Improving isolated and in-context classification of handwritten characters

Vadim Mazalov and Stephen M. Watt

Ontario Research Centre for Computer Algebra
Department of Computer Science
University of Western Ontario
London, Canada

ABSTRACT

Earlier work has shown how to recognize handwritten characters by representing coordinate functions or integral invariants as truncated orthogonal series. The series basis functions are orthogonal polynomials defined by a Legendre-Sobolev inner product. It has been shown that the free parameter in the inner product, the "jet scale", has an impact on recognition both using coordinate functions and integral invariants.

This paper develops methods of improving series-based recognition. For isolated classification, the first consideration is to identify optimal values for the jet scale in different settings. For coordinate functions, we find the optimum to be in a small interval with the precise value not strongly correlated to the geometric complexity of the character. For integral invariants, used in orientation-independent recognition, we find the optimal value of the jet scale for each invariant. Furthermore, we examine the optimal degree for the truncated series. For in-context classification, we develop a rotation-invariant algorithm that takes advantage of sequences of samples that are subject to similar distortion. The algorithm yields significant improvement over orientation-independent isolated recognition and can be extended to shear and, more generally, affine transformations.

Keywords: online handwriting recognition; orthogonal polynomials; Legendre-Sobolev series; transformation-independence

1. INTRODUCTION

We are interested in online classification of handwritten mathematical characters as a foundation for pen-based input of mathematical expressions. Recognition of handwritten mathematics is substantially different from natural text recognition due to limitations of dictionary-based verification methods, two-dimensional structure and large variety of symbols that include different alphabets, digits, operators and special characters. In this context, accuracy of recognition of individual characters is challenging and critically important.

It was proposed earlier[1] to represent an ink sample as a parameterized curve and to approximate the coordinate functions by truncated orthogonal polynomial series. This approach simplifies curve manipulation and feature extraction, as well as facilitates geometric analysis by applying corresponding theoretical domains. It was further shown[2] how to compute a sample's representation online, as the curve is being written, with Legendre polynomials as the basis functions. Later, the Legendre-Sobolev (LS) basis was found to perform better, yielding 97.5% recognition rate[3] with a dataset of samples, most of which were collected as isolated symbols. Although the samples do exhibit certain amount of rotation and shear, it is expected that symbols written in a natural environment are more likely to be distorted in this way. To address the issue, we developed integral invariant methods for rotation- and shear-invariant classification.[4,5]

In the present article, our goal is to improve the already good recognition rates obtained with these orthogonal basis methods. We do this by optimizing the choice of basis functions in two ways: We optimize the free parameter in the inner product definition in each of several settings, and we also optimize the series truncation order.

Further author information:
V.M.:E-mail: vmazalov@uwo.ca
S.M.W.: E-mail: Stephen.Watt@uwo.ca, Telephone: +1 519 661-4244

Additionally, for orientation-independent recognition, we show that considering sequences of nearby characters avoids orientation ambiguities to a large extent.

We find optimal values for use with coordinate functions and with integral invariants. We minimize classification error by investigating the role of the jet scale μ in description of coordinate and invariant functions. We also study whether there exists a dependence between *complexity* of a character and the optimal μ in its recognition. These optimizations are directly applied in the proposed algorithm for distortion-invariant classification, taking advantage of the natural property of human handwriting – writing characters with similar transformation. Experiments are performed for the case of rotation, and a similar setting can be used for shear- and, more generally, affine-independent recognition.

Some work has been done in context-dependent recognition of handwriting, mostly relying on statistical approaches. The context-aware classification of a symbol is often represented in some sort of a joint distribution function of the character and its neighbours. For example, some authors propose[6] to consider substrokes in sets, rather than independently, and encode them in HMMs. To keep the model computationally feasible, a hidden Markov network is used to share states of different HMMs. A similar approach is taken elsewhere,[7] where the authors build trigraph models and share certain parameters between those trigraphs. Context can also be useful when dealing with ambiguous segmentation of handwritten words,[8] where the classification task is represented as an optimization problem in a Bayesian framework by explicitly conditioning on the spatial configuration of the characters. As cited above, the context is typically taken into account for cursive words recognition. We find context useful in a different setting – classification of well-segmented symbols, subjected to certain distortion.

The paper is organized as follows: Section 2 presents the necessary background. Section 3 describes the concepts and experimental methods for improving isolated character recognition via coordinate functions. Section 4 presents a recognition approach for in-context classification of distorted characters. Experimental results are given in Section 5. Section 6 concludes the paper.

