



Digital Capture for Automated Scanner Workflows

Steven J. Simske, Scott C. Baggs
Imaging Systems Laboratory
HP Laboratories Palo Alto
HPL-2004-208
November 11, 2004*

E-mail: Steven.Simske@hp.com, Scott.Baggs@hp.com

user interface,
segmentation,
classification,
zoning, photos,
slides, negatives,
scanning

The use of scanners and other capture devices to incorporate film- and paper-based materials into digital workflows is an important part of "digital convergence", or the bringing of paper-based and electronic documents together into the same electronic workflows. The diversity of captured information—from text and mixed-type documents to photos, negatives, slides and transparencies—requires a combination of document analysis techniques to perform, automatically, the segmentation, classification and workflow assignment of the scanned images. We herein present technologies that provide fast (< 1.0 sec) and reliable (> 95% job accuracy) capture solutions for all of these input content types. These solutions offer near real-time capture that provides automated workflow capabilities to a repertoire of scanning hardware: scanners, all-in-one devices, copiers and multifunctional printers. The techniques used to categorize the documents, perform zoning analysis on the documents, and then perform closed loop quality assurance on the documents are presented.

* Internal Accession Date Only

Published in and presented at DocEng '04, 28-30 October 2004, Milwaukee, Wisconsin, USA

Approved for External Publication

© Copyright ACM 2004

Digital Capture for Automated Scanner Workflows

Steven J. Simske
Hewlett-Packard Labs
Mailstop 85, 3404 E. Harmony Road
Fort Collins, CO, 80528, USA
1.970.898.9561
Steven.Simske@hp.com

Scott C. Baggs
Hewlett-Packard Company
Mailstop 53, 3404 E. Harmony Road, Modular 5
Fort Collins, CO, 80528, USA
1.970.898.9429
Scott.Baggs@hp.com

ABSTRACT

The use of scanners and other capture devices to incorporate film- and paper-based materials into digital workflows is an important part of "digital convergence", or the bringing of paper-based and electronic documents together into the same electronic workflows. The diversity of captured information—from text and mixed-type documents to photos, negatives, slides and transparencies—requires a combination of document analysis techniques to perform, automatically, the segmentation, classification and workflow assignment of the scanned images. We herein present technologies that provide fast (< 1.0 sec) and reliable (> 95% job accuracy) capture solutions for all of these input content types. These solutions offer near real-time capture that provides automated workflow capabilities to a repertoire of scanning hardware: scanners, all-in-one devices, copiers and multifunctional printers. The techniques used to categorize the documents, perform zoning analysis on the documents, and then perform closed loop quality assurance on the documents are presented.

Categories and Subject Descriptors

B.4.2 [Input/Output and Data Communications] Input/Output Devices - *Image display*. G.4 [Mathematical Software] *Algorithm design and analysis, Efficiency, User interfaces*. I.2.10 [Artificial Intelligence] Vision and Scene Understanding - *Intensity, color, photometry, and thresholding, Perceptual reasoning, Shape, Texture*. I.4.1 [Image Processing and Computer Vision] Digitization and Image Capture - *Scanning*. I.4.6 [Image Processing and Computer Vision] Segmentation - *Edge and feature detection, Pixel classification, Region growing, partitioning*. I.4.8 [Document and Text Processing] Scene Analysis - *Color, Object recognition, Shape*. I.5.3 [Pattern Recognition] Clustering - *Algorithms, Similarity measures*. I.5.4 [Pattern Recognition] Applications - *Computer vision, Text processing*. I.7.5 [Document and Text Processing] Document Capture - *Document analysis, Graphics recognition and interpretation, Scanning*.

General Terms

Algorithms, Performance, Design

Keywords

User Interface, Segmentation, Classification, Zoning, Photos, Slides, Negatives, Scanning.

1. CHALLENGES IN DIGITAL WORKFLOWS

Consumer scanning hardware—including desktop scanners, handheld scanners, and all-in-one devices—has shipped with intelligent capture software for the past decade. Historically, much of the consumer-targeted scanning software has been based on a user interface (UI) preview/final mode [1].

