2009 10th International Conference on Document Analysis and Recognition

Off-Line Multi-Script Writer Identification using AR Coefficients

Utpal Garain
Indian Statistical Institute
203, B.T. Road, Kolkata 700108, India
utpal@isical.ac.in

Thierry Paquet
UFR des Sciences, Université de Rouen
76800 Saint-Etienne du Rouvray, France
Thierry.Paquet@univ-rouen.fr

Abstract

The problem of writer identification in a multi-script environment is attempted using a two-dimensional (2D) autoregressive (AR) modelling technique. Each writer is represented by a set of 2D AR model coefficients. A method to estimate AR model coefficients is proposed. This method is applied to an image of text written by a specific writer so that AR coefficients are obtained to characterize the writer. For a given sample, AR coefficients are computed and its L_2 distance with each of the stored (writer) prototypes identifies the writer for the sample. The method has been tested on datasets of two different scripts, namely RIMES containing 382 French writers and ISI consisting of samples from 40 Bengali writers. Modelling of writing styles using different context patterns at different image resolution has been investigated. Experimental results show that the technique achieves results comparable with that of the previous approaches.

1. Introduction

This paper addresses the problem of writer identification based on scanned images of handwriting. Though the research in this field is quite old [1] but in the recent years, there is a renewed interest in this field. The paper in [2] presents a nice survey of the significant approaches so far proposed for the writer identification using offline handwriting. Common among these methods is to use pattern recognition principles, i.e. extraction of a set of features from the handwriting of known writers and then based of these features classify the writer of an unknown sample as one of the known writers.

The features used for this purpose are broadly divided into two categories: texture based features and allograph based features. Among the texture level features contour based features, run-length based features, auto-correlation, etc. have been considered. On the other hand, allograph level features are extracted by segmenting the text into lines, words, characters, graphemes, etc. One significant technique is to cluster the graphemes and generate shape codebook to characterize a writer. Experimental datasets have considered variable number of writers starting from a few writers (e.g. 20-50 writers) up to one thousand writers. Emphasis has been put to investigate the writer identification capability in text dependent and text independent environment.

One prevailing trend as one can notice in these studies is to deal with only one script (mostly handwriting in Roman script) at a time. How does a particular method behave with changing of the script is not well researched. Ability of a method to deal with samples from writers writing in different scripts has also not been studied. The use of the allograph level features requires knowledge in a particular script, i.e. how to segment word into characters or graphemes, etc. and therefore, extension of the method based on allograph level features is not straightforward to tackle multi-script environment where writers may write in different scripts. Use of contour based features also requires knowledge of character contour. On the other hand, the texture based features like run-length, auto-correlation, etc. may tackle a multi-script scenario as the extraction of these features does not need extensive knowledge about a script. Unfortunately, these features perform so poorly that it is difficult to develop any writer identification system based on them [2].

This study extends the research in writer identification to multi-script environment. A non-causal type of two-dimensional model, called the two-dimensional autoregressive (AR) model, is introduced for characterization of writers. Motivation behind using such model comes from its successful application in areas including texture analysis [3]. No script specific knowledge is required to use this model that provides a viable framework to deal with writers writing in more than one script. The rest of the paper presents the proposed method, experimental results and related discussion.

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.222

2. Two-dimensional autoregressive model

A discrete image defined on an $M \times N$ (say, $P = M \times N$) rectangular grid is denoted by $\{x_{ij}\}$ (i = 1, 2, …, M; j = 1, 2, …, N). When each element x_{ij} is a random variable, $\{x_{ij}\}$ is called a discrete random field. In this paper, we deal with a class of random linear equation,

$$x_{ij} = \sum_{(p,q \in D)} \theta_{pq} x_{i-p,j-q} \qquad (1)$$

where D denotes the context region. Normally (but not necessarily) D is represented by a rectangular region as

$$D = \{(p,q) \mid -m \le p \le m, -n \le q \le n, (p,q) \ne (0,0)\} \qquad (2)$$

θ_{pq} is the AR model coefficients and $p \times q$ is the order of the model. So value of each pixel (say, $y = x_{ij}$) is predicted as a linear combination of D neighboring pixels. So in general we can write,

$$y = h\,\theta \qquad (3)$$

where, y is $P \times 1$ dimensional, h is a $P \times D$ dimensional matrix and θ is $D \times 1$ dimensional. So y records value of each of the P pixels in the image. For each of these P pixels, values of the D neighboring pixels are recorded in each row of h.

