ICDAR2013 Competition on Handwritten Digit Recognition (HDRC 2013)

Markus Diem*, Stefan Fiel*, Angelika Garz**, Manuel Keglevic*, Florian Kleber*, Robert Sablatnig*
*Vienna University of Technology, CVL
1040 Vienna, Austria
Email: {diem, fiel, mkeglevic, kleber, sab}@caa.tuwien.ac.at
**University of Fribourg, DIVA
1700 Fribourg, Switzerland
Email: angelika.garz@unifr.ch

***Abstract*—This paper presents the results of the HDRC 2013 competition for recognition of handwritten digits organized in conjunction with ICDAR 2013. The general objective of this competition is to identify, evaluate and compare recent developments in character recognition and to introduce a new challenging dataset for benchmarking. We describe competition details including dataset and evaluation measures used, and give a comparative performance analysis of the nine (9) submitted methods along with a short description of the respective methodologies.**

I. Introduction

Due to the high variability of handwriting, recognition of unconstrained handwriting is still considered an open research topic in the document analysis community. Recognition of handwritten digits has been studied for many years, and several benchmark datasets have been published, such as MNIST [1], USPS [2], Optdigits[1], Semeion[1]. Having analyzed existing databases and benchmarks, we decided to provide a new framework for benchmarking; i.e., a new freely available real world dataset along with objective evaluation measures (overall precision, and per-class F-score, precision, and recall) in order to assess the performance of current digit recognition approaches.

The task of the competition is the recognition of isolated handwritten digits (HDR). In addition to the HDR competition, we proposed a handwritten digit string competition (HDSR), where a string of handwritten digits had to be segmented and recognized. Segmentation of connected handwritten characters is still considered an open research topic in the document analysis community as well. Due to the low number of participants in the HDSR competition, we present the results of the HDR competition only. However, we plan to organize the HDSR competition in conjunction with upcoming conferences.

II. CVL Single Digit dataset

The *CVL Single Digit dataset* is part of the *CVL Handwritten Digit database (CVL HDdb)*, which has been collected mostly among students of the Vienna University of Technology and of an Austrian secondary school; it consists of samples from 303 writers. For the CVL HDdb, 26 different digit strings with varying length were collected from each writer, resulting in a database of 7,800 samples. In order to create the CVL Single Digit dataset, isolated (unconnected) digits were extracted from the CVL HDdb.

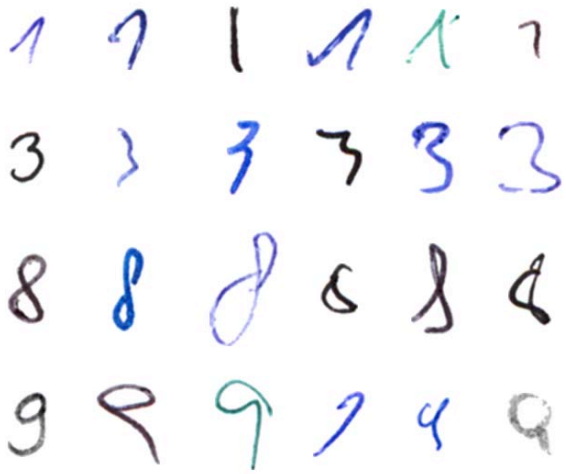

Figure 1. Samples of the HDR evaluation set. Note the high variation of digits among different writers.

To our knowledge, this dataset is the first one to provide files in RGB. In the design process of the database, a uniform distribution of the occurrences of each digit was ensured. For the competition, the images are delivered in original size with a resolution of 300 dpi. Contrary to other datasets, the digits are not size-normalized since in real world cases, differences in a writers' handwriting include variation in size as well as writing style (see Figure 1). The full dataset is available at http://caa.tuwien.ac.at/cvl/research/icdar2013-hdrc/, along with a size-normalized version, which has the potential of becoming a new database for machine learning purposes.

In the following, we describe training, validation and evaluation sets generated for the competition. The images for each set were randomly selected from a subset of writers from the CVL Single Digit dataset. The complete CVL

[1]http://archive.ics.uci.edu/ml/datasets/

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.287

Single Digit dataset consists of 10 classes (0-9) with 3,578 samples per class. For the HDR competition, 7,000 digits (700 digits per class) of 67 writers have been selected as training set. A validation set of equal size has been published with a different set of 60 writers. The validation set may be used for parameter estimation and validation but not for supervised training. The evaluation set consists of 2,178 digits per class resulting in 21,780 evaluation samples of the remaining 176 writers. The evaluation set was published after the evaluation of the submitted methods.

