DOCUMENT IMAGE SKEW DETECTION: SURVEY AND ANNOTATED BIBLIOGRAPHY

JONATHAN J. HULL
Ricoh California Research Center, 2882 Sand Hill Road, Suite 115, Menlo Park, CA 94025, hull@crc.ricoh.com

Algorithms that estimate the angle at which a document image is rotated (called a document's skew) are surveyed. Four broad classes of technique are identified. These include methods that calculate skew from a horizontal projection profile, a distribution of feature locations, a Hough transform, or the distribution of responses from local, directionally sensitive masks. The basic method used by each class of technique is presented and the contributions of individual algorithms within each class are discussed.

1 Introduction

The digital image of a document may be rotated or skewed at an arbitrary angle because of how it was placed on the platen when it was scanned or because of a document feeder malfunction. This results in a skewed image such as that shown in Fig. 1 (b). This represents a skew of only 5 degrees. In fact, a skew of as little as 0.1 degrees may be apparent to a human observer. Thus, a desirable function in a digital photocopier is the automatic detection and correction of skew. Ideally, an input such as that shown in Fig. 1 (b) would produce Fig. 1(a) as output.

A skew detection algorithm is given the digital image of a document and it determines the angle (possibly zero degrees) by which it was skewed when it was digitized. It is assumed that there also exists a method for rotating the image to remove the skew.

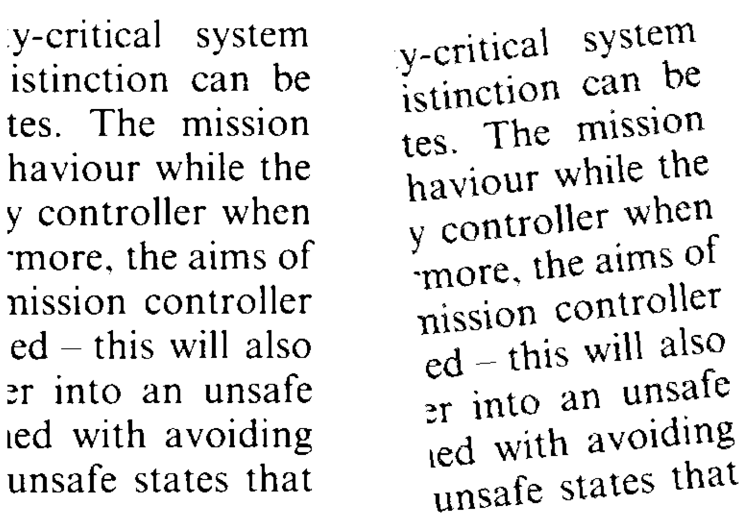
A simple solution for skew detection is to determine the location of at least two corners of the original document [51] and compute the skew angle from them. However, this can be error-prone because of non-linear distortions that occur when pages are not flat on the platen.[16] Also, the entire scan surface may be obscured by the input document or the input may itself have been produced from a skewed original. In either case, deriving the skew angle from the corners or edges of the page is problematic.

This paper surveys skew detection algorithms that derive the skew of a document image. In almost every case, the algorithms that are discussed assume that an input

Appeared in *Document Analysis Systems II*, J..J.Hull, S.L.Taylor, Eds., World Scientific, pp. 40-64, 1998.

document contains some amount of text. Features are often extracted from the text portion of the image that allow the skew to be calculated. This is done because the text is usually structured into lines that are co-linear and aligned with the horizontal (or vertical) axis of the page. Thus, detecting the skew of the text lines provides the skew of the document. Characteristics of the surveyed algorithms are discussed that are pertinent to their commercial implementation. These characteristics include their run-time and algorithm architecture.

The rest of this paper considers several classes of skew detection algorithm. Techniques described in the scientific and patent literature often use either projection profiles on the image data directly, analysis of the geometric distribution of connected components, or a Hough transform. A less thoroughly investigated approach, also considered here, utilizes measurements on the distribution of local features extracted from the image. The general characteristics of each class of technique are discussed and examples of how they are applied to skew detection are presented. The reader is referred elsewhere for an experimental evaluation of skew estimation algorithms.[5]


y-critical system
istinction can be
tes. The mission
haviour while the
y controller when
more, the aims of
nission controller
ed – this will also
er into an unsafe
ed with avoiding
unsafe states that

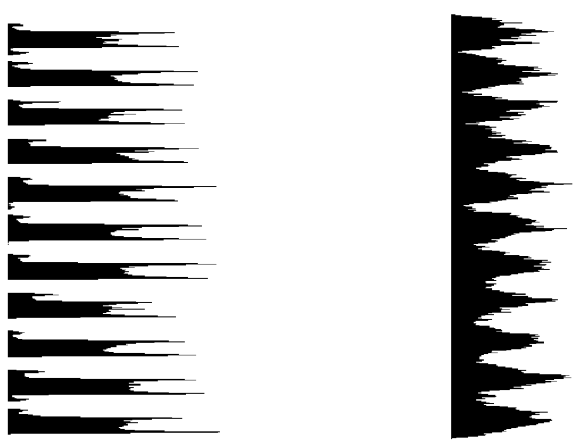
(a) (b)

Figure 1: Original image in correct alignment (a) and skewed by 5 degrees (b).

2 Projection Profile Analysis

A straightforward solution to determining the skew angle of a document image uses a horizontal projection profile. This is a one-dimensional array with a number of locations equal to the number of rows in an image. Each location in the projection profile stores a count of the number of black pixels in the corresponding row of the image. This histogram has the maximum amplitude and frequency when the text in the image is skewed at zero degrees since the number of co-linear black pixels is maximized in this condition. This is illustrated in Fig. 2 which shows the projection profiles calculated from the images in Fig. 1. The peaks in the profile calculated from the unskewed image (Fig. 2 (a)) are taller and more closely spaced than those computed from the skewed image (Fig. 2 (b)).

This characteristic has been used in several algorithms. One way of doing this is to rotate the input image through a range of angles and calculate the projection profile for each one. Then features extracted from each projection profile are compared to determine which one is more peaked. A slightly modified version of this approach is expressed in the pseudo code shown in Fig. 3.