2.1. Estimation of AR coefficients

Next, our problem is to estimate θ. We present this estimation by following a bra-ket notation[1]. Let e denote the error in predicting the value of pixel, y. So we can write,

$$e = y - h\theta \qquad (4)$$

The squared error is defined as,

$$J = \left\{e^2\right\} \qquad (5)$$

So,

$$J = \langle y - h\theta \mid y - h\theta \rangle$$
$$= \langle y \mid y \rangle - \langle y \mid h\theta \rangle - \langle \theta h^T \mid y \rangle + \langle \theta (h^T h)\theta \rangle$$
$$dJ = -2 \langle yh \mid d\theta \rangle + 2 \langle \theta \mid h^T h \mid d\theta \rangle$$

To minimize J, dJ is set to 0, i.e.

$$\langle yh \mid \; = \; \langle \hat{\theta} \mid h^T h \mid$$
$$\Rightarrow \quad (h^T h) \mid \hat{\theta} \rangle = h^T \mid y \rangle \qquad (6)$$
$$\Rightarrow \quad \mid \hat{\theta} \rangle = (h^T h)^{-1} h^T y \rangle$$

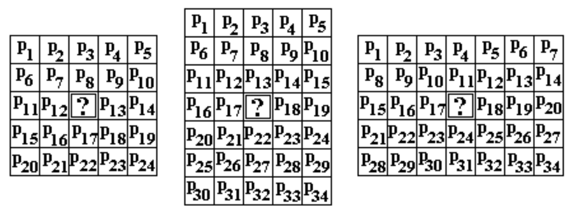
So estimation of θ requires matrix multiplications, transpose and inversion operations. This solution is simply the least mean square solution of equation (3).

[1] Bra-ket notation is a standard notation for describing quantum states in the theory of quantum mechanics composed of angle brackets and vertical bars.

2.2. Implementation issues

If the AR model is applied to each pixel of an image, implementation may require large computation even for a reasonable sized image. Therefore, instead of computing AR model in all pixels in binary images, we compute the model only in black pixels. This drastically reduces the computation requirement. It may be noted that in scanned images of handwritten text only about 3% pixels or less are black.

Regarding the context, we use three different templates of neighboring pixels as shown in fig. 1. Let C_1, C_2, and C_3 denote these three contexts in fig. 1(a), (b) and (c), respectively. Lengths of these three contexts are 24, 34, and 34 bits. Therefore, when they are used in computing AR model they result in coefficients of 24-order, 34-order, and 34-order AR models.



(a) (b) (c)

Figure 1: Different contexts used in estimating AR coefficients: (a) C_1: 5x5 context (b) C_2: 7x5 context and (c) C_3: 5x7 context.

2.3. Writer identification

AR coefficients computed from an image written by a specific writer characterize that writer. Say, there are w writers; each of them contributes one sample. Let $\hat{\theta}_i$ be the estimated AR model coefficients for the i-th writer. For an unknown sample, at first the AR model coefficients are computed. Let $\hat{\theta}$ be the estimated coefficients for this sample. Next, the Euclidean distance between this sample and any of the N samples of the reference database is computed as follows:

$$d(\hat{\theta}, \hat{\theta}_i) = \left\| \hat{\theta} - \hat{\theta}_i \right\|^2 \qquad (7)$$

It is decided that the given sample is written by the j-th writer if $d(\hat{\theta}, \hat{\theta}_j) > d(\hat{\theta}, \hat{\theta}_i) \forall i, i \ne j$.

Low-resolution writer identification: The above identification works at the original image resolution. Further investigation is done at lower resolution of an image. Resolution reduction is done by following the method adopted in JBIG [4]. The basic idea is to group four high-resolution pixels (in every 2×2 block of pixels) into one low-resolution pixel. The problem in this down sampling is to determine the value of that pixel when two pixels are black and the other two are white. JBIG resolves this problem by using a block of 4×4 pixels as shown in fig. 2. In this figure, the pixels marked with '**a**' to '**i**' correspond to higher resolution pixels and '**A**', '**B**', '**C**' and '**X**' are four lower resolution pixels corresponding to the block of pixels. Value of '**A**', '**B**', and '**C**' are already found and the value of '**X**' is to be computed. For this purpose, the JBIG specification provides a table, which is formed by computing the following expression:

$$4\mathbf{e} + 2(\mathbf{b} + \mathbf{d} + \mathbf{f} + \mathbf{h}) + (\mathbf{a} + \mathbf{c} + \mathbf{g} + \mathbf{i}) - 3(\mathbf{B} + \mathbf{C}) - \mathbf{A}$$

If the value of this expression is greater than 4.5, the value of the pixel '**X**' is decided to be 1. Since an image is divided into groups of 4×4 pixels, it should have an even number of rows and columns. If and when necessary *edge conventions* specified in JBIG are used.

