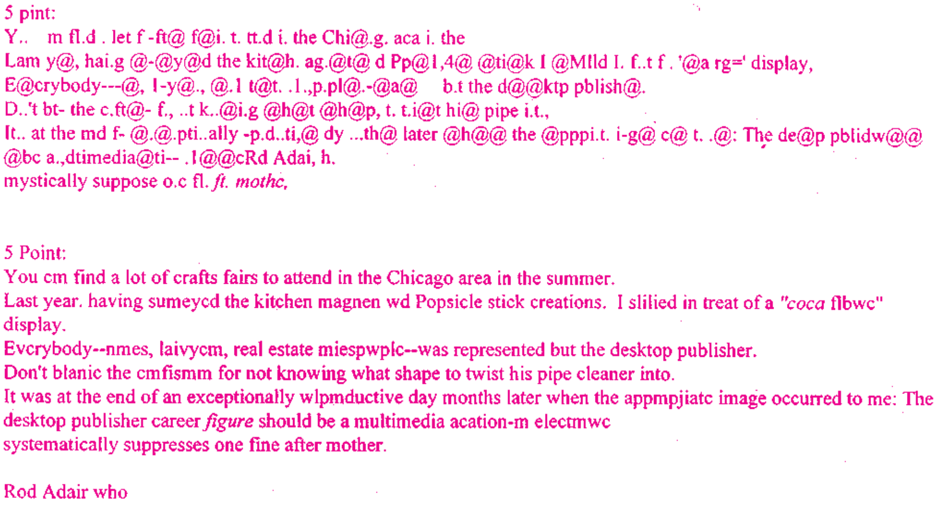


Fig. 8. Gray-level histogram. One of these histograms is made for each 0.2 in^2 area on a page.

ability to capture both text and images at the same time are the result of sophisticated page segmentation technology.

Histograms and Thresholds. HP AccuPage 2.0 deals with different color backgrounds with an adaptive threshold technique, which automatically determines the threshold level at different locations on the page. Instead of scanning the page at a binary setting, HP AccuPage scans using 16 (2^4) levels of gray. This four-bit representation of gray is mapped between 4% reflectance and 74% reflectance (the typical reflectance of a piece of paper) using all 16 levels for valid information. HP Accupage sets a threshold level for each 0.2 in^2 area on the page. The scanned image, along with an optimized threshold for each 0.2 in^2 part of the page, is sent to the OCR engine.

The HP AccuPage 2.0 optimized threshold algorithm samples the page by extracting a histogram of the pixel gray levels. This method sums the number of pixels found at each gray



5 pint:
Y.. m fl.d . let f -ft@ f@i. t. tt.d i. the Chi@.g. aca i. the
Lam y@, hai.g @-@y@d the kit@h. ag.@t@ d Pp@1,4@ @ti@k I @Mlld I. f..t f . '@a rg=' display,
E@crybody---@, l-y@., @.1 t@t. .1.,p.pl@.-@a@ b.t the d@@ktp pblish@.
D..'t bt- the c.ft@- f., ..t k..@i.g @h@t @h@p, t. t.i@t hi@ pipe i.t.,
It.. at the md f- @.@.pti..ally -p.d..ti,@ dy ...th@ later @h@@ the @pppi.t. i-g@ c@ t. .@: The de@p pblidw@@
@bc a.,dtimedia@ti-- .1@@cRd Adai, h.
mystically suppose o.c fl. *ft. mothc,*

5 Point:
You cm find a lot of crafts fairs to attend in the Chicago area in the summer.
Last year. having sumeycd the kitchen magnen wd Popsicle stick creations. I slilied in treat of a "*coca* flbwc"
display.
Evcrybody--nmes, laivycm, real estate miespwplc--was represented but the desktop publisher.
Don't blanic the cmfismm for not knowing what shape to twist his pipe cleaner into.
It was at the end of an exceptionally wlpmductive day months later when the appmpjiatc image occurred to me: The
desktop publisher career *figure* should be a multimedia acation-m electmwc
systematically suppresses one fine after mother.

Rod Adair who

Fig. 9. Results of reading small text (a) with the small text algorithm turned off and (b) with the algorithm turned on.

