

obtain the final surface. Noisy heights will be ignored since the elastic surface will not bend too much. The document image is deskewed (warped back) with the estimated rotation angles. Fig. 4 shows an example. More results can be found at our web page at <http://iris.usc.edu/home/iris/han/tdeskew.html>.

4 Implementation Issues

The functions for decaying with distance and offset angles can be Gaussians, square root, linear, square, cubic. For offset angle decay, the cosine function is also plausible since it decreases from 1 (at 0 degree) to 0 (at 90 degrees). We now use Gaussians for both decays, but it seems that

$$f = 2(d/R)^3 - 3(d/R)^2 + 1 \quad (\text{for } d < R)$$

is a good decay function due to the C^1 smoothness at the window boundary and there is no long tail as for Gaussians. It is also faster to compute than Gaussians.

Embedded small pictures in the texts are tolerable, since a picture contributes only one or a few centroids, and rotations in the picture region can be interpolated by the nearby texts with the elastic surface.

5 Further Discussion

The skew may have a sharp change between two neighboring regions, so the elastic surface may contain a jump or a crease discontinuity. TV can find such discontinuities, since the votes are conflicting along such a boundary. To speed up, the voting can be performed on a coarse grid first, then the grid is refined around the extracted ridge curves for one more refined voting. Besides the rotation, the distance distortion can also be included in the deskewing (warping) transform.

6 Conclusion

We propose using tensor voting for non-uniform deskewing of scanned documents. TV retains the advantages of voting in HT, but overcome drawbacks in the extremely global or extremely local voting schemes.

References

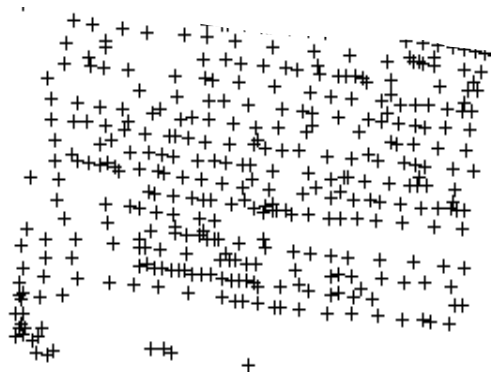
- [1] C. L. Yu, Y. Y. Tang, and C. Y. Suen, Document Skew Detection Based on the Fractal and Least Squares Method, Proc. of 3rd Int'l Conf. on Document Analysis and Recognition, Montreal, pp. 1149-1152, 1995
- [2] G. Guy and G. Medioni, Inferring Global Perceptual Contours from Local Features, Int'l Journal of Computer Vision, vol. 20, no. 1/2, pp. 113-133, 1996
- [3] M. Lee, Using Tensor Voting as a New Framework for Several Computer Vision Problems, PhD thesis proposal, CS Dept., USC, Jan. 1997

Other relevant work includes the method of probabilistic occupancy grids developed by Elfes and Mintzer [10]. Their volumetric space is a scalar probability field which they update using a Bayesian formulation. The results have been used for robot navigation, but not for surface extraction. A difficulty with this technique is the fact that the best description of the surface lies at the peak or ridge of the probability function, and the problem of ridge-finding is not one with robust solutions [8]. This is one of our primary motivations for taking an iso-surface approach in the next section: it leverages off of well-behaved surface extraction algorithms.

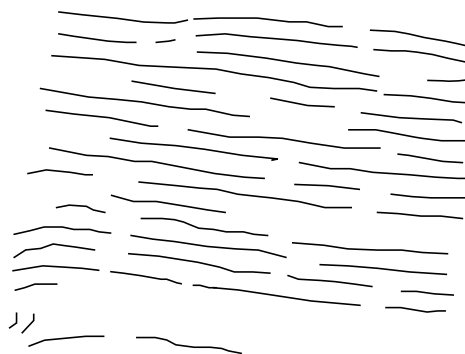
The discrete-state implicit function algorithms described above also have much in common with the methods of extracting volumes from silhouettes [15] [21] [23] [27]. The idea of using backdrops to help carve out the emptiness of space is one we demonstrate in section 4.

