DOCUMENT IMAGE RECTIFICATION USING FUZZY SETS AND MORPHOLOGICAL OPERATORS

Shijian Lu *Ben M. Chen* *C. C. Ko*

Electrical and Computer Engineering Department, National University of Singapore, Singapore, 119260

ABSTRACT

In this paper, we deal with the problem of document image rectification from images captured by digital cameras. The improvement on the resolution of digital camera sensors has brought more and more applications for non-contact text capture. Unfortunately, perspective distortion coupled with resulting images makes it harder to properly identify the contents of captured texts using the traditional optical character recognition (OCR) system. We propose in this work a new technique, which is capable of removing distortion and recovering the fronto-parallel view of text with a single image. Different from reported approaches in the literature, the image rectification is carried out using character boundary and tip point, which are extracted from character strokes based on multiple fuzzy sets and morphological operators. The algorithm needs neither camera calibration nor high-contrast document boundary. Experimental results show our rectification process is fast and robust.

1. INTRODUCTION

Document image processing involves transforming document text into electronic form for further processing. Combined with scanners, the OCR system has been widely used for such a transformation. However, the high-speed non-contact text capture through digital cameras with high resolution is obviously a better choice than scanners for capturing text in the required electric form. Unfortunately, perspective distortion of document image coupled with resulting images brings up a new problem to the traditional OCR system. Similar to skew compensation step required after scanning process, perspective distortion must be removed before the document image is fed to the OCR system.

There are few rectification techniques proposed in the literature for document images captured through digital cameras. In [1], the quadrilateral formed by the borders of document is utilized to get a fronto-parallel view of perspectively distorted text. After the extraction of quadrilateral, a bilinear interpolation is then implemented to construct the rectified document image. The limitation of this algorithm is that it depends heavily on the existence of high-contrast borders between document and background. In another work [2] by Clark as well, the distortion rectification is implemented through the calculation of two vanishing points.

Instead of using document borders, Pilu proposed a new rectification approach [3] based on the extraction of illusory clues. After saliency measure is computed for pairs of neighbouring blobs, a network is transversed over the text and horizontal clues are calculated as the salient linear groups of blobs. Though working well on horizontal clues extraction, the proposed method cannot extract enough vertical ones. And camera calibration is necessary for a full rectification.

Different from the above mentioned methods, we utilize character stroke boundaries and tip points to remove perspective distortion. After extraction of stroke boundaries and tip points using some morphological operators, boundaries of vertical strokes are first identified. Fuzzy sets and aggregation operations are then employed to determine potential horizontal and vertical lines and to rectify distortion. The proposed approach needs neither camera calibration nor assumption of existence of high-contrast borders. It only requires a single document image captured by a digital camera.

2. FEATURE EXTRACTION

2.1 Stroke boundary extraction

Mathematical morphology is a powerful tool for extracting image components that are useful in the representation and description of region shape such as boundaries, skeletons, and the convex hull. In our proposed approach, two types of features including character stroke boundary and tip points are extracted using some morphological operators. The aim of boundary extraction is to identify special ones that indicate the vertical direction of document and extracted tip points can be classified to fit horizontal lines.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

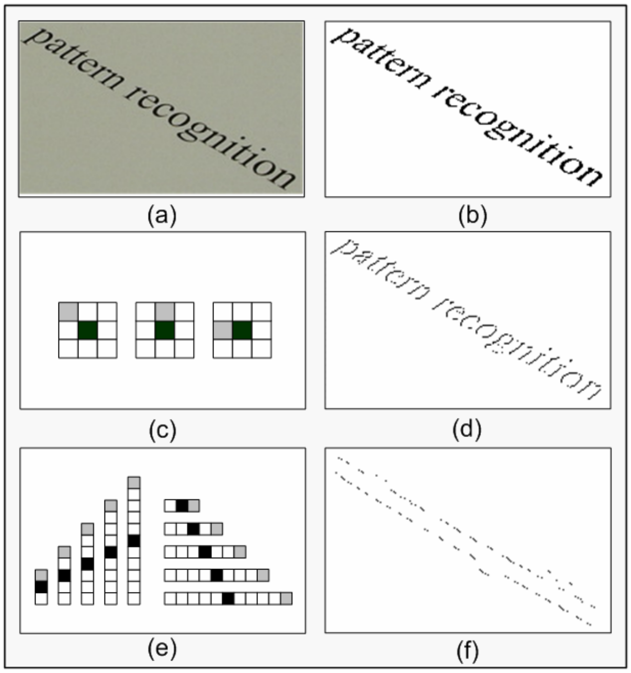

Figure 1: Feature extraction: (a) source image; (b) binarized image; (c) structure elements for boundary extraction; (d) extracted boundaries; (e) structure elements for tip point extraction; (f) extracted tip points