(a) (b)

Figure 2: Horizontal projection profiles calculated from the images in Fig. 1

```
    for theta = -angle to +angle by resolution do {
      for r = 0 to NROWS do
        for c = 0 to NCOLS do
          if (image[r][c] == BLACK)  {
            rotate(r,c,theta,&new_row);
            ++proj_prof[new_row];
          }

      angle_measure[theta] = criterion_function(proj_prof[]);
    }

skew_angle = choose_skew(angle_measure[]);
```

Figure 3: Projection profile analysis algorithm for skew detection.

A complete pass is made over the input image for each possible skew angle. The coordinates for every black pixel are rotated by that angle and the corresponding cell in the projection profile is incremented. Then a *criterion function* is calculated that provides a numerical expression for the peaked characteristic shown in Fig. 2. The choose_skew function is then called to determine the angle that maximized the criterion function. Rotating the coordinates instead of rotating the bitmap saves temporary storage and reduces the computation that would be required for a brute force implementation of this approach.

Postl proposed a method similar to that described above in which the horizontal projection profile is calculated at a number different angles.[42, 43] The directional criterion he used was the sum of squared differences in adjacent cells of the projection profile.

This basic approach was improved by downsampling the image before calculating the projection profile (to reduce computational cost) and using the angle that maximized the variance of the number of black pixels in a scan line as an estimate a document's skew.[11, 12] A novel method for downsampling was used that preserved horizontal structure in the image.

An improvement used a search algorithm that first calculated projection profiles over a sequence of angles that have a coarse resolution.[13] The angle that maximizes the criterion function is used as the center for a finer resolution search for the skew angle. In between the angles that are immediately less than and greater than the coarse maximum, two additional projection profiles and criterion function values are calculated. This recursive subdivision is repeated until the final decision is reached.

Various adaptations of the basic projection profile analysis technique have been proposed that deal with complex input formats without requiring an independent document segmentation method. An example of such an approach includes a method that calculates projection profiles over a number of circular windows. [29, 30] A skew estimation value is accepted only from a window that produces a distribution of criterion function values that is characteristic of a text image. This is done because a reliable measurement of the skew angle of a document are best calculated from a text region.

Further modifications have been proposed that could help solve the problem posed by multiple columns of text. Since lines of text in adjacent columns are often not co-linear, a simple horizontal projection profile taken across the whole page might never exhibit the amplitude and frequency characteristics shown in Fig. 2(a) even if the projection is computed at precisely the correct skew angle.

In one technique a document image is divided into a number of vertical ‘‘swaths’’ and a projection profile is calculated within each one. [4] The rotation of the document is simulated by shifting adjacent projection profiles by an appropriate amount. The sum of the dot product between adjacent projection profiles is then used as the criterion function and the skew is the angle that maximizes the criterion function. A similar approach was used in another method that calculated the cross correlation between vertical scan lines separated by a fixed distance and summed the result over all the columns in the image.[52] Each rotation was simulated by shifting one of the vertical scan lines before calculating the dot product. This perfornamce of a similar technique was improved by calculating the cross correlation over a number of small randomly selected regions.[15] A Monte Carlo sampling technique was used to reduce the effect of noise in the estimation caused by graphics regions. This is needed since the technique is reliable only on text regions.

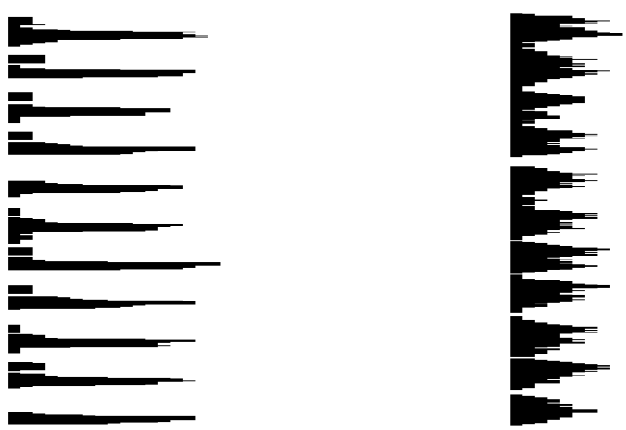
An optical method has also been proposed that essentially calculates horizontal projection profiles over a range of angles. [17] This method uses a rectangular sensor composed of a number of parallel strips of light sensitive material that generate electrical signals proportional to the amount of light reflected on them. The sensor is mechanically rotated and a criterion function is computed that measures the sum of squared differences in the reflected light measured in adjacent rows. The skew of the document is the angle of rotation for the sensor that maximizes this electrically computed criterion.

3 Feature Point Distribution

Another class of technique for document image skew detection reduces the number of memory accesses that are performed in a histogram approach by first extracting the x-y coordinates of features in an image. All subsequent computations are performed on those coordinates. The image itself is not accessed after the feature extraction is completed.

The number of memory accesses and amount of computation saved by feature extraction depends on the ratio of image pixels to number of features and the algorithm that determines skew from the feature point distribution. One method of feature extraction that has been used in several techniques is based on first locating the connected components in the image. These are the groups of connected black pixels that usually correspond to single characters. The x-y coordinates of one representative point for each component (e.g., the centroid) can be used to determine the skew of a document by analyzing their projection profiles at different angles.

The data reduction and retention of information provided by analyzing connected components is illustrated in Fig. 4 where projection profiles of the centroids of the components in Fig. 1 are shown. It can be seen that the distributions are visually similar to those in Fig. 2 that were obtained directly from the image data.



(a) (b)

Figure 4: Projection profiles of connected component centroids calculated from the images in Fig. 1 (a) and (b), respectively.

A pseudo code version of a feature point analysis technique that uses connected component locations is shown in Fig. 5. First all the connected components are located and the x,y coordinates of their centers of gravity (cog) are determined by `find_conn_comp_xy`. This can be done with one pass over the image. Then a projection profile of the rotated y values for the cogs is constructed for a given angle theta. This is repeated for a range of theta values and the one that best describes the skew of the document is determined by analyzing the distribution of criterion function values. Almost the same criterion functions and methods for choosing the criterion value that describes the skew of the document can be used as when the projection profiles were extracted directly from the image data.

```
find_conn_comp_xy( image[], NROWS, NCOLS, &comp[], &ncomps );

for theta = -angle to +angle by resolution do
  for i = 0 to ncomps do {
    rotate(comp[i].center_y, comp[i].center_x, theta, &new_row);
    ++proj_prof[new_row];
  }

  angle_measure[theta] = criterion_function(proj_prof[]);

skew_angle = choose_skew(angle_measure[]);
```

Figure 5: Feature point distribution algorithm for skew detection.

