Localization of Handwritten text in Documents using moment invariants and Delaunay Triangulation

Kandan Ramakrishnan, Arvind KR and AG Ramakrishnan
Department of Electrical Engineering, Indian Institute of Science,
Bangalore 560 012, INDIA
kandan.r@gmail.com,arvind.kr@gmail.com,ramkiag@ee.iisc.ernet.in

Abstract

This paper describes an approach based on Zernike moments and Delaunay triangulation for localization of hand-written text in machine printed text documents. The Zernike moments of the image are first evaluated and we classify the text as hand-written using the nearest neighbor classifier. These features are independent of size, slant, orientation, translation and other variations in handwritten text. We then use Delaunay triangulation to reclassify the misclassified text regions. When imposing Delaunay triangulation on the centroid points of the connected components, we extract features based on the triangles and reclassify the text. We remove the noise components in the document as part of the pre-processing step so this method works well on noisy documents. The success rate of the method is found to be 86%. Also for specific hand-written elements such as signatures or similar text the accuracy is found to be even higher at 93%.

1 Introduction

Most document images invariably consist of a mixture of machine printed elements such as logos, text, barcodes etc and handwritten elements such as address/name texts, signatures, markings etc. There are various applications in which the separation of handwritten text from machine printed text is necessary. Fan et al. [1] have used spatial features and character block layout variance as the prime features for classification of machine-printed and handwritten texts. Guo and Ma [2] have propose a scheme which combined the statistical variations in projection profiles with hidden Markov models (HMMs) to separate the handwritten material from the machine printed text. Imade et al. [3] have extracted the gradient and luminance histogram of the document image and used a feed forward neural network in their system. This is a method to segment a Japanese document into machine-printed Kanji and Kana, handwritten Kanji and Kana, photograph and printed image. Kuhnke et al. [4] developed a method for the distinction between machine-printed and handwritten character images using directional and symmetrical features as the input of a neural network. Arvind et al.[5] have used Horizontal Projection Profiles, Fisher Profiles and Eigen Profiles for separating the printed and hand-written text blocks. Pal and Chaudhuri [6] have used horizontal projection profiles for separating the printed and hand-written lines in Bangla script. In this paper we use Zernike moments, that are insensitive to translation, scale and rotation as the feature for distinguishing the printed and handwritten elements. Then we go on to use the features of the traingles obtained from Delaunay triangulation to reassign the labels assigned to the elements. As a preprocessing step we clean the document image of all the noise components present.

2 Localization of Text Algorithm

2.1 Preprocessing

In this stage we remove the noise elements and the process is described below.

- The area of each connected elements(A_i) in the printed document is determined with the help of Connected Component Analysis (CCA). We also find out the maximum area(A_{max}) and minimum area(A_{min}) in the entire document.
- If one of the following conditions is met then it indicates noise in the document and it is removed.
  - If the value of (A_i- A_{min})/(A_{max} - A_{min}) $<$ 0.002
  - If the aspect ratio of the Bounding Box is greater than 33/1, then it indicates a horizontal line and if it is less than 1/10 it indicates a vertical line and we remove it.
  - Also if the height or width of the connected component is less than 7, then it is a spurious element or a dot, and we remove it.

These thresholds which are used have been chosen empirically.

2.2 Feature Extraction

We introduce Zernike moments [7] as features for discrimination of handwritten text which are invariant to scale, translation and invariance. However, the Zernike moments that are used are only rotation invariant. To obtain scale and translation invariance, the image is first subjected to a normalization process using its regular moments. The rotation invariant Zernike features are then extracted from the scale and translation normalized image.

The geometrical moments for an image f(x,y) are calculated as follows;

$$m_{pq} = \sum_x \sum_y (x)^p (y)^q f(x,y) \tag{1}$$

Since the defined features by means of Zernike moments are only rotation invariant, to obtain the scale and translation invariance, an image must be normalized via image normalization. This scale and translational invariance are obtained by transformation of the image[] to g(x,y) which is given by

$$g(x,y) = f(\frac{x}{a} + \bar{x}, \frac{y}{a} + \bar{y}) \tag{2}$$

where $\bar{x} = \dfrac{M_{10}}{M_{00}}$ and $\bar{y} = \dfrac{M_{01}}{M_{00}}$ and $a = \sqrt{\frac{\beta}{m_{00}}}$, β a predetermined value.

