

Measuring document image skew and orientation

Dan S. Bloomberg, Gary E. Kopec and Lakshmi Dasari

Xerox Palo Alto Research Center
Palo Alto, CA 94304

Abstract

Several approaches have previously been taken for identifying document image skew. At issue are efficiency, accuracy, and robustness. We work directly with the image, maximizing a function of the number of ON pixels in a scanline. Image rotation is simulated by either vertical shear or accumulation of pixel counts along sloped lines. Pixel sum differences on adjacent scanlines reduce isotropic background noise from non-text regions. To find the skew angle, a succession of values of this function are found. Angles are chosen hierarchically, typically with both a coarse sweep and a fine angular bifurcation. To increase efficiency, measurements are made on subsampled images that have been pre-filtered to maximize sensitivity to image skew. Results are given for a large set of images, including multiple and unaligned text columns, graphics and large area halftones. The measured intrinsic angular error is inversely proportional to the number of sampling points on a scanline.

This method does not indicate when text is upside-down, and it also requires sampling the function at 90 degrees of rotation to measure text skew in landscape mode. However, such text orientation can be determined (as one of four directions) by noting that roman characters in all languages have many more ascenders than descenders, and using morphological operations to identify such pixels. Only a small amount of text is required for accurate statistical determination of orientation, and images without text are identified as such.

Keywords: image skew, image orientation, image analysis, scanned image, image morphology

1 Introduction

The lines of text within images of scanned documents can usually be brought into nearly upright orientation with a rotation of a multiple of 90 degrees. This multiple describes the *orientation* of the image. If the required rotation is 0 or 180 degrees, the image is in *portrait* mode; if 90 or 270 degrees it is in *landscape* mode. After such rotation, the textlines (if they exist) usually have a small amount of *skew*, typically less than ± 5 degrees, with respect to the raster lines. If the skew is small, it is possible to treat the problems of orientation and skew determination separately, and that is the approach taken here.

1.1 Why is skew and orientation determination important?

There are a variety of circumstances in which it is useful to determine the text skew and orientation:

- *Improves text recognition.* Many systems will fail if presented with text oriented sideways or upside-down. Performance of recognition systems also degrades if the skew is more than a few degrees.
- *Simplifies interpretation of page layout.* It is easier to identify textlines and text columns if the image skew is known or the image is deskewed.
- *Improves baseline determination.* The textline baselines can be found more robustly if the skew angle is accurately known[7].
- *Improves visual appearance.* Images can be displayed or printed after rotation to remove skew. Multiple page document images can also be oriented consistently, regardless of scan orientation or the design of duplex scanners.

1.2 What are the criteria for a good system?

Various applications have different requirements for skew and orientation determination. Yet we can set down some performance criteria that will satisfy most applications.

For skew measurement, the operations should be fast, with a computation time relatively independent of image content. They should be accurate, with a consistent error less than 0.1 degree to satisfy all applications. The method should not require segmentation of the image to isolate text lines. The measurement should not be degraded by halftones or most graphics (though one can always construct pathological cases). For images with multiple columns, the measurement should be independent of textline alignment between columns. The skew can be determined either locally or globally. And along with an estimate of skew, the method should produce either a confidence measure or an estimate of probable error.

For orientation determination, the computation time should be independent of image content. The operations should not require image segmentation. The operations should also be sufficiently fast that they can be used routinely before character recognition. The presence of a small amount of skew must not affect the result. The method should have sufficient sensitivity that only a small amount of text is required. A confidence value should be returned to differentiate between results that are expected to be accurate and those where the method fails and another method, such as character recognition, is required to find the orientation.

1.3 Previous work on skew determination

In 1987, Baird[1] suggested using bounding boxes of connected components to estimate image skew. The coordinates of a token point, on the bottom center of bounding box, were selected, and a function S_{tokens} of skew angle was computed from these coordinates. Specifically, the function

$S_{tokens}(\theta)$ is the sum of squares of the number of such points computed along a set of lines with angle θ to the raster direction. Baird simulated a vertical shear on the set of points and performed the sums over points with the same y-coordinate. Aside from a constant (independent of θ), the function S_{tokens} is the *variance* of the number of tokens on a line, as a function of the angle. This variance should be a maximum in the direction where the tokens for each textline tend to fall near the same line.

In 1988, Postl[9] described a method for determining skew that contains the basic method we use. Straight lines of the image are traversed at a set of angles relative to the raster direction, and a function $S_\delta(\theta)$ is computed that has a maximum when the scan direction θ is along the textlines. Unlike Baird, who computes tokens from connected components, Postl uses *every pixel in the image*. The function S_δ is similar to Baird’s function. An angle θ is chosen and pixel sums are found along lines in the image at this angle. Instead of squaring the sum of tokens, Postl squares the difference between sums of ON pixels on adjacent lines, and the function S_δ is found by summing over all lines. It can be seen that Postl’s function $S_\delta(\theta)$ is, aside from a constant, just the variance in the *difference* between pixel sums on adjacent lines at angle θ .

More recently, Chen and Haralick[6] used a more involved method that started with threshold reduction[2], applied recursive morphological closings and openings to close up textlines and remove ascenders and descenders, determined connected components, fit the best line to the points in each set of connected components, and estimated a global skew by discarding outlier lines and averaging the remaining ones. When run on a large database, The skew error reported was greater than 0.3 degrees on about 10 percent of the images in the *UW English Document Image Database*[5].

1.4 Plan of the paper

In the next two sections we describe an image-based approach for attaining the skew measurement functionality specified above, and give results when this method is used on a representative set of document images. Following that, we describe a method for determining text orientation and give its performance on the same image database. Results are summarized at the end.

2 Image-based method for skew measurement

Our approach, like Postl’s, computes pixel sums from the image. However, it differs in several ways that improve the speed, accuracy and robustness[3, 4]. Specifically, we use special subsampling to strengthen the “signal” and reduce later computation. To get pixel sums, the image can either be scanned along sloped lines, or vertically sheared and scanned on horizontal lines. Different search strategies, such as linear sweep and bifurcation, can be used to find the maximum of the skew function(s). These can be used hierarchically at different reduction levels, both for efficiency and to avoid missing sharp peaks on the coarse search and to most accurately measure the primary peak on the fine search. It is also important to provide a figure of merit for every skew measurement that is made.