Baird proposed a method similar to that discussed above in which the feature points are the bottom-centers of each connected component. [6, 7] The sum of squares of the values in the projection profile are used as the criterion function and a novel iterative sampling procedure is used to choose the skew angle. The projection profiles are first calculated with a resolution of 0.7 degrees. The peak of the criterion function is then approximated by a least squares procedure. Samples are then taken at a finer resolution around that peak and the location of the peak is re-estimated. This process is continued until the desired resolution of 0.3 degrees is reached.

A similar procedure added the width of each connected component's bounding box to projection profiles taken at several angles.[36] The authors evaluated the performance of four criterion functions and concluded that the simple solution of counting the zeroes in the projection profile provides the best performance (0.1 to 0.2 degree accuracy).

A similar method rotated the x-y values for the center of gravity of each character through several angles.[14] A projection profile is calculated at each angle and the number of entries that are greater than zero is used as the criterion function. The skew angle is chosen as the one that minimizes the number of non-zero entries.

A novel method for detecting feature points was applied to document images compressed in CCITT group 4 (G4) format.[48, 49] The x-y locations of white pass codes in the G4 file are located and projection profiles are calculated from them. White pass codes are often generated in the G4 data when a black run in line N corresponds to no black run in line N+1. The x-y values are rotated through a range of angles and the one that approximates the skew of the document is the one that maximizes the sum of squares of the values in the projection profile. An interesting aspect of this approach is its approximation of connected component location by the location of pass codes. Thus , instead of making a pass over the full resolution image data to locate connected components, a computationally much simpler pass over the compressed image is performed. The speedup is at least proportional to the compression provided by the G4 coding. The author comments that in certain cases the speedup may be significant enough to warrant the compression of an uncompressed data stream even if the compressed data is only used for skew detection.

Another technique has been proposed that is applicable directly to a run-length compressed format.[24]

A number of techniques have been proposed that take advantage of the internal structure of a document and substitute computation in the connected component domain for the repeated rotation and projection of x-y coordinates.

One such technique calculated the nearest neighbor for each component and computed the histogram of the direction of the line segments that connect them.[27] This method capitalizes on the fact that small connected components that represent individual characters are often clustered together along lines of text.

A similar method partitioned connected components (called *segmented blocks (SBs)*) into lines of text by fixed thresholds on the x-start and y-start locations of adjacent SBs.[38] The skew of each line was estimated by a least squares fit to the bottom-left of each SB. The skew of a page was estimated by averaging the skew of each line. A 0.1 degree accuracy was obtained on six test images.

A hierarchical method for skew detection calculated a histogram of the angles between each connected component and its nearest neighbor and used this measure to obtain a gross estimate of a document's skew (with a 0.5 degree accuracy).[37] After the document is segmented into lines of text, the skew of the individual lines is estimated by a least squares fit to the centroids of the individual components. A finer resolution estimate for a document's skew is obtained by applying a criterion function to the histogram of individual line orientations.

A similar hierarchical approach downsampled an image by a factor of eight in both the horizontal and vertical directions before locating connected components.[9,10,18,21] After segmentation into lines of text, the skew of each line is estimated by performing a least *mean* squares fit to the x-y coordinates of the upper left corner of each component in a line. The most frequently occurring angle among all the lines of text in a document is used as the estimate for its skew. A related method has been used to correct for skew within individual text blocks.[54] Another similar technique groups connected components into lines and performs a least *median* squares fit to the components in each line.[47] The skew angle of the document is computed as the median of the gradient of the fit lines.

Another approach preprocessed an image with morphological operations that remove ascenders and descenders and merge adjacent characters.[16] The effect is to smear the image horizontally and produce one connected component for each line of text. A line is then fit to the pixels in each component using a least squares technique. A histogram is constructed of the angles of the lines detected in a document and a search procedure is applied to the histogram to determine the skew of the document. This is necessary since there may be significant disagreement among the angles of the fitted lines. This algorithm places a significant emphasis on morphological operations. While the fitting of lines to pixels in the image domain may be more costly than performing this operation on coordinates extracted from each component, this may be offset if special purpose hardware is used that makes the morphological operations faster than locating connected components.

Another approach to skew estimation first locates columns of text in an input document.[39, 40, 41] The skew of a column is determined by a least squares fit to the centers of patches that span the column horizontally. The skew of a page is determined by averaging the skews of its columns. This technique reduces the computation required to determine skew by calculating the angle of a column as a side effect of its segmentation.

A feature point analysis technique which might work well on images that contain long runs of black pixels (e.g., forms) detected the difference between the minimum x-position of the longest runs on adjacent rows.[8] The average difference was used to calculate the skew angle.

One method combined aspects of the projection profile and feature point analysis methods.[46] Text line baselines were located in projection profiles extracted from 20-pixel wide vertical swaths in a document. A least squares procedure was used to fit a line to such baseline points. This provided an estimate for the skew of the document.

4 Hough Transform Analysis

The Hough transform is a well known technique in computer vision that has been used to detect lines and curves in digital images. It computes the values for the parameters of all the curves of a particular type (e.g., straight lines) that can pass through each black pixel. Votes are then cast for each such curve in a multi-dimensional accumulator array. Each dimension of the accumulator array corresponds to one of the parameters. After the entire image has been processed, the accumulator array is inspected for local maxima. Each such maximum indicates the existence of a curve in the original image given by the corresponding parameter values on the axes.

The ρ–θ parameterization for a line provides a practical application of the transform.[20] It is expressed as:

$$\rho = x \cos \theta + y \sin \theta, \quad (1)$$

where ρ is the distance along the normal from the origin to the line and θ is the angle that the normal to the line makes with the x axis.

In practice, an accumulator array is parameterized by a range of values for ρ and θ. The *x* and *y* values of each black pixel are applied in the above equation over a range of values for θ. The value for ρ that results from Eq.1 is used as an index into an accumulator array, along with the associated value for θ and the contents of the accumulator array at that location are incremented.