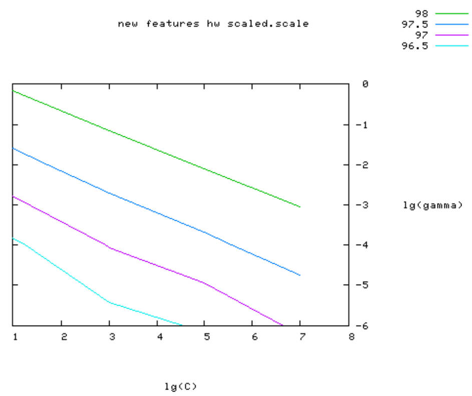


Figure 1. Graph showing the cross-validation results where γ is varied between 1 to 2_{-10}, C is varied between 1 to 2_{+10} and the best pairs of (γ, C) are choosen

Thus g(x,y) is the normalized image function with respect to translation and scaling. The Zernike moments of order n with repetition l for a continuous image function f(x, y) is calculated as

$$A_{nl} = \frac{n+1}{pi} \sum_x \sum_y f(x,y) V^*_{nl}(\rho, \theta); x^2 + y^2 <= 1 \tag{3}$$

where $V_{nl}(\rho, \theta)$ is a set of complex polynomials which form a complete orthogonal set over the interior of a unit circle.

2.3 Classification

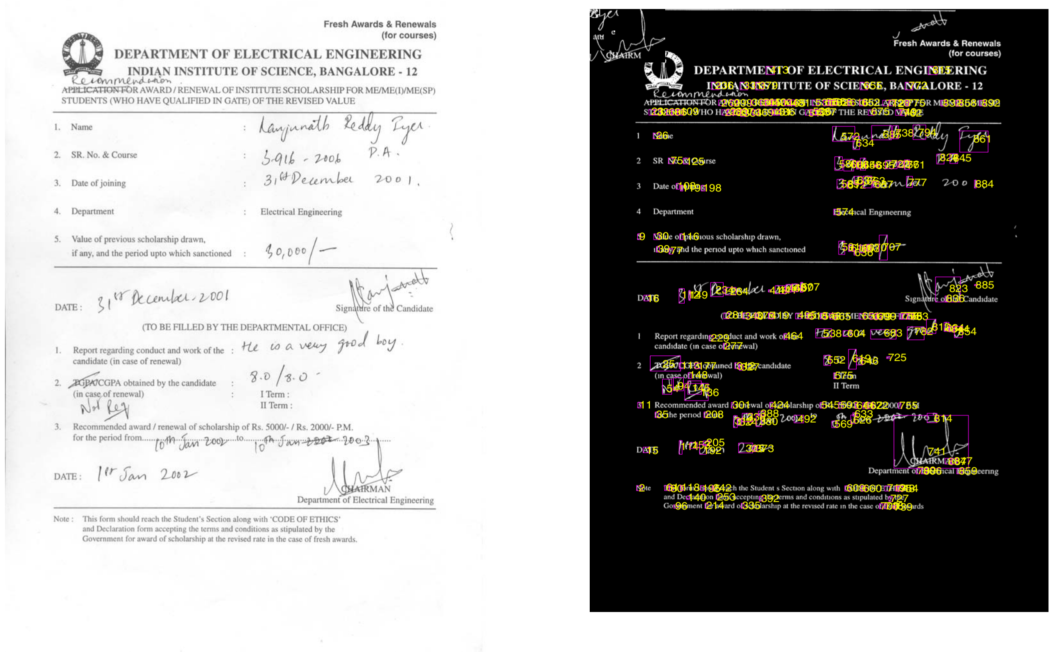
We trained an SVM classifier with Radial Basis Function(RBF) kernel[8] with the invariant moments as the features. Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. C $>$ 0 is the penalty parameter of the error term. The RBF kernel used is given below

$$K(x_i; x_j) = exp(-\gamma(||x_i - x_j||)^2, \gamma > 0 \tag{4}$$

The kernel parameters were chosen by using a five fold cross validation using all the training data samples. The best parameters were found to be C= 64 and $\gamma = 0.125$. Figure 2(a) and 2(b) depict the input image and the output after classification using zernike moments as features. In Figure 2(b), all those elements that have a class value of two, which represent handwritten text are marked by a magenta Bounding Box.

