NISTIR 5726

# Generalized Form Registration Using Structure-Based Techniques

**Michael D. Garris**
**Patrick J. Grother**

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Gaithersburg, MD 20899



NIST

NISTIR 5726

# Generalized Form Registration Using Structure-Based Techniques

**Michael D. Garris**
**Patrick J. Grother**

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Gaithersburg, MD 20899

April 1996

U.S. DEPARTMENT OF COMMERCE
Mary L. Good, Acting Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
Arati Prabhakar, Director

# Generalized Form Registration Using Structure-Based Techniques

Michael D. Garris (mdg@magi.ncsl.nist.gov)

Patrick J. Grother (patrick@magi.ncsl.nist.gov)

National Institute of Standards and Technology
Building 225, Room A216
Gaithersburg, MD 20899

## ABSTRACT

A new method for registering forms has been developed at the National Institute of Standards and Technology. This method automatically estimates the amount of rotation and translation in the image without any detailed knowledge of the form. This is accomplished through the automatic detection of dominant vertical and horizontal structures (lines) commonly found in forms. A general method for rotation estimation and a robust method for translation estimation are presented. Results demonstrate that this technique is extremely tolerant to spurious annotations on the form and scanner noise in the image, and the computational requirements of the utility can be tuned by optionally choosing to process and analyze downsampled versions of the image. All 3,669 Handwriting Sample Forms distributed with *NIST Special Database 19* were successfully registered with the new technique, and using the same code, 255 uniformly laid out IRS tax forms and 500 Census miniforms were also tested and registered. Every type of form contained in the numerous NIST (public) form databases can be registered using this technique. These results also demonstrate how easy it is to set up the computer to register new types of forms, introducing a set-up interface that is much more automated and less tedious than what is currently required to specify new forms for the NIST public domain Form-Based Handprint Recognition System.

**Keywords**: databases, form registration, optical character recognition, public domain, rotation, translation, skew

## 1. INTRODUCTION

This paper presents new work conducted at the National Institute of Standards and Technology (NIST) on improving the state-of-the-art of automated forms processing and the recognition of handprinted information entered on forms. In August of 1994, the NIST Form-Based Handprint Recognition System (a software system that reads the handprint entered onto forms) was released to the public.[1] As of the writing of this paper, over 475 copies of this software have been distributed across 38 countries around the world. The public domain system has proven to be an effective vehicle for technology transfer, intended to provide a working knowledge of the technology, and to demonstrate how optical character recognition (OCR) from forms can be evaluated.

The NIST public domain OCR system was designed to read the handwriting printed on Handwriting Sample Forms (HSF) like the one shown in Figure 1. The recognition system locates the boxes on the form, extracts the handwriting, and recognizes the handprinted characters. To accurately locate the boxes (fields) on the form, the system must account for image degradations such as rotation, translation, and scale introduced by the processes of form printing, replicating, handling, and scanning. The original NIST system was engineered to use spatial histograms to locate predetermined registration points on an HSF form degraded with as much as ±5° of rotation in conjunction with ±1.3cm (0.5 in) of translation in x or y. The registration points (documented and illustrated in Reference 1) include the position of the form's title and the corners of specific boxes. Using a method of Linear Least Squares[2], the *hypothesis* points located on the input image are mapped to a set of *reference* alignment points measured off-line from a prototype form. Estimates on the amount of rotation, translation, and scale are computed and the input form image is deskewed so as to coincide with the prototype form. A zone template of the fields (again measured from the prototype form) is then used to extract the handwriting in the fields on the deskewed form.

This method of form registration was thoroughly tested and shown to work very well on *prototypical* HSF forms. However, when tested all 3,669 HSF forms distributed with *NIST Special Database 19* (SD19)[3], it was discovered that not all the forms could be registered due to variations introduced over three stages of collection. In addition, adapting the public domain system to read other types of forms, while possible, requires a detailed reworking of the specific software module that uses spatial histograms to locate the registration points. This module, which was programmed to look for marks specific to HSF forms, has to be replaced by a customized code that accurately and consistently locates registration features on the new form. NIST has produced a number of other form databases containing Internal Revenue Service (IRS) 1040 tax returns known as *NIST Special Databases 2* (SD2)[4] and *6* (SD6)[5], and databases containing extracts of 1990 Census Long Forms referred to as *NIST Special Databases 11* (SD11)[6] though *13* (SD13)[7]. The process of incorporating these other types of forms into the current public domain system is overly tedious and extremely burdensome. In light of these factors, three goals were embraced for this project: 1.) a new form registration tool should be developed to process *all* the variations of HSF forms in SD19; 2.) the new registration scheme should be general enough to work on all the different form types in the NIST databases; and 3.) the new method should make the integration of new forms into the recognition system much easier.

In meeting these goals, it was determined that the technique should automatically and consistently locate dominant structures within the form *without* any detailed knowledge of the form itself. That way the need to manually customize the registration points on each and every new type of form is avoided. The detection of structures should be quite impervious to annotations on the form, spurious scanner noise, and other sources of noise such as tears in the paper and staple holes. These structures should also be mapped to their ideal position with a *minimal* amount of a priori knowledge as well. Ideally, this mapping information would be obtained automatically by running the structure location code on an image of a prototype form, and the locations of the structures would be stored as the mapping data onto which future form images will be registered.

The following sections describe a new method of form registration that achieves all these goals. Section 2 describes a generic technique for detecting the amount of rotational skew in a document image. In this case the image contains a form, but the detection of rotation works well on most scanned pages containing machine printed text, tables, and figures (especially if the page contains large horizontal lines). Once the image's rotational skew is estimated and removed, a method for detecting translation is applied. The technique, described in Section 3, capitalizes on the fact that, once rotational skew has been removed, the vertical and horizontal lines (prevalent in most forms) are square to the raster grid in the image. The left and right-most dominant vertical lines in the image are located, and the top and bottom-most dominant horizontal lines are located. These data points are then analyzed and mapped to those measured from the prototype form, and translational distances in x and y are computed. Because full page document images are very large (often over 8 million pixels in size), global image analyses and pixel transformations are extremely expensive. Section 4 describes a couple of steps that can significantly reduce the execution time needed to register forms. Section 5 presents the results of applying the new registration method on the various types of forms in the NIST databases. As will be seen, the registration technique does an exceptionally good job at deskewing and aligning HSF forms, tax forms, and the Census form extracts. Conclusions are drawn in Section 6.