The Hough transform has been applied to document image skew detection.[50] This approach used the fact that the highest number of co-linear pixels are on lines that are co-incident with the baseline of the text. This is similar to the characteristic exploited by the projection profile methods. The Hough transform computes very similar information.

A pseudo code expression for the Hough transform method for detecting document image skew is shown in Fig 6.

```
for y = 1 to NROWS do
  for x = 1 to NCOLS do
    if (image[y][x] == BLACK) then
      for theta = -angle to +angle by resolution do {
        rho = x * cos(theta) + y * sin(theta)
        ++accumulator[rho][theta];
      }

    skew_angle = choose_skew(accumulator[]);
```

Figure 6: Hough transform for detecting document skew.

The `choose_skew` function is applied to the accumulator array. A heuristic was proposed that measured the rate of change in the accumulator values for each angle θ.[50] The skew angle was set to the value of θ that maximized this heuristic.

The run time of a similar technique was improved by sampling every twentieth pixel horizontally and vertically transforming them as shown in Fig. 6.[22] The skew angle was chosen by summing over ρ for each value of θ and choosing the angle that maximized the square of the result.

Another Hough technique first applied preprocessing that segmented text into blocks that correspond to paragraphs.[25] The Hough transform was calculated from the edges of the detected blocks.

Preprocessing has also been used to select pixels at the bottoms of connected components and apply the Hough transform to them.[33, 34] This technique was applied to a sub-region of a document that determined to contain mostly text. As such, individual connected components correspond to single characters and the bottoms of the components are aligned on a common baseline parallel to the skew angle of the document.

Other innovative preprocessing has been applied to reduce the computation even further.[28] In this approach the image was down-sampled from 300 pixels per inch (ppi) to 75 ppi and compressed with a run length representation. That is, the termination of each horizontal run of contiguous black pixels was replaced by a count of the number of pixels in the run. The run length counts were used to update the accumulator array in the usual way. The Hough transform was computed for values of θ between -15 and +15 degrees with a resolution of 0.5 degrees since these values bound the skew of most document images. The skew angle was chosen as the value for θ that corresponds to the cell in the accumulator array with the maximum value.

Another preprocessing method was used to reduce the computation of the Hough transform by constructing a piecewise linear approximation (PLA) to the contours of a binary image.[35] Each segment in the PLA was then transformed using the coordinates of its center point. The length of the segment was added to the accumulator array location specified by the distance of the center point from the origin (ρ) and the angle of normal to the segment (θ). That is, each segment results in only one update of the accumulator array. In this way most of the computation is shifted over to the calculation of the PLA. This technique is best suited for application to forms that contain large numbers of horizontal lines. However, it is similar in spirit to some of the feature point analysis methods discussed earlier that fit line segments to lines of text and use a voting procedure on the angles of all the fit lines to determine the skew of a document.

Hierarchical approaches have also been proposed to reduce the computation of Hough transform implementations for skew detection.[32] In one technique a page image was first divided into 20x30 pixel rectangular blocks.[23] If the percentage of black pixels in a block indicated that it might contain text, the Hough transform was calculated from it for θ between -5 and +5 degrees with a resolution of one degree. The angle with the maximum sum over all values of ρ was used as the center for a more detailed analysis in a second pass. The Hough transform was again calculated over a range of -1 to +1 degrees with a resolution of 0.1 degrees around the angle found in the first pass. The skew angle was the value for θ that maximized the sum of its ρ values. Another hierarchical Hough technique was applied to centroids of connected components.[53] This method was demonstrated on a variety of document types including technical articles, postal labels, handwritten text, forms, drawings, and bar codes.

5 Orientation-Sensitive Feature Analysis

Another type of approach to determining the skew of a document detects the presence of local orientation-sensitive features in an image and uses their angles to vote for the skew of the document. Such an approach combines aspects of the feature extraction techniques discussed earlier with the principle behind some of the preprocessing used to reduce computation in the Hough transform methods. That is, by extracting the essential information that represents skew using local characteristics of the image, the processing applied to the extracted information is simplified. In this case, the repeated rotation and projection operations characteristic of a feature point distribution analysis operator are eliminated.

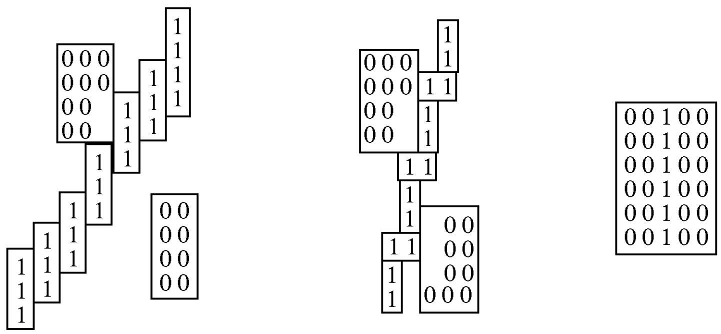
Pseudo code that shows the basic operation of this method is given in Fig. 7. Each entry in a set of masks is applied to the image. If the mask ‘‘matches’’ at that location, a vote is recorded for the angle represented by the mask. After the entire image has been processed in this way, the skew of the document is determined by taking the maximum over the accumulated scores for each angle.

```
for r = 1 to NROWS do
   for c = 1 to NCOLS do
       for each mask in MASKSET do
         if Match ( image[],r,c,mask ) then
           ++angle_measure[ angle( mask ) ];

skew = choose_skew ( angle_measure[] );
```

Figure 7: Algorithm for skew detection with local orientation-sensitive masks.

The basic principle in this type of technique was used in a system that determined the dominant angle at which handwritten text was printed.[45] The three masks shown in Fig. 8 are shifted over all the pixels in an image and the number of times each mask matches is determined. At least one of the ‘1’s in each column must match a black pixel and all of the ‘0’s must match white pixels for the masks to “match.” The masks in Fig. 8 (a), (b), and (c) were designed to match text slanted at angles between 50 and 65 degrees, 65 and 75 degrees, and 90 degrees, respectively.



(a) (b) (c)

Figure 8: Orientation-sensitive masks used to match handwritten text slanted at 50° to 65° (a), 65° to 75° (b), and 90° (from [45]).

The mask that matched the image most frequently provided the skew of the document.