3 Volumetric Integration

(a) scanned document



(b) centroids (25% samples) of connected components



(c) text lines traced from tensor voting result

Other relevant work includes the method of probabilistic occupancy grids developed by Elfes and Mintzer [10]. Their volumetric space is a scalar probability field which they update using a Bayesian formulation. The results have been used for robot navigation, but not for surface extraction. A difficulty with this technique is the fact that the best description of the surface lies at the peak or ridge of the probability function, and the problem of ridge-finding is not one with robust solutions [8]. This is one of our primary motivations for taking an iso-surface approach in the next section: it leverages off of well-behaved surface extraction algorithms.

The discrete-state implicit function algorithms described above also have much in common with the methods of extracting volumes from silhouettes [15] [21] [23] [27]. The idea of using backdrops to help carve out the emptiness of space is one we demonstrate in section 4.

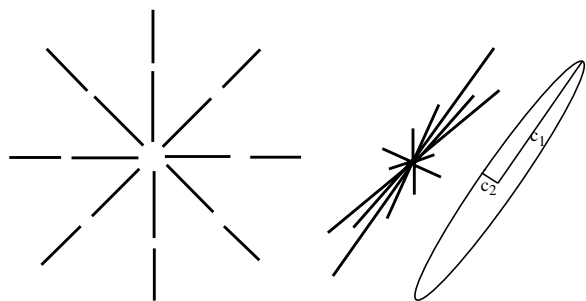
3 Volumetric integration

(d) deskewed document by local warping

Fig. 4. Deskewing by Tensor Voting

Then, we need to figure out which direction is the winner. A simpler way is to discretize the directions into a few bins, and the bin with maximum weight is picked as the winner. If the decaying function is a constant value 1 and we use the above voting, TV becomes exactly the conventional HT. The only difference is that the voting and peak finding is done in the image domain directly, whereas the conventional HT collects votes and find the winner in the transform domain. Of course, TV can also be implemented in the transform domain but seems less efficiently: for any position X in the image, all centroids casts votes along its own sinusoidal curve in the transform domain, and the voting strength is a value in between 0.0 - 1.0 determined by distance (in HT, the strength is always 1). To find the winner direction for the position X , just pick the strength peak on the sinusoidal curve of X in the transform domain.

A more robust way to find the winner direction is to avoid the discrete bin counting, but compute a 2×2 covariance matrix (or scatter matrix, 2nd order moments, ellipse fit) of all votes, and take the eigen vector e_1 corresponding to the larger eigen value c_1 as the winner direction, since this direction is the longer principal axis of the ellipse, as depicted in Fig. 1(b). The value of c_1 reflects the absolute strength, and the eccentricity c_1 / c_2 reflects the relative strength (or certainty, consistency, agreement) of the votes. $c_1 - c_2$ gives a comprehensive measure of the strength (c_2 : length of the shorter axis).



(a) votes cast by a centroid (b) votes at a cell; ellipse

Fig. 1 Tensor Voting Method (Pass One)

Then, the skew angles (winner directions) at all centroids can be used to interpolate the skew angles in the whole image. But the interpolation may not be good enough if the centroids are sparse or the text lines are very curved, since the above voting is based on local linear voting.

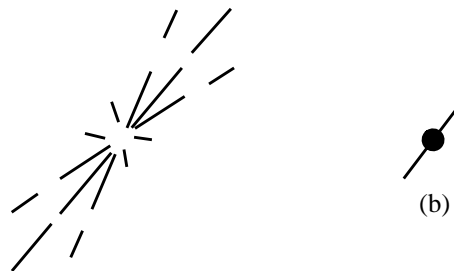
2.2 Pass Two: Voting by Oriented Centroids

Now that we have elected the probable direction at each centroid and known the certainty of the result, a new pass of voting can be performed to obtain better results. For centroids with high eccentricity (> 0.8), it is quite certain that the curve is almost straight here, so an oriented linear pattern is used for voting. For cells along

the same direction, the voting strength is decayed with distance; but for cells at an offset angle, the voting strength is further decayed with the angle. Fig. 2 (a) shows the linear voting kernel for an oriented centroid shown in Fig. 2(b).

For a centroid with low eccentricity, the curve is not very straight here, and we need to use curves to vote for nearby cells. Locally, many circles could pass through this centroid, as indicated in Fig. 3 (a). For a cell P , the voting strength is obtained by decaying with distance and offset angle. The direction (tangent) at P is estimated by the tangent of the circle at P , as shown in Fig. 3(b).