2. PRELIMINARIES

The main idea of the classification methods we consider is to represent some functions in terms of an orthogonal basis and to use distance-based classification in the coefficient space.[3] The functions represented may be coordinate functions of an ink trace parameterized by time, arc length or affine arc length, or they may be integral invariants of these coordinates. In each case, the basis functions are degree-graded polynomials constructed to be orthogonal with respect to a Legendre-Sobolev (LS) inner product

$$\langle f, g \rangle = \int_a^b f(\lambda)g(\lambda)d\lambda + \mu \int_a^b f'(\lambda)g'(\lambda)d\lambda. \tag{1}$$

We call these basis functions *Legendre-Sobolev (LS) polynomials*. Recognition is based on the distance to the convex hulls of k-nearest neighbors (CHNN_k) in the truncated series coefficient space.

The inner product (1) has one free parameter, μ, which may be assigned any non-negative value. Because this parameter determines the relative weight of the coordinates and their derivatives (i.e. the weights in the jet space), we call μ the *jet scale*. In earlier work, $\mu = 1/8$ was taken as a suitable value.

The rotation-independent[4] and shear-invariant[5] algorithms compute special functions from coordinates. These functions are invariant to certain transformations and therefore describe curves in terms of values that remain relatively constant, even when samples are rotated or sheared on large angles. In rotation-independent classification,[4] we consider the first two invariants[9]

$$I_0(\lambda) = \sqrt{X^2(\lambda) + Y^2(\lambda)} = R(\lambda),$$
$$I_1(\lambda) = \int_0^\lambda X(\tau)dY(\tau) - \frac{1}{2}X(\lambda)Y(\lambda)$$

where $X(\lambda)$ and $Y(\lambda)$ are the coordinate functions.

Even though integral invariants maybe sufficiently discriminating for certain subsets of handwritten samples (e.g. digits or Latin characters) in our dataset, we find I_0 and I_1 alone not to give satisfactory performance on
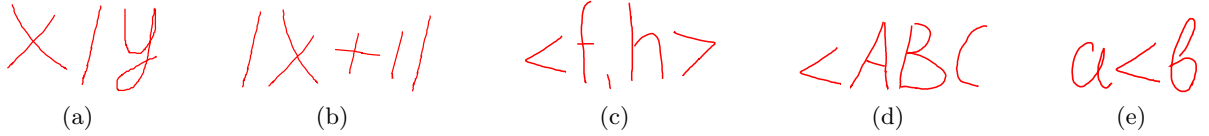
(a)   (b)   (c)   (d)   (e)

Figure 1. Distorted characters: (a) division vs. (b) modulus; (c) angle bracket vs. (d) angle vs. (e) less than

the whole range of classes. To improve recognition, we select the top T classes with the distance to CHNN_k in the space of coefficients of approximation of integral invariants with LS polynomials. To find the final class among these, we solve the following minimization problem for each class C_i among the top T:

$$\min_{\phi} \mathrm{CHNN}_k(XY(\phi), C_i),$$

where $XY(\phi)$ is the rotated or sheared image of the test sample curves X and Y, and $\mathrm{CHNN}_k(X, C)$ is the distance from a point X (in the Legendre-Sobolev space) to the CHNN_k in class C.

3. IMPROVING ISOLATED SYMBOL CLASSIFICATION

3.1 Evaluation of the Jet Scale and the Degree of Approximation

It is easily seen that the jet scale parameter, μ, in the LS inner product has an impact on recognition rate. We would like to better understand this dependency in order to optimize this parameter. For each value of μ considered in the experiments, LS polynomials are generated, orthogonal with respect to the corresponding inner product.

For these experiments we use a dataset of about 50,000 isolated handwritten mathematical symbols, identical to that described earlier.[3]

Coordinate Functions   To optimize μ for coordinate functions, we consider recognition of the original samples in our dataset without additional distortion. The coordinate functions of samples are approximated with LS polynomials for different μ. We test values of μ in the range from 0 to 0.10 with the step of 0.002 and from 0.10 to 0.20 with the step of 1/64. Values outside this range give substantially worse results. Samples are classified with the distance to the CHNN_k in the space of the coefficients of coordinate functions.[3]

