SKEW DETECTION, SKEW CORRECTION, AND SHIROREKHA EXTRACTION OF LINE SEGMENTED OFFLINE HANDWRITTEN HISTORICAL MARATHI DOCUMENTS WRITTEN IN MODI SCRIPT

Parag Anil Tamhankar[1], Satish Ramesh Kolhe[2]

Abstract- The motivation for this work comes from the fact that there are large number of documents written using MODI script, and very few number of people can interpret and/or comprehend it. As of now, very less work is been reported on this script recognition. Skew detection and correction greatly simplifies character segmentation from a line. Shirorekha extraction plays a key role in subsequent character recognition phase. This work deals with skew detection, skew correction, and shirorekha extraction of line segmented offline Handwritten Historical Marathi Documents written in MODI Script. Broadly, the work is divided into three phases, viz., 1) skew detect, 2) skew correct, and 3) segmentation. In the first phase, skew angle is detected. In the second phase, skew is corrected using affine transformation and the angle obtained in the first phase. In the third phase, a novel algorithm is designed and developed for shirorekha extraction. The results obtained using this work have an accuracy of more than 90% although all the methods used in this paper works on spatial domain of the image.
Keywords – MODI Script, skew detection, affine transformation, skew correction, shirorekha extraction.

1. INTRODUCTION

The Modi script was used to write the Marathi language spoken in the Indian state of Maharashtra. It originated as a cursive variant of the script during the 17th century CE. Modi was used until the 1950's when Devanagari replaced it as the written medium of the Marathi language. Typical of South Asian scripts, a "consonant" Modi letter is in fact the consonant followed by the inherent vowel /a/. To represent a vowel other than /a/, additional strokes called matras are added to the basic letter.

Usually the same stroke form is used for the same vowel across all letters of the script, but in the case of Modi (and a few other South Asian scripts as well) different stroke forms exist to represent the same vowel in different letters. In the following example you can see how the vowels /a:/ and /u/ have different stroke forms for the letters /k/, /c/, and /gh/. [1].

Learning Modi-Script is useful to Academician, Historian, Researcher and Legal experts and also for knowing more about cultural, heritage preservation [2].

The motivation for this work comes from the fact that segmentation and recognition of MODI script characters is quite demanding, for the various reasons listed as follows: 1) Very small difference in the representation of characters like 'ja' and 'na', 2) In a MODI script document, if some name appears several times, then MODI script equivalent of the word 'uprokt' is used, 3) same representation is used for the characters 'laa' and 'ka', thereby, requiring semantic analysis of the text as well, 4) few MODI script characters are represented using Devnagari script, 5) sometimes a word spans across multiple lines, 6) no delimiter is used in between two successive words, and several others.

The layout of the paper is as given below:

Section II cites the earlier work done on image enhancement, binarization, segmentation and shirorekha extraction. Preprocessing steps are given in section III. The crux of the paper is described in section IV. Results obtained are discussed in section V. Section VI illustrates conclusion and future research directions. Finally the paper ends with acknowledgment and references.

2. PREVIOUS WORKS

This section illustrates the work completed earlier on image enhancement, binarization, segmentation and shirorekha extraction.

Reference [3] eliminates bleed-through effect present in the double-sided document images using wavelet transform. It also works well with single-sided document images, albeit with little lesser accuracy. Strokes of faint edges are preserved using a global classifier called as flow field.

A spatial domain noise removal filter is implemented in [4] which makes use fuzzy logic approach. The filter works in several iterations, with every iteration consisting of two phases and an adaptive fuzzy membership function. In the first

[1] Department of Computer Science, MES Abasaheb Garware College, Pune, Maharashtra, India
[2] School of Computer Sciences, North Maharashtra University, Jalgaon, Maharashtra, India

phase, it computes derivative along 8-neighbors of every pixel using fuzzy rules. In the second phase, this derivative is used to reduce noise from the image.

A local binarization method is proposed in [5] which works with small portions of an image, thereby, using different local threshold for every part of the image. It works well with various images having uneven brightness. It makes use of decision rules constructed using Support Vector Machine (SVM).