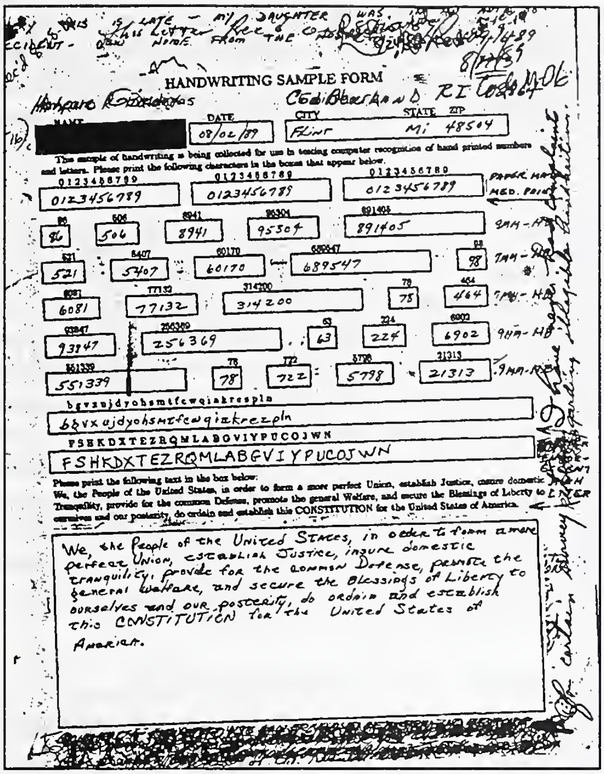
## 2. ROTATION ESTIMATION

We have employed a technique for rotation estimation that is similar to a technique originally described by Postl.[8] A description of a variant of this method and a discussion of the issues relevant to skew estimation for machine printed documents can be found in Reference 9. The method is simple and effective, though, in its naive implementation, not particularly efficient.

Given a binary image with dimensions $w$ and $h$, we can estimate the global rotational skew, $\theta_m$, by maximizing the skew function

$$S(\theta) = \sum_{i=1}^{n} e^{p_i/q_i} \tag{1}$$

where $p_i$ is the sum of the black pixels on the $i^{\text{th}}$ parallel line trajectory (ray) inclined at angle $\theta$, and the expected number of black pixels $q_i$ is obtained by dividing the total number of black pixels in the image by the image height. We have used only those $n$ rays that intercept the vertical axis at the left edge of the image and are inclined horizontally at $\theta$ degrees. The locations of the coordinates on each ray are determined using Bresenham's line drawing algorithm[10], and adjacent rays are merely vertically offset versions of one another.



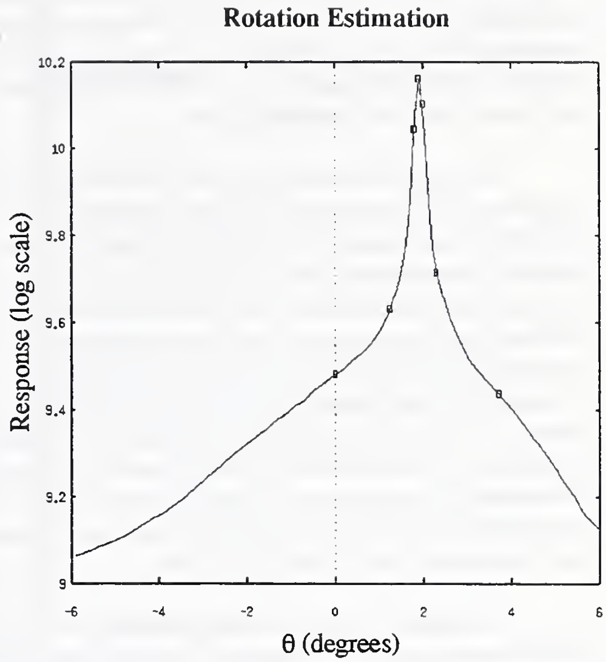




Figure 1. A noisy Handwriting Sample Form (left) containing rotational skew, and its skew function $S(\theta)$ response (right) plotted at 0.1° intervals.

Rotation estimation proceeds by selecting trial angles and constructing the sum of the exponentials of the directed ray occupancies according to Equation (1) and determining the angle $\theta_m$ that gives maximum skew function response. For images with prominent horizontal structure, the function $S(\theta)$ contains a strong resonance at the global skew angle. The method works very well if the image contains ruled horizontal lines or rows of machine printed text. For more general images the response may degrade, becoming less peaked, choppy or, in some instances, flat. Given suitable images, we seek the position of a maximum in the skew functions. The naive method is to evaluate $S(\theta)$ at many points in a global search. Given that $S(\theta)$ is expensive to compute for typical document images (containing greater than $10^6$ pixels), a traditional one-dimensional optimization algorithm that requires fewer function evaluations is applied. Given an initial bracketing interval, Brent's method[11] returns the position of the maximum to within a specified tolerance, and in our application, requires one sixth of the number of function calls. For an N×M image the cost of each skew function evaluation is approximately $O$(NM). It is therefore advantageous to subsample the image prior to the optimization search. Downsampling and other efficiency issues are discussed in Section 4.

A noisy HSF form is displayed on the left in Figure 1. This form has been artificially superimposed with annotations and scanner noise collected across 500 different HSF forms and then rotated at approximately 2°. The graph on the right plots on a log scale the response $S(\theta)$ for $-6° \leq \theta \leq +6°$. The points on the curve correspond to angles selected by the Brent optimization algorithm, in which case, only 7 angles were required to evaluate $S(\theta)$ and locate the maximum response. The maximum point selected by the optimization algorithm is at 1.9°, which corresponds very well to rotation known to be in the actual image. The noisy HSF form with its rotational skew removed is shown in Figure 2.