2.1 Subsampling the full resolution image

The time to perform the computation varies directly with the number of pixels in the image. A reduction by 2x in each direction decreases the time by a factor of 4. Thus, there is a significant incentive to perform the calculation at the greatest possible reduction. As we will see, a reduction in image size, particularly in the width, also increases the error. Image reduction can be performed by subsampling, but results are better if subsampling is preceded by filtering to reduce the randomness in the result by maintaining some of the structure that exists at higher resolution. We have found two different types of filters that give similar results. The first is the threshold reduction operation[2], that is implemented as a cascade of 2x reductions, each with a threshold of 1. Thus, for each 2x reduction, the image is tiled into 2×2 pixel regions and the output pixel is ON if any of the pixels in the corresponding 2×2 cell are ON. The second type of reduction operation begins by tiling the image into $N \times N$ pixel cells, where $N = 2^n$. In each tile, one of the N rows of pixels is chosen arbitrarily. If any pixel in that row is ON, the output pixel is ON; otherwise it is OFF. This is an example of a method of filtering and subsampling, which we call a *textured reduction*, that preserves horizontal structure in the image, and is extremely fast for $N \geq 8$. Variants of this operation can be used to extract different types of texture from a binary image in the process of subsampling. For example, thin horizontal lines can be reliably extracted in a reduced image by sampling pixels along some column of the cell rather than a row.

2.2 The skew functions

Here, we define the skew functions $S(\theta)$ and $S_\delta(\theta)$ introduced above, and illustrate some of their properties. The binary image has been scanned into a set of horizontal *raster lines*, and reduced, typically by 2x, 4x or 8x. The next step is to construct the sum function. We can either shear the image vertically and sum along horizontal lines, or scan the image along sloped lines, summing the number of ON pixels in each *scanline*. We will describe the process using the latter approach, but both methods are equivalent. Let the image dimensions be $w \times h$ pixels. The sum for the i^{th} sloped line at angle θ is denoted

$$s(i, \theta) = \sum_{\substack{\text{pixels in } (i, \theta) \text{ scan} \\ j \in w}} p(i, j) \quad (1)$$

where lines near the top and bottom of the image may not traverse the full width, and $p(i, j)$ is 1 if the pixel is ON and 0 if it is OFF. This line sum is raised to a power that is greater than 1 (typically, 2), and then accumulated for all scans, of which there are approximately h :

$$S(\theta) = \sum_{\substack{\text{scanlines} \\ i \in h}} s^2(i, \theta) \quad (2)$$

Every pixel in the image is sampled, and the result, within a constant independent of the angle θ , is equal to the variance of the number of pixels on scanlines at angle θ . The variance is maximized

when the scans are parallel to the textlines. The left hand side of Fig. 1 shows a typical profile of S vs. θ .

The shape of the peak is typical of many text images, and can be described by a simple model. Suppose the image is idealized as a single column of aligned textlines. Each textline is a rectangle of width w_l and height h_l , with the property that the probability that a pixel within the rectangle is ON is a constant, and all pixels outside the rectangle are OFF. The normalized angles $\bar{\phi}$ are found by dividing the actual skew angle by the scale-independent parameter h_l/w_l , and the sum function \bar{S} is normalized to the value at zero skew, which is proportional to $h_l w_l^2$. It is assumed that $h_l/w_l \ll 1$, and the analysis proceeds in two regions: $\bar{\phi} \leq 1$ and $\bar{\phi} > 1$. In the first, the longest scan through a line is of length w_l , whereas in the second, the longest scan is shorter than w_l . The normalized sum as a function of $\bar{\phi}$ is easily shown to be

$$\bar{S}(\bar{\phi}) = \begin{cases} 1 - \bar{\phi}/3 & \bar{\phi} < 1 \\ 1/\bar{\phi} - 1/(3\bar{\phi}^2) & \bar{\phi} \geq 1 \end{cases} \quad (3)$$

This function is shown on the right hand side of Fig. 1. The break point between linear and tail occurs at an angle $\bar{\phi} = 1$, corresponding to a skew angle equal to h_l/w_l . The sum skew function from real images is flattened at the top, due to the presence of ascenders and descenders which smear the response near zero angle.

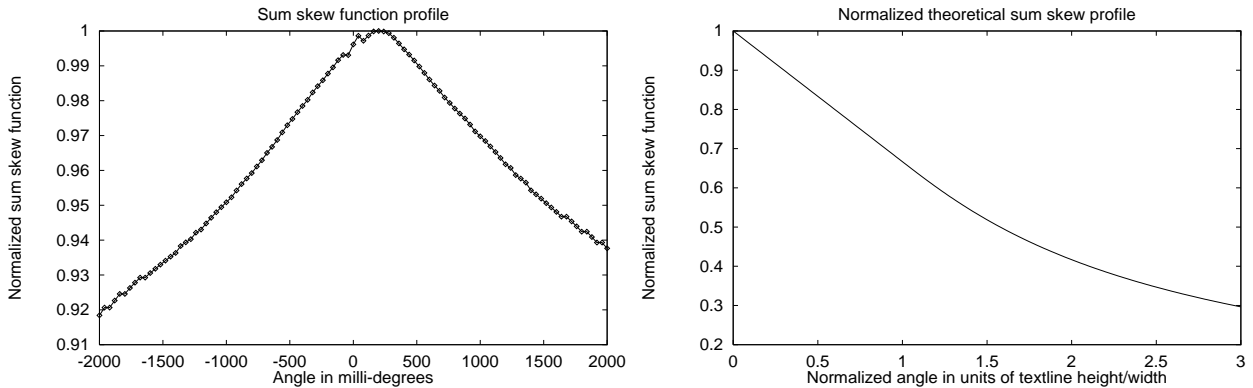


Figure 1: On the left, the measured sum skew profile $S(\theta)$ from a typical scanned image, A03K in the UW database. The angle is measured in milli-degrees. A large amount of isotropic background “noise” is evident from the scale. On the right, a simple approximation to a normalized sum skew function \bar{S} vs. normalized angle $\bar{\theta}$, given in (3). The angle is normalized to the ratio of the textline (height /width).