III. METHODS AND PARTICIPANTS

Seven (7) research groups have participated with nine (9) methods for Handwritten Digit Recognition (HDR). Two groups submitted two different algorithms each. In the following, brief descriptions of the respective submissions are given. The order of appearance is alphabetical.

1. François Rabelais: Université François Rabelais, Laboratoire d'Informatique (O. Razafindramanana, F. Rayar, G. Venturini)

First pre-processing of the isolated digit image is done. The image is cropped to the bounding box of the digit, deleting the white border, and is surrounded with a 1-pixel margin. It is then magnified to have a final size of 128×128. Finally a skew and slant normalization is done. The black pixels are considered as input data points, and reduce the amount of points by keeping $10\,\%$ of them using k-means algorithm.
The Delaunay Triangulation is built on the input data points. Then, $(1 - \alpha^*)$ % of the higher valued sorted triangles are pruned with respect to a *local heterogeneity measure*. α^* is the proportion associated to the maximum curvature index of the distribution of the sorted triangles regarding the proposed measure.
A *multilevel* static uniform zoning is computed at K distinct orders. For each cell of a grid, two values are appended: *(i)* the number of input elements in the cell, *(ii)* the average of input elements in the neighborhood of the cell. Both the centers of gravity of the triangles and the black pixels within the cell are input elements. This eventually produces a feature vector FV^k. Multilevel feature extraction is the appending of the vectors $FV^k, \forall k \in [1..K]$.
A SVM classifier (*libSVM*) with a RBF kernel is trained and then used for the prediction.

2. Hannover: Hochschule Hannover (K.-H. Steinke, M. Gehrke)

The numerals are normalized, binarized and slope corrected. The feature vector is composed of three methods: first, number of black pixels in each row and column; second, lengths of 12 probes in different directions from different positions; and third, normalized central moments. The numerals are classified with a nearest-k-neighbor classifier [3].

3. Jadavpur: Jadavpur University, CMATER (N. Das, A. Roy, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri)

The system is based on a Fuzzy-Entropy-based feature selection strategy over a combination of Quad-tree-based longest-run [4], [5] and convex-hull-based [6] feature sets with SVM classifiers. Initially, 239 features (84 from Quad-tree-based longest-run features and 155 from convex-hull-based features) were extracted. Using the Fuzzy-Entropy-based feature selection strategy over the features, 190 feature set was found to be good for validation set. The system is finally developed using the selected 190 features.

4. Orand: ORAND S.A. (J. M. Saavedra, J. M. Barrios)

The approach is based on the combination of four descriptors which allow exploiting three different characteristics of image digits. The method consists of three general stages: (1) pre-processing (2) feature extraction, and (3) classification. For the pre-processing, a thresholding operation using Otsu's method is applied. For the feature extraction, three different characteristics of the digits are exploited: (1) the stroke orientations, (2) the relation between background and foreground, and (3) the contour. For the case of stroke orientations, the descriptor is based on the HOG approach [7]. In particular, the image is divided into 2×2 regions. For each region a histogram of orientations using 32 bins is computed. Then, the descriptor is produced by concatenating the region histograms.
For characterizing the relation between background and foreground, a descriptor based on concavities [8] is used. For each background pixel a 4-bit code yielded by searching for a foreground pixel in four directions is computed, if a foreground is found the corresponding bit is set to 1, in other case to 0. The descriptor is a histogram of the occurrences of the codes. Two kinds of directions are used. The first one uses directions with respect to the closest four neighbors (north, south, east, and west), the second one uses the diagonal directions which yields two 16-size descriptors.
For characterizing the contour of the digits, horizontal profiles are used. To this end, the image is resized to 40×40 pixels. Then a thinning operation following the Zhang and Suen approach [9] is applied. The profile with respect to the left and right side is computed yielding an 80-size descriptor. Finally, the digit descriptor is the concatenation of the four described descriptors yielding a 240-size descriptor.
For classification, a multi-class SVM classifier using a RBF kernel is used, the cost parameter is set to 6, and the gamma parameter is set to 1.4.