Reference [6] deals with image enhancement and binarization. It works well with images having varying degree of brightness. It also uses local binarization threshold by calculating background surface using interpolation and upsampling.

Reference [7] recovers the lost character shapes from binarized Japanese documents using Geodesic morphological approach, Quasi Euclidean distance transform, and Anisotropic smoothing.

A probabilistic context model is developed in [8] using Markov Random Field for the restoration of degraded document images. The approach present is independent of the script used in the text image.

A local adaptive image binarization technique is presented in [9] which makes use of Canny's edge detector and an adaptive contrast map. A new algorithm is developed for the estimation of edge width using histogram of original image and Corresponding binary text stroke edge image.

Reference [10] also deals with image binarization using Gaussian filter and Canny's edge detector. In the first stage, it performs global thresholding and then in the second stage, it makes use of local thresholding.

Gurumukhi script character segmentation work is described in [11] using Water Reservoir approach. It works well even though the adjacent characters are touching each other.

Reference [12] performs skew detection and correction using Radon transform and Hough transformation. It works in case of printed as well as handwritten documents and with non-uniform skew angle.

Reference [13] also deals with skew detection and correction for Urdu script documents. It uses statistical moments and image centroid.

Two different approaches are explained in [14] for skew angle detection. The first uses centroid of the selected words for skew detection, whereas, the later one uses morphological dilation and thinning for the same. The output of each of these approaches is later presented as an input to Hough transform.

Reference [15] presents a new algorithm for extraction of lines from documents having several skew angles a hypothetical water flow technique and morphological erosion.

Writing-style independent shirorekha Identification work is carried out in [16] using perceptive vision and a line segment extractor which is based on Kalman filtering on Bangla script document images.

3. PREPROCESSING STEPS

RGB Image Acquisition is done at a spatial resolution of 300 dpi. Then, the RGB image is converted to gray-scale, following which Image is resized to 1024x1024 pixels. Image binarization is carried out using a global threshold obtained using [17]. Image padding by 15x15 size window is carried out to take care of characters touching the border of image. Finally, line Segmentation work is accomplished using [18].

4. THE PROPOSED METHOD

4.1. Problem Statement

Let $f = f(v)$ be an output image obtained using the work done in [18], where $f = [i, j]^T$. The goal is to detect the amount of skew angle, correcting it using affine transformation, and to extract shirorekha from the line.

4.2 Mathematical Formulation

Let point p with coordinates (a, c_1) be the first nonzero pixel while scanning columns from left to right. This point p is considered to be the start of shirorekha. Similarly, point q with coordinates (b, c_2) be the first nonzero pixel while scanning columns from right to left. This point q is considered to be the end of shirorekha as shown in the given diagram.
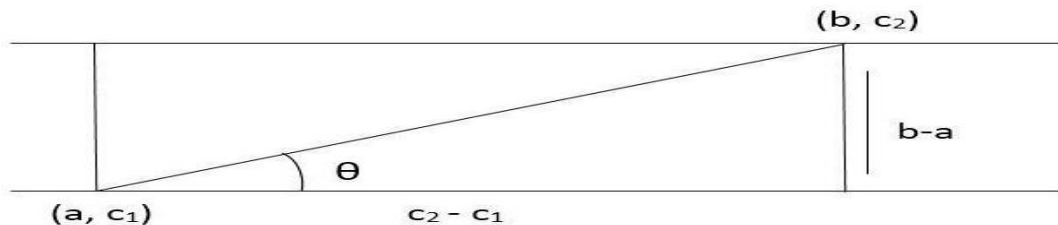
The skew angle θ of the shirorekha from the horizontal axis is given by

Therefore,   tan   =   2 − 1

                         = arctan   2 − 1

                    $0 \leq \theta \leq$

                                                            through angle θ using affine transformation. An affine
The skew is corrected by rotating shirorekha about point p   2
transformation is a transformation of Euclidean plane which is an isometry (norm/distance preserving transformation).
Basically, translation, rotation, and reflection are affine transformations. Affine transformations
which has exactly one fixed point are orthogonal transformations, viz., rotation and reflection. Orthogonal
transformations are linear and hence given by an orthogonal matrix. A matrix A is said to be orthogonal if
 . The determinant of orthogonal matrix is ±1. If determinant is +1, then it is a rotation, else it s reflection.   .   =
                                                            .   =
Let =   " is a rotation of 2-D Euclidean plane. Therefore, #! det   = 1.