In addition to this intrinsic width, the sum skew function is subject to bias away from the true textline orientation. The most serious situation is where the text is composed of multiple columns, and where the textlines are not in parallel conjunction between columns. The result depends on specific alignment. With relatively uncorrelated textlines, the sum skew function will be very broad with a poorly defined peak. The largest bias is introduced when the textlines in adjacent

columns are correlated with constant offset. For two such columns, the bias (error in the peak location) is given by the ratio of the offset to the width of one column. Halftone images contribute to an isotropic background “noise” of the sum skew function, but typically have little effect on the location and shape of the peak.

The differential skew function $S_\delta(\theta)$ is typically defined as the sum of squares of differential counts on adjacent lines:

$$S_\delta(\theta) = \sum_{\substack{\text{scanlines} \\ i \in h}} (s(i, \theta) - s(i - 1, \theta))^2 \quad (4)$$

In contrast to the sum skew function $S(\theta)$, the width of the differential function can be much less than the textline asperity ratio. Fig. 2 shows typical differential skew profiles for two images. On the left, the image is a single column of perfectly aligned text. On the right, the scanned image has multiple non-aligned text columns.

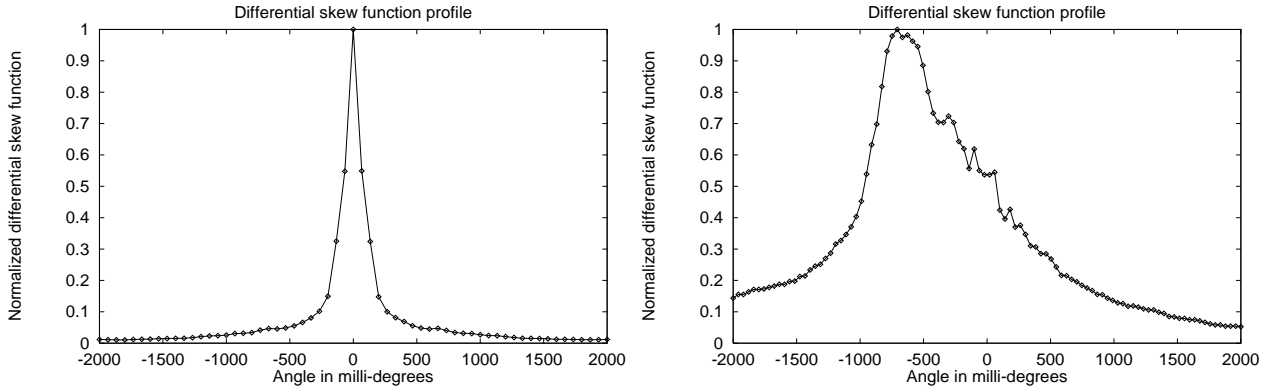


Figure 2: Measured differential skew profiles $S_\delta(\theta)$ from two images. The angle is measured in milli-degrees. The profile on the left is from an unscanned image with a single column of text, LOOK, in the UW database; that on the right is from a somewhat problematic scanned image, A006.

Most of the contribution to S_δ occurs at the baseline, topline and x-height lines of each textline. For perfectly aligned text, the angular width of the differential skew profile is expected to be about two pixels divided by the width of the textline, and the location of the maximum can be determined much more accurately. Signal degradation and bias introduced by various correlations between textlines in multiple columns is typically much smaller than for the sum skew function. For example, if textlines are located in random positions in different columns, the effect is to reduce the signal and broaden the width because textlines in each column effectively contribute separately, but no bias is introduced in the peak position. The worst case for bias is when textlines in adjacent columns are correlated with a constant offset of a few pixels; here, a bias in the peak position of several pixels divided by the column width can be introduced. For larger offsets, the peak at zero angle should be larger than the biased peak. This textline alignment contribution to bias can be

reduced by performing the image analysis on a portion of the image expected to contain mostly one column, such as the left-hand side.

The profile on the right in Fig. 2 is from an image scanned with nonuniform skew: the skew is about twice as large at the top as at the bottom. This accounts for some of the broadening of the peak, including part of the tail to the right of the peak. Broadening further out is due to interactions between textlines in adjacent columns, for which multiple relative alignments are in evidence. Dithered image regions contribute essentially nothing to $S_\delta(\theta)$, and have no effect on the shape or location of the peak.

2.3 Search strategies

We describe two strategies for finding the maximum of a skew function: *sweep* and *bifurcation*. The *sweep* method evaluates the skew function at a sequence of equal intervals between end points that are assumed to encompass the maximum, and a quadratic fit is made to the largest value and its two neighbors in order to estimate the peak location. The *bifurcation* method is best used when the location of the peak is known approximately, e.g., within ± 0.5 degree. The value of the skew function is evaluated at the center and end points of the initial interval. (The value at the center should be the largest.) Then an iterative process is begun where two more measurements are taken angles half way between the central peak and its two neighbors. Of these five values we choose the largest and its two adjacent values, which now delineate the intervals to be bifurcated, and the process repeats. The termination condition is when the difference in peak angles between two successive iterations is sufficiently small. Convergence is rapid because the intervals are halved at each iteration.

The most successful methods are hierarchical, using a coarse sweep in the first stage and a fine bifurcation or sweep in the second. Because the skew function widths can be much less than a degree, it is best to use a higher order of image reduction for the coarse sweep to broaden the peaks. The peak can also be broadened in the coarse sweep by using the sum skew function, but the differential should always be used to guide the fine bifurcation step.

2.4 Degree of confidence

It is important to provide a measure of confidence with any skew measurement. We are not interested in error bars giving the likely range of the exact value. Such a range is typically known a priori for these techniques, as will be shown in the next section. Our concern is that there exist a small fraction of “pathological” images, that by virtue of their intrinsic nature or imperfections in the scan process itself, are able to bias the skew functions, and it is these outliers that we wish to flag. For a hierarchical search, we generate two profiles of skew function vs. angle, and these can be analyzed for irregularities.

Indications of trouble in the coarse sweep profile, when the differential skew function is used, are:

- Large noise-to-signal ratio. Define the noise as the average background, calculated from the profile but not including the peak and its two neighbors. Define the signal as the difference between the peak and the average background. This ratio should be small.
- Large fluctuations in background about the average. Normalize the variance to the mean, again omitting the peak and the two samples adjacent to the peak.
- Large values of the skew function far from the peak. Calculate a power of the skew function, e.g. 3, weighted by the absolute value of the angle from the peak.