This preview/final mode of document scanning requires user feedback—the document placed on the scanner is scanned, often at a low (< 75 pixels/inch, or ppi) resolution, and displayed in a "live preview" mode so that the user can work with the document quickly. Once the preview is available to the user, the user can rapidly create, adjust, and save elements of documents—called regions or "zones". Adjustments include sharpening, altering the resolution and scaling, lightening and darkening, hue and saturation changing, mirroring, rotating, moiré minimizing, and—in the case of binary documents—thresholding. Saving initiates the "final" scanning mode, where the changes made to the document in "live preview" are applied during the final scan. The output of a single clicked region in schema-driven XML [2] is shown in Figure 1.

Complementing the preview/final motif is fully automatic document scanning. Here, there is no UI, and many documents can be scanned one after the other with automatic content understanding technologies—such as zoning analysis, optical character recognition (OCR), and archival—being applied. Automated modes require higher accuracy, more complete quality assurance (QA), and often multiple parallel algorithms to reduce the severity and occurrence of errors. Our team has focused extensively on zoning analysis as a key differentiator, while OCR and archiving have typically been achieved through third-party partnerships.

The move to automation is exemplified by the currently shipping scanning application, which automatically draws a selection rectangle around the object on the scanner in preview mode. It does not *require* user feedback unless the correct object was not found or the user had something else in mind (e.g. only wants a part of the page we found). The emphasis is on the "ease of use" and therefore the need for full automation with high accuracy. In fact, if the user launches a scan using a button on the scanner the default is to not show a preview, but send the region(s) that analysis finds directly to an imaging application. This naturally imposes the following design constraint—if the software is not 100% certain about what to crop to, include more rather than less and ensure the user has the full image in her end application.

```

<?xml version="1.0"?>
<ZoningSet
xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xmlns="http://pstd.hpl.hp.com/zoning.xsd">
  <n_regions>1</n_regions>
  <XResolution>75</XResolution>
  <YResolution>75</YResolution>
  <width>638</width>
  <height>875</height>
  <ZonedRegion>
    <region_type>DRAWING</region_type>
    <region_modality>BW</region_modality>
    <bbox>
      <xmin>88</xmin>
      <xmax>257</xmax>
      <ymin>47</ymin>
      <ymax>199</ymax>
    </bbox>
    <polygon>
      <n_vertices>5</n_vertices>
      <vertex>
        <xcoord>88</xcoord><ycoord>199</ycoord>
      </vertex>
      <vertex>
        <xcoord>257</xcoord><ycoord>199</ycoord>
      </vertex>
      <vertex>
        <xcoord>257</xcoord><ycoord>47</ycoord>
      </vertex>
      <vertex>
        <xcoord>88</xcoord><ycoord>47</ycoord>
      </vertex>
      <vertex>
        <xcoord>88</xcoord><ycoord>199</ycoord>
      </vertex>
    </polygon>
  </ZonedRegion>
</ZoningSet>

```

Figure 1. XML describing (using the schema `pstd.hpl.hp.com/zoning.xsd`) a region created by a single click.

Scanner users are currently—in general—more interested in automatic workflows than in preview/final modes. In order to provide higher-accuracy workflows that are either fully automatic or provide consistent accuracy to improve throughput, *the repertoire of technologies described here are planned for the entire 2005 scanning product line in all scanning modes*—flatbed, ADF (automatic document feeder), APF (automatic photo feeder) and Transparent Materials Adapter (TMA).

An automated approach that also allows consumer ease of use requires software more sophisticated than those in which slides, negatives and even multiple photos must be placed in pre-defined portions of the scanbed for proper capture. For currently-shipped scanning software, scanners with complex lids (that is, with ADF and/or TMA built in) typically provide a white cover to mask the lid complexities. This is clunky and confusing—and often the consumer throws out the cover sheet upon setting up the hardware. Given this, our goal has been to automate capture processes so that the user can send the images they want directly to their end application with little or no "upfront" rules about how and where to place the images in/on the scanning device. Additionally, these technologies allow HP to ship lower end scanners without the white cover resulting in a lower cost scanner and no user issues using the cover.

