Analysis of Document Skew

Dan Bloomberg

1 Issues with document skew

Scanned documents can be expected to have some document skew. A few tenths of a degree is typical for cut sheet paper that is fed with an automatic document feeder, and it is barely noticeable. However, for scans from bound volumes or pages manually placed in a feeder or on a platen, one degree of skew is common and two degrees is not unusual.

There are three main reasons for removing the skew. The first is appearance. Anything more than about 0.25 degrees (0.004 radian) is quite noticeable. The second is that it is important to remove skew if any analysis is to be done on the page. The presence of skew, and particularly more than 0.01 radians, complicates the analysis of page elements such as text columns. The third is that the performance of symbol-based compression on multipage documents is badly degraded by random skew of 0.01 radian or more, because the same characters are placed in different equivalence classes due to skew.

Printed documents typically have a single, global skew angle, and removal of skew by image rotation is simple once the skew angle is known. So the issue is how to find skew.

Some complications that should be kept in mind as we discuss different methods are:

- Pages with halftone images can have many tiny dots, often covering an appreciable fraction of the page
- Pages with multiple columns often have no alignment between text lines in the different columns.
- Pages scanned from bound documents often have part of the opposing page appearing at a skew angle differing by several hundredths of a radian.

For many applications, it is important to find the global skew in a time that is a small fraction of the processing time for the scanning or for other computation, such as OCR. Thus, a typical engineering requirement is to find the skew angle in a fraction of a second.

2 Methods for determining document skew

2.1 Use all the pixels

Many papers have been published on how to determine the skew of printed documents. They can be divided into two groups: those based on the projection profiles of markers derived from the image, and those based on projection profiles of all the pixels within binary images. A projection profile is a function that is the sum of either markers or ON pixels along a set of parallel lines, with the parallel lines taken through the image at a spacing of approximately one pixel. The easiest projection profile to compute is the horizontal one, taken along the raster lines. In theory, the lines can be at any angle θ with the horizontal (raster) axis.

Some early papers (ca. 1990) tended to eschew computation on large binary images. Under the impression that it would take on the order of a minute to estimate the skew angle, they would spend perhaps 10 seconds to find the connected components, and choose a marker from each component. Typically, you might choose a marker from the bottom of the bounding box for each connected component. Then for characters that don't have descenders, the marker would approximate the baseline of the line of text at that character. From this typically sparse set of markers, it was relatively easy to find the best global skew angle such that the markers tended to line up.

However, since it is possible to compute the skew angle using *all the pixels in the binary image* in a small fraction of a second, such methods are preferred over those using a selection of markers. Further, where there are a very large number of connected components in a halftone image, the marker method fails for two reasons: the connected component calculation is slow and most of the resulting markers are noise. And finally, there are problematic images, described above, for which use of markers will fail.

2.2 Use projection profiles

The use of the *Hough transform* for finding the projection profiles is common. The Hough transform is a transform from (x, y) pixel space to (θ, ρ) projection space, where θ is the angle with respect to the x-axis that gives the projection direction and ρ is the coordinate along the line perpendicular to the projection direction that passes through the origin. Thus, ρ parametrizes the location of the projection sum. The basic equation for the Hough transform of an image $f(x, y)$ is:

$$p(\rho, \theta) = \int\int f(x, y)\delta(\rho + x \sin\theta - y \cos\theta)dx\, dy \tag{1}$$

This is easily interpreted. The Dirac delta function has meaning only within an integral. Integrating over a delta function extracts the value(s) of the function $f(x, y)$ where the argument of the delta function goes to zero; i.e., at $\delta(0)$.

For each value of ρ and θ, $\delta(0)$ gives you a straight line that satisfies

$$y = g(x) = \rho/\cos\theta + x \tan\theta \tag{2}$$

so that the double integral integrates (or sums, in the discrete case) all values of $f(g(x), x)$ along this line. The line in (2) is at angle θ with the x-axis, has a slope $\tan\theta$, and intercepts the y-axis at $\rho/\cos\theta$. Draw the diagram and verify this.

This can be re-paramaterized into a single integral over the straight line:

$$p(\rho, \theta) = \int f(\rho\sin\theta + s\cos\theta, \rho\cos\theta + s\sin\theta)\, ds \tag{3}$$

If there is a global skew angle, it might be hoped that the Hough transform has the largest set of peaks in ρ for some specific value of skew angle θ. This idea is quantified below using the variance of the projection values. We have described the Hough transform for three reasons: (1) it does give you the projection profiles, which are necesary for the computation of skew angle, (2) it is usually mentioned in the literature, and (3) it is equivalent to a very important transform, the *Radon transform*, that is used for finding tomographic (slice) solutions of inverse problems, such as densities, in medical imaging.

