



**HAL**  
open science

# Multi-Criteria Handwriting Quality Analysis with Online Fuzzy Models

Damien Simonnet, Eric Anquetil, Manuel Bouillon

► **To cite this version:**

Damien Simonnet, Eric Anquetil, Manuel Bouillon. Multi-Criteria Handwriting Quality Analysis with Online Fuzzy Models. *Pattern Recognition*, 2017, 69, pp.310-324. 10.1016/j.patcog.2017.04.003 . hal-01515397

**HAL Id: hal-01515397**

**<https://hal.science/hal-01515397v1>**

Submitted on 27 Apr 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Multi-Criteria Handwriting Quality Analysis with Online Fuzzy Models

Damien Simonnet, Eric Anquetil and Manuel Bouillon

*INSA Rennes, UMR IRISA, Campus de Beaulieu, F-35042 Rennes, France*

---

## Abstract

*IntuiScript* is an innovative project aiming at the development of a digital workbook providing feedback during the handwriting learning process for children from three to seven years old. In this context, the paper presents a method to analyse handwriting quality that responds to the expectations of the *IntuiScript* educational scenario: on-line and real time feedback for children, an automatic detection of children mistakes guiding the pedagogical progression, and a precise analysis of children writing saved to help teacher to understand children writing skills. The presented method introduces a *multi-criteria* architecture to analyse handwriting quality based on three different aspects: *shape*, *order* and *direction*. The validation of the proposed approach is done on a realistic dataset collected in preschools and primary schools with 952 children. Results show a positive feedback of children and teachers about the use of tactile digital devices, and a significant improvement of the performances of the *multi-criteria* architecture compared to the previous analyser. The ground truth has been annotated by experts with different levels of confidence. Specific evaluation metrics are introduced to deal with confidence annotations.

*Keywords:* Handwriting Analysis, Pen-Based Interaction, Educational System

---

## 1. Introduction

Jolly *et al.* [1] have presented a comparative study about the acquisition of the training handwriting between digital devices and paper. Results show a significant improvement of children trained on the digital device compared to children trained on paper, notably in terms of fluency (decreasing of the in-air time, and stopping time). In this context, this paper tackles the problem of handwriting quality analysis. More precisely, this work takes place in an innovative project *IntuiScript*<sup>1</sup> founded by the French government to take advantages of the digital devices in primary schools. It aims at developing a

---

<sup>1</sup><http://intuiscript.com>

10 digital workbook to help teachers and children from three to seven years old during the handwriting learning process. This work contributes to the *IntuiScript* educational scenario with block letters on three main aspects:

- it allows children to work in autonomy with on-line and real time feedback;
- it proposes automatically pedagogical exercises that are adapted to children difficulties based on the automatic analysis of children writing;
- 15 • it provides a precise off-line analysis of children writing (*i.e. order, direction, shape*) to help teachers to understand children writing skills and difficulties.

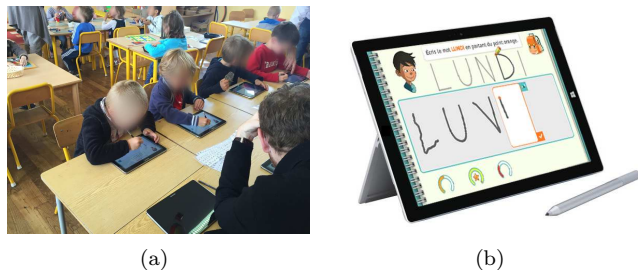


Figure 1: First in-class experiment (a) of the *IntuiScript* project with tablet tactile devices (b) about the analysis of week word letters.

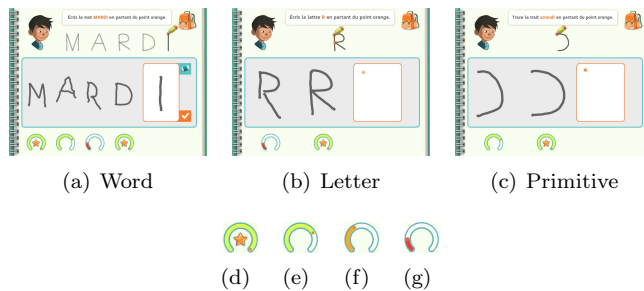


Figure 2: Hierarchical methodology that enables children to be autonomous by first analysing letters of a word (a), and then working on letters (b) and primitives (c) if it is required. On-line feedback is given with a colour scale indicator drawn under each gesture with the following colouring code: very good (green with yellow star) (d), good (green) (e), average (orange) (f) and incorrect (red) (g).