As a particular example of this, we herein describe the challenges involved in automating intelligent capture with the HP 4670 scanner, which has a see-through design with the consequence that the background for images captured can be virtually anything. Solving this problem is a superset of solving the problem of ever-changing automatic document feeder (ADF) and automatic photo feeder (APF) footprints that are inevitable with changing scanner product lines. ADF/APF footprints often show the mechanical parts (belts, guides, etc.) used to move sheets across the scanbed, adding additional background detection challenges that must be overcome for intelligent capture.

2. TYPES OF DIGITAL CAPTURE

2.1 Text

Text is historically the most important information in a document, allowing OCR and ICR (handwriting) technologies to provide re-purposing, summarization, categorization and indexing of the captured material. Text-based documents, however, have become less important to the low-end, consumer devices such as scanners and digital cameras. Instead, the focus for many scanner vendors has shifted to photos and related region types such as (film) negatives, slides and transparencies. Even in the capture of photos, however, text regions are identified for a variety of purposes: (1) possible meta-data about the photos; (2) to provide QA for any cropping that takes place; and (3) to allow the user to capture mixed-type documents as easily as "photos only". Fortunately, text regions, on documents with traditional layouts, can be disambiguated from photos based on classical page analysis techniques [3]. Text regions are also identified by their the density of sub-threshold (black) pixels and the repetitive nature of text rows (using, for example, projection profile analysis) when skew-corrected.

2.2 Photo

Photos are a specialized region type, and they are of increasing importance as scanner purchasers focus more on photography. Also, because the identification of photos is more general than that of slides and negatives (which are, after all, specialized photos), we present here some of the methods to simultaneously address performance and accuracy of photo-find.

Identifying the content can be done from the top-down and from the bottom-up. The top-down method is generally useful when we do not know in advance any salient information about the content scanned, or when multiple items may be scanned simultaneously (in other words, fully-automatic zoning analysis). The bottom-up method (e.g. click and select [4]) is generally faster, and often more accurate, when the user has selected a particular workflow (fax, copy, email, archive, etc.) or when a (fast) pre-analysis of the content has been performed.

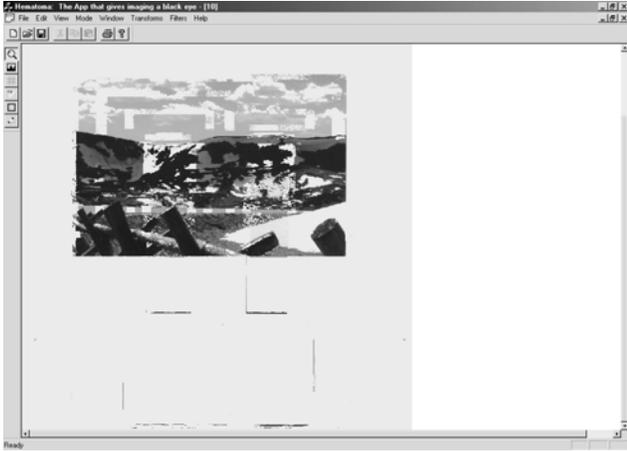


Figure 2. Effect of masking (XOR) an ADF background (here achieved via whitening out the two TMAs and a photo door that are built into the lid) to enable top-down photo find. Note that the 3D features (and incumbent shadows) of the ADF leave gaps in the photos and some lines in the extra-photo space of the image which are readily filtered out.

Using the *top-down method*, full zoning (segmentation and classification) analysis of the scanned image is performed, matching the background to templates [5] or other formatted descriptions of the scanner lid/cover and then determining the boundaries (segmentation) and types (classification as photos, drawings, line art, text, slides, negatives, etc.) of the foreground regions—that is, what the user intended to scan. Figure 2 shows the application of a similar technology (whiting out the ADF lid complexities) to photo find. After whiting out the lid + photo combined scan, any narrow runs of non-white data are removed (this eliminates the lines below the photo). Next the remaining pixels are smeared [3] and regions are formed. In Figure 2, only one (5"x3.5") region is formed.

