

Integration of Contextual Information in Online Handwriting Representation

Sara Izadi and Ching Y. Suen

Concordia University, Montreal, Canada H3G 1M8
s_izadin@cs.concordia.ca, suen@cs.concordia.ca

Abstract. Robust handwriting recognition of complex patterns of arbitrary scale, orientation and location is yet elusive to date as reaching a good recognition rate is not trivial for most of the application developments in this field. Cursive scripts with complex character shapes, such as Arabic and Persian, make the recognition task even more challenging. This complexity requires sophisticated representations and learning methods, and comprehensive data samples. A direct approaches to achieve a better performance is focusing on designing more powerful building blocks of a handwriting recognition system which are *pattern representation* and *pattern classification*. In this paper we aim to scale up the efficiency of online recognition systems for Arabic characters by integrating novel representation techniques into efficient classification methods. We investigate the idea of incorporating two novel feature representations for online character data. We advocate the usefulness and practicality of these features in classification methods using neural networks and support vector machines. The combinations of proposed representations with related classifiers can offer a module for recognition tasks which can deal with any two-dimensional online pattern. Our empirical results confirm the higher distinctiveness and robustness to character deformations obtained by the proposed representation compared to currently available techniques.

1 Introduction

Handwriting is a natural way for humans to communicate and exchange information. Online handwriting recognition is used in the context of user interfaces for computing devices. For a variety of small devices such as hand-held computers and personal digital assistants (PDAs), keyboard-based data entry is not convenient. The ease of using a digital pen for writing also suits some larger devices such as tablet PCs and smart boards, which are designed for easy note taking in business, oee, clinic, and industry environments. The recognition of different forms of pen input such as mathematical equations and graphics, or of entire page layouts have attracted a lot of research attention. Interest in online handwriting recognition has increased due to market demand for both improved performance and for extended supporting scripts for those devices.

Handwriting recognition has still remained an active area of research to date [3, 5]. This is because of the following main challenges: (1) low recognition rate

in general and only a reasonable recognition rate on highly restrictive classes of patterns which do not represent real handwriting data; (2) lack of enough data sets, especially for scripts such as Persian and Arabic; and (3) long recognition time (a few seconds). The main sources of difficulty in the recognition of handwriting patterns are variation and variability. Variation refers to the unique way of writing of each individual, while variability addresses the changes of a specific individual in producing different samples at different times under different conditions and moods. In online systems, variations in the number and the order of strokes introduce more variations in handwriting and should be taken into account. While not all of these subject-specific parameters are feasible to consider and carry useful information about the character identity, it is not still known which part of this information provides a parsimonious representation for recognition. There exist huge variability in Persian/Arabic scripts due to the facts that a letter can appear in up to four different forms: isolated, initial, medial, and final; there exist a large allographic variety for letters and combination of letters; and ane transformation includes main shape distortion and complementary part distortion.

Recognition methods try to tackle the variability problem to gain more accuracy. However, this may result in a long recognition time. Some techniques focus on the classification aspect of the recognition task by using a combination of classifiers, while others emphasize on data representation in order to obtain descriptive feature vectors. There has been relatively little work on feature representation which interact with direct shape of a character.

In this paper, we aim at improving time and accuracy of recognition of online isolated Arabic characters. We investigate novel approaches that target the challenges involved. This include improving the recognition accuracy of online isolated Arabic characters on existing databases by designing new features and improving the recognition speed using a suitable choice of classifiers. The main contributions of this paper is the design of two novel feature representations: relational histogram and relational context representations for isolated characters, and their corresponding extraction methods which suites two important supervised learning methods of neural networks (NNs) and support vector machines (SVMs) respectively. The idea behind our new feature representations is to consider measurements of local and global features which capture information related to the character shape.

2 Background

A typical online recognition system consists of three main building blocks: preprocessing, pattern representation and classification. In order to reduce the noise introduced by the digitizing device, preprocessing is performed. Patterns representation includes design and extraction of representative features. Features are most relevant information for classification purposes, in the sense of minimizing the within-class pattern variability while enhancing the between-class pattern variability. Features should be as invariant as possible to the expected

distortions. Selecting the feature extraction method has been stated as the most important factor in achieving high recognition performance [16]. Features can be broadly divided into high-level (global) and low-level (local) features. High-level features extract a property related to the topology of the pattern using the whole sequence of trajectory points. Some of the frequently used global features for a stroke include: number of ascenders, descenders, cusps, loops, mass center, start or end points, and bounding box information. Low-level features are features which are calculated on a local area of the pattern, i.e. the trajectory points in a certain vicinity of a point for online handwriting. Local features (point features) are assigned to each point along a script trajectory. Some typical local features include: pen directional features, pen coordinate features, digital curvature, cumulative curve length, point aspect ratio, pen pressure, and ink related features such as point density. Global features may provide more descriptive information about a character than local features, usually for a higher computational cost. Global features are more powerful but less robust. The results of high level feature extraction tend to be highly erroneous due to large shape variations in natural cursive handwriting, especially among different writers [6].