20 The main objective of the *IntuiScript* project is to offer an advanced digital writing experience at school by using tablet and tactile digital devices (with finger touch and stylus). The validation of the research project is based on experiments performed in schools from half a day to a month with more than

thousand of French primary school students in Brittany. This paper presents results of the first experimentation conducted in 20 preschools with 952 children. During a two hours session, groups of seven/eight children as illustrated in Figure 1 and Figure 2 were drawing the block letters of the day of the week and having automatic feedback with a colour scale indicator reflecting the correctness of written letters.

The analyse of the handwriting quality is related to legibility (*shape*) and ductus (*order*, *direction*) aspects. During the early learning stage of the writing, it is very important to have strong constraints on the ductus to acquire a correct handwriting coordination. However, during their handwriting learning process, children acquire their own handwriting style in which there are tolerances for ductus constraints, and the legibility becomes the main aspect evaluated. The legibility score decreases continuously with the deformation of a symbol. On the opposite, scores related to the ductus correspond to a binary decision, where educational experts do not always agree on a single writing convention, but think that children must have rules defining uniquely how to write correctly a symbol. That is why the French Ministry of Education defines documents that describe a writing convention which is a recommendation for teacher as illustrated in Figure 3. In this case, the gesture correctness is related to the notion of *elementary strokes* (e.g. horizontal/vertical/slanted line, half-circle) where a correct *order* and *direction* correspond respectively to a correct sequence and direction of each *elementary stroke*. Therefore, the system presented in the paper allows to return an adapted feedback corresponding to children skills and balanced between legibility and ductus aspects.

In this paper, three main criteria are used to analyse the handwriting quality: *shape*, *order* and *direction*. Similarly to [2, 3, 4, 5], the proposed approach evaluates the handwriting quality with regards to reference models using *intra-class* and *inter-class* scores (i.e. the similarity to the analysed model and the difference to other models, see Figure 4). Compared to the previous approach [5], specialised features and classifiers are introduced for each criterion.

This work presents a *multi-criteria* approach that analyses *shape*, *order* and *direction* aspects of children handwriting. It is also seen as an extension of [5] that analyses symbols with an evolving fuzzy inference system [6] and HBF49 [7] features. More precisely, this paper contains five main contributions by introducing:

- a *multi-criteria* approach combining specialised classifiers (*shape*, *order* and *direction*) about legibility and ductus, with hierarchical dependencies (e.g. *order* and *direction* depend on *shape*) dedicated to specific problems, see Section 4;
- a confidence score meaningful for teachers and children depending on two aspects: the distance to the theoretical model and the confusion with other models, see Section 4;
- a methodology to check the order correctness based on the identification

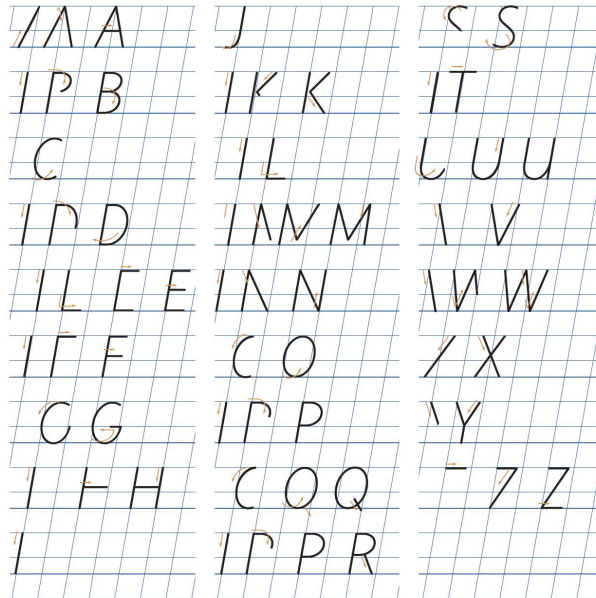


Figure 3: Recommendation of the French Ministry of Education relatively to the ductus of block letters (extracted from the document entitled ‘Modeles d’écriture scolaire. Document d’accompagnement. Police de caracteres cursives’ (June 2013)).

of elementary strokes represented in a model as median strokes (see Sections 5.2.2 and 6.2.2);

- features to characterise finely shape and direction (fuzzy histogram of direction and orientation), see Sections 5.2.1 and 5.2.3;
- 70 • feedback of teachers and children about the introduction of tactile digital devices in the learning writing process. A database from in-class experiments has been created with a ground truth and made publicly available for comparison with future works on handwriting *analysis*.