All these measures must be constructed with normalization factors that remove dependencies on image size and the number of ON pixels. They can then be combined using weighting factors, chosen empirically, to subtract from an initially “perfect” confidence level.

For the fine bifurcation step, we can calculate profiles for both skew functions with no extra work, and the location of peaks can be compared. (The differential peak is found by the search; the sum peak must be fit from data collected.) If there is a significant difference, this is a warning of bias. The bias from multiple columns is typically larger in the sum function, but other sources of bias, such as text from adjacent pages in a book scan or non-affine distortions introduced by the scanner, may also be present.

3 Experimental results of skew measurement

Performance of the skew functions was measured using 979 scanned images and 168 unscanned synthetic (and perfectly aligned) images in the *UW English Document Image Database*. For each of the scanned images, an estimate of the actual image skew is provided in the database, with an expected error of from 50 to 100 milli-degrees.

We evaluate our skew measurements in two different ways. In the first, we measure the *intrinsic* variance by comparing measurements made on the same image rotated through known angles. Such measurements will probably not detect a systematic bias in the measurements away from the correct skew direction. In the second evaluation, we compare the results directly against the “ground-truth” estimates provided with the UW database. The results are different in the two cases, and it is important to assess the contributing factors.

3.1 Measurement on unscanned images

It is a simple matter to measure the skew of these images. However, the skew functions are so accurate that they typically give the exact result, 0, which is not very useful. The reason is that it is difficult to measure skew near 0 degrees, because one needs to consider lines with a finite slope (or equivalently, images that have nonzero vertical shear). It is important to shear about the left edge of the image, and not the center, because at the center a minimum shear angle of $2/w$ (radians) is required before any pixels are moved, whereas at the edge the minimum shear angle is $1/w$. For an

image of width 2550 pixels, reduced by 2x, this minimum angle is about 50 milli-degrees. In order to measure the performance of the skew functions more accurately than this, for each image in the database we generate 10 rotated images, at angles up to ± 2.0 degrees. On this set of 11 images we measure the skew at reductions of 2x, 4x and 8x, and fit the measurements for each set to the best straight line. (We could also compare with the expected angle; the results are essentially the same). Then for each image we have an rms error and 11 absolute errors. These are plotted separately in Fig. 3, where the y-axis gives the cumulative histograms (as a percentage) of measured skew angles on the 168 unscanned images. The plots give the the measured RMS and absolute skew errors.

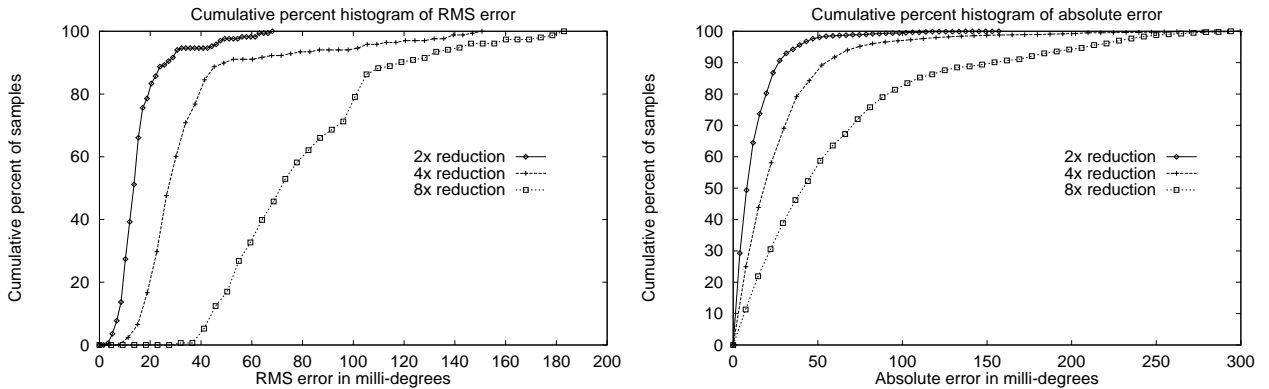


Figure 3: Cumulative percentage histogram of unscanned documents (zero skew) with measured rms and absolute skew errors, as a function of angular error. These are derived from 168 images in the UW image database. The curves are for measurements on images reduced by 2x, 4x and 8x. 40 histogram bins are used in all plots.

We call these the *intrinsic* errors, in that they measure statistical fluctuations between different measurements on the same image, including errors in rotating the image. It should be noted that both the RMS and absolute error scale approximately with the reduction factor. The magnitude of the intrinsic RMS error at the 90th percentile is for these synthetic images is

$$E_{synth}^{int}(90\%) \approx 10 * reduction \quad (\text{milli-degrees}) \quad (5)$$

The image width is 2550 pixels, so 10 milli-degrees of skew corresponds to about 1/2 pixel of vertical shear over the full width of the image. This is at the *resolving power* limit of this method.

3.2 Measurement on scanned images

Assessment of skew measurement performance on scanned images is problematic. The initial problem is that the actual image skew is not known. The approach taken with the UW database was to measure the skew of each image manually. Unfortunately, observation of many of the scanned images indicates that there is no single global skew. The most common scan artifact is

that the skew at the top and bottom of the image differ, often by two hundred milli-degrees or more. There are also a few “outliers” where the measured angle is far from correct. Significant measured bias is occasionally induced by text fragments with non-conforming skew imaged from adjacent book pages.

As a result, we have measured intrinsic errors, as with the unscanned images, and also compared results with the “ground truth” from the UW database. In analogy with Fig. 3, Fig. 4 and Fig. 5 give the the RMS and absolute intrinsic errors, again generated by rotating each image at angles of $-2.0, -1.6, \dots, 2.0$ degrees and fitting the measured skew to the best straight line. The zero crossing of this line is the best estimate of skew of the un-rotated image.