In Perasian/Arabic online recognition, rule-based methods are the most widely used classification methods [14, 1, 4, 1]. These systems all share a similar pattern representation approach, which is to segment a stroke into some geometric primitives, and their recognition part is based on a set of fuzzy rules or variations of the traditional fuzzy theory. Neural networks have been used for isolated Persian and Arabic characters [9], however, these systems showed lower performances than the one-nearest-neighbor (1-NN) clustering method using the same features [12]. Besides, the neural networks classification becomes computationally expensive as the dimensions of the data increase, and the initial conditions change. The use of support vector machines (SVMs) classifiers has gained increasing attention in recent years due to excellent recognition results in various pattern recognition applications. However, this type of classifiers has never been used for online Arabic handwriting recognition. There has been some recent attempts for using SVMs in learning online Arabic handwriting recognition, but the choice of effective features is mandatory for boosting the performance of SVMs.

3 Relational Histogram Representation and Neural Network Classifiers

In this section we present our first proposed representation, *Relational Histogram*(RH). This representation is an adaptation of the shape context idea [2]. Shape context is a descriptor that has been proposed for measuring similarities in the shape matching framework. For a shape with N boundary points, a set of $N - 1$ vectors originating from each boundary point to all other sample points expresses the configuration of the entire shape relative to that point. A coarse histogram of the relative coordinates of those remaining $N - 1$ points is defined as the shape context of that particular point. The representation and its matching method is computationally expensive. This is because a point correspondence

must be formed between the points of each two shapes in order to measure the similarity between them. Shape context in its original form cannot be applied directly in online handwriting recognition since all computations need to occur in real time. The following algorithm presents our idea for using this concept in online handwritten data as follows:

Algorithm 1. Relational Histogram Feature Extraction

INPUT: A set of re-sampled trajectory points S
 OUTPUT: Feature vector V
 Select an arbitrary set of reference points R
 Select an arbitrary set of r-bins: r_1, r_2, \dots, r_n and θ -bins: $\theta_1, \theta_2, \dots, \theta_k$
for all $r \in R$ **do**
 Initialize($r - bins, \theta - bins$)
 for all $s \in S$ **do**
 Compute $dist(r, s)$
 Compute $angle(r, s)$
 Assign($r - bins, \theta - bins, r, s$)
 Update($r - bins, \theta - bins$)
 end for
end for
 $V = count(r - bins, \theta - bins, r)$
Return V

We normalize the character's trajectory points and denote them by a set P . Let R be a set of reference points. Our feature extraction method selects adequate features of the relational type. In the first part of this algorithm, we select an arbitrary and fixed set of points R . This set of reference points must not be outside the bounding box. The points in R may capture some interrelationship structures (for instance, symmetrical corners of the bounding box) or be random. The algorithm re-samples the normalized representation of the character P , as a set of equi-distanced points S . The surrounding area of each reference point that falls in the bounding box is divided into bins according to the distance and angle of each bin with respect to the reference point. The values of all the bins are initialized to zero for each reference point. Then, all the points in S are described from the view of each reference point and are placed into the corresponding bins. This is done by computing the distance and angle between the pair of points, $dist(r, p)$ and $angle(r, p)$, and updating the corresponding bin that this pair can be mapped into. After this step, the number of points in the bins provided by all the reference points will give a compact representation of the character. Using this system makes the descriptor more sensitive to differences in nearby points.

Figure 1 shows the log-polar histogram bins for computing the features of the Arabic character S , pronounced as "seen". We used the center point of a character bounding box as reference point in a log-polar coordinate system. The diagram has 5 bins in the tangential direction and 12 bins in the radial direction, yielding a feature vector of length 60. We capture the global characteristics of the character by this feature vector. In addition, we use a directional feature to

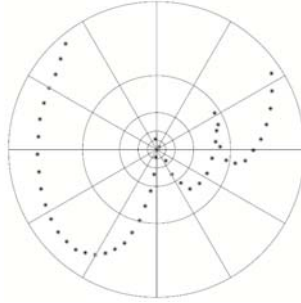


Fig. 1. Diagram of the log-polar bins around the center of the bounding box of the Arabic character “Seen” used for relational feature computation

extract the local writing directions as: $Arctan((x_i - x_{i-1}) + j(y_i - y_{i-1}))$ for $i = 2 : N$. This allows for spatio-temporal data representation and augments the visual realism of the shape of a character.