The rest of the paper is organised as follows. First, related work is presented  
75 in Section 2. The educational scenario associated to the *IntuiScript* project and an overview of the approach are detailed in Section 3 and 4. Analysis-based features and classifiers are respectively described in Section 5 and 6. Then, Section 7 presents experiments and results. Sections 8 and 9 conclude this paper.

## 80 2. Related Work

In the literature, handwriting quality is related to legibility and kinematic [8]. The former corresponds to letter *shape* and its associated readiness. The latter

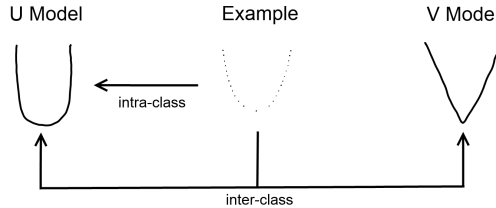


Figure 4: Analysis of an example U. The analysis approach is based on two complementary aspects: intra-class (*i.e.* the similarity of the example with the analysed model:  $U$ ) and inter-class (*i.e.* the difference with other models indicating by a confusion degree of other classes, *e.g.*  $V$ ) scores.

concentrates on the writing process (*e.g.* *order, direction, fluidity*) that must be efficient, as writing is a fundamental skill that is necessary for learning and using  
 85 knowledge. More precisely, the writing process requires three central skills: a complete visual representation of each letter, a recognition of *elementary strokes* in letters, and the ability to reproduce a letter as a sequence of *elementary strokes* respecting the direction [9].

There are two typical applications resulting from the handwriting analysis:  
 90 medical and education systems. Guinet *et al.* [10] focus on the kinematic aspect (*i.e.* velocity, duration, fluency, pauses) to detect handwriting pathologies. Jolly *et al.* [11] analyse the handwriting velocity to identify developmental coordination disorders. Accardo *et al.* [12] study handwriting kinematics of children on digital tablets with a various number of features (*e.g.* number of strokes per  
 95 letter, peak velocities, and pen lift duration) and show that the handwriting is related to four kinematic domains: velocity, spatial arrangement, automation and motor planning plan. Hu *et al.* [13] present an educational system that identifies three types of errors for Chinese characters: stroke production errors (*i.e.* stroke reversal, concatenation of separate strokes, broken strokes), stroke  
 100 sequence errors (*i.e.* wrong sequencing of components, wrong stroke sequences in a component) [14] and stroke relationship errors (*e.g.* relative length and position) [15]. Falk *et al.* [3] use five primitives (legibility, form, alignment, size and space) to quantify handwriting proficiency of children. More generally, approaches for medical systems such as [10, 11, 12] use mainly the kinematic aspect  
 105 by opposition to educational systems [13, 3] that pay more attention to legibility features, and use simple kinematic features (*e.g.* velocity and acceleration).

The legibility of the writing is often evaluated with low-level features related to the letter recognition task that makes difficult to evaluate quantitatively the correctness of a criterion (*e.g.* shape in [4]). Therefore, specific confidence scores  
 110 (*e.g.* a confidence based on a modified quadratic discriminant function [16]) are often used to assess quantitatively this correctness. Following this aspect, the

use of a reference model to compute confidence values is similar to applications related to the signature analysis [17, 18] where the aim is to evaluate if the handwriting of the signature corresponds to its owner. This underlines an important aspect of this paper which is the difference between recognition and analysis tasks. The recognition consists in identifying a symbol independently of its quality and writing style. On the opposite, the analysis evaluates the quality of a symbol known by comparing it to a reference model, corresponding in this paper to teacher's expectations about handwriting.