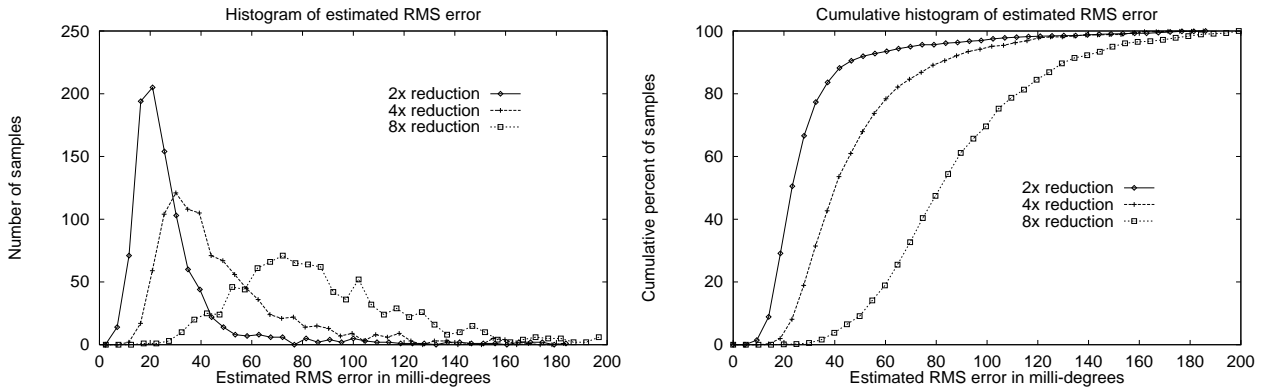


Figure 4: Histogram and cumulative histogram of rms errors, inferred from a linear least squares fit to measurements on each of 979 scanned images in the UW image database. The images were scanned at 300 dpi, and the curves are for reductions of 2x, 4x and 8x. 40 histogram bins are used in all plots.

In Fig. 4, we show the histogram of intrinsic estimated RMS errors on the left, and the cumulated histogram on the right, for the 979 images in the UW database. We have cut off the histograms at 200 milli-degrees, eliminating a very small number of images. In each, the three curves are for 2x, 4x and 8x reduction. As with the unscanned images, the RMS error scales approximately with the reduction factor. The magnitude of the RMS error at the 90th percentile is about

$$E_{scan}^{int}(90\%) \approx 20 * reduction \quad (\text{milli-degrees}) \quad (6)$$

The image width is 2550 pixels, so 20 milli-degrees of skew corresponds to about 1 pixel of vertical shear over the full width of the image. This is about twice the intrinsic error of skew measurements as on the set of un-scanned images.

In Fig. 5, we show the histogram (and cumulative histogram) of intrinsic estimated *absolute* error in skew measurement. The magnitude of the absolute error at the 90th percentile is similar to that of the RMS error.

Fig. 6 gives the histogram and accumulated histogram of absolute differences between the UW “ground truth” and our measured skew values, on a set of 979 scanned images. The three

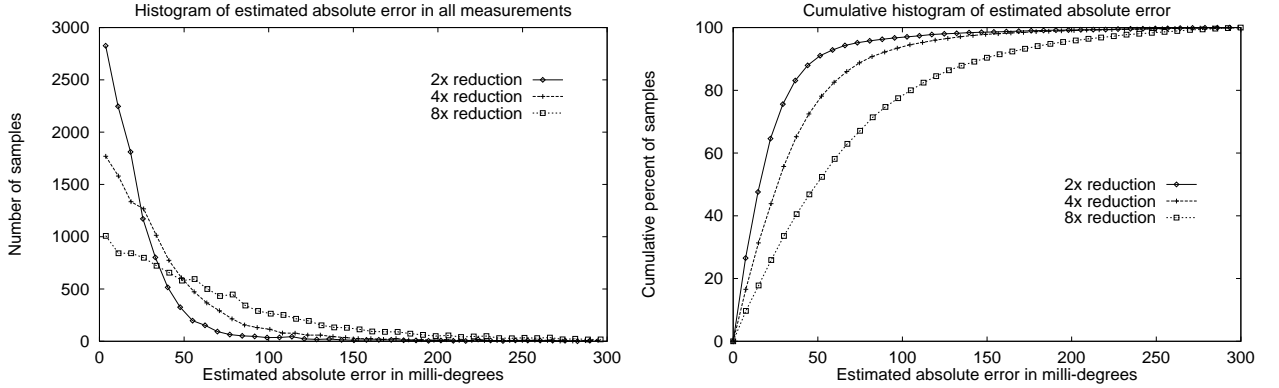


Figure 5: Histogram and cumulative histogram of absolute errors, inferred from a linear least squares fit to 11 measurements on each of 979 scanned images in the UW image database. The images were scanned at 300 dpi, and the curves are for reductions of 2x, 4x and 8x. 40 histogram bins were used in both plots.

curves in each plot are for skew measurements using reductions of 2x, 4x and 8x. Although we believe that the error in our measurement increases with image reduction from 2x to 8x, this is not apparent, particularly for the half of the scanned images with differences under 0.1 degree. This is contrary to expectation. If the errors in both our measurements and the “ground-truth” are normally distributed with variances σ_s^2 and σ_{gt}^2 , respectively, then the absolute value of the difference should be distributed as the convolution of these two gaussians. This is another gaussian, with a variance σ_{diff}^2 given by

$$\sigma_{diff}^2 = \sigma_s^2 + \sigma_{gt}^2 \quad (7)$$

If our measurements have errors given by σ_s that increase with reduction, this should be reflected in an increase in σ_{diff} , which is *not* strongly evident in Fig. 6. One possible explanation is that the width σ_{gt} is considerably larger than σ_s , at least for 2x and 4x reductions. This is an area that needs further exploration. Ultimately, we would like a better estimate of accuracy. At 2x reduction, and for the 90th percentile absolute error, there is a large difference between the intrinsic error of 40 milli-degrees and the value of about 250 milli-degrees in Fig. 6.

On a Sparc10 for a full-page image and using vertical shears, hierarchical skew measurement typically takes about 2.5 sec at 2x, 1.0 sec at 4x, and 0.3 sec at 8x reduction. Most of the time is spent doing vertical shears. The alternative method, summing along sloped lines, can be done significantly faster.