Integral Invariants   To study the impact of μ on integral invariants, we consider characters with unknown orientation. The whole collection of original samples is rotated by an angle α between $\pi/9$ and 2π. All multiples of $\pi/9$ are tested. For each angle, I_0 and I_1 are computed for original and transformed samples. The invariants are then approximated with LS series for different values of $\mu \in (0, 0.2]$ with the step of 0.002. For each value of μ, the average maximum approximation error with respect to angle is found as

$$\omega = \frac{1}{n} \sum_{k=1}^{n} \max_{ij}(c_{ij} - c_{ij}^{\frac{k\pi}{9}}) \tag{2}$$

where c_{ij} is the j-th coefficient of the i-th original sample, and $c_{ij}^{\frac{k\pi}{9}}$ is the corresponding coefficient of the sample, rotated on angle $\frac{k\pi}{9}$.

Figure 2. Characters from the training dataset

Complexity of Handwritten Characters   We consider the possibility that the optimal value of μ may depend on the nature of the characters to be recognized. To understand this, we take the notion of a sample's complexity as

$$\eta = \sum_{i=1}^{d}(X_i^{1/i} + Y_i^{1/i}),$$

where X_i and Y_i are normalized coefficients of approximation of the sample with orthogonal polynomials. Coefficients of higher degree are typically greater for "complex" characters – characters that contain large number of loops and/or amount of curvature.

Degree of Approximation   The degree of the truncated series, d, regulates how well curves are approximated. In general, higher degree polynomials provide lower error. Sometimes, however, higher order approximation of equidistant nodes may cause extreme oscillation at the edges of an interval (Runge's phenomenon). To find the optimal degree, we evaluate the recognition error, the maximum absolute and the average relative approximation error depending on d. The approximation errors are computed similar to the way, as shown in subsection 3.1, but instead of coefficients we compare original and approximated coordinates of samples.

4. IMPROVING IN-CONTEXT INVARIANT CLASSIFICATION

Context-Dependent Recognition   There are two main approaches to recognition of handwritten mathematics: symbol-at-a-time and formula-at-a-time. Even though comprehensive semantic and syntactic verification of math is quite challenging, studies suggest that context can play an important role in accurate classification and grammatical information can be an asset.[10] Moreover, it has been shown[11] that n-grams provide useful information in a mathematical setting. These facts suggest that contextual information should be taken into account, especially considering large number of similar-shaped symbols that appear ambiguous on their own. Figure 1 shows typical challenges that arise in classification of individual samples, but which are resolved by considering context.

Algorithm   To improve classification of transformed characters, we propose to recognize a set of n samples at a time with the assumption that the characters in the set are transformed by approximately the same degree, see Algorithm 1. This assumption is justified, since samples written by a person are subject to similar distortions. Moreover, we find that symbols in our dataset exhibit various degrees of rotation and shear. Such initial transformations incorporate noise to the model and reflect real-life handwriting. We consider the case of rotation. Shear may be handled similarly. The algorithm is applied to a sequence of n characters rotated on a random angle $\gamma \in [-\beta, \beta]$. The value of ϵ_α, given in (3), can be interpreted as the error likelihood in recognition of the sequence distorted by an angle α. The value is derived from the observation that while the distance to the closest class is decreasing and sum of the distances to the closest p classes is increasing, the possibility of a recognition error is declining. Therefore, in the last step the algorithm finds and returns the angle γ of transformation that yields the least error likelihood of the whole sequence.

Complexity Analysis   As has been shown,[2] the coefficients of a d-dimensional approximation can be computed in online time $OL_n[O(d), O(d^2)]$, where $O(d)$ is the time complexity as each new point is observed and $O(d^2)$ is the cost at pen up. Sample normalization is performed in linear time. It was shown[5] how to compute each coefficient of approximation of I_1 in $O(d^2)$. Distance from a point to a CHNN_k is theoretically computed in

---

Algorithm 1 In-context rotation-invariant recognition

Input: A set of n rotated test samples and an angle β of the maximum possible rotation of the samples.
Output: A set of n recognized samples and the angle γ of rotation of the samples.