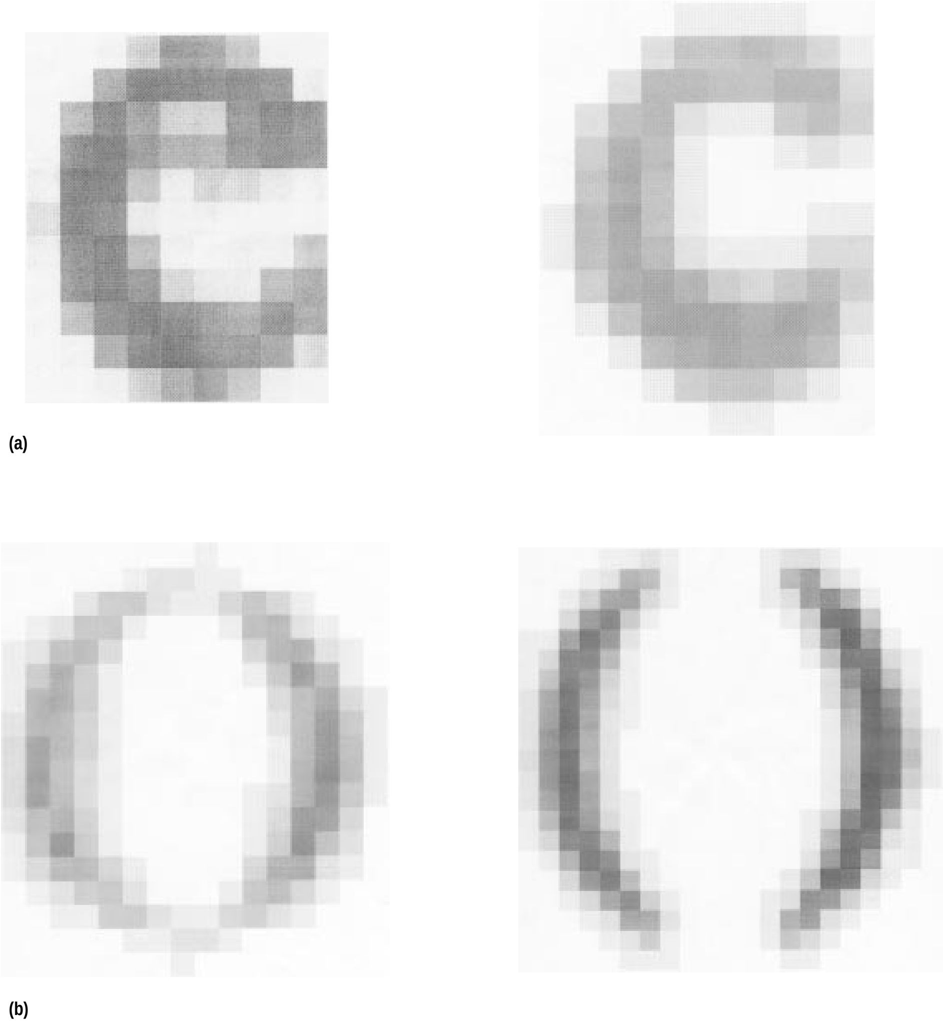


Fig. 10. Examples of doing upsampling with one bit per pixel. (a) If the threshold is too low an e starts to look like a c. (b) If the threshold is too high, an o turns into a pair of parentheses.

level. For four-bit data this results in a list of 16 gray levels that each contain the number of pixels found for a particular gray level. For pages that contain black text on white background the histogram will exhibit a strong bimodal distribution (see Fig. 8). The black text pixels will occupy the lower gray levels and the white background pixels will occupy the upper gray levels. Pages that contain text on colored backgrounds will also have a bimodal distribution, but not as well segmented because the background pixels will be some intermediate gray value instead of white.

The histogram allows HP AccuPage 2.0 to identify the gray levels that represent the text pixels, which should be preserved, and the gray levels that represent background pixels, which should be removed. By selecting a gray value that bisects the foreground (text) and background peaks, HP AccuPage 2.0 can convert the scanned page into a binary format that preserves the text as black and removes the background by assigning it to white. This preserves the text characters while removing the background pixels.

Small Text. HP AccuPage 2.0 is also able to recover small point-size text. The algorithm in HP AccuPage takes the four-bit grayscale information at 300 dpi and performs intelligent interpolation to provide more accurate character recognition. This algorithm takes a small amount of additional processing time, but it is able to improve recognition of text point sizes between five and seven point. Fig. 9 demonstrates the results of a scan with and without the algorithm provided in HP AccuPage 2.0.

One approach to doing this scaling might be to scan in binary (1 bit per pixel) and then upsample to 600 dpi. However, since the fine details would already have been lost, the upsampling would only provide larger versions of the same distorted characters. When closely inspected, these small character details, both the strokes and the transitions, appear as intermediate gray levels (see Fig. 10). If the threshold is too low, these fine details will be converted to black binary pixels, turning an e into a c (Fig. 10a). If the threshold is too high then the fine stroke of the small text will be converted to white pixels, converting an o to look like a pair of parentheses (Fig. 10b).