To facilitate the feature extraction, document image is first binarized using adaptive local thresholding method proposed in [4]. Figure 1(a) shows the source image and Figure 1(b) shows the binarized one. As the aim of boundary extraction is to get a group of vertical line segments, we use some customized structure elements shown in Figure 1(c) to extract stroke boundaries, where darker pixels refer to origins of structure elements. The composite morphological operation is:

$$BDY_l = \left[A \veebar \left(A \Theta B\right)\right] \cap \left\{\left[A \veebar \left(A \Theta C\right)\right] \cup \left[A \veebar \left(A \Theta D\right)\right]\right\} \qquad (1)$$

where $A\Theta B$, $A\Theta C$ and $A\Theta D$ refer to erosion operations on binarized image A using structure elements B, C and D as shown in Figure 1(c). And symbol $\veebar$ represents *XOR* operation.

Figure 1(d) shows the extracted left side stroke boundaries. Another pair of structure elements, which is the reflection of ones shows in Figure 1(c), can be used to extract the right side boundaries.

2.2. Tip point extraction

Similar to the character boundary extraction, the extraction of tip points is implemented using a group of customized structure elements as well. Figure 1(e) shows ten structure elements that are combined to extract tip points near top line position. And the reflection of ten structure elements shown in Figure 1(e) can be combined to extract tip points near base line position. Figure 1(f) shows the extracted tip points. The composite operation is:

$$BDY_t = \left\{\left[A \veebar \left(A \Theta B_1\right)\right] \cap \cdots\cdots \cap \left[A \veebar \left(A \Theta B_5\right)\right]\right\} \cap \left\{\left[A \veebar \left(A \Theta C_1\right)\right] \cap \cdots\cdots \cap \left[A \veebar \left(A \Theta C_5\right)\right]\right\} \qquad (2)$$

where $B_1 \cdots B_5$, $C_1 \cdots C_5$ are ten structure elements shown in Figure 1(e). The symbol Θ represents the erosion operation and symbol $\veebar$ refers to the *XOR* operations.

3. FEATURE PROCESSING

3.1. Vertical boundary identification

Vertical boundary of character such as the left side boundary of character "b" and right side boundary of character "d" refers to the boundary that indicates the vertical direction of corrected document image. In our proposed approach, multiple properties of extracted boundaries are utilized to separate vertical ones.

Connected Component Analysis is firstly implemented and the boundary image is labeled as multiple foreground regions with each connected boundary as one region. Three properties of these foreground regions are calculated. The first property is the region size, which is equal to the number of pixels within the region. The second one is the region orientation and it can be calculated as the slope of fitted line using least square method. And the last property is the region linearity, which describe the "straightness" of region and can be calculated as:

$$dist = \frac{1}{n}\sum_{i=1}^{n} d\left(p_i, l\right) \qquad (3)$$

where n is the region pixel number, l refers to the fitted line using least square method, and function d calculates distance between pixel p_i and fitted line l.

Three fuzzy sets are constructed to characterize the size, linearity and pose properties. As larger size is preferred for vertical boundaries extraction, we choose Zadeh's S-function $S(x;\ a,\ b,\ c)$ [5] as membership function. Parameter a is the minimal size of all extracted boundaries and parameter c is chosen as 1.5 times of average size of boundaries.

The second property is linearity property. As we prefer to the straightness of extracted region, we use the complement of S-function as membership function. Two parameters are chosen as the minimal distance and two times of average distance, where distance refers to the normalized point to line distance calculated using equation (3).

For the pose property, we calculate the average slope of all fitted lines. And the region with slope closest to average slope is assigned a highest value. Therefore, we use bell-shaped function $\pi(x, b, c)$ as membership function, in which parameter c is taken as the average boundary slopes and parameter b is half of parameter c

Consequently, each boundary corresponds to a triple, which is composed of size, linearity and pose properties. Figure 2(a) shows boundary distribution within a frame defined by size, linearity and pose as three axes where each point represents an extracted boundary.