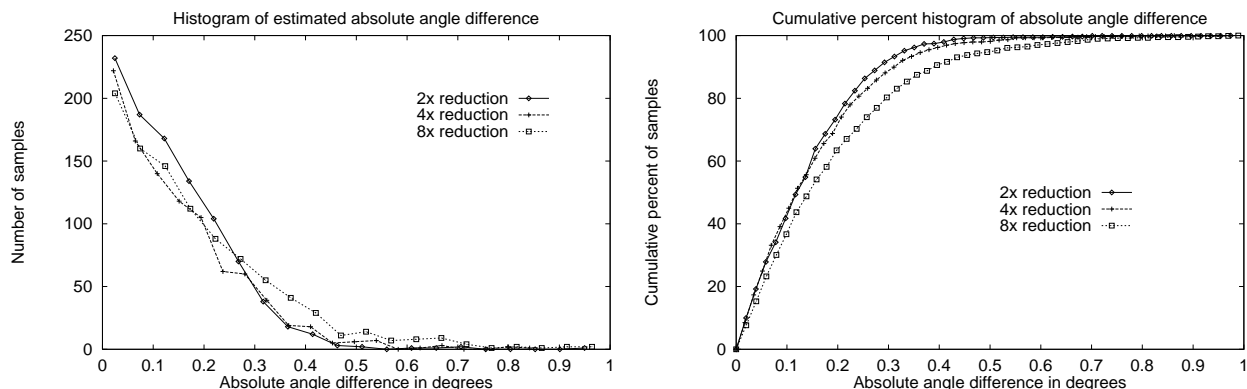


Figure 6: Histogram and cumulative percent histogram of the absolute value of the difference between skew measurements and the UW "ground-truth" skew. In each plot, the full set of skew measurements were taken at 2x, 4x and 8x reduction. 20 bins were used for the histogram, and 50 bins for the cumulative histogram.

4 Image-based method for orientation measurement

We next describe a method for determining the orientation of an image[8] that satisfies the criteria stated at the outset. The method relies on the preponderance of ascenders over descenders in languages with roman characters. Tests reported here are only on English, where the ratio of ascenders to descenders is typically about 3:1.

4.1 Extracting the orientation signal

Morphological operations can be used to identify separately those pixels belonging to ascenders and descenders, and for landscape and portrait modes. Because of the statistics involved, only a small amount of text is necessary to make this determination. The classifier should give a confidence level for any result, and a threshold can be applied, below which the result is "unknown". For example, images without text should always return an "unknown" orientation.

The first step is to simplify the image structure and reduce further computation, while leaving the ascenders and descenders salient. We reduce the image by 4x (using a cascade of 2x threshold reductions with threshold level 1), and then dilate the image with a horizontal structuring element (SE) that is sufficiently large to merge characters within a word and words within a textline at the x-height level. We choose a horizontal SE of width 7, which leaves ascenders and descenders protruding as blocks above and below the merged x-height part. Some ascenders and descenders are joined in the process, but the results are nevertheless satisfactory. The joining can be prevented by opening the result with a sufficiently large horizontal SE to remove the dilated ascenders and descenders, leaving only a solid region spanning the the x-height part, and then *OR*-ing this result with the reduced image before dilation to restore the original ascenders and descenders.

the problem

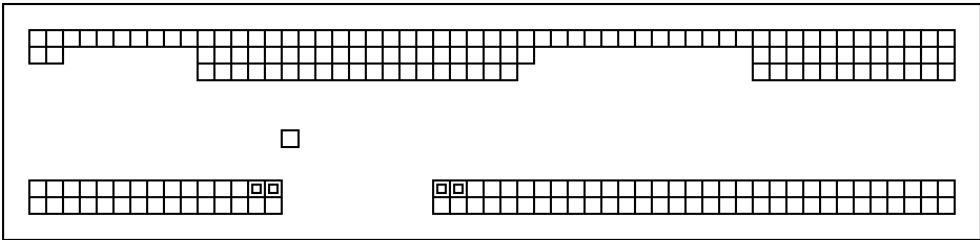
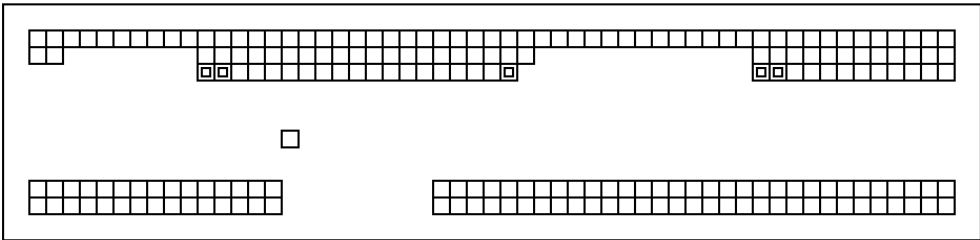
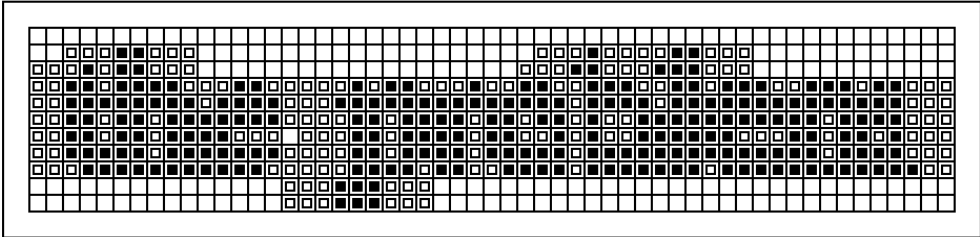


Figure 7: Example showing steps involved with the extraction of pixels associated with ascenders and descenders. An image fragment is shown in the top frame.

Fig. 7 shows the sequence of operations used for counting the ascenders and descenders in a fragment of a text image. The fragment shown has both ascenders and descenders. The top frame shows the image scanned at full 300 dpi resolution, and displayed at about 50 pixels/inch. The second frame shows the fragment after reduction by 4x, displayed at about 12 pixels/inch. Three different types of pixels are shown: dark pixels are ON in the reduced image, gray (frame) pixels are turned on by a horizontal dilation of width 7, and light pixels are those that remain OFF after dilation.

To identify ascenders and descenders, hit-miss transforms are applied using four different SEs. These elements are shown in Fig. 8. The two SEs on the left are for counting pixels on right and left sides of ascenders; the two SEs on the right are for descenders. The solid dot represents a hit, the circle is a miss, and other pixels are “don’t care”. The plus mark within one of the misses is the center of the structuring element, the location in the output of the hit-miss transform relative to the successful match.