5. Paris Sud: University of Paris Sud, Linear Accelerator Laboratory and Computer Science Laboratory & CNRS (F. Dubard, B. Kégl)

The images are pre-processed following the MNIST setup of Yann LeCun (getting rid of color, down-sampling to 20×20 resolution and placing the images on a 28×28 grid by

centering their center of gravity[2]). Then AdaBoost.MH [10] is used. The base classifiers were Hamming trees over Haar filters [11]. The Hamming tree algorithm and the detailed description of the particular AdaBoost.MH implementation is available in the Appendix of the documentation of the multiboost software [12]. The training was done on the provided training set and the hyperparameters were validated on the provided validation set. The chosen classifier has 47,642 trees of 4 nodes (5 leaves) each. In each boosting iteration 100 random Haar filters are tested, chosen uniformly from the possible geometries.

6. Salzburg I: University of Salzburg, Institute of Computer Science (C. Codrescu, C.L. Badea)

The Finite Impulse Response Multilayer Perceptron (FIR MLP), a class of temporal processing neural networks, is a multilayer perceptron where the static weights (synapses) have been replaced with finite impulse response filters [13], [14] [15]. Hereby, the FIR neuron represents a model for spatio-temporal processing.

First the color images are transformed to 8 bit gray-scale images. Each of these gray-scale images is resized to 20 $\times$ 20 pixel images by preserving their aspect ratio and their center of mass is computed. Each scaled image is positioned by their center of mass in the center of a 28 $\times$ 28 pixel image. Each pixel value has been normalized into the range [-1,1].

For the experiments a neural network framework written in java has been further developed [16]. Fully and partially connected neural networks are created, respectively. For the last type only the output layer is fully connected. For the networks initialization and training some suggestions from [17] have been adapted to the FIR MLP. The output layer consists of neurons with linear transfer function; all other neurons are sigmoidal units. Over the entire training process affine deformations of the input patterns are generated by using uniform distributed random values in the range: [-20, 20] degrees for the rotation angle, [-0.2, 0.2] for shearing and [0.8, 1.2] for scaling. As training algorithm online temporal back-propagation were used and the mean squared error was minimized. In the validation phase, after computing the neural network response for a given input pattern, the output neuron with the maximum activation was set to 1 and all others to 0. After this step a comparison with the one-of-ten representation of the digits is done to perform the classification.

This method uses one partially connected FIR MLP with four layers.

7. Salzburg II: University of Salzburg, Institute of Computer Science (C. Codrescu, C.L. Badea)

The description of this method is similar to the previous (*Salzburg I*), but uses an ensemble of four FIR MLP partially and fully connected with four layers.

[2]http://yann.lecun.com/exdb/mnist/

8. Tébessa I: University of Tébessa, LAMIS (A. Gattal) & University of Sciences and Technology Houari Boumediene, LCPTS (Y. Chibani)

In this work, the objective is to improve the performance of a recognition system based on combining different pertinent structural features from the digits. This method is conducted with combination two structural features without uniform grid method, background features [18] of 14 components and foreground features [19] of the skeleton from 4 components. For the three remaining methods (classic features, ridgelet transform and foreground features), the image was divided into four regions by using uniform grid method [20]. This method was applied for each region of 17 components. Generally, the global features vector is composed of 86 (17 $\times$ 4 + 14 + 4) components. The recognition module is based on the SVM multi-class approach using the one-against-all implementation. SVM and RBF kernel parameters are fixed to $C = 10$ and $\sigma = 8$.

9. Tébessa II: University of Tébessa, LAMIS (A. Gattal, C. Djeddi) & University of Sciences and Technology Houari Boumediene, LCPTS (Y. Chibani)

This method is based on multi-scale run length features [21] which are determined on the binary image taking into consideration both the black pixels corresponding to the ink trace and the white pixels corresponding to the background. The probability distribution of black and white run-lengths has been used. There are four scanning methods: horizontal, vertical, left-diagonal and right-diagonal. The runs lengths features are calculated using the grey level run length matrices and the histogram of run lengths is normalized and interpreted as a probability distribution. The method considers horizontal, vertical, left-diagonal and right-diagonal white run-lengths as well as horizontal, vertical, left-diagonal and right diagonal black run-lengths extracted from the original image. To compare two documents, the Manhattan Distance Metric is used. The algorithm proposed for these applications mainly includes classic features, the ridgelet transform, background features and foreground features (contour, skeleton) and this method is based on multi-scale run length features, completing the system by a multi-class SVM classifier based on approach one-against-all.