Combining three fuzzy sets, an aggregation set is constructed to separate the vertical boundaries. In our proposed approach we utilize the compensatory operators. The aggregation operation arises as a sort of combination of "pure" logical AND and OR connectives providing the required mechanism of compensation. The operation defined in Zimmermann and Zysno [6] is expressed as:

$$(A \otimes B)(x) = (1-\gamma)[(A \cap B)(x)] + \gamma[(A \cup B)(x)] \quad (4)$$

where $\cup$ and $\cap$ are union and intersection operations, whereas γ stands for the compensation factor $\gamma \in [0,1]$, indicating where the actual operator is located between *AND* and *OR*.

With constructed aggregation set, the vertical boundaries can be separated as α cut of aggregation set. Parameter α must determined with care. If it's too small, some false boundaries will be selected. And if it's too big, some vertical boundaries will be missed. Figure 2(b) shows the extracted vertical boundaries with α as 0.7.

3.2. Tip point tracing

Tip point refers to the uppermost and lowermost points of character. The aim of tip point tracing is to classify all extracted tip points shown in Figure 1(f) into groups and ideally, each group fits to one desired straight line. In our proposed approach, point to point and point to line distance constraints are utilized to classify tip points. Figure 2(c) shows the tracing process.

In our proposed approach, we exploit both point to point and point to line distance constraints to classify the extracted character tip points. For example, in Figure 2(c), Points 1-6 are six points classified to the same group. Line L is straight line fitted based on all classified tip points, which include points 1-6 in this example. The target of the tip point tracing process is to search for next point candidate that is nearest to last classified point (point 6) and at the same time satisfies two distance constraints.

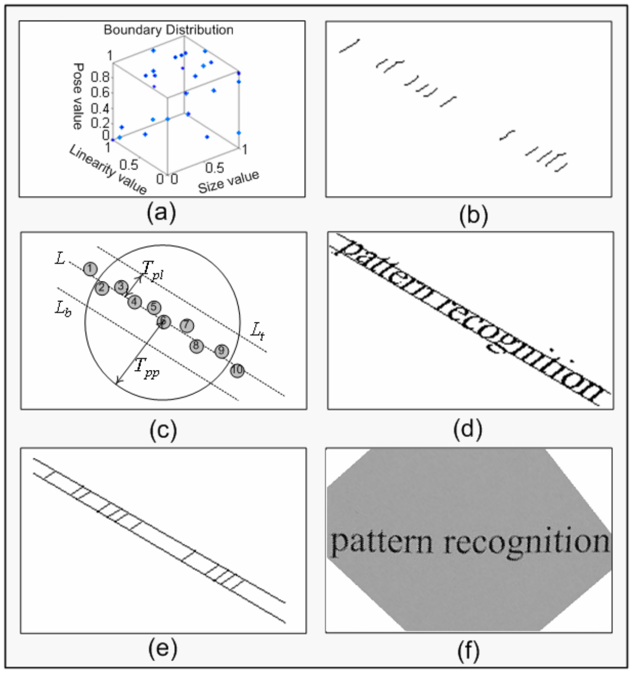

Figure 2: Feature processing: (a) boundary distribution based on size, linearity and pose properties; (b) extracted vertical stroke boundaries; (c) tip point tracing; (d) fitted top and base lines; (e) two sets of fitted lines (f) rectified image

In Figure 2(c), parameter T_{pp} refers to point to point distance threshold. All tip points that satisfy this constraint must fall within the circle defined by the last classified point, i.e., Point 6 in this example, as its center and T_{pp} being its radius. Parameter T_{pl} refers to the point to line distance threshold. Accordingly, all tip points that satisfy this second constraint must lie within the band defined by two parallel lines L_b and L_t, which are parallel and with the same distance, T_{pl} to Line L. As we can see, the nearest point to the last classified point, Point 6, which satisfies the above two distance constraints is Point 7.

Two thresholds T_{pp} and T_{pl} must be chosen with care. The point to point threshold should be large enough to span the blank between words and the point to line threshold should be small enough to exclude tip points that belong to other text lines. We determine both thresholds based on the length of identified vertical stroke boundaries.

Generally, the number of groups classified is not ideally two times but more than that of the number of text lines. This may result from the noise or from some special characters such as "b", "q" and "g" whose tip points deviate far from the top or base line. We propose to use size filter to separate the desired point groups and fit horizontal lines using least square method. Figure 2(d) shows the fitted top line and base line.