## 3. TRANSLATION ESTIMATION

It was stated in the introduction that the new registration method should be able to detect and locate dominant structures within a form without any detailed knowledge of what is in the form. Taken literally, this requirement is impossible, but with a few reasonable assumptions the task becomes achievable. Every form to be registered is assumed to contain one or more dominant horizontal lines and one or more dominant vertical lines, and the configuration of these dominant lines is assumed to be fixed across all the forms of a given type. By dominant, we mean that these structures are significantly longer in contiguous length than any other information on the form including machine printed titles and instructions, handprint responses and spurious annotations, and other sources of relatively unpredictable noise. Many forms satisfy this assumption as they contain multiple vertical and horizontal rules (straight lines) that demarcate regions on the form. This assumption can be easily met on other forms by simply "picture-framing" the form inside one all-encompassing box.

These dominant horizontal and vertical lines are the structures that are to be automatically detected within each image. The ability to detect these structures is greatly enhanced as a result of being able to estimate the rotation within the image (described in Section 2). Once detected, the original image is rotated so that any rotational skew is removed. The dominant lines in the image are now relatively aligned with the raster grid, so to locate the dominant horizontal lines in an image, all the horizontal runs (contiguous sequences of black pixels) in the de-rotated image are computed, and the location of each run's left-most pixel is stored along with the run's pixel length. The runs are sorted first in ascending order on their y-coordinates, and secondly (where multiple runs exist on the same row) they are sorted in descending order according to their length. The top-$n$ runs in length from each row are then summed together and stored in a histogram, one histogram accumulator (bin) for each row in the image. In this application, the top 3 longest runs in each row are summed together. The dominant horizontal lines in the image can be easily detected by analyzing the resulting histogram values. The rows in the image containing long segments of horizontal lines stand out in comparison to those lines containing many small runs generated by machine printed text, handprinted responses, annotations, staple holes, and other types of noise.

An example of one of these run-based histograms is shown in Figure 2. The rotational skew in the noisy HSF form on the left side of the figure has been automatically detected and removed. Plotted on the right side of the figure are the run-based histogram values derived from the form. Notice that the peaks in the histogram directly correspond to the top and bottom edges of the rows of boxes on the form and that the annotations and noise are effectively suppressed and easily ignored.

For registration purposes, the top-most and bottom-most dominant horizontal lines are selected. These are chosen by searching the histogram for the maximum bin value, and then starting at the top and searching down the bins, the first value found to exceed 50% of the maximum value is determined to be the top hypothesis coordinate. By starting at the bottom and searching up the bins, the first value found to exceed the same threshold is determined to be the bottom hypothesis coordinate. The image is rotated 90° on its side and the whole run length and histogram process is repeated to find the left and right dominant vertical lines in the image. The left, right, top, and bottom structures in the image are now compared against the prestored reference coordinates measured from the original prototype form. By comparing these values, it can be estimated how much translation in x and y is needed to align the current form with the prototype form. Only one vertical and one horizontal structure need be detected to compute these x and y distances respectively. However, having two vertical and horizontal coordinates adds redundancy and a significant amount of robustness to the process.

First, the left and right hypotheses are analyzed to compute the horizontal (x) translation. The error between the hypothesis and reference left coordinates are computed, as is the error between the right pair. If the difference between the two pair's errors is within an acceptable tolerance (in this case 6mm or 0.25in) then both points are used to calculate translation; otherwise, the pair with the larger error is rejected. If both points are used, then the average of the two errors is used to represent the horizontal translation in the image, else the smallest error of the two points is used. This error comparison process is then repeated on the top and bottom hypotheses to compute the vertical (y) translation. The redundancy has a number of benefits. If one of the structures is incorrectly detected, in other words some other line in the image was mistakenly selected, then the corresponding coordinate is determined to be in error and not used in calculating the translation. At times, there may be enough noise in the image to preclude the detection of a structure in the run-based histograms, but rarely will both structures be missed. Plus, by having two correctly detected structures, taking the average of errors between the two helps compensate for scale distortions in the image. Upon calculation, if either of the horizontal or vertical translations exceed an upper limit (in this case 2.54cm or 1.0in) then the entire form registration is rejected.

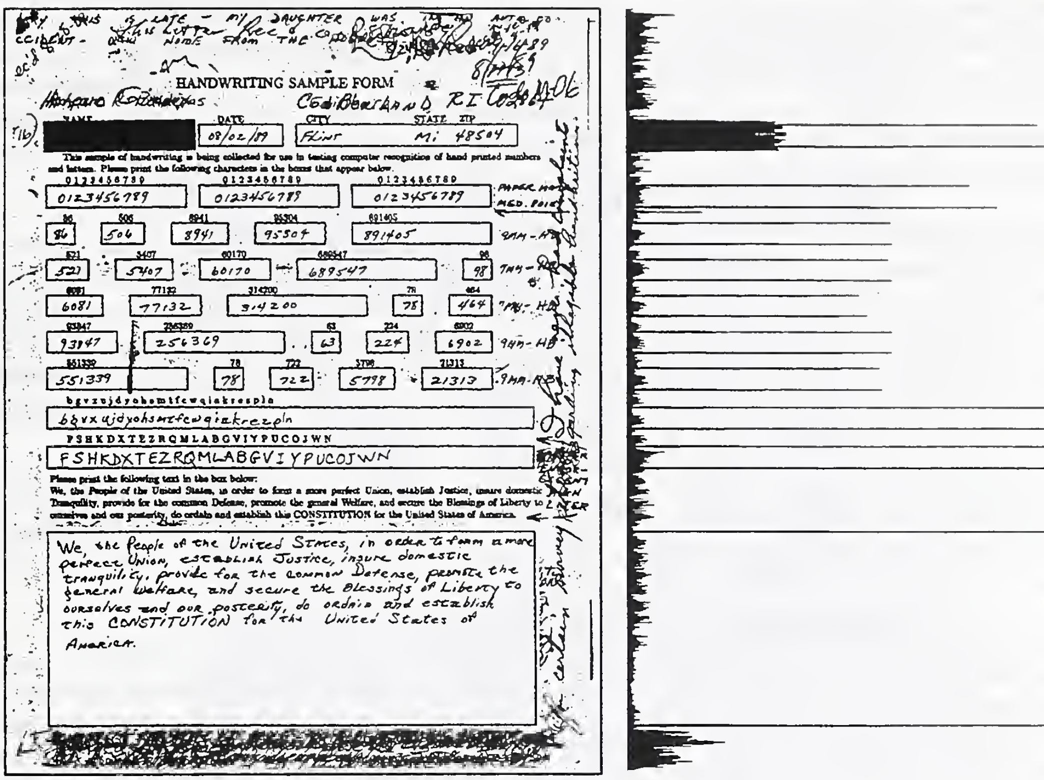



Figure 2. Run-based histogram of the de-rotated noisy HSF form.