Using the *bottom-up method*, the content regions are defined by searching for their boundaries. For example, captured photos, slides and negatives have in common the following characteristics:

- (1) They are rectangular and their edges usually show as edges against any background
- (2) They have characteristic sizes (4"x6", 3.5"x5", 10x15 cm, 35 x 23.3 mm)
- (3) They often appear in known locations, such as against the side of the scanner, or in the region where the slides or negatives are inserted.

Because of (1), photos can be searched for using projection profiles at the predominant angles of skew. These skew angles are determined by using a modified Hough transform in which only the edges of large regions in the scan are sampled. To reach this point, the following steps are required:

[A] Threshold the image. This is found from the best fit of two Gaussian populations to the histogram [6]. Although several other thresholding techniques were found to work similarly [7-9], we chose a modified form of [6] based on minimum mean squared error from the empirically (human) determined optimum

threshold over a test set of 100 documents. That is, the method chosen had the lowest probability of "catastrophic" failures that prevent downstream processing from successfully analyzing the captured content.

[B] Form edge map [10]. For efficacy and high-accuracy, edges can be defined as the residual pixels after thresholding a variance map, where each pixel at (i,j) in this map is represented by its variance V:

$$V_{ij} = |4 P_{ij} - P_{i-1,j-1} - P_{i-1,j+1} - P_{i+1,j-1} - P_{i+1,j+1}|$$

More sophisticated edge techniques can be deployed for higher accuracy, but we found the variance map (97% accuracy, 0.1 sec processing time) to provide a good trade-off in performance and accuracy. The edges are then filtered to reduce false positives. Two such filtering techniques require the pixels to also be above the threshold in [A] or to have directional correlation with nearby edge pixels.



Figure 3. Finding a photo against a belt-driven ADF (light boundary). Many line candidates arise from the belt artifacts in the background, but the correct photo is cropped to in the *HP Photo & Imaging Gallery*.

[C] Determine the step size between edge pixels. We target 1500-3000 separate pixels to compute the skew angle (which has been shown to be sufficient on our document corpus to determine skew angle within 0.25 degrees 99% of the time), and so for example if there are 8000 edge pixels, we choose a step size of 4 and use only 2000 separate pixels in the skew detection. An angle increment of 0.25 degrees is used, and the angles between the centroids of the separate pixels are accumulated.

[D] Compute the 10 most likely angles of skew (from the angles with highest number of entries in [C]) and generate projection profiles (arrays of the percent of edge pixels in each successive skewed line cutting across the image) at each angle to find line candidates. We chose 10 based on experiments on a document corpus of 100 "difficult" scans—while 4 candidates provided the correct skew candidate 95% of the time, we had to extend to 9 angles to get the entire set of 100 correct.

[E] Form all potential rectangular candidates and assign a fitness value to each based on the size and percent of the rectangle outline that is edges (better rectangular candidates will tend to have better edges). The relative fitness values were determined empirically using a modified steepest descent procedure.

[F] Sequentially eliminate overlapping rectangles based on cost until no remaining rectangles overlap.

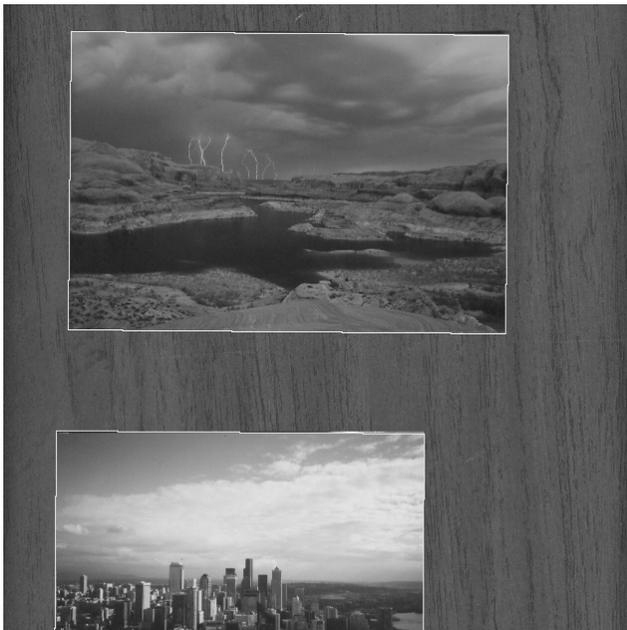


Figure 4. Photo find against a grainy table top. Light boundaries indicate the photos that are automatically cropped to in the HP Photo & Imaging Gallery.