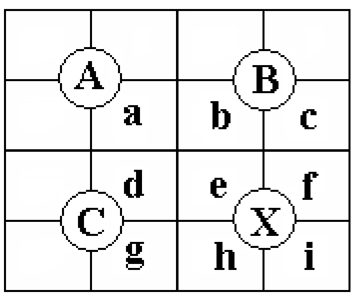


Figure 2: Resolution reduction.

AR model is computed in the same way as it is done at the original image resolution. All the three context patterns as shown in Fig. 1 are also used at lower resolutions. Performance of AR coefficients in identifying writers is presented in the next section.

3. Experimental results

Two datasets are used to conduct the experiment. First dataset is known as RIMES [5]. The training set used for writer identification task consists of 382 writers each contributed a letter containing reasonable amount of text. The letters are written in French. The test set used here consists of 100 samples. Test samples contain smaller (one-third or less) amount of text than that of the training samples. Each of the writers contributing these test samples has also given training samples. Therefore, ideally there should not be any rejection while identifying the writer for a given test sample. The second dataset has been developed at the Indian Statistical Institute (ISI), India. ISI developed dataset consists of samples from only 40 native Bengali writers. Each writer has contributed two samples; each containing different text. One sample per writer is used in the training set and another sample forms the test dataset. Both the training and test samples are of comparable size. Samples contain about 200 words or more. Earlier, this dataset was used for handwriting recognition purpose [6].

At first, writer identification performance is tested on RIMES dataset. Gray images are converted into binary ones. Results are shown in Table-1. Effectiveness of using three different contexts as shown in fig. 1 are investigated separately. It is observed that bigger contexts outperform smaller one, e.g. 34-order AR coefficients perform better than 24-order coefficients. Identification results in lower resolution are inferior as reported in Table-1. Therefore, further experimental results are only reported on the original image resolution, i.e. 300dpi. Interesting to note that top-1 results are not impressive (at best 57%) but the accuracy rapidly increases to a significant level (at best 97%) when top 10 choices are considered. To compare this performance with an existing technique, we consider the study in [7]. The identification method in [7] uses allograph level features and when tested on RIMES, it achieves an accuracy of 73% when top choice is considered. However, the accuracy is increased to only 84% when top 10 choices are taken into account.

This reveals the potential of the proposed method for writer identification. Unlike an allograph-based technique (e.g. a grapheme-based method [7]), it does not use any knowledge about the script but shows a power of identification that is comparable with that of a technique using allograph-level features. Script independence of the proposed method is verified on ISI Bengali dataset. Table-2 reports identification results on ISI Bengali dataset. Compared to RIMES, identification accuracies are slightly better for ISI dataset. Two important aspects can be attributed for this: dataset consists of only 40 writers and the test samples are of considerable sizes (text contains 200 or more number of words). A test sample containing adequate handwritten data helps in properly estimating the AR coefficients. In case of RIMES samples, training samples contain enough handwritten text but handwritten contents in test samples are quite small. Moreover, number of writers in RIMES is 382.

The capability of the method in handling multi-script environment is tested by mixing the RIMES and ISI samples together. Therefore, number of writers in this mixed dataset becomes 422 (382 French and 40 Bengali writers). Test set contains 140 samples (100

French and 40 Bengali). The identification results are reported in Table-3. Accuracies are 61% and 95% corresponding to the consideration of only the top choice and the top 10 choices. This clearly shows that multi-script handling capability of the method. The identification performance is comparable to the results obtained for a single script.

Next, the results obtained using the three context patterns are integrated through voting method. Table-4 presents the results obtained after combining the results achieved by three different pixel templates. It is noticed that identification accuracies are improved due to this combination. When individual results are integrated, the accuracy is increased by 1% to 2% at different number of top choices.

Table-1: Writer identification results on RIMES using different context patterns at different resolutions.

Context Type (Refer Fig. 1)	Image Resolution	Recognition Results (% correct) on Rime Dataset # Writers: 382, # Test samples: 100					
		Top 1	Top 2	Top3	Top4	Top 5	Top 10
C_1	300 dpi	55	68	70	70	75	90
	150 dpi	50	59	62	63	64	72
C_2	300 dpi	57	66	75	79	79	92
	150 dpi	47	58	62	64	65	69
C_3	300 dpi	57	62	70	75	77	97
	150 dpi	48	55	57	59	68	73

Table-2: Writer identification results on ISI dataset using different context patterns at original resolution.