Local orientation-sensitive masks have also been used to detect the general orientation of a page of English language text.[19] In this technique, the image was first smeared horizontally with a 1x5 morphological dilation. Then an erosion was performed with a set of masks designed to detect ascenders and the number of on pixels are counted. The same operation was performed with a set of masks designed to detect descenders, The number of ascender pixels was compared to the number of descender pixels to determine whether the page was right side up, upside down, or aligned at 90°.

A pair of ascender masks are shown in Fig. 9. These masks are must be applied in parallel. That is, both masks in each pair are used at the same time.

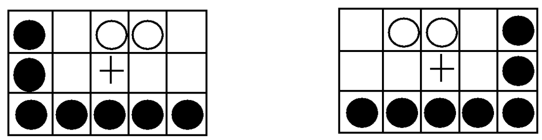


Figure 9: Masks used in morphological erosions to detect ascenders (a) and descenders (b).[19]

6 Discussion and Conclusions

Over 50 published works on at least 25 different methods for document image skew detection were surveyed. These included approaches that calculated projection profiles at different angles directly from image data, methods that calculated projection profiles from image features, and algorithms that used the Hough transform. Another class of technique extracted features with local, directionally sensitive masks.

The performance of most methods reported in the literature range up to a 0.1 degree accuracy. While it is arguable whether this fine a resolution is needed in a digital copier application, at least a resolution of 0.2 to 0.3 degrees should be achieved.

Only preliminary efforts have been conducted in comparative performance evaluation.[5] Further work in this area could help show the strengths and weaknesses of individual algorithms.

References

1. H. K. Aghajan, B. H. Khalaj and T. Kailath, "Estimation of skew angle in text image analysis by slide-subspace-based line detection," Machine Vision and Applications 7, 4 (1994), 267-276.

   A signal processing solution for skew estimation from document images. Document skew detection is mapped onto a parameter estimation problem in sensor array processing. The authors point out that their algorithm (known as *SLIDE*) "introduces a perfect mathematical analogy between the problem of estimating angles of multiple lines in an image and the problem of determining the directions of arrival of planar electromagnetic waves impinging on an antenna array." The computational and storage advantages of their technique over Hough transform methods are pointed out.

2. H. K. Aghajan, B. J. Khalaj and T. Kailath, "Estimation of skew angle in text image analysis by sensor array processing techniques," SPIE Conference on Character Recognition Technologies, vol. 1906, San Jose, CA, February 1-2, 1993, 49-60.

   Conference version of SLIDE[1] technique.

3. H.K. Aghajan and T. Kailath, "Estimation of skew angle in text image," Stanford University, U.S. Patent 5,583,956, December 10, 1996.

   Patent for application of SLIDE[1] technique to skew detection.

4. T. Akiyama and N. Hagita, "Automated entry system for printed document s," Pattern Recognition 23, 11 (1990), 1141-1154.

   A document image that contains vertically printed text is first divided into "swaths" of fixed height. A vertical projection profile is calculated within each swath and the angle of phase shift between projection profiles in adjacent swaths is estimated and used as the skew angle of the document.

5. A. D. Bagdanov and J. Kanai, "Evaluation of document image skew estimation techniques," SPIE Conference on Document Recognition, v. 2660, San Jose, CA, January 29-30, 1996, 343-353.

   Four skew detection algorithms were implemented and tested on 460 page images randomly selected from about 100,000 available images. Algorithms designed by Baird,[6] Postl,[42] and two techniques proposed by Nakano et al.[36] were used. Preliminary results in the development of a general purpose methodology for performance evaluation are presented.

6. H. S. Baird, “The skew angle of printed documents,” Proceedings of the Society of Photographic Scientists and Engineers, Rochester, New York, 1987, 14-21.

Each connected component is approximated by its bottom-center point. Horizontal projection profiles of these points are constructed at a number of different angles and the angle that maximizes the sum of squares in the projection profile is chosen as the document skew. Since a brute-force implementation of this approach would be computationally prohibitive to achieve the desired accuracy (2 minutes of arc), the author presents a clever approach that uses successive approximations at finer resolutions.

7. H. S. Baird, “Apparatus and method for skew control of document images,” AT&T Bell Laboratories, U.S. Patent 5,001,766, March 19, 1991.

Patent for the Baird skew detection method[6]. C source code is included.

8. L. L. Barski, “Skew detector for digital image processing system,” Eastman Kodak Company, U.S. Patent 4,866,784, September 12, 1989.

After a horizontal smearing step, the start position (x-location) of black runs on adjacent rows is determined. Then the offset distance between start positions of the longest runs on adjacent rows is accumulated. An offset distance is added into an accumulator only if it is below a threshold. The average offset distance (or equivalently the average offset angle) is used as the skew angle of the document. The standard deviation of the estimate is compared to a threshold to improve reliability.

9. G. Bessho, K. Ejiri and J. F. Cullen, “Fast and accurate skew detection algorithm for a text document or a document with straight lines,” Proceedings of the SPIE - The International Society for Optical Engineering, Conference on Document Recognition 2181 (February 9-10, 1994), 133-140.

An image is first down-sampled by a factor of eight in both the horizontal and vertical directions. A line is then fit to the upper left corners of each connected component within a line of text. The angle of the fit line is added to an accumulator.The maximum value of theta in the accumulator is used as the skew of the document.

10. G. Bessho and K. Ejiri, “Method and apparatus for detection of a skew angle of a document image using a regression coefficient,” Ricoh Co. Ltd,, U.S. Patent 5,563,403, October 8, 1996.

Patent related to [9].

11. D. S. Bloomberg and G. Kopec, “Method and apparatus for identification and correction of document skew”, Xerox Corporation, U.S. Patent 5,187,753, February 16, 1993.

Determine the variance of the number of black pixels in a scanline as an image is rotated through a number of angles θ. The skew angle corresponds to the angle θ that maximizes the variance.

12. D. S. Bloomberg and G. Kopec, "Method and apparatus for identification of document skew," Xerox Corporation, U.S. Patent 5,355,420, October 11, 1994.

    Patent related to [11].

13. D. S. Bloomberg, G. E. Kopec and L. Dasari, "Measuring document image skew and orientation," Document Recognition II (SPIE vol. 2422), San Jose, CA, February 6-7, 1995, 302-316.