After all of the rectangular regions (which can also be the page boundary for a document) are found, the residual pixels can be analyzed to determine what the background is. This can be used to determine different scanner lid backgrounds and/or inserts automatically, which may be valuable in optimizing the downstream image processing. After photos, text documents, mixed documents, slides, negatives, etc. have been identified and/or cropped to, it is often the case that the residual pixels do not match any known background (especially if no background templates exist). In this case, an extra step is performed. Any rectangular regions previously identified in the scanned image are removed and zoning analysis is performed on the remaining pixels. If a high probability of important content (text, photos, drawings, etc.) being leftover is obtained, a statistical signal is sent to the calling software suggesting that a larger cropped area (or indeed the entire scanbed) be used instead of the currently cropped set. This also allows the management of regions that overlap each other (i.e. a photo over a white background over a

scanbed can be sequentially cropped to). Examples of photo finding using steps [A]-[F] are shown in Figures 3-4. In each case, analysis of the backgrounds revealed the lack of text or distinct photo/drawing regions.

At each stage of the analysis, relevant statistics on the confidence of each decision made are computed. Skew, crop, photo find and region analysis output all self-report these statistics, allowing the calling workflow software to decide when to use and when to disregard the analysis output.

2.3 Slides and Negatives

Slides and negatives provide a simpler analysis problem than photos, since they are surrounded by at least some width of black boundary on two or more sides. Using this information, the process of finding these two region is greatly sped up by first defining and convexing the black (sub-threshold) regions, and then forming regions from the non-black pixels within these regions. When the slide or negative region boundaries are smaller than 35 mm by 23.3 mm, they are adjusted to this size by maximizing the variance of the rows and columns of pixels. This is necessary since slides and negatives are often very dark. Examples of slide and negative identification are shown in Figures 5-6.



Figure 5. Results of slide finding algorithm (9 different examples shown). Rectangular boundaries (light) are snapped to standard size.

2.4 Background

As mentioned above, distinguishing the foreground from the background is crucial for intelligent capture. Many commercial OCR engines, however, simply snap all backgrounds to white, which can result in a "ransom note" effect on complex document layouts—that is, the snapped white may itself be different from the background of the page. Fast and accurate automatic capture

of scanned information relies on the identification of both the content itself and the background against which it was scanned.



Figure 6. Results of slide and negative finding. Rectangular boundaries (yellow) are snapped to standard size. Note the partial and missing negatives are not selected.

As mentioned above, the release of the HP ScanJet 4670 "see-through vertical" scanner (Figure 7) this past year both intensified the need for and stimulated the development of "background-proof" technologies that allow intelligent capture of content while filtering out the background. This is because the 4670 can be placed in virtually any direction—on the wall, facing up, or facing down—to capture images.



Figure 7. HP ScanJet 4670.

The normal use of the 4670 is with the HP Photo & Imaging Gallery (Figure 8). Figure 8 shows the automatic cropping of a photo on the scanner. The photo can be automatically sent to a file, another application, or emailed.

As a consequence of its see-through design, the background for photos captured by the 4670 can be virtually anything, and so capture software had to be able to address this problem while still capturing the images. Fortunately, solving this problem is a superset of solving the problem of ever-changing automatic document feeder (ADF) and automatic photo feeder (APF) footprints. ADF/APF footprints often show the mechanical parts (belts, guides, etc.) used to move sheets across the scanned (e.g. Figure 3), and a software solution is substantially more cost-

effective than a hardware solution. Thus, techniques for handling diverse backgrounds allow the effective capture of photo, document, negative, slide and transparency content. Also, rectangular region capture is a simplified form of both shape-templated and quadrilateral finding techniques also performed by our team. This allows trade-offs between robustness (quality) and performance (time for analysis) to be made in our software. We discuss these trade-offs in the next section in light of the new focus on "fast automation" for scanning workflows.