## 4. EFFICIENCY ISSUES

Throughout this process, a number of global image analyses and image pixel transformations are performed. A full page image scanned at 11.8 pixels per millimeter (ppm) or 300 pixels per inch (ppi) will typically contain over 8 million pixels; therefore, these analyses and transformations (even when cleverly coded) are very expensive and become the performance bottlenecks in the system. Global analyses include the summing of pixels along skew trajectories for rotation estimation and the calculation of runs for translation estimation. Image pixel transformations are performed to de-rotate the image after rotational skew is detected, to rotate the image 90° on its side to locate left and right structures, and to align the image once translation estimates are computed. A couple of steps can be taken to reduce the computational burden of these operations.

The routines used to detect rotational and translational skew can be performed on downsampled images. This directly cuts down the number of pixels used in the image-based operations. In our tests, downsampling by a factor of 4 and 8 were used. The downsampling used does not throw away (or skip) rows and columns in the image; rather, it aggregates rows and columns in order to condense the image. This is accomplished by sliding a non-overlapping square window across the image the size of the desired reduction factor in pixels, and if at any point the window contains one or more black pixels (logical-OR), then a black pixel is written to the reduced output image. This method of downsampling causes a binary blurring which helps preserve the structures within the image for later detection. The trade-off with downsampling is that structures detected at the lower resolutions are more *loosely* defined as they are mapped back to the original (higher) resolution of the image. In other words, as the amount of downsampling is increased, both the accuracy of rotation estimation and the pinpointing of structures in the original image decrease. So, depending on the application, a little error in the detection of structures may be tolerable, whereas in other applications it may not.

For the purpose of comparison, all the timings reported in the paper were generated on a Sun Microsystems SPARCstation 2 running SunOS 4.1.3 with an 80MHz Weitek CPU and 64Mb of internal memory.* A typical (non-multiple of 90°) rotation on an 20.8×27.9 cm (8.5 × 11 in) page scanned at 11.8ppm has 2560×3300 pixels and requires 7 seconds. When reduce by a factor of 4, the image is 640×825 pixels and requires 0.4 seconds. The image reduced by a factor of 8 is 320×413 pixels and only requires 0.1 seconds, an increase in speed of 70 times over the original image. These improvements are large compared to the time of 0.8 seconds needed to do the downsampling by a factor of 4 and 0.4 seconds to downsample by a factor of 8.

Using downsampling within the new form registration scheme involves taking the input image of the form and reducing it by a factor of 4 or 8. This condensed image is then analyzed (computing $S(\theta)$ along multiple skew trajectories) and the amount of rotation in the image is calculated. The condensed image (rather than the original image) is then rotated to remove any rotational skew. The de-rotated image is then analyzed (computing two run-based histograms with a 90° rotation in-between to locate left and right structures) and the translation of the image is calculated. The translation parameters are scaled back up to the resolution of the original image, and then in a single transformation, both the rotational and translational skew in the original image is removed.

A dramatic reduction in time is achieved by minimizing the operations conducted on the full-scale image. To register an HSF form (20.8×27.9 cm page scanned at 11.8ppm) without any downsampling requires 29 seconds. The same form when downsampled by a factor of 4 requires 6.3 seconds, whereas using a factor of 8 only requires 4.4 seconds. The trade-offs of using downsampling are visually illustrated in the next section.

Another step that can be used to decrease the burden of image pixel transformations capitalizes on the fact that the images of the forms being processed are binary (pixels are black or white). Taking this into consideration, only the transformation addresses of black pixels need be computed. On many forms, as little as 10% of the entire page may be comprised of black pixels, so the majority of the image (the white space) is ignored. A transformation such as rotation is a continuous (or analog) operation on the pixels. However, the pixels in the image are represented on a discrete grid, therefore addresses of rotated pixels must be rounded off causing a certain amount of error in the operation. To minimize the effect of this error, pixels are typically *pulled* from the original image into the rotated image. For each pixel in the output image, addresses are computed to the input image and pixels values at these locations are copied. In so doing, an address is computed for every pixel in the output image. In order to consider only the black pixels in the image, the operation of pulling must be replaced by *pushing*. For each black pixel in the input image, compute the destination address in the output image. The benefit is that all the white space in the image is ignored in the transformation. The trade-off is that the discrete rounding error is no longer minimized causing small amounts of white speckle noise in the image.

An 11.8ppm HSF form rotated at 3° using the optimal-coverage pulling algorithm requires 7 seconds. The same image rotated using the pushing algorithm requires only 1.8 seconds. This represents almost a factor of 4 speed up. The feature extraction and classification algorithms used in the NIST recognition system are by design tolerant to small amounts of speckle noise, therefore the improvements in speed gained by the rotational push far out-weigh the trade-offs of increased noise in the image. If this source of noise is of great consequence to your specific application, then other more error-reducing rotations can be used.

## 5. RESULTS

The new form registration method was tested on the different types of forms in the NIST databases. In each case, a blank prototype form was processed and the locations of its left, right, top, and bottom structures were automatically detected and stored. Successive images of forms were then registered using the new technique, and the resulting images were stored to disk. If the form registration is successful, then the registered images will all coincide with each other. To visually check this, multiple registered forms were logically ORed together and the resulting composite image was inspected. In all cases, the results were quite pleasing.

* Specific hardware and software products identified in this paper were used in order to adequately support the development of the technology described in this document. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply the equipment identified is necessarily the best available for the purpose.