This paper presents a handwriting analyser for educational systems which uses static and dynamic examples. One important aspect is that a dynamic example - showing to children how to reproduce block letters - is a specific feature only possible with the use of digital devices. The proposed approach makes a step forward by introducing a finer analysis based on four criteria: *multi-criteria*, *shape*, *order* and *direction*. Firstly, a *multi-criteria* score summarises in real-time legibility and kinematic aspects for children, according to the teacher expectations. Secondly, a hierarchical approach presents for teachers the legibility aspect (*shape*) followed in case of the correctness by a kinematic analysis (*order* and *direction*). Finally, a fundamental criterion to assess the quality of the cursive writing is the *fluidity* [19]. However, it will be considered later with the second step of the *IntuiScript* experiments related to the cursive analysis, and is therefore outside of the scope of this paper.

### 3. IntuiScript Educational Scenario

This paper investigates the handwriting quality analysis for its use in an educational system to give an on-line feedback to children and off-line feedback to teachers. The on-line feedback helps children by providing a simple score to measure their improvements and to work in autonomy. The off-line feedback is composed of a set of more detailed feedback that can be used by teachers. In-class experiments presented in this paper target the validation of this handwriting analysis methodology. The educational scenario and digital workbook design will be evaluated at the end of the *IntuiScript* project with several months of experiment and more than 1000 children. Nevertheless, this project is based on a user-centered design approach that includes several cycles of conception followed by experiments. Therefore, feedback of children and teachers related to this experiment will be used to improve the next version of the education scenario. The rest of this section presents the current educational scenario proposed by *IntuiScript* educational experts for learning words based on block letters.

The scenario illustrated by Figure 2 is based on three hierarchical steps adapted to the skills and difficulties of each child. It is important to note that the automatic adaptation of pedagogical exercises is only possible with an on-line handwriting analysis, such as the one presented in this paper. First, based on examples containing rules about *shape*, *order* and *direction*, children have to write each letter of a given word and an on-line feedback is returned (see Figure 2(a)). In the application, examples correspond to dynamic animations

that can be repeated as many times as children want. Then, if the system detects some incorrectly letters drawn, it proposes to children to work on individual letters (see Figure 2(b)). An on-line feedback allows them to measure autonomously their progress. Finally, if the system detects that the handwriting is still incorrect, an exercise with letter primitives (e.g. horizontal/vertical/slant line, loop) is given with the same approach than individual letters (see Figure 2(c)).

#### 4. Multi-Criteria Gesture Analysis

The handwriting *analysis* task is fundamentally different from the *recognition* task that focuses on the recognition of a letter independently of its quality and its writing style. The input of this system illustrated in Figure 5 corresponds to a gesture acquired from a digital device (a) that can be decomposed in *strokes* (delimited by pen up) (b), *elementary strokes* (delimited by singularity points) (c) and *median strokes* (median segments of *elementary strokes*) (d).

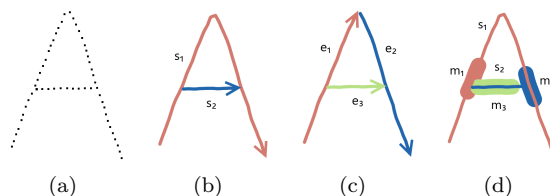


Figure 5: The original gesture from a digital device (a), and its decomposition in *strokes* (delimited by pen up) (b) and *elementary strokes* (c) with the drawing direction and order. *Median strokes* are represented by thicker lines in (d).

In the *analysis*, the letter is already known but the aim is to evaluate its quality by comparing it against a reference model under various criteria that can be interdependent (e.g. *order* depends on *shape*). Two complementary measurements are related to this process: an *intra-class* measure describing the closeness to the model of the expected letter, and an *inter-class* measure which quantifies its discriminatory quality with respect to other symbols.

The *analysis* system depicted in Figure 6 is based on a *hierarchical* approach with three criteria: *shape*, *order* and *direction*. Each criterion can be used by teachers during the remediation to identify errors and children difficulties. A refined *multi-criteria* confidence gives to children a score related to legibility and ductus aspects.

More precisely, the *elementary-analyser* presented in Figure 7 consists in extracting features that are used by *discriminative* and *generative* models to compute *inter-class* and *intra-class* scores. These scores are then merged with the *fusion* model to give a final confidence score with regards the symbol analysed. In the rest this paper, the denomination *analysis-based classifier* refers to a set of *discriminative*, *generative* and *fusion* models.