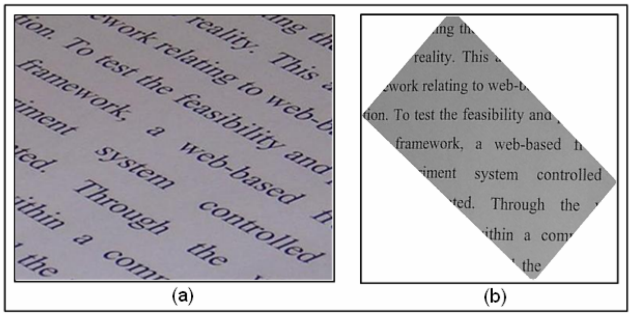

Figure 3 Rectification result: (a) distorted document image; (b) rectified document image

4. PERSPECTIVE DISTORTION CORRECTION

With the extracted vertical boundaries and fitted horizontal line, the feature map composed of some quadrilaterals can be constructed. Generally the endpoints of vertical boundaries can be chosen as the quadrilateral vertexes. But for the characters such as “b”, “q” and “f”, the boundary endpoints deviate far from the base line and top line, so they can’t be used for map construction. In our proposed approach, we fit the vertical lines based on extracted vertical boundaries and determine the quadrilateral vertexes as intersection of fitted horizontal and vertical lines. Figure 2(e) shows the constructed quadrilateral map.

The characters within the source quadrilaterals can be counted based on the connected component analysis. The blank between words is treated as a character as well, which can be detected based on the centroid distance between adjacent characters. Generally, character width-to-height ratio lies between 0.5:1 and 1:1. Therefore, we determine this ratio as 1:1 because blank between adjacent characters is calculated into the character width. With the approximated character width-to-height ratio and the counted character number, the target quadrilateral can be constructed.

For each pair of quadrilaterals, a rectification homography can be calculated with four pairs of quadrilateral vertexes based on four point algorithms [7]. Since tip points should be ideally mapped to a horizontal line, we choose the optimal homography as the one that minimizes the variance of y coordinate of rectified tip points. The rectification result with optimal homography is shown in Figure 2(f).

5. SUMMARY

We have tested document images captured using different digital cameras. The experiment results show that our proposed rectification technique can deal with images captured from different and distances and viewpoints. The sample image used in Sect.2, 3, and 4 contains only one text line. The reason of using this document image is to simplify description and save space. In fact, the rectification performance of document images with multiple text lines is better than that of ones with just one text line. Figure 3(a) shows one perspectively distorted document image with multiple text lines and Figure 3(b) shows the rectification result.

Characters with small size may affect the extraction of vertical boundary. Besides, they may affects target quadrilateral construction because connected characters make it hard to determine height-width ratio. In our proposed approach, average size of character is calculated based on connected component analysis. And if the average size is less than a pre-defined size threshold, nearest neighbor interpolation is utilized to enlarge the document image before feature extraction.

The proposed approach can be extent to correct geometric distortion as well. Preliminary experiment has been done on captured image where text lies on a rubbed paper surface. Based on the approximation that each character lie within a small plane patch, source & target mesh of document surface can be constructed and geometric distortion can be corrected using 2-pass image warping algorithm [8].

6. REFERENCES

[1] P. Clark and M. Mirmhedi, “Location and recovery of text on oriented surfaces,” *SPIE conference on Document Recognition and Retrieval VII*, pp. 267-277, 2000.

[2] P. Clark and M. Mirmhedi, “Rectifying perspective views of text in 3D scenes using vanishing points,” *Pattern Recognition* vol. 36, no. 11, pp. 2673–2686, 2003.

[3] M. Pilu, “Deskewing Perspectively Distorted Documents: An Approach Based on Perceptual Organization,” Technical Report, HPL-2001-100, Hewlett Packard, May 2001.

[4] L. O’Gorman, “Binarization and multithresholding of document images using connectivity,” *CVGIP: Graphical Models and Image Processing*, vol. 56 no. 6 pp. 494–506, 1994.

[5] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision,” *Cambridge University Press*, June 2000.

[6] H. J. Zimmermann, P. Zysno, “Latent connectives in human decision making”. *Fuzzy Sets and Systems,* vol. 4, pp. 37-51, 1980.

[7] L.A. Zadeh, “Calculus of Fuzzy Restrictions,” *Fuzzy Sets and Their Application to Cognitive and Decision Processes*, Academic Press, New York, pp. 1-39, 1975.

[8] G. Worlberg, “Digital Image Warping,” IEEE Computer Society Press, Los Alamitos, California, 1990.