IV. EVALUATION

Contributions were accepted as binaries. The input is a RGB image (300 dpi, not size-normalized); the required output is the recognized ASCII character. First and second guess are evaluated.

Precision is employed as performance measure, contributions are ranked upon it. Precision p is computed by

$$p = \frac{tp}{tp + fp}$$

with tp being the sum of true positives (a true positive is an element where the class label c_i of class i equals the assigned class label a_i), and fp being the sum of false positives ($c_i \neq a_i$). For the precision p including the second guess, tp is defined as the sum of all elements whose first or second prediction equals the class label.

Figure 2 shows the performance of the methods submitted measured in precision and precision including second guess. The test was conducted on the HDR evaluation set which consists of 21,780 isolated handwritten digits. *Salzburg II* performs best with precision $p = 97.74\,\%$, and $p = 99.33\,\%$ including second guess, respectively. In total, six methods had a precision $p > 90\,\%$, out of which three methods correctly recognized more than $95\,\%$ of the handwritten characters. The greatest improvement (by $6.22\,\%$) regarding the second guess is achieved by *Hannover*. Table I shows the participating methods sorted by their precision.

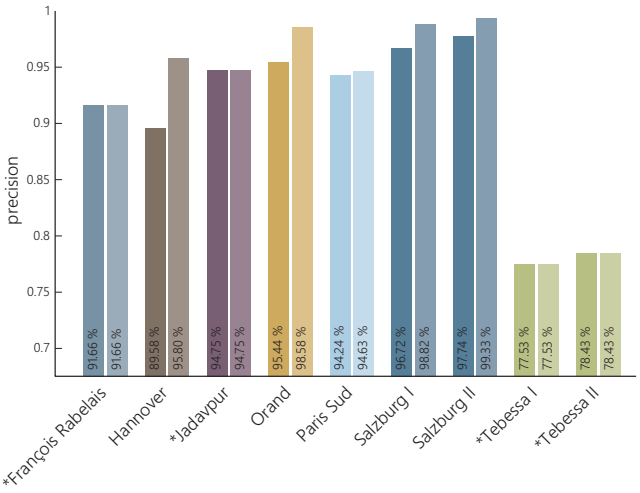

Figure 2. Precision and precision including the second guess. Methods marked with * did not provide a second guess.

	Precision	Precision 2nd
Salzburg II	97.74 %	99.33 %
Salzburg I	96.72 %	98.82 %
Orand	95.44 %	98.58 %
Jadavpur	94.75 %	-
Paris Sud	94.24 %	94.63 %
François Rabelais	91.66 %	-
Hannover	89.58 %	95.80 %
Tébessa II	78.43 %	-
Tébessa I	77.53 %	-

Table I
PRECISION AND PRECISION INCLUDING SECOND GUESS. THE TEST WAS CONDUCTED ON THE HDR EVALUATION SET WHICH COMPRISES 21,780 SINGLE DIGITS.

Furthermore, F-score, precision, and recall are calculated for each class in order to allow for drawing conclusions about the nature of errors and class confusions. For these error measures, true positives tp_i, false positives fp_i, and false negatives fn_i of a given class i are defined by:

$$\begin{aligned} tp_i &\ldots \langle a_i, c_i \rangle \\ fp_i &\ldots \langle a_i, c_{j \neq i} \rangle \\ fn_i &\ldots \langle a_{j \neq i}, c_i \rangle \end{aligned}$$

where $i, j \in 0...n$ and $n = 9$ to represent all digit classes. To illustrate the definitions, a confusion matrix with corresponding labels is given in Table II.

	a_0	a_1	$\cdots$	$\mathbf{a_i}$	$\cdots$	a_n
c_0				fp		
c_1				fp		
$\vdots$				$\vdots$		
$\mathbf{c_i}$	fn	fn	$\cdots$	tp	$\cdots$	fn
$\vdots$				$\vdots$		
c_n				fp		

Table II
CONFUSION MATRIX WITH $n = 9$ TO COVER ALL DIGIT CLASSES; a_i ARE PREDICTIONS OF CLASS i, AND c_i ARE THE TRUE CLASS LABELS.