    An input image is downsampled by operators that "reduce the randomness in the result by maintaining some of the structure that exists at higher resolution." The authors comment that this step preserves horizontal structure in the image. After the reduction, horizontal projection profiles are calculated at a number of different angles. Then two measurements are computed on the projection profiles: the sum of squared values in the projection profile and the sum of squared differences between adjacent values. These are computed over a range of angles and a search is performed to locate the angle which maximizes them. This paper includes extensive experimental results on the University of Washington CDROM database.

14. R. H. Britt, "Optical chararacter reader with skew correction," Lundy Electronics and Systems, Inc., Charlotte, N.C., U.S. Patent 4,941,189, July 10, 1990.

    The X-Y coordinates of the center of each character are obtained. The coordinates are rotated through a set of angles and the horizontal projection profile is calculated. The number of different "horizontal lines" that occur at each angle are determined and the angle that *minimizes* the number of horizontal lines in the document is chosen as its skew angle. The number of horizontal lines is equal to the number of bins with non-zero entries in the projection histogram. Thus, the centers in a single horizontal line must be co-linear.

15. A. Chaudhuri and S. Chaudhuri, "Robust detection of skew in document images," IEEE Transactions on Image Processing, v. 6, no. 2, February, 1997, 344-349.

    Method closely related to Yan's.[52] A cross correlation function is computed between vertical slices in a document and skew is estimated by finding a peak in the cross-correlation function. The effect of graphics regions on the skew estimate is reduced with a Monte Carlo technique that estimates skew from a number of small randomly selected regions of the document. Only the result from a subset of them are used in determining the skew of the document.

16. S. Chen and R. M. Haralick, “An automatic algorithm for text skew estimation in document images using recursive morphological transforms,” ICIP-94, Austin, Texas, November, 1994, 139-143.

An input image is downsampled to 100 ppi and a recursive opening transform (ROT) and recursive closing transform (RCT) are applied. They are designed to remove the ascenders, descenders, and overfills from each character. The desired result is one blob for each text line. The connected components in the resulting binary image are located and a least squares procedure is used to fit a line to each blob. A search procedure is then used to determine the skew of the document from the angles of the lines fit to each blob.

17. P. Chevillat and H. R. Schindler, “Arrangement for determining the optimum scan angle for printed documents,” International Business Machines Corp., U.S. Patent 4,338,588, July 6, 1982.

A mechanical-optical method for skew detection. An 8mm by 12mm sensor is described that is composed of a number of parallel strips of light sensitive material that generate electrical signals proportional to the amount of light reflected on them. Four sums (A, B, C, and D) are calculated from every fourth stripe and the value $(A-C)^2 + (B-D)^2$ is computed. This provides a single voltage that measures the amount of light projected onto the sensor at an angle parallel to the stripes. The sensor is connected to a drive motor that rotates the sensor in a plane parallel to that of an imaged document. When a drop is detected in the voltage output by the skew detection circuit (i.e., the peak in the projection profile sum of squared differences has been located), the angle indication is locked and can be read out by other circuitry.

18. J. F. Cullen and K. Ejiri, “Skew detection and correction of a document image representation,” Ricoh Corporation, U.S. Patent 5,452,374, September 19, 1995.

Patent related to [21].

19. L. Dasari and D. S. Bloomberg, “Rapid detection of page orientation, Xerox Corporation,” U.S. Patent 5276742, January 4, 1994.

A document image is first dilated by a general purpose structuring element to smooth local variations. Structuring elements are then applied that detect ascenders and the number of “on” pixels are counted (N_a). Other structuring elements detect descenders and the number of “on” pixels are counted (N_d). Tests on functions of N_a and N_d determine whether the image should be rotated by 90 or 180 degrees. Otherwise, the orientation is unchanged.

20. R. O. Duda and P. E. Hart, “Use of the Hough transform to detect lines and curves in pictures,” Comm. of the ACM 15, 1 (January, 1972), 11-15.

Presents the $\rho-\theta$ parameterization for the Hough transform.

21. K. Ejiri, J. Cullen, M. Tachikawa and K. Kojima, “Fast skew correction of a printed document”, Proceedings of the 12th Annual Conference of the Institute of Image Electronic Engineers of Japan, 1993, 45-48.

Skew estimation performed with a least squares fit to connected components.

22. G. S. D. Farrow and C. S. Xydeas, “Detecting skew in digitised images,” Int. Computers Ltd., London, European Patent App. 485,051, May 13, 1992.

A straight-forward Hough transform method. Every twentieth pixel horizontally and vertically is sampled. If that pixel is black, the accumulator array is updated. The skew angle is estimated by computing an “energy” function at each angle as the sum of squared values over ρ for each value of θ. The maximum of the energy function over theta provides the skew angle.

23. G. S. D. Farrow, M. A. Ireton and C. S. Xydeas, “Detecting the skew angle in document images,” Signal Processing: Image Communication 6, 2 (May, 1994), 101-114.

A multiple-pass Hough transform approach with an accuracy of 0.1 degrees. A page image is first divided into 20x30 rectangular blocks and the percentage of black pixels in each block is determined. If that percentage is between 5% and 25%, the block is considered to be non-noise and the Hough transform is calculated from it with a resolution of one degree over the range plus or minus five degrees. The angle with the maximum response in the transform space is used as the center for a further Hough transform analysis at plus or minus one degree with a resolution of 0.1 degrees.

24. B. B. Fast and D. R. Allen, “OCR image pre-processor for detecting and reducing skew of the image of textual matter of a printed document,” U.S. Patent 5,594,817, January 14, 1997.

A skew detection algorithm applicable directly to run-length compressed data.

25. Y. K. Ham, H. K. Chung, I. K. Kim and R. H. Park, “Automated analysis of mixed documents consisting of printed korean alphanumeric texts and graphic images”, Optical Engineering 33, 6 (June, 1994), 1845-1853.

A constrained run-length algorithm (CRLA) is used to segment a page of text into paragraph-like blocks. A Sobel operator is used to find edges of the blocks and the Hough transform is applied to the detected edges. It is assumed that the document is skewed less than 10 degrees. The skew angle is determined by detection of a local maximum in the ρ–θ space.

26. M. Hase and Y. Hoshino, “Periodic characteristics in printed document,” Transactions of Japanese Institute of Electronics and Communications Engineers, v. J65-D, No. 2, 298-299.