Context Type (all at 300 dpi)	Recognition Results (% correct) on Bengali Dataset # Writers: 40, # Test samples: 40					
	Top 1	Top 2	Top3	Top4	Top 5	Top 10
C_1	72.5	77.5	82.5	85	85	95
C_2	75	82.5	87.5	87.5	90	97.5
C_3	75	80	82.5	85	87.5	100

Table-3: Writer identification results on RIMES+ISI dataset.

Context Type (all at 300 dpi)	Recognition Results (% correct) on Mixed Dataset # Writers: 422, # Test samples: 140					
	Top 1	Top 2	Top3	Top4	Top 5	Top 10
C_1	58.6	70.1	71.4	72.1	75	88.6
C_2	60.7	69.3	76.4	79.3	79.3	90.7
C_3	60.7	65.7	71.4	75.7	77.1	95

Table-4: Writer identification results on RIMES+ISI dataset after classifier combination.

Recognition Results (% correct) on Mixed Dataset # Writers: 422, # Test samples: 140					
Top 1	Top 2	Top3	Top4	Top 5	Top 10
62.1	70.7	77.9	80.7	81.4	96.4

4. Conclusions and future work

A writer identification method using two-dimensional auto-regression model is presented. The potential of the method is demonstrated by applying it to two publicly available datasets. The proposed method contributes substantially for writer identification in a multi-script environment. The capability of the method to work in an environment is verified by using a mixed dataset of samples written in two different scripts.

Several others aspects would be investigated in future extension of this study. For example, only the Euclidean distance is used at present to find the best match. Performance of a cosine distance would be an interesting investigation. Moreover, the AR coefficients are at present used to characterize the writers. The estimated error map for each writer can also be considered as likelihood criteria of the model fitting to the writers.

The results reported in this paper do not consider any rejection. However, using the distance measure (the Euclidean or the cosine distance) a rejection criteria can be found and its performance can be tested in future. The RIMES dataset does provide separate datasets (validation as well as test) for this investigation, i.e. writer identification with rejection.

The choice of a right context pattern for computing the AR model is another issue that needs further investigation. Here it is observed that higher order AR coefficients perform better than lower order coefficients. However, finding a right context pattern maximizing the identification accuracy may need addition of more sophisticated technique that can be investigated in future extension of the present study.

The working principle of the proposed method indicates another significant capability of the approach. The size of the text content in RIMES test samples is significantly smaller than that of in ISI test samples. Irrespective of this difference, the comparable performances obtained in these datasets shows that the method can implement a way of progressive writer identification. The system need not go through the entire content of the test sample at a time. It may provide an approximate result after reading some portion of the test sample and then the result would be gradually refined as more and more content of the test sample are processed. Viability of such a framework requires further extension of the proposed method.

Acknowledgement: The authors sincerely thank the developers of the RIMES database for freely sharing the resource with the authors.

References

[1] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification—The State of the Art," Pattern Recognition, vol. 22 (2), pp. 107-131, 1989.

[2] M. Bulacu and L. Schomaker, "Text-Independent Writer Identification and Verification Using Textural and Allographic Features," in IEEE Trans. Pattern Anal. Mach. Intell. Vol. 29(4), pp. 701-717, 2007.

[3] K. Deguchi, "Two-dimensional Auto-regressive Model for Analysis and Sythesis of Gray-level textures," in Proc. of the 1st Int. Sym. for Science on Form, General Ed. S. Ishizaka, Eds. Y. Kato, R. Takaki, and J. Toriwaki, pp. 441-449, Tokyo, Japan, 1986.

[4] ITU-T recommendation T.82, "Information Technology - Coded representation of picture and audio information-Progressive bi-level image compression," March 1993.

[5] E. Grosicki, M. Carré, J. Brodin, and E. Geoffrois, "RIMES evaluation campaign for handwritten mail processing,", in Proc. of ICFHR, 2008.

[6] A. Lemaitre, B.B. Chaudhuri, and B. Coüasnon, "Perceptive Vision for Headline Localisation in Bangla Handwritten Text Recognition," in Proc. of ICDAR, pp. 614-618, 2007.

[7] A. Bensefia, T. Paquet, and L. Heutte, "A Writer Identification and Verification System," Pattern Recognition Letters, vol. 26 (10), pp. 2080-2092, 2005.