We have used neural networks as the learning classifier method when trying our RH feature. This feature magnifies the differences between similar characters and improves learning in NNs. We train a Multi-Layer Perceptrons (MLP) classifier through a conjugate gradient method for classification using a three-layer network layout with 100 nodes in the hidden layer. Classification is obtained by adjusting the weights during the training using gradient descent optimization (back propagation method), where we calculate the derivatives of the cost function and find the weights in an iterative way. The evaluation of this system is presented in section 5.

4 Relational Context Representation and Support Vector Machine Classifiers

We call our second representation *relational context* (RC) since this representation uses the inclusion of the character context as the relative pairwise distances and angles. The idea is to use all pairs of points on the trajectory of a digital character to capture as much contextual information about the shape of a character as possible. The length of the RC feature for a trajectory of N sample points is $N(N - 1)$. We re-sample all characters with an equal number of points in order to make the feature vectors of the same length. We also preserve the order of the trajectory samples. Local relevance is captured by the representation of any single component of the character’s shape feature vector. Neighborhood components capture the relationship between different parts of the character’s shape. Preserving these relationships provides invariance with respect to different writing styles. In different writing styles, character points may have absolute information that is different from independent data points, however, the relative interrelationships on a set of points might not be different. Therefore, a small local change in the total shape of the character does not cause a large change

in its representation. The global shape of the character is captured through all pairwise relationships between any two components. Different levels of details in the character's description can be selected by increasing or decreasing the trajectory point density. Figure 2 depicts some of these features for letter *Y*, pronounced *ye* in Arabic scripts. In this figure, the pairwise relationships are illustrated only for point P_j .

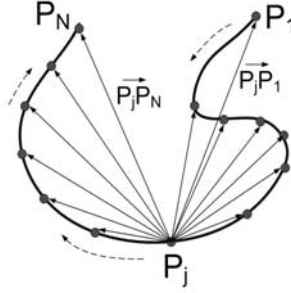


Fig. 2. Arabic letter “*ye*” in relative context feature representation

To extract RC feature, we use Algorithm 2. The length of the proposed RC feature can be potentially long for some very long trajectories with many points. Therefore, this feature, especially when used in fine level, may not be the best match with classifiers such as NNs in terms of required network complexity. However, for SVM classifiers, the size of the feature vector is not usually a practical constraint. The length is not a concern for the execution time since the number of descriptor data points, in the character level, is almost always very short, and the features extraction is of low-order polynomial. In our experience with RC, the number of support vectors required is low. As a consequence, the ultimate system will be very fast and practical with respect to memory constraints.

Algorithm 2. Relational Context Feature Extraction

INPUT: A set of re-sampled trajectory points S

OUTPUT: Feature vector V

for all ordered pairs of data points $(a, b) \in \text{combination}(|S|, 2)$ **do**

 Compute $\text{dist}(a, b)$

 Compute $\text{angle}(a, b)$

end for

$V = (\text{dist}(a, b), \text{angle}(a, b))$

Return V

The second supervised learning method we used for our recognition system with RC representation is a SVM classifier. A SVM was trained using RC feature

on the currently largest database of isolated Arabic letters. The non-linear mapping is performed by using kernel functions. In order to extend SVM to be applied to a multi-class classification task, we have used one against all technique [10].

5 Empirical Results

Quantitative analysis of our feature representations is presented in this section with respect to recognition accuracy and time. In our evaluation studies, we used preprocessing operations in three steps: first smoothing, then de-hooking, and finally point re-sampling. We then normalize all characters to have the same heights, while their original aspect ratios remain unchanged, to achieve scale invariance.

In Table 1, the accuracy of our RH feature in an MLP-NN-based classifier for isolated Arabic characters is compared with the previous research on the same database in a NN-based system. It should be noted that in all tables, our results are reported as averages over 10 independent runs over 3-fold of the data. In each experiment, the dataset is randomly partitioned into disjoint training and test sets, using a fixed proportion (33% test and 66% training). The results are compared with the best previously existing results in the literature, on the same database, by Mezghani et. al. The recognition time for each character sample using the NN-based system is 23s with our relational feature approach, while this is 26s for the experiments in [12].