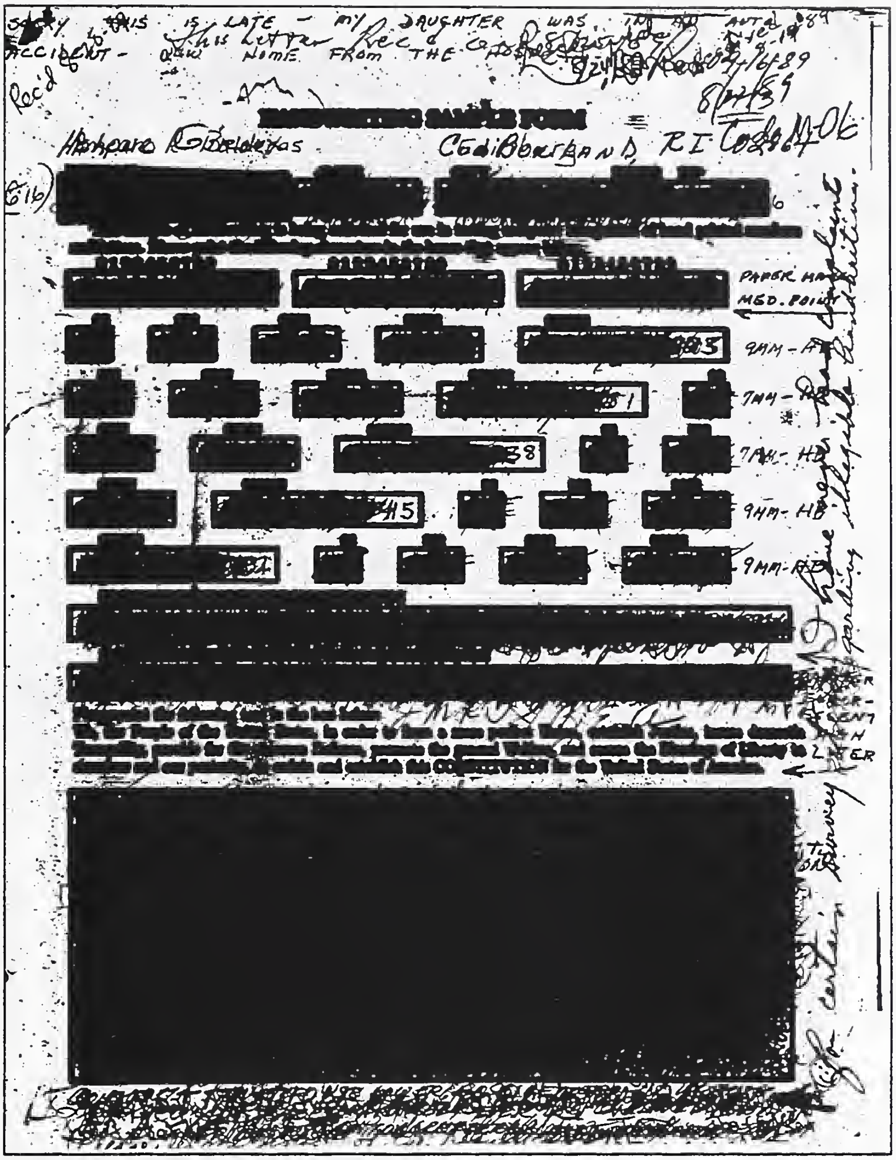


Figure 3. Composite image of 500 registered HSF forms using a downsampling factor of 4.

The first set of forms tested were the HSF forms from SD19. As anticipated, the left, right, top, and bottom structures detected on the HSF forms corresponded to the outer perimeter of the boxes on the page. All the forms were registered and blocks of approximately 500 forms were ORed together. Figure 3 shows the registration results from the set of 500 HSF forms called hsf_2 from SD19 overlaid (ORed together). In this example, a downsampling factor of 4 was used to detect rotation and locate structures within the image. Notice the relatively tight correspondence across the 500 forms. The shapes of letters and words comprising the title and instructions on the form are still somewhat distinguishable. Keep in mind that it only requires one instance of a black pixel across the 500 forms to turn a pixel black in the composite image. Notice all of the handprinted annotations collected across the set of forms. Numerous people wrote notes in the top, left, and bottom margins. A number of writers also ran out of room when writing the Preamble to the U.S. Constitution, so they completed their response by writing in the bottom margin of the form. The amount of annotations and other sources of scanner noise in this image testifies to the tolerance and robustness of the new registration method.