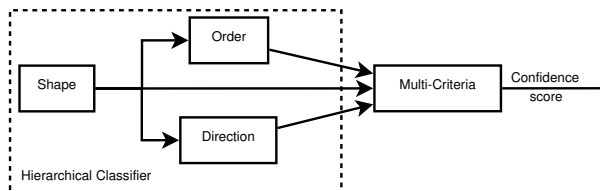


Figure 6: Multi-Criteria Analyser : Overview of the system composed of three sub-criteria: *shape*, *order* and *direction* based on hierarchical dependencies.

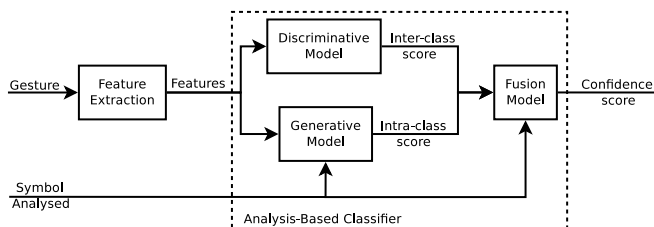


Figure 7: Workflow of the elementary-analyser that gives a confidence about a gesture (or symbol). This analyser is used in *shape*, *order* and *direction* analysers.

The two following sections describe *analysis-based* features and classifiers specific for each criterion, followed by the evaluation in Section 7 comparing the results with this approach obtained against an approach using a *recognition-based* methodology [5].

## 5. Analysis-Based Features

The analysis is based on a set of specific features to characterise different aspects of an handwriting letter, and used as input for an *analysis-based* classifier.

In the previous approach [5], features for each criterion were a subset of a set of *recognition-based* features : HBF49. In this paper, a new set of dedicated *analysis-based* features are defined for each criterion. The rest of this section presents these two approaches.

### 5.1. A baseline with HBF49 features

HBF49 [7] is a generic set of features designed for handwriting symbols recognition. More precisely, it is composed of *dynamic* features that depend on the writing process (*e.g.* starting and ending positions, proportion of downstrokes trajectory, angle of the initial vector, inflexions), and *visual* features that focus on the appearance of the written results (*e.g.* 2D histogram of point, k-perpendicularity, k-angle). This set of features captures multiple criteria: *shape*, *order*, *direction* and *global* as depicted in Table 1.

Table 1: Description of *HBF49* features [7] with regards *global*, *shape*, *order* and *direction* criteria. Each number refers to the features identifier described in [7]. Features 6 and 7 characterise two criteria: *order* and *direction*.

| Type    | Criteria  | Features   | Number |
|---------|-----------|--|--------|
| Dynamic | Order     | HBF49 (1-4): starting and ending points position   | 4      |
|         |           | HBF49 (6-7): first point to last point vector      | 2      |
|         | Direction | HBF49 (6-7): first point to last point vector      | 2      |
|         |           | HBF49 (9-10): angle to initial vector              | 2      |
|         | Global    | HBF49 (5): first point to last point vector (norm) | 1      |
|         |           | HBF49 (8): closure                                 | 1      |
|         |           | HBF49 (11-12): inflexions                          | 2      |
|         |           | HBF49 (13): proportion of downstrokes trajectory   | 1      |
|         |           | HBF49 (14): number of strokes                      | 1      |
| Visual  | Direction | HBF49 (19): average direction                      | 1      |
|         |           | HBF49 (24-27): absolute angle histogram            | 4      |
|         | Shape     | HBF49 (28-31): relative angle histogram            | 4      |
|         |           | HBF49 (15): bounding box diagonal angle            | 1      |
|         |           | HBF49 (18): deviation                              | 1      |
|         |           | HBF49 (32-40): 2D histogram                        | 9      |
|         | Global    | HBF49 (41-47): hu moments                          | 7      |
|         |           | HBF49 (48-49): convex hull                         | 2      |
|         |           | HBF49 (16-17): trajectory length                   | 2      |
|         |           | HBF49 (20-21): curvature and perpendicularity      | 2      |
|         |           | HBF49 (22-23): k-perpendicularity, k-angle         | 2      |

Table 2: *RECOGNITION-BASED* features of the previous work [5] used for *Shape*, *Order* and *Direction* criteria.