Table 1. Performance comparison of different recognition systems

Performance	1-NN	SOM [12]	RH-Tangent [8]
Recognition Rate	Ref	Ref-1.19%	Ref+4.22%
Training Time	-	2 hrs	7.5 min
Recognition Time	526 s	26 s	23 s

Our RC representation achieved a higher performance with a small number of support vectors, shown in Table2. The recognition time therefore, is remarkably low. According to our experimental results, it takes only 15ms for our system to recognize a character, while this time is reported in [12] as 26s, which is orders of magnitude higher. This confirms an improvement in our method for speeding up the recognition process, and therefore proves the suitability of our approach for real time applications.

We are also interested to compare our features with previously used ones in the literature. The experimental settings (the dataset and the number of samples used in training and test sets), and the data preprocessing methods may differ among different researchers. Therefore, to have a fair comparison, we implemented these features and conducted experiments in the same setting as our proposed features. Measuring the representational power of a feature by using the same classifier is a standard approach in feature selection methods

Table 2. Experiments on isolated Arabic letters dataset

Approach	Accuracy	Settings	time
K.MAP [12]	94.6%	67% train; 33%test 432 sample/class	26 s
Bayes [13]	90.19%	67% train; 33%test 528 sample/class	-
RC-SVM	99.01%	67% train; 33%test 528 sample/class	0.015 s

Table 3. Features used for Arabic datasets. Each symbol is described by 6 points.

feature	dimension	error%
directional	10	2.7
positional	12	3.7
directional+positional	24	2.2
zone (3x3)	9	34.5
zone (10x10)	100	17.2
RC	30	0.91

[11, 7, 17]. We used SVM for these comparisons. Table 3 presents the results for Arabic characters. The combination of the directional and positional features shows the best performance on all datasets. However, the error rate by using our RC feature is still lower than other features, given the same data, the same preprocessing, and the same classifier. The results also show that the worst performances belong to zone features for all datasets. Such a low performance might be explained by the fact that a small number of points represent each symbol. However, even using a 30-point representation for each symbol, 3×3 zone features resulted in a high error rate of 17.2%. In this table, we also present the length of each feature vector. As we can see, the length of our RC feature is manageable in terms of the introduced complexity for all classifiers in the level of stroke recognition.

6 Discussion

The strength of the novel features presented in this paper can be summarized qualitatively according to several important properties: time complexity, uniqueness, performance accuracy, flexibility, ease of implementation, and memory issue. The recognition time is considered as a measure of performance since the training happens offline in a supervised classification. We demonstrated that the recognition time with SVMs is considerably short and the whole system with RC is very fast. Using a decreased resolution of RC features, even NNs are capable of providing faster results than the ones previously reported in the literature on the same data set [8].

Preserving all relationships between character points provides invariance with respect to similarity transformation, which could be decomposed into translation, scaling and rotation. Similarity transformation invariant is one kind of integral invariants. We can set the integral domain to be a segment of a curve [15]. Supposing each point p_i has coordinate $(x_i, y_i) = (r_i \times \cos\theta_i, r_i \sin\theta_i)$, we can express the three invariant geometric primitives in the plane with polar coordinates. We keep $(\cos^2\theta_i + \sin^2\theta_i)$ which is always 1 in the formula for a quick transformation to xy coordinates: $R(i) = (x_i^2 + y_i^2)^{1/2} = r_i(\cos^2\theta_i + \sin^2\theta_i)^{1/2}$; $A(O, i, j) = 1/2(x_i y_j - x_j y_i) = 1/2 r_i r_j \sin(\theta_j - \theta_i)$; and $Dp(O, i, j) = (x_i, y_i)(x_j, y_j) = r_i r_j \cos(\theta_j - \theta_i)$. In different writing styles, character points may have different absolute information of independent data points, however, the relative interrelationships on a set of points might not be different. Global shape of the character is also captured through all pairwise relationships between any two components.

Our proposed representations, in particular the RC feature, provide very well separated clusters with high margins in the feature space. The RC feature has the ability to incorporate a fair distribution of weights among all local and global features to be used in discriminating a character. We conjecture that RC feature combined by our choice of SVM kernel make the recognition system tolerant to Gaussian noise. This is an area of future investigation.