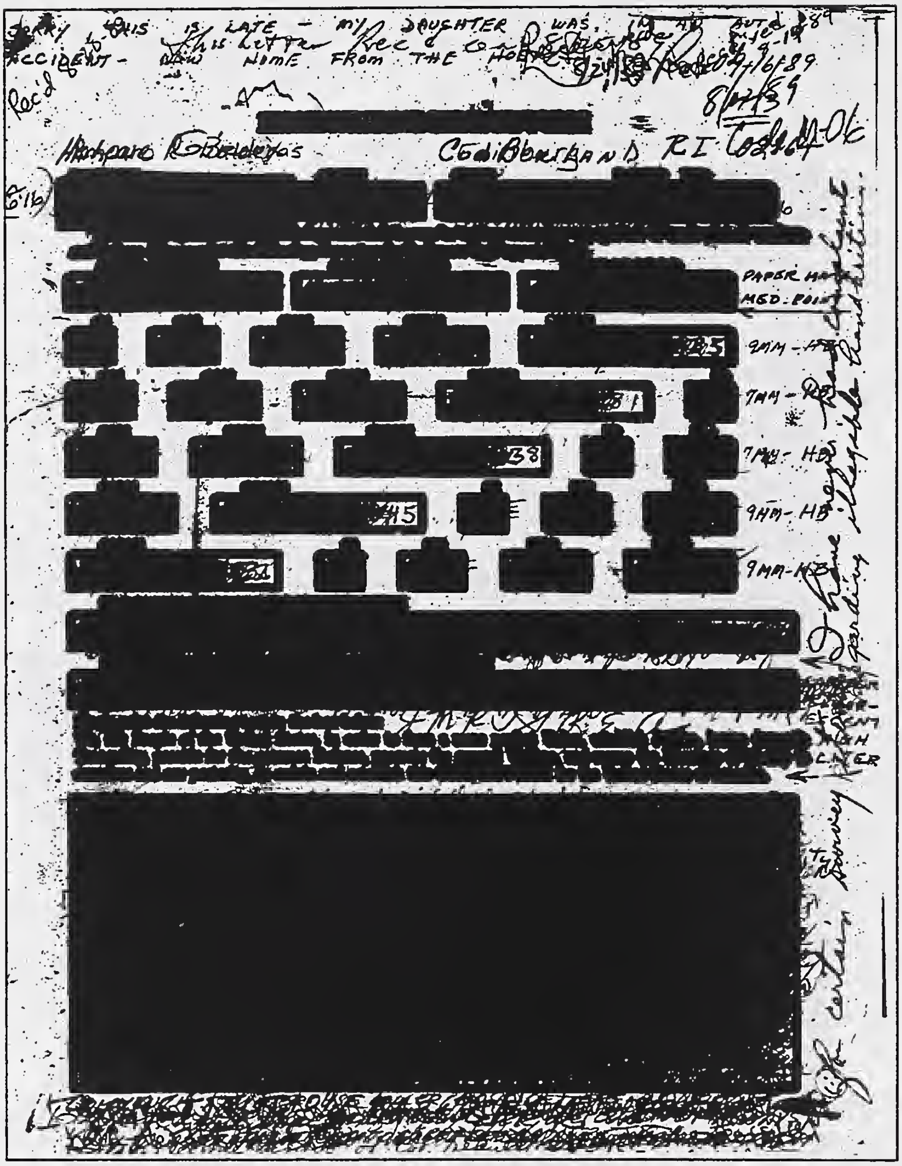


Figure 4. Composite image of same 500 registered HSF forms using a downsampling factor of 8.

The image displayed in Figure 4 shows the registration results from the same set of forms used in Figure 3, only a downsampling of 8 was used in this case. The effects of increased downsampling can be seen by comparing the images in these two figures. By using a factor of 8, the resulting composite image is more blurred as a result of poorer correspondence across the 500 forms. The words (and certainly the characters) in the title and instructions are no longer distinguishable in the second image, but these registration results required less time to compute, and it is believed that the correspondence in the second image is good enough to isolate the handprint in the boxes on the form. Upon inspection, it was determined that the all the HSF forms in SD19 were successfully registered using a downsampling factor of 8 (even with all the variations in the form due to the stages of collection).

The next set of images tested were from SD6, a database of computer synthesized 1040 tax returns. The composite shown in Figure 5 was generated from 98 front pages of the IRS 1040 form. These forms were registered using the same code that was used to register the HSF forms. The left structure detected corresponded to the vertical line along the left edge of the right-most column of money fields (to the left of the line numbers 7 through 31) on the bottom

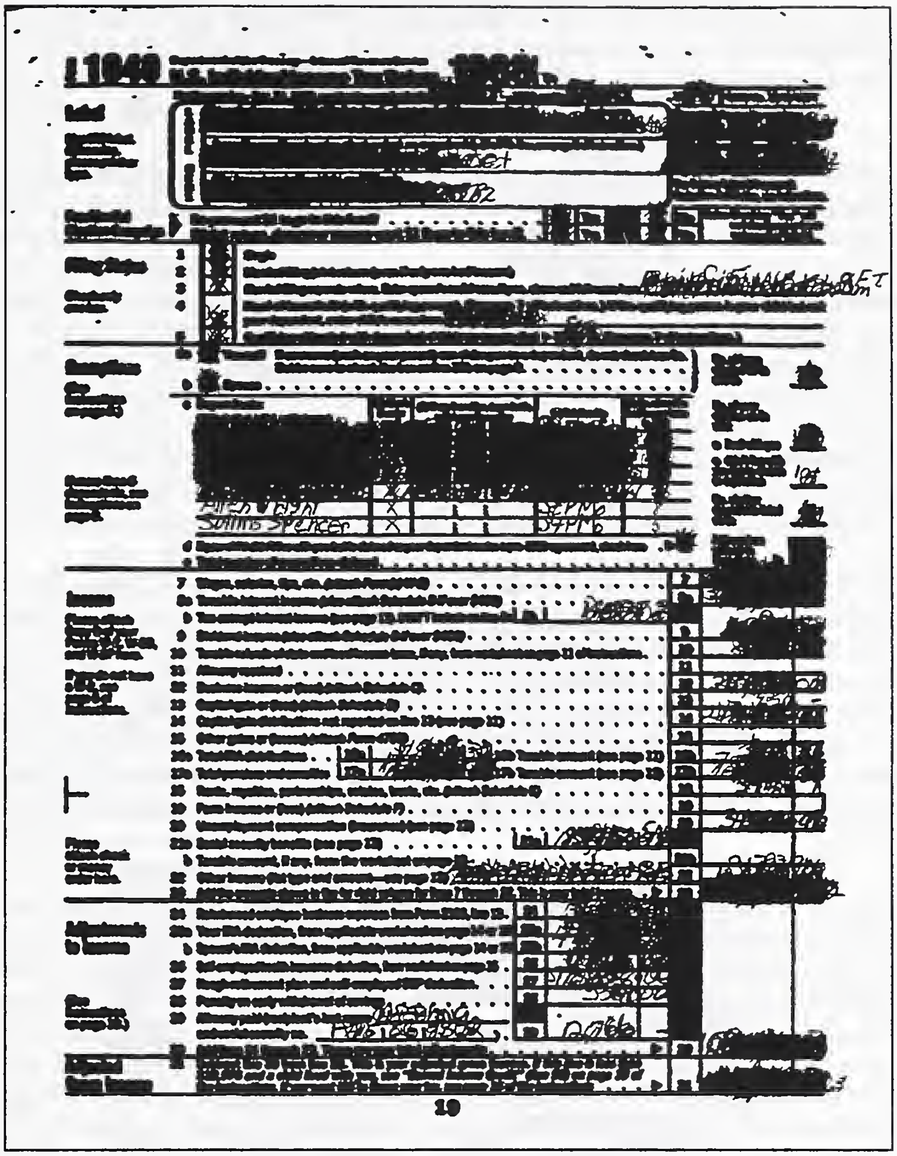
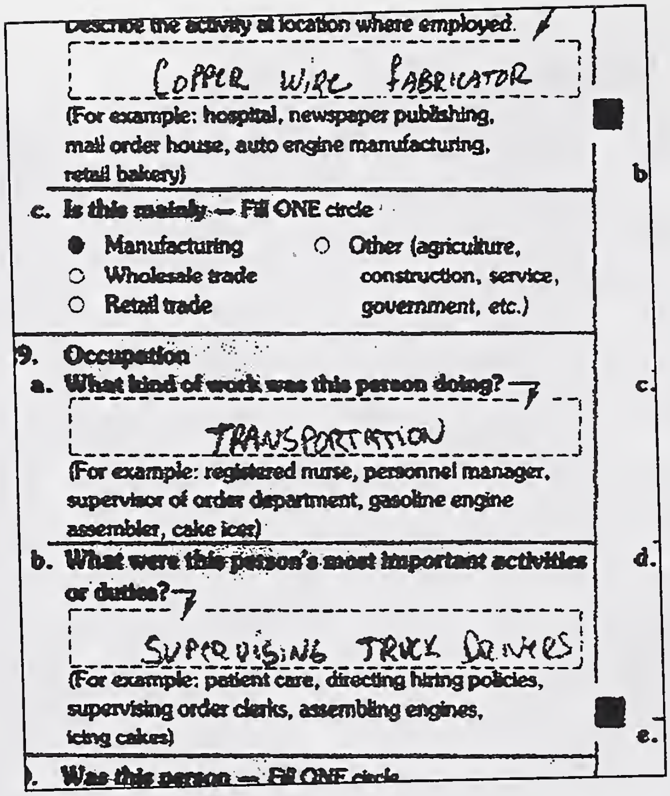