  for $i = 1$ to n **do**
    Approximate coordinate functions $X_i(\lambda)$ and $Y_i(\lambda)$, parameterized by arc length, with LS polynomials up to degree d, as was described in Section 2

$$c^i_{xy} = (X_{i0}, ... X_{id}; Y_{i0}, ... Y_{id}).$$

    Normalize the sample with respect to position by ignoring the 0-order coefficients X_{i0} and Y_{i0}, and with respect to size by dividing each coefficient by the norm $\sqrt{\sum_{j=1}^{d}(X_{ij}^2 + Y_{ij}^2)}$.
    Approximate I_0 and I_1 with LS polynomials, yielding

$$c^i_{II} = (I^0_{i0}, ... I^0_{id}; I^1_{i0}, ... I^1_{id}).$$

    With Euclidean distance between vector c^i_{II} of the test sample and analogous vectors of training characters: Find T closest CHNN_k. These T classes serve as candidates for the i-th sample in the sequence.
  end for
  for $\alpha = -\beta$ to β by step of 1 degree **do**
    Compute

$$\epsilon_\alpha = \prod_{i=1}^{n} \frac{D^1_{i\alpha}}{\sum_{j=1}^{p} D^j_{i\alpha}} \tag{3}$$

    where $D^j_{i\alpha}$ is the Euclidean distance to the j-th *closest* CHNN_k among the candidate classes T for the sample i in the sequence, rotated by angle α, and p is a parameter to be evaluated. Distance D is computed in the space of coefficients of LS polynomials of coordinate functions.
  end for
  Find $\epsilon_\gamma = \min_{-\beta \leq \alpha \leq \beta} \epsilon_\alpha$
  return n and γ.

$O(d^4)$. It performs much faster in practice, however, because at each recursive call the dimension often drops by more than one.[4]
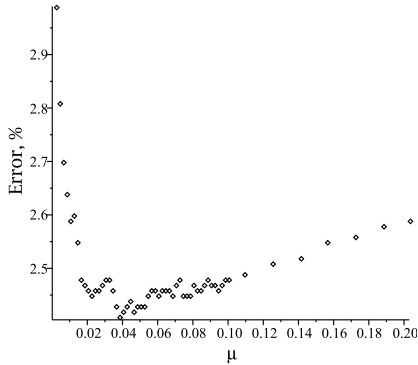

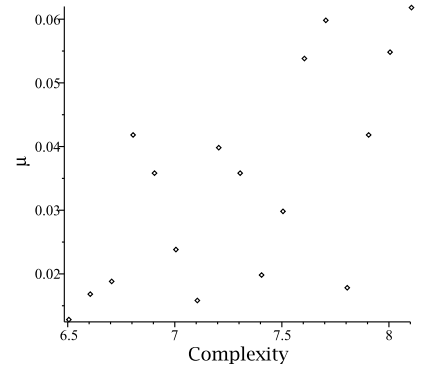Figure 3. Recognition error of non-transformed characters for different values of μ

Figure 4. The optimal values of μ for samples with different complexity

Experimental Setting The testing set is represented in InkML[12] format and contains 50,703 characters in 242 classes. It includes UNIPEN[13] and LaViola[14] databases, as well as symbols collected at the Ontario
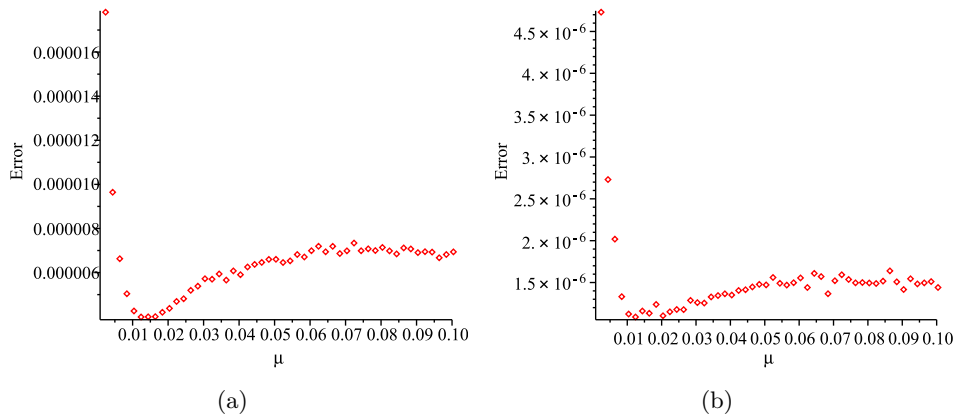
Figure 5. Average maximum error in coefficients of (a) I_0 and (b) I_1 depending on μ

Research Centre for Computer Algebra. Class labels incorporate the number of strokes (therefore, single-stroke and double-stroke "7" are considered as different classes). Examples of the characters from the dataset are given in Figure 2. Further details of the experimental data set have been given earlier.[3]

The experiments are performed with 10-fold cross-validation. The model is trained with non-transformed samples. For the recognition phase, sequences of n characters are taken from the dataset and each sequence is rotated by a random angle $\gamma \in [-\beta, \beta]$.