2.4 Reclassification

We use Delaunay triangulation[9] to reassign the labels that have been already assigned to the misclassified elements by the nearest neighbour classifier as shown in the figure. We briefly give the definition of Delaunay triangulation. Delaunay triangulation of a set of non-degenerate vertices V is defined as the unique triangulation with empty circles, i.e., no vertex lies inside the circumscribing circle of any Delaunay triangles, as follows:



(a) Input Document

(b) Labelling of the Document after Zernike moment is done

Figure 2.

$$DT(V) = (p_i.p_j, p_k)\epsilon V^3, B(p_i.p_j, p_k) \cap V \setminus (p_i.p_j, p_k)\epsilon\phi \tag{5}$$

where B($p_i; p_j; p_k$) is the circle circumscribed by the three vertices $p_i; p_j; p_k$ that form a Delaunay triangle. It is seen that after Delaunay triangulation is carried out on printed text/handwritten text these regions have the following features:

- The height of the triangles in a printed text region are not longer than a threshold as compared to the height of the handwritten text.
- Most triangles in the printed text have their shortest side in the direction of the text line and the longest sides link the point pairs between two adjacent text lines while the handwritten text have their shortest side inclined at an angle to the text line.

Based on the above features we re-classify the text regions based on the similarity of the delaunay triangles with its neighbours. If a particular element has similar features with more than 50% of its neighbouring elements and the element in consideration is labelled differently then it is given the same label of the neighbours. The reclassification is done as per the following algorithm:

- Delaunay triangulation is done on the document with the centroid points as shown in Figure 3(a).
- Now, let us consider a centroid point P(x,y). From the point P a number of traingles originate and the point P is associated with other centroid points P_1, P_2 and P_n by

the Delaunay triangles. These points P_1, P_2 and P_n are said to be the neighboring points of P.

- We now compare the label of each neighbouring point with the label of the centroid point P(x,y). If the label is not the same, then we find the difference between the heights of the element defined by the neighboring centroid point and the height of the element defined by the reference point P(x,y). This is done for the point P with all its neighbouring points.
- If with more than half of the neighbouring points the diference in height is less than a threshold value then we reassign the label i.e 1 as 2 or 2 as 1. Let us assume the number of centroid points with similar height as C_i and the total no of neighbouring points as N_i for the centroid point i then if the value of $\frac{C_i}{N_i} * 100 > 50$ - the label is reassigned else the label remains the same. This is then done for all the centroid points.
- This means that if the centroid point P, has different text as compared to the neighboring points, then the height difference will be greater than 7 pixels and no reclassification is required. However if the centroid point P, has similar text as compared to the neighboring points and is misclassified we compare the height feature of the triangles. If the feature is same(i.e if the difference in height is less than 7 pixels), with more than 50% of the neighboring points then the point P(x,y) is given the label of the neighboring points and is re-classified. If the point P has the same label as that of its neighbors then the above steps are not required and the above algorithm is done for the next point.

Figure 3(b) shows the image of the document with the handwritten elements localized within a magenta colored bounding box after the labels are reclassified.