| Criteria      | Features                  | Number |
|---------------|---------------------------|--------|
| Shape [5]     | HBF49 (15, 16, 17, 32-40) | 11     |
| Order [5]     | HBF49 (1-7)               | 7      |
| Direction [5] | HBF49 (13, 24-31)         | 9      |

Some set of *analysis-based* features, presented in Table 2, were build in [5] by selecting a subset of HBF49 features for each criterion. However, HBF49 has been designed to recognise handwritten symbols and not to analyse specifically the correctness of a criterion with regards to a reference model. For instance, HBF49 does not contain any precise features to characterise *order* and *direction*. *Order* features used in [5] (*i.e.* HBF49 (1-7) ) capture starting and ending position but are not robust enough to detect order changes in *elementary strokes*. Similarly, *direction* features [5] (*i.e.* HBF49 (13, 24-31) ) focus on the proportion of down-strokes trajectory and the global orientation that is not sufficient to characterise changes of direction related to *elementary strokes*.

## 5.2. A new set of analyse-based features

The set of new *analyse-based* features introduced in this paper is sum up in Table 3 and described as follows where each criterion must be independent.

### 5.2.1. Shape

The proposed *shape* features consist firstly in selecting HBF49 features that do not depend on order and direction (*e.g.* 2D histogram of points [32 – 40]),

Table 3: Features used for the four criteria based on HBF49 [7] and specific analysis-based features.

| Criteria  | Reuse of HBF49 [7] features                      | New Analysis-based features  | Number |
|-----------|--|--|--------|
| Shape     | HBF49 (15, 18, 20, 21, 24-27, 32-42, 44, 48, 49) | Fuzzy Histogram of Orientation O(1-48)   | 69     |
| Order     |  | HBF49 (15, 18, 20, 21, 24-27, 32-42, 44, 48, 49) features applied on each <i>elementary stroke</i> | 21     |
| Direction |  | Fuzzy Histogram of Directions D(1-48)  | 48     |

absolute angle histogram [24 – 27], hu moments [41 – 42, 44], hull area [48, 49]). Secondly, a novel set of features is added to capture finely shape changes because most of HBF49 features are global except the 2D histogram of points that lacks of precision (*e.g.* a square and its included circle have very close 2D histogram of points but shapes are fundamentally different). These orientation features presented in Figure 8 are based on a  $4 \times 3$  grid in which a four bins histogram - associated to the range  $[0, \pi]$  - with a fuzzy quantification is built for each cell. Consequently, 48 features are added, and the set of *shape* features is composed of 69 features. The use of the fuzzy theory [20, 21] in these features allows to be more robust to evaluate the quality of shape drawn by children. During the learning writing process, children writing can have a stroke positioning less precised that adults and the use fuzzy theory in features allows to capture with less training samples a description of the orientation of a given model. First, non-fuzzy histograms of orientation or direction are computed in each cell. Then, these histograms contribute to the fuzzy histogram of the closest cells with a fuzzy voting. This voting depends on the contributions of each segment in a cell. As illustrated in Figure 8(d), the fuzzy weight is related to the distance between the segment and closest cells (4 among the 9 around the segment). The distance is expressed as an exponential function of the Manhattan distance between the gravity centre of a cell and a segment. More precisely, the equation (1) defines the contribution of a segment to the fuzzy weight associated to a cell  $(i, j)$  where  $g_{seg}$  and  $g_{i,j}$  are respectively the gravity centre of the segment considered and the cell  $(i, j)$ , and  $\mathcal{C}_{cells}$  corresponds the set of closest cells to the current segment.

$$w(g_{seg}, g_{i,j}) = \frac{\exp(-\|g_{seg}g_{i,j}\|_1)}{\sum_{(k,l) \in \mathcal{C}_{cells}} \exp(-\|g_{seg}g_{k,l}\|_1)} \quad (1)$$

### 5.2.2. Order

The correct *order* of a gesture corresponds to the correct order of the associated *elementary strokes* (see Figure 5(c)). The identification of the correct *order* is based on the median part of strokes (see Figure 10) which is a part of *elementary stroke* that is invariant to direction changes. During the learning process, median strokes are computed from a letter fully decomposed in *elementary strokes* as presented in Figure 12. The classification stage, described