For completeness, we note that the Radon transform is usually defined by using the angle ϕ (with the x-axis) made by *the perpendicular from the origin to the projection line*, rather than the angle made by the projection line itself. (As before, ρ is the distance from the origin to the projection line.) The angle ϕ is related to θ by

$$\phi = \theta + \pi/2 \tag{4}$$

Then in terms of ϕ, the standard forms for the Radon transform are

$$p(\rho, \phi) = \int\int f(x, y)\delta(\rho - x\cos\phi - y\sin\phi)dx\, dy \tag{5}$$

$$p(\rho, \phi) = \int f(\rho\cos\phi - s\sin\phi, \rho\sin\phi + s\cos\phi)\, ds \tag{6}$$

We use the definitions in (1) and (3) because they give the skew angle in terms of θ, the angle of the projection with the x-axis.

2.3 Use the variance of the projection profiles

Intuitively, if we take the projection profiles at the global skew angle of the text (i.e., parallel to the text lines), we expect that some projection values will have large values and others will have small values, depending on whether the projection runs through a text line or between lines. Thus, we should maximize the *variance* of the projection values p. If we have N projections for each angle θ, then the expectation value of the projection is

$$< p >= (1/N)\sum_{\rho} p \tag{7}$$

and the variance, which measures the deviation from the average is

$$\begin{aligned}\Delta p(\theta) &= (1/N)\sum_{\rho}(p- < p >)^2 &(8)\\ &= (1/N)\sum_{\rho}(p^2 - 2p < p > + < p >^2) &(9)\\ &= < p^2 > - < p >^2 &(10)\end{aligned}$$

The expectation value of the projection, $< p >$, given in (7), is proportional to the sum of all foreground pixels, and hence is a constant, independent of the projection angle θ. Consequently, it suffices to determine the angle θ that maximizes $< p^2 >$ alone in (10), as the $< p >^2$ term gives the same additive constant to each value of $\Delta p(\theta)$.

It may be worth noting that other, more complicated, computations have been advocated. One of these is the auto-correlation function of the projections. This gives a relatively large oscillating signal when the projection direction is aligned with the text lines. However, it is silly to calculate this function because it takes far more work to compute than the variance, and once you have it, you still need to reduce it to a single number that can be maximized as a function of θ. And that single number can be no more indicative of alignment than the variance.

2.4 Use the variance of the projection profile derivative

The method advocated here is described in detail in ”Measuring document image skew and orientation,”, D. S. Bloomberg, G. E. Kopec and L. Dasari, SPIE Conf 2422: Document Recognition II, pp. 302-316, San Jose, CA, Feb. 1995. It can be found online at *www.leptonica.com/recent-pubs.html*.

There are several problems with using the maximum of the variance to find the global skew angle:

- The peak (variation with θ) is quite broad. The half-angle is roughly the x-height divided by the width of the text line. With a broad peak, it is difficult to find the center accurately.

- The method tends to fail when there are multiple columns of text. Often with multiple columns, the text lines do not line up across columns. With a random distribution of text line positions in each column, the variance in the number of pixels is greatly reduced.

- The method tends to fail when the scan includes some of a second page. The effect is similar to situation with unaligned text lines in multiple columns.

- In some images the signal-to-noise level is small. For images with large halftone regions, little of the contribution to $< p^2 >$ comes from the text lines. The halftone regions can add a significant amount of noise to the remaining signal, further complicating the task of identifying the peak.

All these problems are solved by using a *differential signal from the projection profile*. Here's the prescription: *Take the difference in projection values between neighboring values of ρ, square them, and sum over all values of ρ.* Mathematically, denote the individual projections $p(\rho, \theta)$ as $p_i(\theta)$, where the index i runs over all values of ρ. Then we are computing a signal $S(\theta)$

$$S(\theta) = \sum_i (p_i(\theta) - p_{i-1}(\theta))^2 \tag{11}$$

This has maximum contributions for projections (values of i) that run along the text baselines and x-heights. It has essentially no contribution from projection pairs that run through either text or between the text, or through large regions of halftone noise. It works well for multiple columns, regardless of the relative text line alignment in the columns. And the peak is very sharp, having a width in radians of approximately 2/(text line width in pixels).

2.5 Other ways to compute projection profiles

Computation of the Hough or Radon transforms in libraries such as Matlab are not optimized to handle large binary images. Further, the values of θ to be computed must be specified in advance. We discuss search algorithms for maximizing $S(\theta)$ in the next section. Here, we discuss the calculation of the projection profiles for a given value of θ.