Given the previously defined true positives tp_i, false positives fp_i, and false negatives fn_i; precision p_i, recall r_i, and F-score F_i of a class i are defined as:

$$p_i = \frac{tp_i}{tp_i + fp_i}$$
$$r_i = \frac{tp_i}{tp_i + fn_i}$$
$$F_i = \frac{tp_i}{2tp_i + fp_i + fn_i}$$

Figure 3 shows the precision p_i of the participating methods for each digit class (0-9). The plot shows that all methods have a high precision when recognizing the digit `1` (six methods achieve their respective highest precision). At the same time, four methods attain their lowest recall for `1` (see Figure 4); i.e., `1` is recognized best, however, at the same time most digits are falsely classified as `1`.

According to the evaluation, recognizing the digit `9` is most difficult; a mean F-score of $83.2\,\%$ (see Figure 5) is achieved. In most cases (1,510 in total), `9` is confused with `1`.

V. CONCLUSION

We presented the results of the competition on handwritten digit recognition. Along with brief descriptions of the participating methods, we provided details on the evaluation and ranking methods. The CVL Single Digit dataset has been introduced, which is freely available at http://caa.tuwien.ac.at/cvl/research/icdar2013-hdrc/. 303 writers have contributed to this dataset resulting in 7,000 digits for training, a validation set of equal size, and an evaluation set consisting of

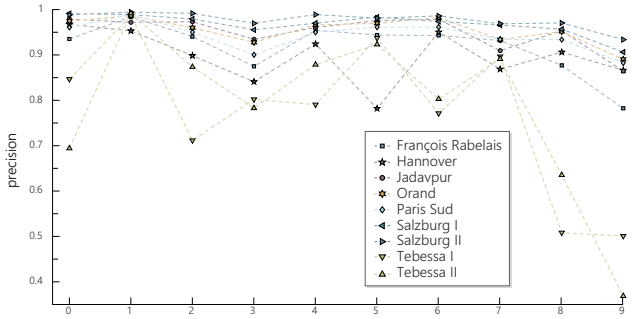

Figure 3. Precision of all digit classes (0-9) of all participating methods

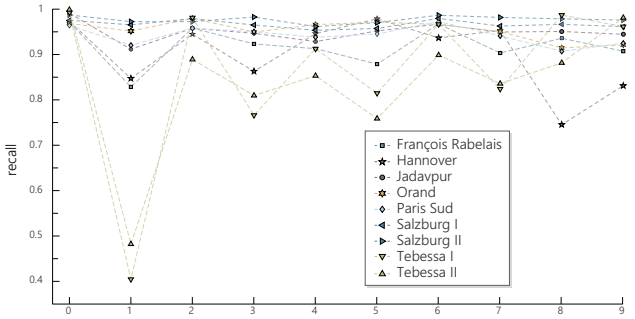

Figure 4. Recall of all digit classes (0-9) of all participating methods

21,780 digits. A uniform distribution of occurrence of each digit was ensured. The evaluation of the competition was based on the precision of the first and second guess of each method. Seven (7) research groups participated with nine (9) different methods in the contest. The best performance for both precision measures (97.74 % and 99.33 %, respectively) was achieved by Salzburg II, submitted by C. Condrescu and C.L. Badea from the University of Salzburg, Institute of Computer Science.

ACKNOWLEDGMENTS

We want to thank all authors who participated in this contest. Additionally we would like to thank all individuals who have contributed to the CVL Handwritten Digit database. The research was funded by the Austrian Science Fund (FWF) project P23133, and the Swiss National Science Foundation (SNF) project CRSI22 125220.