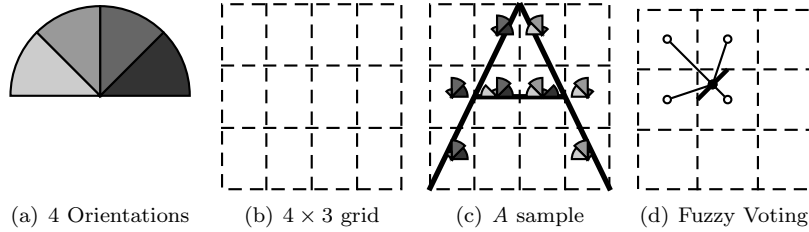


Figure 8: Shape Features: Fuzzy Histogram of Orientation  $O(1-48)$ . (c) illustrates the feature computation on the letter A based on a  $4 \times 3$  grid (b) on which a four bins histogram associated to four orientations (a) is computed. The fuzzy voting (d) is based on the contribution of each segment to the closest cells (4 among the 9 around the segment) as expressed in equation (1).

in Section 6.2.2 and Figure 13, searches the sub-part that matches the best each *elementary stroke* of the learnt model. The features used on the median  
 230 strokes are global shape features that are invariant to the direction (see Table 3). These features do not include fuzzy histogram of orientations because *elementary strokes* are simple shapes that can be discriminated with spatial position features (e.g. 2D histogram of points) and the global orientation features (e.g. absolute angle histogram). Consequently, 21 features are used in the set  
 235 of *order* features.

### 5.2.3. Direction

To capture finely local direction changes, a fuzzy histogram of direction based on a  $4 \times 3$  grid is introduced in Figure 9. For each cell, a four bins histogram - associated to the range  $[\frac{-\pi}{2}, \frac{3\pi}{2}]$  - with a fuzzy quantification is  
 240 built. Consequently, 48 features are added and constitute the set of *direction* features.

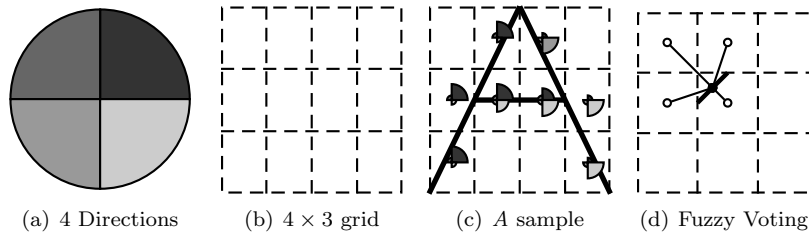


Figure 9: Direction Features: Fuzzy Histogram of Direction  $D(1-48)$ . (c) illustrates the feature computation on the letter A based on a  $4 \times 3$  grid (b) on which a four bins histogram associated to four directions (a) is computed. The fuzzy voting (d) is based on the contribution of each segment to the closest cells (4 among the 9 around the segment) as expressed in equation (1).

The next section describes the different classification systems used with these input features.

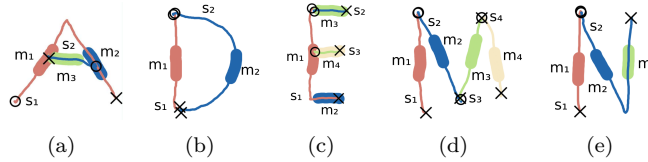


Figure 10: Median strokes coloured and numbered from the first to the fourth with red, blue, green and brown. Strokes are coloured with the same colour code.

## 6. Analysis-based Classifiers

245 In the previous approach [5], a single classifier (evolving fuzzy inference system [6]) was used with specific feature sets for each criterion.

In this paper, in addition of specifics features presented in the previous section, a specific classifier is introduced for each criteria. First, the *elementary analyser* described in Figure 7 is presented. Then, specifics classifiers using the *elementary analyser* for *shape*, *direction* and *order* criteria are described. 250 Finally, a *multi-criteria* analyser corresponding to Figure 6 is explained.

In the rest of this paper, a gesture is defined as a set of strokes as illustrated in Figure 5(b). More formally, each stroke is represented as a set of points  $x_g = \cup_{i=1}^{N_p} p_i$  where a point  $p_i$  is represented by its  $x$  and  $y$  coordinates and time values. 255  $\mathbf{G}$  represents the set of gestures that is currently analysed.  $g$  represents one gesture or class of the set  $\mathbf{G}$ .