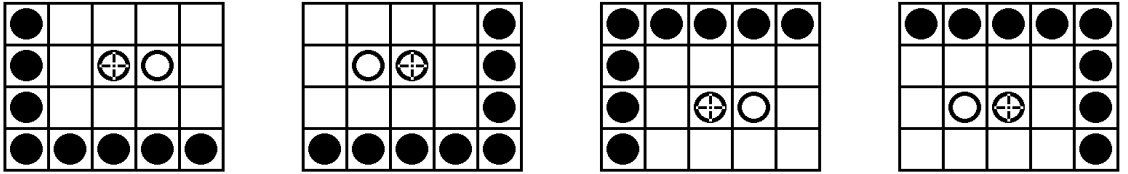


Figure 8: Structuring elements used for measuring orientation. They extract pixels from the right and left of ascenders, and the right and left of descenders, respectively.

The third and fourth frames in Fig. 7 show the results of the hit-miss operations, and are also displayed at a resolution of 12 pixels/inch. The third frame shows the five pixels that are extracted by either of the hit-miss SEs for ascenders, and the fourth frame shows the four extracted descender pixels.

These pixels are counted, and the statistical significance of the difference between up and down pixels is evaluated using the law of large numbers: the expected variance in each of these numbers is proportional to their square root. The probability that the two populations can be distinguished (i.e., that the distributions do not overlap) can be estimated from the square root of the sum of the individual variances. Form

$$\sigma_o = \sqrt{N_{up} + N_{down}}/2 \quad (8)$$

Then we define a *normalized orientation signal* as the difference between the number of ascender and descender pixels, expressed as a multiple of σ_o :

$$\bar{S}_{orient} \equiv |N_{up} - N_{down}| / \sigma_o = 2 |N_{up} - N_{down}| / \sqrt{N_{up} + N_{down}} \quad (9)$$

To determine the orientation signal from landscape orientation, start with the reduced image. Either use vertical dilation and hit-miss transform by SEs rotated by 90 degrees on the reduced

image, or rotate the image by 90 degrees and repeat the operations described here.

4.2 Angular dependence of the orientation signal

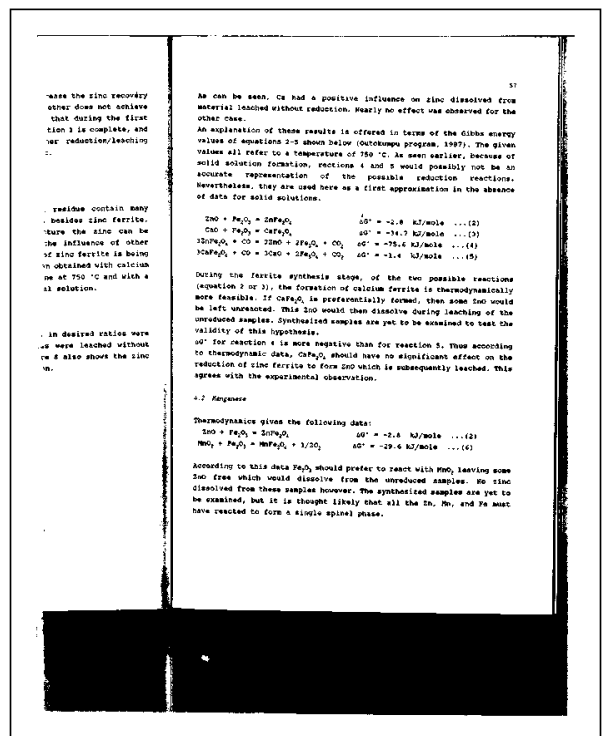
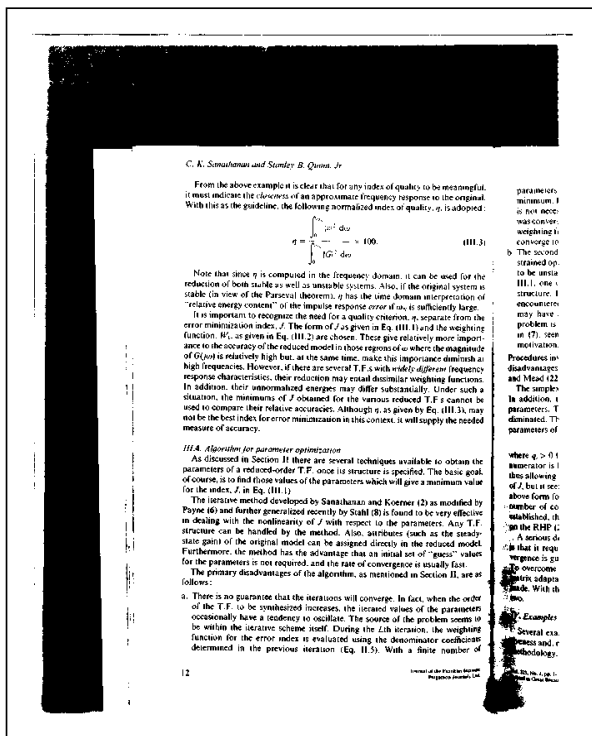


Figure 9: Two images, A03F and A00F, from the UW image database.

The normalized orientation signal \bar{S}_{orient} typically is greater than 4 between ± 20 degrees of the correct orientation. Fig. 9 shows two images, A03F and A004 from the UW database. In Fig. 10, we show the normalized orientation signal from versions of these images that have been rotated between ± 20 degrees.

Both profiles demonstrate reliable determination of image orientation within the range of angles shown. The profile on the left is typical, showing a single broad peak. The double peak on the right is not typical, deriving from a broad *minimum* in the number of ascender counts in the ± 10 degree range. Typically, the ascender counts go through a broad *maximum* there. Both images have essentially flat descender counts within this ± 10 degree range of angles.