Figure 8. Auto-crop to photo in HP Photo & Imaging Gallery.

These technologies also work better than competitive algorithms when complex lid types are "whited out" (as in Figure 2) as they are less sensitive to residual noise left behind in the template-subtracting procedure. One approach is to allow the user to automatically capture any type of content with minimal effort and without customized placement of material—the technologies described here make such an approach possible.

2.5 Cropping

Finding photos, slides and negative are specialized forms of crop. Cropping of documents—similar to the finding of photos, slides and negatives—can be performed from the "top down" or from the "bottom up". From the top down, crop is determined from the boundaries of the largest rectangle found. From the bottom up, crop is determined from the smallest rectangle that incorporates all of the regions found by zoning analysis. If the top-down crop fully encloses the bottom-up crop, a degree of QA has also been assured.

3. ZONING ANALYSIS

Zoning analysis—the determination of region type (classification), location (segmentation) and logical association (clustering, including the formation of columns, and ordering)—underlies both the preview/final mode and the automatic mode of scanning. Classification consists of determining the region type, and the set of region types varies considerably depending on the destination. The purpose of classification is to define the output type (common values are 1-bit binary, 8-bit grayscale, and 24-bit color) and the proper final resolution to scan the region (e.g. 300 ppi for Roman text, 200 ppi for photos, and 400 ppi for fine line drawings). Segmentation and clustering are used to provide the

minimum possible set of regions for clarity, while providing enough regions to allow proper region order and to prevent the misclassification of regions. Zoning analysis tools are the starting point for all of the region finding technologies discussed here. Zoning analysis technologies that we have developed include the following three broad types: (1) fully-automated zoning analysis, (2) user-directed zoning analysis, and (3) multi-engine zoning analysis.

Zoning analysis is used in every digital image capture workflow described above, even when a different method (as in section 2.2) is used for the primary region definition. In general capture, zoning analysis is used to distinguish text regions (which are scanned in the final pass at 300-400 pixels per inch, but are binarized) from photo regions (which are scanned in the final pass at 150-200 ppi, but with full—e.g. 24-bit—bit depth). After identifying photos, slides and negatives, zoning analysis is used to analyze the residual captured data to make sure no photos, drawings, text areas, etc., are lost.

3.1 User-Initiated

User-initiated zoning analysis is fundamental for the preview/final motif. Click and select [4], for example, allows the user to click anywhere on the preview image, and zoning analysis software generates from the "bottom up" a classified and segmented region at or near the point clicked. Other UI tools which enhance the preview/final workflow include adaptive region typing and clustering [11] and the use of the familiar highlighter motif for UI-based capture [12]. In the preview/final mode, it is important to provide a relevant preview to the user as soon as possible. To accommodate this, low resolution (typically 30-100 ppi) content analysis is performed for all of the analysis technologies described here. In fact, very low resolution (30-50 ppi) zoning analysis is performed to provide a "quick estimate" of the region zoning, so that the user can instantly start working on the preview image [13].

3.2 Automated

Fully-automated zoning analysis [3, 14-16] provides a complete description of all of the regions on the scanned image. The simplest region set is {text, non-text}; whereas most scanner software has shipped with the following repertoire of identified regions: text, table, photo/image, black and white (1-bit, 8-bit) drawing, and color (24-bit) drawings. Other region types (equation, handwriting, graphs, and business graphics) are mapped to one of these. This is more specialized than most publicly-available zoning engines [3, 14-16]. In fact, several other region types are identified. They are mapped directly to this set; for example, equations and handwriting are mapped to black and white drawings, while graphs and business graphics are mapped to color drawings. Historically, slides and negatives have also been mapped to photos. In addition to region classification, a region manager is required to determine how to cluster regions. Regions can, for example, be allowed to overlap other regions—this means that bounding box and polygonal descriptions of the regions are not sufficient for describing the region. Instead, a full scan line segment description is used throughout. This allows any Z-order depth; for example, text within a rectangular border over a blue background on a white sheet of paper has Z-order 3 (the paper is Z-order 0 by convention). Due to these concerns, fully-automated zoning analysis typically requires 2-3 seconds of

processing time per page. This adds some delay for the user, but is much faster than, for example, OCR (5-15 seconds/page).