                    &   +   &   =   1,
                    +   !   =   0   ,
                    #!   = 1, & + !& = 1)*   *+   ,
This   !   ! − = 1.   & + &   =
      gives   So,
                    Therefore,   ∃   -.   /
                    /
!   =   *-   =   !,
#!   = −.   ! = -0#□ = −.

Hence the rotation matrix R is given by cos   sin   .
                              − sin   cos   "
Now translation of Euclidean plane is not a linear transformation, therefore, it cannot be expressed by matrix multiplication.
As, affine transformation of any type must be function composition of translation and orthogonal transformation. To express
any affine transformation by matrix multiplication, homogeneous coordinates are required. The need for homogeneous
coordinates comes from the fact that a nonlinear system in 2-D (which is
;
difficult to solve) becomes linear (thus easily solvable) in 3-56D. To perceive96 a93-dimensional object in 2-dimensional
space,6a scaling9 factor6 h is 9added as a 3rd coordinate, $4 = 7,8, h$: $7, 8$ ∈ &, ∈ }.

Now   7,   8,>   ~   7@,8@,>@   0AA   7   =   7@   ,   8   =   8@   ,
#! $> = >@h$ is an equivalence relation on 4. The quotient space 4/~ is hyperplane in 4 with a scale $h = 1$.

7   +   6

                    1   0   , 7, 19

Therefore, translation by vector   is

written as C8 +   D = C0 1   D . E8F

1   0   0 1 1   *-   -0#   0

Similarly, The rotation matrix R in homogeneous coordinates can be expressed as C−-0#   *-   0D

                                    0   0   1

5. METHODOLOGY

The work is classified into three different phases, viz., 1) skew detect, 2) skew correct, and 3) shirorekha extraction. In the first phase, skew angle is detected using an angle θ as illustrated in the previous section. This angle θ determines the amount of skew present in the line. In the second phase, The skew is corrected by rotating the line in counterclockwise direction by an angle θ using affine transformation as explained in the previous section. In the third phase, a new algorithm is developed. This algorithm is based on the observation that only the rows representing shirorekha will have 200+ streak of consecutive foreground pixels. The algorithm finds such streak by keeping track of current largest streak of foreground pixels and the largest streak found so far along each row and remembers the newly found largest streak if current streak of foreground pixels is larger than the previously found largest streak along that row. It also remembers the length of the longest streak found so far in that row. It then identifies and extracts the shirorekha by scanning all the rows from left to right and by putting all the rows into background which contains 200 or more consecutive foreground pixels.

6. EXPERIMENTAL RESULTS

Figure 1 and 4 depict input images obtained after performing preprocessing steps as shown in section 3.

Figure 2 and 5 shows the results obtained after performing skew detection and skew correction using affine transformation on figure 1 and 4 respectively

Output of shirorekha extraction performed on figure 2 and 5 is shown in figure 3 and 6 respectively.

Fig. 1: Input Image 1

Fig. 2: Result of Skew Correction

Fig. 3: Result of Shirorekha Extraction

Fig. 4: Input Image 2

Fig. 5: Result of Skew Correction

Fig. 6: Result of Shirorekha Extraction

7. CONCLUSION AND FUTURE PROSPECTS

The results obtained using this work have an accuracy of more than 90% although all the methods used in this paper works on spatial domain of the image. The shirorekha extraction algorithm is based on the observation that only those rows containing shirorekha will have 200+ consecutive foreground pixels. Hence, such rows are put into the background, thereby, eliminating the shirorekha. Out of 216 samples, skew detection and correction was done accurately for 203 samples, giving an accuracy of 93.98%, whereas, shirorekha was accurately extracted for 195 samples, providing accuracy of 90.27%.

The future work involves use of frequency domain approach for improving the results obtained for shirorekha extraction, as algorithm also partly puts the foreground into the background wherever letter modifiers are involved. Furthermore, work will be carried out on character segmentation, feature selection/extraction followed by character recognition phase.