Another approach would be to tune the adaptive threshold algorithm to work well with small text. This method would produce well-formed small text at 300 dpi. The problem with this is that OCR applications are tuned for standard-sized text, which ranges from 8 to 14 points. As the text strays from those point sizes, the character recognition accuracy will generally deteriorate. This is especially pronounced when the point size is reduced.

The approach HP AccuPage 2.0 uses is to scan the page in grayscale at 300 dpi. The gray pixel data preserves the details so that after upsampling, the characters contain no distortions that might decrease the OCR accuracy. After scaling the page up to 600 dpi, HP AccuPage converts the grayscale value to binary, typically using its adaptive threshold algorithm.

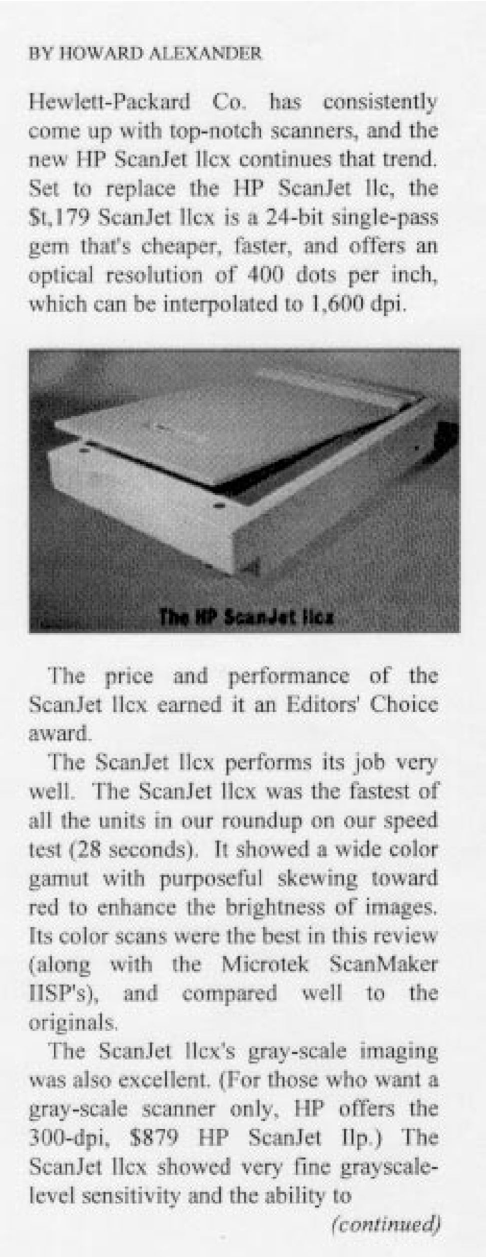
The algorithms for decomposing the page and separating pictures and text are another key aspect of HP AccuPage 2.0. Since some of these algorithms are in the patent application process, we chose not to describe them in this article.

Conclusion

HP AccuPage 2.0 is able to capture text and high-quality images from the scanned page. Many different image processing transforms are required to identify the image areas and retain their quality based on four-bit data input. Fig. 11 shows the capabilities of the algorithm to reproduce a page consisting of text and picture information. This capability is useful not only for OCR applications but also for document management and convenience copy applications in which the best possible rendering of the page for binary output is desired. These techniques ultimately make scanning more valuable for our customers.

Acknowledgments

We would like to thank Sue Arment, Ken Crow, Mimi Dong, Lynn Formanek, Bob Gann, Hal Hansen, Jose Moncivais, Tom Schoenleber, Christine Torrez, and Oscar Zuniga for their help with the HP AccuPage 2.0 product.



BY HOWARD ALEXANDER

Hewlett-Packard Co. has consistently come up with top-notch scanners, and the new HP ScanJet IIcx continues that trend. Set to replace the HP ScanJet IIc, the $1,179 ScanJet IIcx is a 24-bit single-pass gem that's cheaper, faster, and offers an optical resolution of 400 dots per inch, which can be interpolated to 1,600 dpi.

The HP ScanJet IIcx

The price and performance of the ScanJet IIcx earned it an Editors' Choice award.

The ScanJet IIcx performs its job very well. The ScanJet IIcx was the fastest of all the units in our roundup on our speed test (28 seconds). It showed a wide color gamut with purposeful skewing toward red to enhance the brightness of images. Its color scans were the best in this review (along with the Microtek ScanMaker IISP's), and compared well to the originals.

The ScanJet IIcx's gray-scale imaging was also excellent. (For those who want a gray-scale scanner only, HP offers the 300-dpi, $879 HP ScanJet IIp.) The ScanJet IIcx showed very fine grayscale-level sensitivity and the ability to

(continued)

Fig. 11. The result of using HP AccuPage 2.0 to scan in a portion of the document shown in Fig. 5.