For binary images, one has two choices:

- compute the sum of ON pixels along parallel lines at angle θ, or
- do a vertical shear on the image that corresponds to the angle θ, and then compute the sum of ON pixels horizontally along the raster lines.

Either of these methods can be optimized, but it is easier to shear the image and then count ON pixels in the raster lines. Note that it is more efficient to do a single vertical shear than to actually rotate the image, and that the vertical shear and the rotation have the same effect in putting pixels initially distributed on a line at angle θ onto a horizontal raster line. Shearing about one edge of the image, and not about the center, doubles the ability to make small "rotations."

The vertical shear is efficiently implemented using rasterops. The basic rasterop takes a rectangular block of a source image and does some logical operation with the existing pixels, at the new location, of the destination image. For the vertical shear, the rectangular block of the source image is moved up or down and just written into the destination image. The number of blocks moved increases linearly with the angle θ, and the width of the block in pixels is approximately $1/\theta$. Then the sum of ON pixels on a raster line can be done very quickly using an 8-bit or 16-bit table lookup.

2.6 Using linear and binary search

If the absolute value of the image skew is not expected to exceed some value θ^{max}, a search needs to be done between $-\theta^{max}$ and $+\theta^{max}$. A typical value for θ^{max} is about 5 degrees. There are two general approaches to the search for the optimum angle θ:

- Sweep at equal intervals, followed by interpolation for the peak.
- Binary search with interval halving until the peak is located.

Either method can be used, and both methods can be combined. However, when using the variance of the differential profile sums, the peak is very narrow. More than about 0.01 radians from the actual skew angle, the signal is essentially flat. Therefore, a binary search starting at large angles is likely to miss the peak.

There are many possible methods for doing the search. Among them, the following will work:

- Linear sweep at equal intervals. When the peak is determined to be between two values of θ, do another linear sweep between these endpoints, using smaller increments. This can be followed by interpolation for the final value.
- Linear sweep at equal intervals. When the peak is determined to be between two values of θ, do a binary search in θ with interval halving until the peak is located.
- Do a binary search using the variance of projection values (not the variance of *differential* projection values), which has a sufficiently wide peak that it is likely to be found by interval halving. Refine the search using differential projection values.

The interval halving method can be done as follows. Suppose you are searching in the interval of width Δ, from θ_1 to $\theta_1 + \Delta$. You already have the values at θ_1, $\theta_1 + \Delta/2$, and $\theta_1 + \Delta$. You then find the values at the intermediately spaced angles $\theta_1 + \Delta/4$ and $\theta_1 + 3\Delta/4$. From these five points you find the value of θ with the largest signal and choose the two nearest values of θ within the interval to it. These three values of θ then constitute a new interval of size $\Delta/2$, and the process is repeated.

3 Evaluating the result

Suppose you use one of these methods and apply it to an image that has little or no text. Perhaps you have analyzed a line drawing or a halftone image. You have found a skew angle, but how do you know it is correct, or even has any meaning? You must evaluate it for validity.

This is not difficult to do. You evaluate some function of the result you get and test it against a threshold. The purpose may be to get a number, or simply to get a boolean output (valid/invalid). It is difficult to get a number here that represents a known statistic on the image for skew, such as the probability that it is correct within some error range, so we won't worry about that. (However, for the determination of orientation – up or down – one can easily give a statistical weight to the result. This is described in the paper referenced above.)

To evaluate the validity of the skew angle, you can compute any or all of the following:

- Get the ratio of the peak height to the average value.
- Get the ratio of the peak height to the height of the next largest peak.
- Find the width (say, half width at half maximum) of the peak.

As a general rule, if the peak is narrow, and the peak height is much larger than both the average background and the height of the next-highest peak, you can assign a very high confidence to its value. Things get fuzzier when one of these conditions is not obtained. For example, a second peak of significant size can represent part of another page scanned at a different angle (as often is found when a bound volume is scanned). However, even a very small amount of text, such as one or two text lines, can provide a signal that is sufficiently robust for determining the skew angle to within ± 0.003 radians.

The accuracy of the angle determination is related to the width of the peak in $S(\theta)$. For an ideal image and scan, the angle can be determined to approximately $1/L$, where L is the length of the text lines in pixels. For a text line of width 1500 pixels, this gives an uncertainty in the skew angle of less than 0.001 radians (about 1/17 degree), far below the ability of the eye to detect.

The peak in $S(\theta)$ can be degraded by various factors, such as very short text lines, multiple columns of text with random inter-column alignment, curved baselines from scanning out-of-focus near bindings, etc. But in most situations, the differential variance method is sufficiently good to provide a reference standard for skew of any scanned document.