User-directed zoning analysis precludes the delay of 2-3 seconds for fully-automated zoning analysis, and thus a preview screen is presented as soon as the (low resolution) scan is completed. This is of benefit to the users, since they can see what they've scanned more quickly. Also, click and select, slider and highlighting tools are familiar to the user and allow the selection of specific content. Most regions so selected are calculated in less than a second, and so the "analysis time" is largely transparent to the user.

3.3 Multi-Engine

Multi-engine zoning analysis consists of two distinct approaches: (1) where a low-resolution (placeholder) analysis is performed to reduce the delay in preview, and later corrected by a higher-resolution analysis; and (2) when multiple zoning engines are used together to improve the overall quality of the analysis. We have found that regions differently classified by two engines, for example, are 15 times more likely to be misclassified in statistical combination than regions classified the same by two engines. This allows us to direct user attention to these problem regions, and make their editing that much more efficacious.

4. RESULTS

Software for skew detection, crop and photo finding has been shipped with the HP ScanJet 4670 and other 2003 product line scanners after assessing the performance, memory footprint and accuracy. On a challenging 50-file test set for combined (simultaneous) auto-cropping and photo finding, our performance was 0.94 sec/page using our benchmark system (a 1.13 GHz Pentium III laptop with 768 MB of RAM) for evaluation, with 94% accuracy. In the remaining 6% of the files, the additional quality assurance tests (zoning analysis outside the boundaries of the defined regions and/or values of the confidence statistics) allowed these workflows to inform the user of the errors and make simple corrections. Overly-aggressive cropping (and loss of user-intended material) did not occur. Two such results are shown in Figures 3-4, in which non-standard sized photos are found against a complex ADF background.

On the other end of the performance spectrum, Figure 5 shows the results for a small suite of slides, for which the accuracy is 100% and the analysis time is negligible—13 msec/file mean. Figure 6 shows the results for a suite of 27 negatives, for which the accuracy (100%) and performance (16 msec/file mean) both exceed specifications.

Recently, we have extended the photo finding specifically to zoning multiple photos with potentially differing skews on the scanned. Competitors mainly provide "cookbook" solutions to this problem, wherein photos are specifically arranged and unskewed. To enable random placement of the photos on the scanned, background training is performed on the image. Background candidates are identified from the intensity and saturation (for color images) histograms of the image. Background candidates are weighted by their ubiquity (distribution on the scanned), their quantity (percent of the scanned pixels) and the number and rectangularity of the regions enclosed. The best background is selected from these candidates. Figure 9 shows the results of background training for a scanned with three images on it.

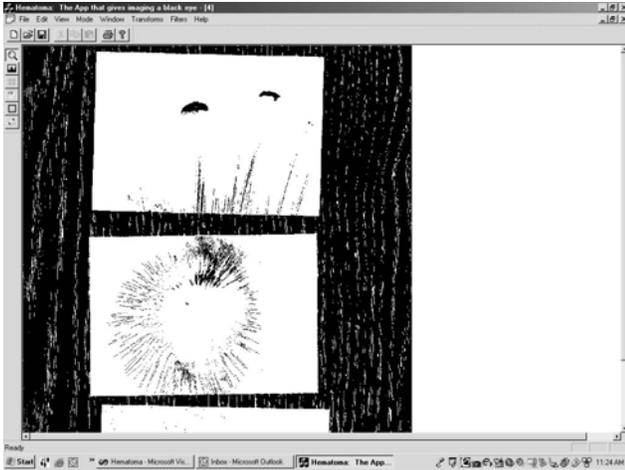


Figure 9. Results of background training, same desktop as Figure 4. The black pixels fall within the range of the background range for intensity and saturation. Three distinct photos are then easily extracted from the white pixels.