Figure 5. Composite image of 98 registered front pages of the IRS 1040 form using a downsampling factor of 4.

half of the page. The right structure corresponded to the right-most vertical line running through these same fields. The top structure was the top-most horizontal line on the form (just below the title), and the bottom structure was the bottom-most horizontal line (just above the page number). One hundred of these forms were processed using a downsampling factor of 4, and even though the original forms contained a significant amount of rotational skew and translation, 98 registered successfully and their correspondence in the figure is quite good. Other tax forms (not shown here) were also tested, including 100 second pages of the 1040 form and 55 Schedule A forms. Using the downsampling factor of 4, all the second pages of the 1040 form and all but one of the Schedule A forms registered successfully. It should be noted that the 3 IRS forms that did not register successfully at a reduction factor of 4, were retested and successfully registered when no downsampling was used.

The last type of forms tested were extracts from the 1990 Census Long Form, called *miniforms*. These forms are distributed in SD11 through SD13, and an example of these forms is shown in Figure 6. In order to set up the registration on this type of form, a prototype form was deskewed and it structures automatically located using the run-based histograms. The top-most horizontal line (just above question c.) was selected as the top reference structure, and



Describe the activity at location where employed.

COPPER WIRE FABRICATOR

(For example: hospital, newspaper publishing, mail order house, auto engine manufacturing, retail bakery)

c. Is this mainly — Fill ONE circle

● Manufacturing
○ Wholesale trade
○ Retail trade
○ Other (agriculture, construction, service, government, etc.)

9. Occupation

a. What kind of work was this person doing?

TRANSPORTATION

(For example: registered nurse, personnel manager, supervisor of order department, gasoline engine assembler, cake icer)

b. What were this person's most important activities or duties?

SUPERVISING TRUCK DRIVERS

(For example: patient care, directing hiring policies, supervising order clerks, assembling engines, icing cakes)

. Was this person — Fill ONE circle

Figure 6. A Census miniform containing rotational skew.

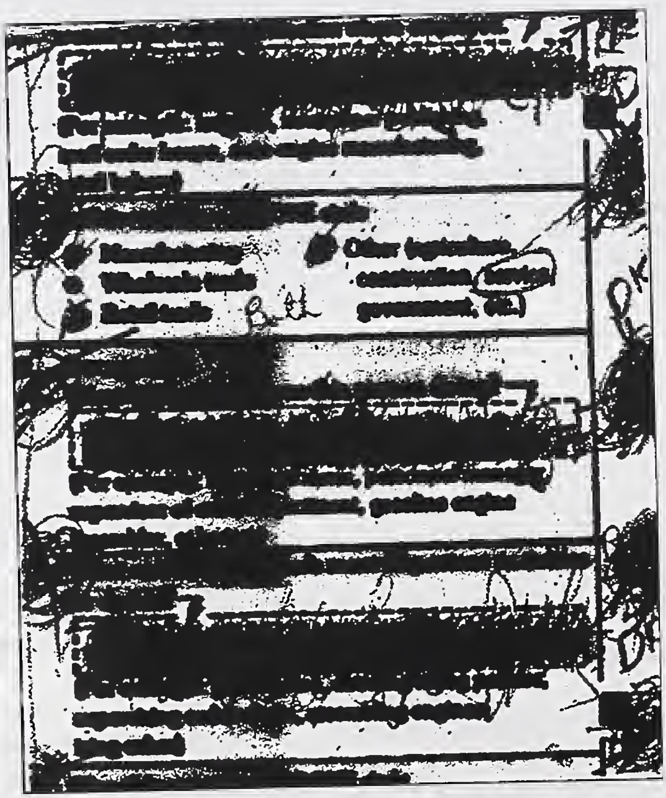


Figure 7. Composite image of 100 registered Census miniforms using no downsampling.

the bottom-most horizontal line (just below question b.) was selected as the bottom reference structure. The vertical line extending down the right side of the form was automatically selected to be both the left and right reference structures in the image, as this is the only dominant vertical structure on the form. This didn't cause the registration algorithm any problems, but it does decrease redundancy and therefore make registration very sensitive to the correct location of this structure.