REFERENCES

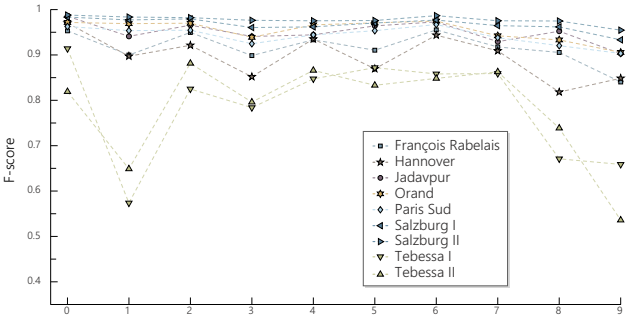

Figure 5. F-Score of all digit classes (0-9) of all participating methods

[1] J. Hull, "A database for handwritten text recognition research," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 550–554, 1994.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] K.-H. Steinke and B. Mund, "Datamining in Herbarbelegen," in *Wismarer Wirtschaftsinformatik-Tage*, 2010.

[4] N. Das, J. M. Reddy, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, "A Statistical–Topological Feature Combination for Recognition of Handwritten Numerals," *Applied Soft Computing*, vol. 12, no. 8, pp. 2486–2495, 2012.

[5] N. Das, R. Sarkar, S. Basu, M. Kundu, M. Nasipuri, and D. K. Basu, "A Genetic Algorithm based Region Sampling for Selection of Local Features in Handwritten Digit Recognition Application," *Applied Soft Computing*, vol. 12, no. 5, pp. 1592–1606, 2012.

[6] N. Das, S. Pramanik, S. Basu, P. Saha, R. Sarkar, M. Kundu, and M. Nasipuri, "Recognition of Handwritten Bangla Basic Characters and Digits using Convex Hull based Feature Set," in *International Conference on Artificial Intelligence and Pattern Recognition*, 2009, pp. 380–386.

[7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.

[8] L. Oliveira, R. Sabourin, F. Bortolozzi, and C. Suen, "Automatic Recognition of Handwritten Numerical Strings: A Recognition and Verification Strategy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1438–1454, 2002.

[9] T. Y. Zhang and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.

[10] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[11] M. Jones and P. Viola, "Fast Multi-View Face Detection," *Mitsubishi Electric Research Lab TR-20003-96*, vol. 3, 2003.

[12] D. Benbouzid, R. Busa-Fekete, N. Casagrande, F.-D. Collin, and B. Kégl, "MultiBoost: A Multi-Purpose Boosting Package," *Journal of Machine Learning Research*, vol. 13, pp. 549–553, 2012.

[13] A. D. Back and A. C. Tsoi, "A Time Series Modeling Methodology using FIR and IIR Synapses," in *Workshop on Neural Networks for Statistical and Economic Data*, 1990, pp. 187–194.

[14] C. Badea, "Chapter 3 - FIR Neuron Modeling," in *Approximate Dynamic Programming for Real-Time Control and Neural Modeling*, F. Ionescu and D. Stefanoiu, Eds. Steinbeis Edition, 2004.

[15] E. Wan, "Finite Impulse Response Neural Networks with Applications in Time Series Prediction," Ph.D. Dissertation, Stanford University, CA, 1993.

[16] C. Codrescu, "Temporal Processing Applied to Speech Recognition," in *Intelligent Systems Design and Applications*, 2004.

[17] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient BackProp," in *Neural Networks: Tricks of the Trade*, ser. Lecture Notes in Computer Science, G. B. Orr and K.-R. Müller, Eds. Springer Berlin Heidelberg, 1998, vol. 1524, pp. 9–50.

[18] P. R. Cavalin, A. de Souza Britto, F. Bortolozzi, R. Sabourin, and L. E. S. Oliveira, "An Implicit Segmentation-based Method for Recognition of Handwritten Strings of Characters," in *ACM Symposium on Applied Computing*, 2006, p. 836.

[19] L. E. S. Oliveira, "Automatic Recognition of Handwritten Numerical Strings," Ph.D. Dissertation, École de Technologie Supérieure Université du Québec, 2003.

[20] J. T. Favata and G. Srikantan, "A Multiple Feature/Resolution Approach to Handprinted Digit and Character Recognition," *International Journal of Imaging Systems and Technology*, vol. 7, no. 4, pp. 304–311, 1996.

[21] C. Djeddi, L. Souici-Meslati, and A. Ennaji, "Writer Recognition on Arabic Handwritten Documents," in *International Conference on Image Analysis and Processing*, vol. 7340, 2012, pp. 493–501.