5. EXPERIMENTAL RESULTS

5.1 Isolated Symbol Classification

Coordinate Functions Figure 3 shows the error rate for recognition using coefficients of the X and Y coordinate functions. An error rate of approximately 2.4% is reached for $\mu = 0.04$ and therefore this value is taken as the optimum for approximation of coordinate functions.

Optimal μ for Characters with Different Complexities We find that the optimal μ value is not strongly correlated with the complexity of characters. On the other hand, the recognition error is correlated with the complexity. Samples with small complexity $4 \leq \eta \leq 4.5$ (most of which are linear symbols such as "-") have 0% classification error for most of values of $\mu \in (0, 0.1]$. Recognition error is increasing with the increase of complexity and reaches 5.8% for samples that have the maximal value of $\eta \approx 8.2$ in our dataset, such as "g". The optimal values of μ for recognition of samples with different complexities are shown in Figure 4. Results of Spearman and Kendall tau-a correlation tests between complexity and μ are respectively: $\rho_{\mu,\eta}(13) = 0.52, p = 0.047$ and $\tau_{\mu,\eta}(13) = 0.38, p = 0.053$.

Integral Invariants Figure 5 shows the average maximum error of the coefficients of integral invariants with respect to different rotation angles as explained in section 3.1. The optimal value of μ, giving minimal error for I_0 and I_1, is found to be 0.012. Thus, for robust rotation-independent classification, each invariant should be approximated with the obtained μ. This value is preferable, since it provides the highest degree of invariance.

Evaluation of Degree of Approximation The recognition error, the maximum absolute approximation error and the average relative approximation error are presented in Table 1. We find degree 12 to be the optimum for recognition of symbols in our collection.

It is interesting that the recognition error starts to increase for $d > 12$. A similar trend applies to the maximum absolute and average relative errors. This confirms that higher order approximation may not be the optimal choice. On one hand it may lead to the Runge's phenomenon and on the other hand it may cause overfitting.

Table 1. The recognition error, the maximum approximation error and the average relative error for different degrees of approximation d, $\mu = 0.04$

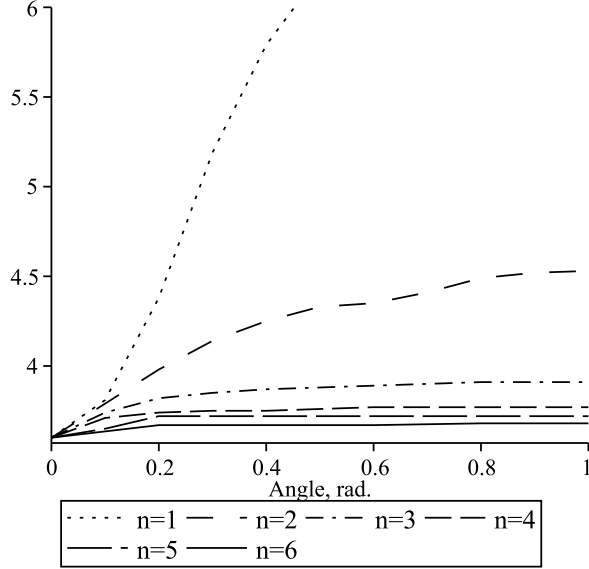| Degree of approximation	9	10	11	12	13	14	15
Recognition error %	2.57	2.49	2.46	2.43	2.44	2.45	2.46
Maximum approximation error	707	539	539	484	475	494	500
Average relative error $(\times 10^{-3})$	1.9	1.6	1.4	1.2	1.1	1.0	1.2

Figure 6. Recognition error (%) for different size of context n and different angles of rotation (in radians)

5.2 In-Context Classification

There are 3 parameters that in-context recognition rate can depend on, see Algorithm 1: the number p of closest classes in computation of error likelihood, the rotation angle, and the size n of the set of characters. To evaluate p, we fix the parameter $n = 3$ and perform classification for values p of 2, 3 and 4. We find that p has almost no effect on recognition error, and therefore, we take $p = 3$ and continue the experiments.