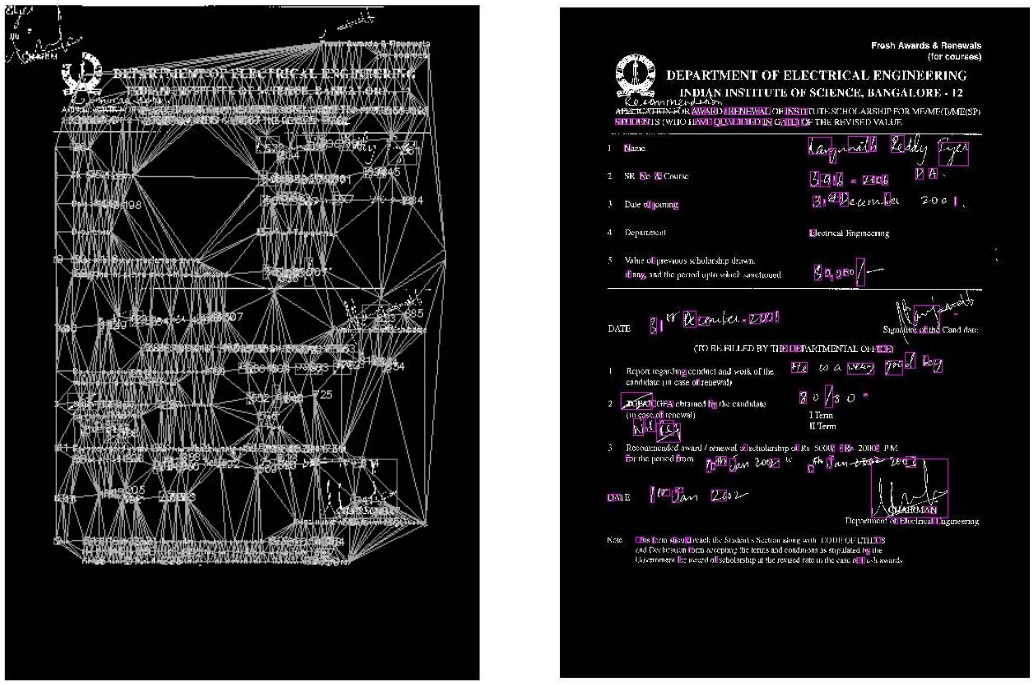
3 Experimental Results

3.1 Data Description

The training data consists of 30000 patterns in all, out of which 20000 patterns are machine printed and the remaining are handwritten elements including both cursive and block handwriting besides signatures, dates and address locations. Our test data consists of 150 English document images, scanned at 200 dpi and stored in 1-bit depth monochrome format. These documents contain handwritten elements, signatures, logos along with free-flowing text paragraphs. Our testing also contains documents which have only signatures as part of their handwritten components. It is observed that we obtain a higher accuracy on such type of documents.

3.2 Accuracy Calculation

Table (1) shows the classification accuracy using the proposed method. It observer that the accuracy value is higher when the document consists of handwritten text such as signatures or similar text.



(a) Document on which Delaunay Triangulation is carried out

(b) Document with the Handwritten elements marked in a magenta bounding box

Figure 3.

Table 1. Classification accuracy

	No of Blocks	Correctly Classified	% Accuracy
Handwritten text	1,678	1,441	85.85%
Handwritten text which consist of Signatures	360	332	92.22%

4 Conclusion

We have described a method which is invariant and robust for discrimination of handwritten text in printed documents. The miscalssification accuracy is reduced with the use of Delaunay triangulation to reclassify the text. This is found to give us good and accurate results. The text can appear in any part of the document, even overlapping the machine printed text and is still classified accurately.

There are certain limitations to this method. It fails in some specific cases when the handwritten text has features similar to machine printed text. Also in some cases when the handwritten text is in blocks, it extracts only part of the handwritten text. We also have not tried this algorithm on other fonts except English so this method can be further explored in such situations.

References

[1] K. Fan, L. Wang and Y. Tu, Classification of Machine-Printed and Hand-Written Texts Using Character Block Layout Variance, Pattern Recognition, Vol. 31, No. 9, 1998, pp. 1275-1284.

[2] Jinhong K. Guo and Matthew Y. Ma, Separating handwritten material from machine printed text using hidden markov models, in Proceedings of the International Conference on Document Analysis and Recognition, 2001, pp. 439443.

[3] S. Imade, S. Tatsuta and T. Wada, Segmentation and Classification for Mixed Text/Image Documents Using Neural Network, Proceedings of 2nd ICDAR, 1993, pp.930-934.

[4] K. Kuhnke, L. Simonicini and Z. Kovacs-V, A System for Machine-Written and Hand-Written Character Distinction,Proceedings of 3rd ICDAR, 1995, pp. 811-814.

[5] Arvind.K.R, Peeta Basa Pati, A.G.Ramakrishnan Horizontal Projection Profile for extracting printed text paragraphs from document images IEEE- International conference on Signal and Image Processing (ICSIP), India, 2006.

[6] U. Pal and B. Chaudhuri, Automatic Separation of Machine-Printed and Hand-Written Text Lines, Proceedings of 5th ICDAR, Bangalore, India, 1999, pp. 645-648.

[7] M. R. Teague, ”Image analysis via the general theory of moments”, J. Opt. Soc. Amer., vol. 70, pp. 920-930, Aug. 1980.

[8] Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[9] F. Davoine, et al., Fractal image compression based on Delaunay triangulation and vector quantization, IEEE Trans. Image Process. 5 (2) (1996) 338346