8. ACKNOWLEDGMENT

9. REFERENCES

[1]   http://www.ancientscripts.com/modi.html; Accessed as on 07th January, 2018.
[2]   https://cdac.in/index.aspx?id=lu_modi_script; Accessed as on 07th January, 2018.
[3]   Moghaddam, Reza Farrahi, and Mohamed Cheriet. "A variational approach to degraded document enhancement." Pattern Analysis and
[4]   Van De Ville, Dimitri, Mike Nachtegael, Dietrich Van der Weken, Etienne E. Kerre, Wilfried Philips, and Ignace Lemahieu. "Noise reduction by fuzzy image filtering." Fuzzy Systems, IEEE Transactions on 11, no. 4 (2003): 429-436.
[5]   Chou, Chien-Hsing, Wen-Hsiung Lin, and Fu Chang. "A binarization method with learning-built rules for document images produced by cameras." Pattern Recognition 43, no. 4 (2010): 1518-1530.
[6]   Gatos, Basilios, Ioannis Pratikakis, and Stavros J. Perantonis. "Adaptive degraded document image binarization." Pattern recognition 39, no. 3 (2006): 317-327.
[7]   Shirai, Keigo, Yuta Endo, Akihito Kitadai, Shingo Inoue, N. Kurushima, Hajime Baba, Atsuyori Watanabe, and Masaki Nakagawa. "Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology." In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pp. 1285-1289. IEEE, 2013.
[8]   Banerjee, Joydeep, Anoop M. Namboodiri, and C. V. Jawahar. "Contextual restoration of severely degraded document images." In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 517-524. IEEE, 2009.
[9]   Su, Bolan, Shijian Lu, and Chew Lim Tan. "Robust document image binarization technique for degraded document images." Image Processing, IEEE Transactions on 22, no. 4 (2013): 1408-1417.
[10] Biswas, Barun, Ujjwal Bhattacharya, and Bidyut B. Chaudhuri. "A Global-to-Local Approach to Binarization of Degraded Document Images." In Pattern Recognition (ICPR), 2014 22nd International Conference on, pp. 3008-3013. IEEE, 2014.
[11] Kumar, Munish, M. K. Jindal, and R. K. Sharma. "Segmentation of isolated and touching characters in offline handwritten gurmukhi script recognition." International Journal of Information Technology and Computer Science (IJITCS) 6, no. 2 (2014): 58.
[12] Kapoor, Rajiv, Deepak Bagai, and Tara Singh Kamal. "A new algorithm for skew detection and correction." Pattern Recognition Letters 25, no. 11 (2004): 1215-1229.
[13] Ramteke, R. J., Imran Khan Pathan, and S. C. Mehrotra. "Skew angle estimation of urdu document images: a moments based
[14] approach." International Journal of Machine Learning and Computing 1, no. 1 (2011): 7.
[15] Nandini, N., K. Srikanta Murthy, and G. Hemantha Kumar. "Estimation of skew angle in binary document images using hough transform." World Academy of Science, Engineering and Technology 18 (2008): 44-49.
[16] Basu, Subhadip, Chitrita Chaudhuri, Mahantapas Kundu, Mita Nasipuri, and Dipak Kumar Basu. "Text line extraction from multi-skewed handwritten documents." Pattern Recognition40, no. 6 (2007): 1825-1839.
[17] Lemaitre, Aurélie, Bidyut Baran Chaudhuri, and B. Couasnon. "Perceptive vision for headline localisation in bangla handwritten text recognition." In Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, vol. 2, pp. 614-618. IEEE, 2007.
[18] De Silva, D. V. S. X., W. A. C. Fernando, H. Kodikaraarachchi, S. T. Worrall, and A. M. Kondoz. "Adaptive sharpening of depth maps for 3D-TV." Electronics Letters 46, no. 23 (2010): 1546-1548.
[19] Tamhankar, Parag A., Krishnat D. Masalkar, and Satish R. Kolhe. " Segmentation of Overlapping, Skewed, and Touching Lines in Handwritten MODI Script." International Journal of Computer Science & Applications 14, no. 2 (2017).