A two dimensional Fourier transform technique (as reported in [9]).

27. A. Hashizume, P. S. Yeh and A. Rosenfeld, “A method of detecting the orientation of aligned components,” Pattern Recognition Letters 4 (1986), 125-132.

Given the connected components detected from a binary image, the histogram of angles between nearest neighbors is calculated. The direction along which the components are aligned is determined from a smoothed local maximum in the histogram.

28. S. C. Hinds, J. L. Fisher and D. P. D’Amato, “A document skew detection method using run-length encoding and the Hough transform,” Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, 1990, 464-468.

An input image is downsampled to 75 ppi and reduced to horizontal and vertical “bursts.” Every black pixel horizontally adjacent to another black pixel is replaced with a white pixel except the one furthest right which is replaced by the length of the run. An analogous process is done in the vertical direction. Page orientation (horizontal or vertical) is first determined by inspection of the frequency of runs of length 1-4. The Hough transform accumulator is then incremented by burst values between 1 and 25 pixels. This emphasizes text up to 24 points in size. The skew angle is determined by finding the (ρ, θ) accumulator cell that has the largest value. The Hough transform is computed for θ between ±15° (or between 75° and 105°) with a resolution of 0.5 °.

29. Y. Ishitani, “Document skew detection based on local region complexity,” Proceedings of the Second International Conference on Document Analysis and Recognition, Tsukuba Science City, Japan, October 20-22, 1993, 49-52.

The number of black-white transitions in each horizontal scan line are calculated as the input image is rotated through a number angles. For each angle, the variance of the number of black-white transitions is determined. The angle that maximizes the variance is the skew angle. Because this method works well only in text areas, the author applies it over a number of *circular* windows. The skew angle is only calculated from windows that produce variance distributions typical of text regions. An accuracy to 0.12 degrees is reported.

30. Y. Ishitani, “Document skew detection/control system for printed document images containing a mixture of pure text lines and non-text portions,” Toshiba, U.S. Patent 5,506,918, April 9, 1996.

Patent related to [29].

31. K. Koga, K. Nakashima, K. Marukawa, Y. Shima and H. Fujisawa, “A method of skew detection of bar-shaped figure in binary image,” Proceedings of the 1992 IEICE Fall Conference 6 , 220.

Reference reported in [9].

32. Y. Kurosu, Y. Yokoyama, K. Nishikawa, H. Masuzaki and M. Fujinawa, “Method for determining the amount of skew of image, method for correcting the same, and image data processing system,” Hitachi, Ltd., U.S. Patent 5,181,260, January 19, 1993.

    A generic hierarchical method for reducing the runtime of a skew detection algorithm. Skew is first measured in a narrow range around one degree. Only if no skew is detected in this range is skew measured in a wider range of plus or minus ten degrees.

33. D. S. Le, G. R. Thoma and H. Wechsler, “Automated page orientation and skew angle determination for binary document images,” Pattern Recognition 27, 10 (October, 1994), 1325- 1344.

    Document skew is estimated from a square sub-region of a document. The bottom pixels (y value the same and maximum within the component) of the connected components in the selected square are input to the Hough transform. Document skew is estimated by finding a local peak in Hough space. Tested on 250 page images. Accuracy of 0.5 degrees reported.

34. D. X. Le and G. R. Thoma, “Document skew angle detection algorithm,” SPIE Conference on Visual Information Processing, v. 1961, August, 1993, 251-262.

    Conference version similar to [33].

35. Y. Lee, “Method of detecting the skew angle of a printed business form,” Eastman Kodak Company, U.S. Patent 5,054,098, October 1, 1991.

    A Hough transform technique. The contours in a binary image are traced and reduced to a piecewise linear approximation (PLA). Each segment in the PLA is transformed using only its center point. The length of the segment is added to the accumulator array location given by the obtained values of ρ and θ. The skew of the document is determined by detecting peaks in the accumulator array. An advantage of this approach is that only one cell of the accumulator array is accessed for each segment in the PLA.

36. Y. Nakano, Y. Shima, H. Fujisawa, J. Higashino and M. Fujinawa, “An algorithm for skew normalization of document images,” Proceedings of the 10th International Conference on Pattern Recognition, Atlantic City, New Jersey, 1990, 8-13.

    The connected components in a document are located and the bounding boxes that meet unspecified size constraints are processed. The width of each bounding box is added to a projection profile taken at several directions. The “sharpness” of each projection profile is measured and the angle that produces the sharpest projection profile is taken as the skew angle of the document. Four measures of histogram sharpness are evaluated. The authors prefer counting the number of zeroes in the projection histogram. Results show that an accuracy of 0.1 to 0.2 degrees can be obtained. This technique will perform better as the amount of text in the document is increased.

37. L. O'Gorman, "The document spectrum for page layout analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence 15, 11 (November, 1993), 1162-1173.

    The five nearest neighbors to each connected component in an input image are located. The angles between nearest neighbors are accumulated in a histogram and a gross estimate of skew is obtained by determining the angle that has the maximum value in the histogram. This is accurate to a resolution of 0.5 degrees. Further processing in the docstrum system segments an image into lines and the skew of each line is determined by a regression fit to the centroids of the components within the lines.

38. M. Okamoto, H. M. Twaakyondo and H. Nishizawa, "Skew detection, skew normalization and segmentation of document images using segmented block code," Journal of the Faculty of Engineering, Japan, 1988, 9-18.

    A document image is reduced to its "segmented block (SB)" coded format. This is similar to calculating the bounding boxes of connected components. The SBs of an image that correspond to isolated characters are determined with tests that use fixed thresholds on the distances between the x-start and y-start points of adjacent SBs. The thresholds are set such that the image is segmented into lines of text. The skew of each line is estimated by a least squares fit to the bottom-left point of each SB.

39. T. Pavlidis and J. Zhou, "Page segmentation by white streams," Proceedings of the First International Conference on Document Analysis and Recognition, Saint-Malo, France, September 30 - October 2, 1991, 945-953.

    Skew angle is calculated during document segmentation. A document image is segmented into "columns" by merging the results of a series of short vertical projections from over the document. The centers of a set of projections that span the same column are determined and the least squares fit to those centers is calculated. The slope of the best fit line is taken as the skew of that column. The skew of the page is determined by averaging the skew values of the columns.