With the fixed value of p, evaluation is performed depending on n and the rotation angle, see Figure 6. A significant reduction in error rate is achieved compared to the results reported earlier,[4] which are equivalent to $n = 1$. The major improvement is obtained if sequences of length $n = 3$ are recognized, rather than 1 or 2. For example, with rotation of 1 radian, $n = 3$ gives an error rate of 3.75% versus 8.2% reported previously. For more accurate classification and/or depending on an application, higher n can be used.

6. CONCLUSION

We have investigated several methods for improvement of orthogonal-series based character recognition, both in the case of isolated characters of known orientation and sequences of characters of unknown orientation. We have found (1) an optimal range of values for the jet scale for coordinate basis functions, (2) that this optimal value of μ, to a first approximation, does not depend on the complexity of the characters tested, (3) optimal values for the jet scale for the integral invariants I_0 and I_1, used for transformation-independent recognition, and (4) the optimal degree of the approximating series. In addition, we developed an in-context rotation-invariant algorithm that yields substantially better results than isolated recognition and can be extended to other transformations.

REFERENCES

[1] Char, B. W. and Watt, S. M., "Representing and characterizing handwritten mathematical symbols through succinct functional approximation," in [*Proc. ICDAR*], 1198–1202, IEEE Computer Society (2007).

[2] Golubitsky, O. and Watt, S. M., "Online stroke modeling for handwriting recognition," in [*Proc. CASCON '08*], 72–80, ACM, New York, NY, USA (2008).

[3] Golubitsky, O. and Watt, S. M., "Distance-based classification of handwritten symbols," *International J. Document Analysis and Recognition* **13**(2), 113–146 (2010).

[4] Golubitsky, O., Mazalov, V., and Watt, S. M., "Orientation-independent recognition of handwritten characters with integral invariants," in [*Proc. Joint Conf. ASCM 2009 and MACIS 2009*], *COE Lecture Notes* **22**, 252–261, Kyushu University, Japan (Dec. 2009).

[5] Golubitsky, O., Mazalov, V., and Watt, S. M., "Toward affine recognition of handwritten mathematical characters," in [*Proc. International Workshop on Document Analysis Systems, (DAS 2010)*], 35–42, ACM Press, Boston, USA (June 9-11 2010).

[6] Tokuno, J., Inami, N., Matsuda, S., Nakai, M., Shimodaira, H., and Sagayama, S., "Context-dependent substroke model for hmm-based on-line handwriting recognition," in [*Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*], 78 – 83 (2002).

[7] Bianne, A.-L., Kermorvant, C., and Likforman-Sulem, L., "Context-dependent hmm modeling using tree-based clustering for the recognition of handwritten words," in [*Proc. of the Document Recognition & Retrieval XVII*], (2010).

[8] Wang, J., Neskovic, P., and Cooper, L., "A probabilistic model for cursive handwriting recognition using spatial context," in [*Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*], **5**, v/201 – v/204 Vol. 5 (march 2005).

[9] Feng, S., Kogan, I., and Krim, H., "Classification of curves in 2d and 3d via affine integral signatures," *Acta Applicandae Mathematicae* **109**, 903–937 (2010). 10.1007/s10440-008-9353-9.

[10] Smirnova, E. and Watt, S. M., "Context-sensitive mathematical character recognition," in [*Proc. IAPR International Conference on Frontiers in Handwriting Recognition, (ICFHR 2008)*], 604–610, CENPARMI Concordia University, ISBN 1-895193-03-6, Montreal, Canada (August 19-21 2008 2008).

[11] Watt, S. M., "An empirical measure on the set of symbols occurring in engineering mathematics texts," in [*Proceedings of the 2008 The Eighth IAPR International Workshop on Document Analysis Systems*], 557–564, IEEE Computer Society, Washington, DC, USA (2008).

[12] Watt, S. and Underhill, T., "Ink markup language (InkML)." `http://www.w3.org/TR/InkML/`. (valid on November 15, 2011).

[13] Guyon, I., Schomaker, L., Planiondon, R., Liberman, M., and Janet, S., "Unipen project of on-line data exchange and recognizer benchmarks," in [*Proc. 12th International Conference on Pattern Recognition (ICPR 1994)*], 29–33, IAPR-IEEE, Jerusalem, Israel (1994).

[14] LaViola Jr., J. J., "Symbol recognition dataset," tech. rep., Microsoft Center for Research on Pen-Centric Computing. (valid on November 15, 2011).