7 Conclusion and Future Work

We introduced in this paper new types of feature representations referred to as relational histogram and relational context. The strength of these representations resides in capturing contextual information related to the general shape of the characters. We successfully deployed our representations in neural networks and support vector machines classifiers. This was the first time that SVMs are used for recognition of online Arabic characters. Both systems showed superior results, compared to the state of the art methods for Arabic character recognition on the same database. The relational context feature not only outperforms the other approaches in recognition rate, but also provides a significant improvement in recognition time compared to the best existing results in the literature. The measured recognition speed makes the system a potentially perfect choice for the real-time applications. Future work includes the use of more sophisticated kernels for SVMs and extending our study to other series of recognition tasks such as the recognition of gestures, mathematical symbols, signatures, or any other shape or drawing. We also plan to check the ability of the proposed systems for sub-words and a limited dictionary size word recognition.

References

1. Baghshah, M.S., Shouraki, S.B., Kasaei, S.: A novel fuzzy classifier using fuzzy LVQ to recognize online persian handwriting. In: ICTTA 2006: The 2nd Information and Communication Technologies From Theory to Applications, April 2006, vol. 1, pp. 1878–1883 (2006)

2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 24(4), 509–522 (2002)
3. Dinesh, M., Sridhar, M.K.: A feature based on encoding the relative position of a point in the character for online handwritten character recognition. In: *ICDAR 2007: Proceedings of the 9th International Conference on Document Analysis and Recognition*, pp. 1014–1017 (2007)
4. Halavati, R., Shouraki, S.B.: Recognition of Persian online handwriting using elastic fuzzy pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)* 21(3), 491–513 (2007)
5. Han, S., Chang, M., Zou, Y., Chen, X., Zhang, D.: Systematic multi-path HMM topology design for online handwriting recognition of east asian characters. In: *ICDAR 2007: Proceedings of the 9th International Conference on Document Analysis and Recognition*, pp. 604–608 (2007)
6. Hu, J., Rosenthal, A.S., Brown, M.K.: Combining high-level features with sequential local features for on-line handwriting recognition. In: *ICIAP 1997: Proceedings of the 9th International Conference on Image Analysis and Processing*, London, UK, vol. 2, pp. 647–654. Springer, Heidelberg (1997)
7. Huang, B.Q., Kechadi, M.-T.: A fast feature selection model for online handwriting symbol recognition. In: *ICMLA '06: Proceedings of the 5th International Conference on Machine Learning and Applications*, Washington, DC, USA, pp. 251–257. IEEE Computer Society, Los Alamitos (2006)
8. Izadi, S., Suen, C.Y.: Incorporating a new relational feature in Arabic online handwritten character recognition. In: *VISAPP '08: Proceedings of the Third International Conference on Computer Vision Theory and Applications*, PortugalMadeira, Portugal, January 2008, vol. 1, pp. 559–562. INSTICC - Institute for Systems and Technologies of Information, Control and Communication (2008)
9. Klassen, T.J., Heywood, M.I.: Towards the on-line recognition of arabic characters. In: *IJCNN 2002: Proceedings of the 2002 International Joint Conference on Neural Networks*, May 2002, vol. 2, pp. 1900–1905 (2002)
10. Kressel, U.H.G.: Pairwise classification and support vector machines. In: *Advances in kernel methods: support vector learning*, pp. 255–268 (1999)
11. Liwicki, M.M., Bunke, H.: Feature selection for on-line handwriting recognition of whiteboard notes. In: *IGS 2007: The 13th Conference of the International Graphonomics Society*, pp. 101–105 (2007)
12. Mezghani, N., Mitiche, A., Cheriet, M.: A new representation of shape and its use for high performance in online Arabic character recognition by an associative memory. *International Journal on Document Analysis and Recognition (IJDAR)* 7(4), 201–210 (2005)
13. Mezghani, N., Mitiche, A., Cheriet, M.: Bayes classification of online arabic characters by gibbs modeling of class conditional densities. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 30(7), 1121–1131 (2008)
14. Randa, S.A.M., Elanwar, I., Rashwan, M.A.: Simultaneous segmentation and recognition of Arabic characters in an unconstrained on-line cursive handwritten document. *International Journal of Computer and Information Science and Engineering (IJCISE)* 1(4), 203–206 (2007)
15. Sluzek, A.: Using moment invariants to recognize and locate partially occluded 2d objects. *Pattern Recognition Letters* 7, 253 (1988)

16. Trier, O., Jain, A., Taxt, T.: Feature extraction methods for character recognition - a survey. *Pattern Recognition* 29(4), 641–662 (1996)
17. Verma, B., Ghosh, M.: A neural-evolutionary approach for feature and architecture selection in online handwriting recognition. In: *ICDAR '03: Proceedings of the Seventh International Conference on Document Analysis and Recognition*, Washington, DC, USA, pp. 1038–1042. IEEE Computer Society, Los Alamitos (2003)