40. T. Pavlidis and J. Zhou, "Page segmentation and classification," CVGIP: Graphical Models and Image Processing 54, 6 (November, 1992), 484-496.

    An expanded version of [39]. Skew detection algorithm is described in detail.

41. T. Pavlidis and J. Zhou, "Page segmentation with tilt compensation," Research Foundation of SUNY, U.S. Patent 5,285,504, February 8, 1994.

Patent related to [39,40]. C source code included.

42. W. Postl, “Detection of linear oblique structures and skew scan in digitized documents,” Proceedings of the 8th International Conference on Pattern Recognition, Paris, France, October, 1986, 687-689.

    Brief conference version of [43]. The description in the patent is more thorough than the conference paper.

43. W. Postl, “Method for automatic correction of character skew in the acquisition of a text original in the form of digital scan results,” Siemens AG, U.S. Patent 4,723,297, February 2, 1988.

    Two skew detection algorithms are described. The first technique uses a projection profile calculated over a range of angles from the whole image. A “directional criterion” is calculated at each angle to measure how well the corresponding angle describes the skew of the image. The directional criterion for the projection profile is the sum of squared differences between adjacent cells in the projection profile. Apparently the angle that maximizes the directional criterion is the skew. The second method for skew detection uses a Fast Fourier Transform (FFT). The coefficients of the power spectrum are calculated and stored in a buffer (spectrum memory). A directional criterion for each of a number of angles is then calculated. The angle that maximizes the directional criterion is the skew angle of the image.

44. M. Rondel and G. Burel, “Cooperation of multi-layer perceptrons for the estimation of skew angle in text document images,” Proceedings of the Internation Conference on Document Analysis and Recognition, Montreal, Canada, August 14-16, 1995, 1141-1144.

    Two neural networks are used. The first outputs a rough estimate of the skew that is used to initialize the weights of the second network. The second network computes the skew of the document. This technique uses a formulation similar to the angle of arrival theory solution proposed by Aghajan et al.[1,2,3]

45. A. Rundle, “Optimum scan angle determining means,” International Business Machines, Inc., U.S. Patent 3,831,146, August 20, 1974.

    A series of masks oriented at different angles (50 degrees, 75 degrees, and 90 degrees are shown) are compared to the image and the number of times each mask matches is computed. The angle associated with the maximum counter value is the skew of the document. The detected skew angle is used to deflect the scanning beam of a flying spot scanner on a subsequent “recognition scan.” This invention seems intended for handwritten characters that have relatively large amounts of skew.

46. A. Schlang, “Text line bounding system,” Litton Systems, Inc., U.S. Patent 4,558,461, December 10, 1985.

A document image is divided into a number of 20-pixel-wide vertical swaths. A binary projection profile is computed for each swath. The projection profile with the most black-white transitions is the start for a search that identifies points on the baseline of text rows. When enough potential baseline points are accumulated from the document image, a least squares procedure is used to fit a line to them. The slope of the line is used as the skew of the document.

47. R, Smith, “A simple and efficient skew detection algorithm via text row accumulation,” Proceedings of the Internation Conference on Document Analysis and Recognition, Montreal, Canada, August 14-16, 1995, 1145-1148.

Connected components are filtered and character-size components are segmented into lines. A least median of squares fit is performed to the components in each line. The median gradient of the fit lines is used as an estimate for the skew angle of the page.

48. A. L. Spitz, “Skew determination in CCITT group 4 compressed document images,” Proceedings of the Symposium on Document Analysis and Information Retrieval, Las Vegas, Nevada, March 16-18, 1992, 11-25.

The white pass codes in a CCITT group 4 fax image are detected. White pass codes are generated when a black run in line N corresponds to no black run in line N+1. Black components typically generate at least one such pass code. Aliasing effects may generate “false” pass codes but since these are often aligned with the skew of the text anyway, they improve algorithm performance. Then a version of Baird's alignment measure is used. The x,y locations of the pass codes are rotated a number of times and projected onto the vertical axis. The rotation that maximizes a given statistical measure provides the skew angle. The use of CCITT group 4 images avoids the need to locate connected components.

49. A.L. Spitz, “Determination of image skew angle from data, including data in compressed form,” Xerox Corporation, U.S. Patent 5,245,676, September 14, 1993.

Patent related to [48].

50. S. N. Srihari and V. Govindaraju, “Analysis of textual images using the Hough transform,” Machine Vision Applications 2 (1989), 141-153.

A Hough transform method for skew detection. The Hough transform is computed at all angles of θ between 0 and 180 degrees. A heuristic measures the rate of change in accumulator values at each value of θ. The skew angle is set to the value of theta that maximizes the heuristic.

51. M. Yamada, “Image processing system,” Canon, U.S. Patent 4,802,229, January 31, 1989.

The four corners of a sheet of paper on a copier platen are detected. The skew angle is determined from the corner points with an arc tangent lookup table. The resolution of skew detection is 0.225 degrees.

52. H. Yan, “Skew correction of document images using interline cross-correlation,” CVGIP: Graphical Models and Image Processing 55, 6 (November, 1993), 538-543.

A binary image is represented as ones for the background and zeroes for the foreground. The cross-correlation between two vertical scan lines given a fixed shift s in Y is calculated at a fixed distance d between scan lines. A scalar value R(s) is determined for each shift in Y by summing over X. The value of the shift s that maximizes R(s) is used to calculate the skew of the document. This method was tested on two images and accuracy to 0.2 degrees was reported if the input image contained primarily text.

53. B. Yu and A. K. Jain, “A robust and fast skew detection algorithm for generic documents,” Pattern Recognition, 29, no. 10, October, 1996, 1599-1630.

A hierarchical Hough transform method applied to centroids of connected components. Demonstrated on images of technical articles, postal labels, handwritten text, forms, drawings, and bar codes.

54. C. L. Yu, Y. Y. Tang and C.Y. Suen, “Document skew detection based on the fractal and least squares method,” Proceedings of the Third International Conference on Document Analysis and Recognition, Montreal, Canada, October 14-16, 1995, 1149-1152.

A two stage approach to skew correction. Whole-page skew is corrected first. The skew of individual text blocks is then estimated with a least squares method that fits lines to text within the blocks.