Nonetheless, the registration performed successfully across a test set of 500 miniforms scanned directly from microfilm (not paper). The composite image shown in Figure 7 is comprised of the first 100 registered forms in this set. These images were digitized at 7.9ppm (200 ppi) with 624×744 pixels, so they are much smaller in size than the HSF and tax forms; also the widths of the lines in the image are rather thin and quite jagged; therefore, no downsampling was performed prior to estimating the amount of rotation and translation in the image. On average, a census miniform took 2.2 seconds to be registered. Notice the reasonably good correspondence of the words and the vertical and horizontal lines on the forms. There tends to be a lot of annotation on these forms as people circled and checked off the different question numbers on the form. The original images (because they were scanned from microfilm) also contain a significant amount of *pepper* (black speckle) noise, which contributes to the blotchiness in parts of the composite image. Two of the 500 forms were rejected by the registration process as a result of their structures not being reliably detected. Upon inspection, these forms were determined to be scanning failures in which case the forms were shifted far enough to the right during scanning that the vertical line on the right side of the form was clipped.

Given the success of the new registration method on different types of forms, it was determined that the technique should be tested on an extremely difficult problem that would push the limits of the generalized approach. NIST has a private image database of *real* 1988 IRS 1040 tax returns. The forms in this database are representative of the variation of forms currently seen at IRS tax processing centers. This sample is much different than the *synthesized* tax forms distributed in SD2 and SD6, in which single templates of each form type were perturbed with varying amounts of rotational and translational skew. The real forms vary significantly as a result of relatively relaxed specifications on the layout and printing of current tax forms. The topology of instructions and fields on the forms remain fairly consistent, but the size and demarcation of the fields vary, and (most significantly to the new registration method) the location and the number of rules on the forms vary.

To test the registration, the front pages of 100 writers' 1040 forms were studied. A prototypical form for extracting reference structures was selected, but it was a difficult choice due to the numerous variations in the sample. Using these reference structures, the remaining forms were registered using the new method. Upon inspection, the registration was determined to have performed quite poorly. A small subset of forms that were similar to the prototype form did register successfully, but the remaining 1040 forms either failed registration or had significant errors. It may be possible to categorize the population of 1040 forms into consistent subsets and utilize prototype forms for each subset, but many of the differences between the forms are subtle and would be very difficult to detect. This demonstrates the new technique's reliance on the assumption that the configuration of the dominant lines are fixed across all the forms of a given type. Without adherence to rigid form design specifications[12], there can be no reliable form prototype; and without a reliable form prototype, this generalized form registration method will likely fail.

## 6. CONCLUSIONS

A new method for registering forms has been presented. This method automatically estimates the amount of rotation and translation in the image without any detailed knowledge of the form by detecting dominant structures (vertical and horizontal lines) commonly found in forms. Results demonstrate that this technique is extremely tolerant to spurious annotations on the form and scanner noise in the image, and the computational requirements of the utility can be tuned by optionally choosing to process and analyze downsampled versions of the image. All 3,669 HSF forms distributed with SD19 were successfully registered. Using the same exact code, 255 uniformly laid out IRS tax forms and 500 Census miniforms were also tested and registered. Over 4,400 forms were tested with only two automatically rejected due to scanning errors; the rest were correctly registered using varying levels of downsampling. In fact, every type of form contained in the numerous NIST (public) form databases can be registered using this technique. These results also demonstrate how easy it is to set up the computer to register a new type of form. A prototype form is run through the rotation estimation, de-rotated accordingly, and the left, right, top, and bottom dominant structures on the form are automatically located and stored as future reference points. This introduces a set-up interface that is much more automated and dramatically less tedious than that currently required to enter new forms into the NIST public domain OCR system. Plus, if the form being introduced to the system does not have the necessary vertical and horizontal structures already embedded within it, the form can be simply "picture-framed" within a bounding box to make this registration scheme work. As demonstrated by these results, the new form registration tool achieves all the immediate, intermediate, and longer range goals set forth at the beginning of this project. A new release of the NIST Form-Based Handprint Recognition System incorporating the source code for this new technique is expected as a result of this work.

## REFERENCES


1. M. D. Garris, J. L. Blue, G. T. Candela, D. L. Dimmick, J. Geist, P. J. Grother, S. A. Janet, and C. L. Wilson, "NIST Form-Based Handprint Recognition System," Technical Report NISTIR 5469 and CD-ROM, National Institute of Standards and Technology, July 1994.
2. M. D. Garris and D. L. Dimmick. Evaluating Form Designs for Optical Character Recognition. Technical Report NISTIR 5364, National Institute of Standards and Technology, February 1994.
3. P. J. Grother, "Handprinted Forms and Character Database, *NIST Special Database 19*," Technical Report and CD-ROM, National Institute of Standards and Technology, March 1995.
4. D. L. Dimmick, M. D. Garris, and C. L. Wilson, "Structured Forms Database, *NIST Special Database 2*," Technical Report and CD-ROM, National Institute of Standards and Technology, December 1991.
5. D. L. Dimmick and M. D. Garris, "Structured Forms Database 2, *NIST Special Database 6*," Technical Report and CD-ROM, National Institute of Standards and Technology, September 1992.
6. S. A. Janet, "Census Miniform Training Database 1, *NIST Special Database 11*," Technical Report and CD-ROM, National Institute of Standards and Technology, August 1995.
7. S. A. Janet, "Census Miniform Test Database, *NIST Special Database 13*," Technical Report and CD-ROM, National Institute of Standards and Technology, August 1995.
8. W. Postl, "Method for automatic correction of character skew in the acquisition of a text original in the form of digital scan results," U.S. Patent 4 723 297, Feb. 1988.
9. D. S. Bloomberg and G. Kopec, "Measuring document image skew and orientation," SPIE Conference on *Document Recognition II*, San Jose, Feb. 1995.
10. D. Hearn and M. P. Baker, *Computer Graphics*, pp. 58-61, Prentice-Hall, 1986.
11. Brent, *Numerical Recipes*, W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, pp. 283-286, Cambridge University Press, 1989.
12. M. D. Garris and D. L. Dimmick, "Form Design for High Accuracy Optical Character Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to be published.