5. DISCUSSION

The "photo find" technology was shipped with the HP ScanJet 4670 scanner in September 2003. This product included automatic identification of photos (Figures 3-4), as well as skew detection and smart cropping as discussed above. In the 2004 product line, this analysis set has been extended to include automatic multiple photo detection for slides and negatives as well as photos—both through parallel photo find technologies and through combined template+zoning analysis techniques. The 2004 product line also ships the first release of the slide and negative finding technologies. These latter technologies will also be adopted by the all-in-one (scanner + copier + printer) product line, since the algorithms involved are not tied to any specific layout.

Future plans are to use these technologies to automate capture processes on the enterprise-level scanning devices, including copiers and multifunctional printers (MFPs). Workflows for copiers and MFPs can be more readily automated with intelligent capture techniques, even for largely text-based processing.

6. ACKNOWLEDGMENTS

The authors thank many helpful teammates—especially Patty Lopez, Jian Fan, Jeff Lee, Archie Carrington, David Auter, Jay Veazey, Virgil Russon and Ted Neff—for making these technologies possible for customers over the past decade. Finally, we thank Ellen Tappan for helpful suggestions on the paper and its organization.

7. REFERENCES

[1] Santos, F.E. "A scanning system in which a portion of a preview scan image of a picture displaced on a screen is selected and a corresponding portion of the picture is scanned in a final scan," U.S. Patent 4,837,635, June 6, 1989.

[2] Simske, S. "The Use of XML and XML-Data to Provide Document Understanding at the Physical, Logical and Presentational Levels," In Proc. of the ICDAR99 Workshop on Document Layout Interpretation and its Applications, Sept. 1999.

[3] Lee, J.P., Lopez, P.D., and Simske, S.J. "Click and select user interface for document scanning," U.S. Patent no. 6,151,426, Nov. 21, 2000.

[4] Simske, S.J., Carleton, J.M. and Lesser, R.R. "Digital imaging device with background training," U.S. Patent 6,683,984; Jan. 27, 2004.

[5] Wahl, F.M., Wong, K.Y. and Casey, R.G. "Block segmentation and text extraction in mixed/image documents," Computer Vision Graphics and Image Processing, Vol. 2, pp.375-390, 1982.

[6] Kittler, J. and Illingworth, J. "Minimum error thresholding," Pattern Recognition, Vol 19, no. 1, pp. 41-47, 1986.

[7] Otsu, N. "A threshold selection method from gray level histograms," Pattern Recognition, Vol 9, no. 1, pp. 62-66, 1979.

[8] Kurita, T., Otsu, N. and Abdelmalek, N. "Maximum likelihood thresholding based on population mixture models," Pattern Recognition, Vol 25, no. 10, pp. 1231-1240, 1992.

[9] Kittler, J., Illingworth, J. and Föglein, J. "Threshold selection based on a simple image statistic," Comp. Vision Graph. Image Proc., Vol 30, pp. 125-147, 1985.

[10] Wilkinson, M.H.F. "Optimizing edge detectors for robust automatic threshold selection: coping with edge curvature and noise," Graph. Models Image Proc., Vol 60, pp. 385-401, 1998.

[11] Simske, S.J. and Lee, J.P. "System and method for manipulating regions in a scanned image," U.S. Patent 6,263,122, Jul. 17, 2001.

[12] Simske, S.J. and Lesser, R.R. "User interface high-lighter function to provide directed input for image processing," U.S. Patent 6,385, 351, May 7, 2002.

[13] Simske, S.J. and Russon, V.K. "Document analysis system and method," U.S. Patent 6,674,901; Jan. 6, 2004.

[14] Revankar, S.V. and Fan, Z. "Image segmentation system", U.S. Patent 5,767,978, January 21, 1997.

[15] Shi, J. and Malik, J. "Normalized cuts and image segmentation," IEEE Trans Pattern Analysis Machine Intelligence, Vol 22, no. 8, pp. 888-905, 2000.

[16] Zramdini, A. and Ingold, R. "Optical font recognition from projection profiles," Electronic Publishing, Vol 6, no. 3, pp. 249-260, Sept. 1993.