After the second pass voting, we compute the winner direction and the strength at each cell. The text lines (curves) can be found by tracing the ridges (peak curves) in strength with guide from the tangents. To obtain smooth curves, B-splines can be fitted.



(a) votes cast by an oriented centroid shown in (b)

Fig. 2 Pass Two Voting with Lines



(a) votes cast by an oriented centroid shown in (b)

Fig. 3 Pass Two Voting with Curves

3 Skew Angle Interpolation

The tangent at each point on the ridge curves provides the skew angle of the text lines at this point. Computing the skew angles for all positions in the image is a surface interpolation (or fitting, reconstruction) problem. Just imagine the angle value as a height value, and the problem is to fit a smooth surface to these landmark heights. For uniform skew, the surface is a horizontal flat plane (same height everywhere). A rectangular polynomial surface, or a triangular spline surface can be used for fitting. We start with a flat plane as the initial surface, then deform the elastic surface with the landmark heights to

Non-uniform Skew Estimation by Tensor Voting

Song Han, Misuen Lee, and Gérard Medioni

Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, CA 90089
han@iris.usc.edu, medioni@iris.usc.edu
(213) 740-6442 (tel), (213) 740-7877 (fax)

Abstract

We apply a perceptual grouping concept for document deskewing. The method is similar to Hough Transform, but each voter casts its votes in a neighborhood instead of the whole domain; and the winning policy is based on tensors, which are more sophisticated than scalars and vectors. The voters are centroids of connected components, and skew angles are estimated along the skewed text lines by tracing the ridges (arbitrary curves of peaks) in the voting domain, and the complete skew angle field is obtained by elastic surface interpolation.

1 Previous Work

In unskewed documents, the texts should go along horizontal lines, but scanned documents may have undergone a rotational skew. To estimate the skew angles, most methods rely on some kinds of voting. The projection profile method collects votes of all black pixels along a direction (parallel straight lines), and computes the variance of the histogram. Many directions are tested and the direction with the maximum histogram variance reveals the text lines. The nearest neighbor clustering method links each connected component to its closest neighbor, and each line segment gives an angle. When all segments vote, the peak of the histogram is taken to be the skew angle. In Hough transform (HT) method, the centroids or the bottom midpoints of all connected components act as voters, each voting along a sinusoidal curve in the 2D voting domain and casting one vote at each cell, and finally the peak is found from the voting domain to specify the skew angle.

The voting scheme is robust to outliers, noise and errors. But the voting is performed for the whole image, and can estimate only one skew angle. Yu et al. [1] detected multi-skews by fitting a straight line to each text line.

However, such least square fitting is sensitive to outliers, and also relies on nearest neighbor clustering of characters and words into text lines, which fails for narrow spacing, large fonts, blurred letters, or some languages. In the non-uniform skewing effects existing in copying and scanning, the text lines deform to unknown curves, instead of straight lines. To use Hough transform, there must be a pre-defined equation or table, and thus unknown curves cannot be detected. We were inspired by a perceptual grouping method in the computer vision and pattern recognition literature [2], and based on it, we propose a new document deskewing method, which performs tensor voting (TV) in a medium-range window, and the skewed text lines (unknown arbitrary curves) are found by tracing the peak curves (or ridges) from the voting domain. The local voting (nearest neighbor clustering, least square fitting) and the global voting (Hough transform) can be regarded as the extreme cases of our method.

2 Tensor Voting Method

Given a scanned binary image of a document, we first compute the connected components, and remove the too small or too large components, then the centroids of the components are used for tensor voting as follows.

2.1 Pass one: Voting by Centroids

In the image domain, we overlay a voting domain grid. Since we do not know the orientations (tangents) at the centroids, the text line could pass through a centroid along any direction. Locally the skewed text line can be regarded as a straight line first, so each centroid casts votes along all directions as shown in Fig. 1 (a), but more generalized than the HT, the voting weight (or strength) decays with distance, as depicted by the length of the segments. Due to the decaying, each centroid casts votes at cells only within a window.

After all centroids have voted, each cell in the whole domain collects a bundle of votes as shown in Fig. 1(b).

This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by Topographic Engineering Center of the U.S. Army.