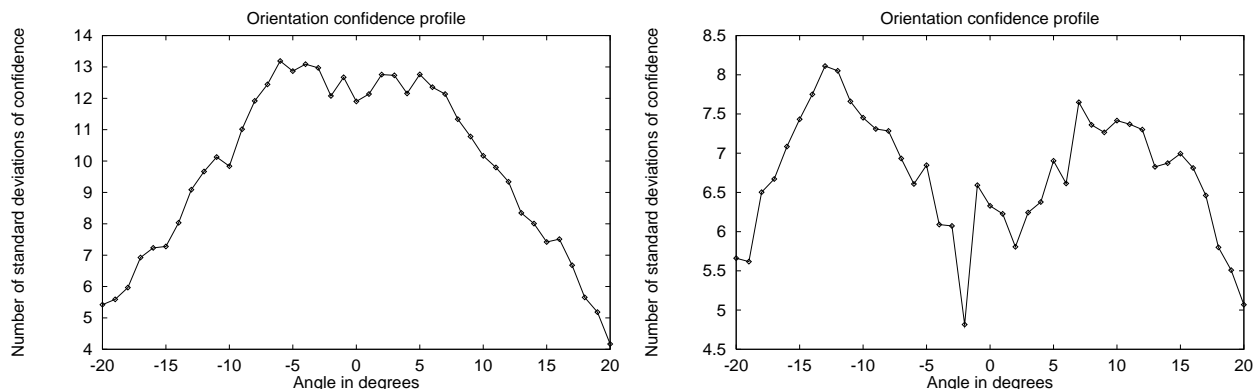


Figure 10: The normalized orientation signal S_{orient} for two images, A03F and A00F, given in units of σ_o , as a function of rotation angle.

4.3 Orientation results on UW image database

We set the threshold on the normalized orientation signal at 3.0. Then on the full UW image database, the following results are obtained:

<i>Orientation</i>	<i>Number found</i>	<i>Number wrong</i>
portrait upright	936	1
portrait upside-down	0	0
landscape up-to-left	2	0
landscape up-to-right	0	0
below threshold	41	-

Six images in the database are in landscape mode. Of these six, the program identified two as landscape, three as uncertain, and one incorrectly as portrait. Orientation measurement on this image, A002, was the only error made on the entire set of 979 images. The image is shown in Fig. 11, and the reason for the failure is evident: the landscape table is fairly weak in ascenders and descenders ($\bar{S}_{orient} = 2.7$), while there is a considerable amount of portrait text from the adjacent page ($\bar{S}_{orient} = 4.6$). Such errors, while expected to be very rare, could be prevented by choosing an interior rectangular region, well away from the edges, to perform the analysis. In about 4 percent of the images, the largest signal for the four orientations was below the chosen limit of 3.0, and most of these were also correctly determined. On a Sparc10, it takes about 0.25 seconds to check for all four orientations on a standard 2550 x 3300 pixel image. This time is independent of image content.

5 Summary

We have presented efficient, accurate and robust methods, of great simplicity, for measuring document image skew and orientation. These methods use as many of the pixels of the scanned image

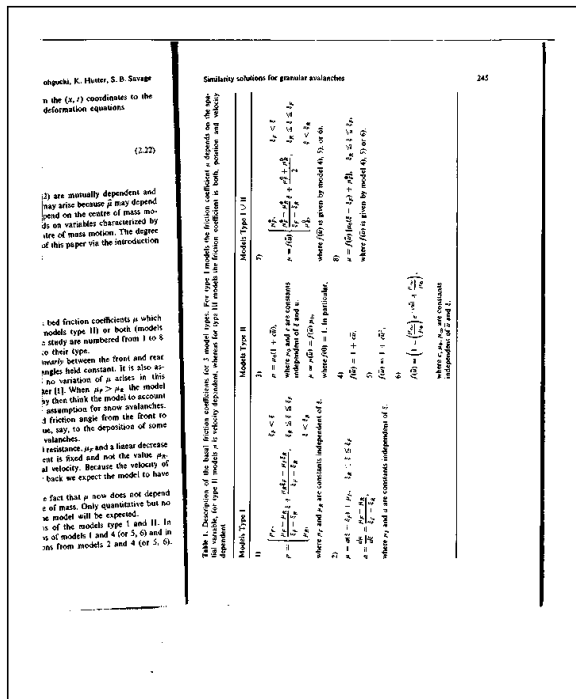


Figure 11: Image A002, which fooled the orientation program into choosing portrait mode. This was the only orientation that was incorrectly measured.

as are required for accuracy. The results are accompanied by confidence measures; in the case of orientation measurement, the result *is* the confidence. On a Sparc10, the orientation measurement takes about 0.25 seconds, and for most applications, sufficient accuracy in skew measurement can be achieved in less than 1 second.

The methods were used on a large database of images to acquire good statistics for evaluation. Problems were noted in evaluating the performance of skew measurement on scanned documents. It is advantageous to simulate real-world scanned images, but these come with a significant uncertainty in the skew angle and various scanner distortions. We distinguish between *intrinsic* measurement errors, that derive from statistical fluctuations and uncertainties in measuring angles, and actual errors that include skew function bias from the “true” skew. For many images, however, there is no global “true” skew, particularly at the level of uncertainty represented by the intrinsic error.

To keep things in perspective, we note that most problems are in measurement of very small skew errors, often under 100 milli-degrees, and document skews of this size are not visually apparent.

References

- [1] H. S. Baird, “The skew angle of printed documents,” *Proc. SPIE Symp. on Hybrid Imaging Systems*, Rochester, NY, 1987, pp. 21-24.
- [2] D. S. Bloomberg, “Image analysis using threshold reduction,” *SPIE Conf. on Image Algebra and Morphological Image Processing II, Vol. 1568*, San Diego, CA, July 1991, pp. 38-52.
- [3] D. S. Bloomberg and G. Kopec, “Method and apparatus for identification and correction of document skew,” *U.S. Pat. 5,187,753*, Feb. 16, 1993.
- [4] D. S. Bloomberg and G. Kopec, “Method and apparatus for identification of document skew,” *U.S. Pat. 5,355,420*, Oct. 11, 1994.
- [5] S. Chen, M. Y. Jaismha, J. Ha, I. T. Phillips and R. M. Haralick, “UW English Document Image Database – (I) Manual,” *Reference Manual*, 1993
- [6] S. Chen and R. M. Haralick, “An automatic algorithm for text skew estimation in document images using recursive morphological transforms,” *ICIP-94*, Austin, TX, November 1994, pp. 139-143.
- [7] F. Chen, D. S. Bloomberg and L. Wilcox, “Spotting phrases in lines of imaged text”, *SPIE Conf. on Document Recognition II, Vol. 2422*, San Jose, CA, Feb. 1995.
- [8] L. Dasari and D. S. Bloomberg, “Rapid detection of page orientation,” *U.S. Pat. 5,276,742*, Jan. 4, 1994.

- [9] 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. Pat.* 4,723,297, Feb. 2, 1988.