### 6.1. Elementary Analyser

Supervised learning algorithms use a *generative* and *discriminative* analysis to compute *intra-class* and *inter-class* scores (see Figure 4). These algorithms 260 take as input a set of labelled data  $(x_{c_g, k})_{k \in \llbracket 1, N_g \rrbracket}$  where  $N_g$  is the number of samples available for a gesture  $g$ . More precisely, each sample is represented by a set of features.

The *generative* algorithm builds a prototype  $\mathbf{M}$  that is used to compute an *intra-class* score, also called *absolute confidence*. The *discriminative* algorithm 265 generates a model  $\mathbf{MD}$  that discriminates samples among classes using the prototype  $\mathbf{M}$  and features extracted. The *inter-class* score resulting from this process is also called *relative confidence*. These algorithms are incremental algorithms to let teachers expend models with their own drawing. The first builds a prototype as a set of clusters and the second is a classifier based on a fuzzy 270 inference system. The rest of this section described in details each algorithm of the *elementary analyser* as follows:

#### 6.1.1. Generative Model

For each gesture (or class)  $g \in \mathbf{G}$ , a cluster  $c_g$  is represented by a mean vector  $\mu_{c_g}$  and a covariance matrix  $\Sigma_{c_g}$ . As in [22], an incremental process is used to compute the mean vector and the precision matrix  $\Phi_{c_g} = \Sigma_{c_g}^{-1}$  that are respectively expressed by equations (2) and (3) where  $x_{c_g, k}$  is the  $k$ -th

sample relative to the gesture  $g$ ,  $\alpha_k = 1/k$ ,  $\Psi_{c_g,k} = x_{c_g,k} - \mu_{c,k}$  and  $\lambda_{c_g,k} = \Psi_{c_g,k}^T \Phi_{c_g,k-1} \Psi_{c_g,k}$ .

$$\mu_{c_g,k} = (1 - \alpha_k) \cdot \mu_{c_g,k-1} + \alpha_k \cdot x_{c_g,k} \quad (2)$$

$$\Phi_{c_g,k} = \frac{\Phi_{c_g,k-1}}{1 - \alpha_k} - \frac{\alpha_k \cdot \Phi_{c_g,k-1} \Psi_{c_g,k}^T \Psi_{c_g,k} \Phi_{c_g,k-1}}{(1 - \alpha_k) \cdot [1 + \alpha_k(\lambda_{c_g,k} - 1)]} \quad (3)$$

Finally, the prototype resulting from the supervised clustering is  $\mathbf{M} = (\mu_{c_g}, \Phi_{c_g})_{g \in \mathbf{G}}$  where the Mahalanobis distance  $d_m$  quantifies the closeness of a sample  $x$  to the cluster  $c_g$ . The *absolute confidence* described in Section 6.2.1 is based on this distance.

$$\forall g \in \mathbf{G}, d_m(x, c_g) = \sqrt{(x - \mu_{c_g})^T \Phi_{c_g} (x - \mu_{c_g})} \quad (4)$$

### 6.1.2. Discriminative Model

An evolving fuzzy inference system named *Evolve* [6, 22] is used in addition of the *generative* model to build decision boundaries as linear functions of the prototype  $\mathbf{M}$  and input features. This process, which is equivalent to a neural network, is particularly efficient to improve the discriminatory power of the system by increasing the precision of the class boundaries between prototypes. The model resulting from this learning process is defined as:

$$\mathbf{MD} = (\mu_{c_g}, \Phi_{c_g})_{g \in \mathbf{G}}, E_c \quad (5)$$

where  $E_c$  represents the linear functions.  $(s_g)_{g \in \mathbf{G}}$  corresponds to a vector of the membership degrees relatively to all classes in  $\mathbf{G}$  and is used to compute the *relative confidence* described in Section 6.2.1.

### 6.1.3. Fusion Model

The fusion model combines results of the *generative* and *discriminative* models with a conjunction between two comparative metrics: the *absolute confidence*  $c_a$  which is the closeness to a cluster model resulting from the *generative* model, and the *relative confidence*  $c_r$  which is the confidence to not confuse gestures resulting from the *discriminative* model.

This paper makes a contribution by introducing a probabilistic version of the confidence defined in equation (6) where  $x_g$  represents a gesture to analyse as a  $g \in \mathbf{G}$ .

$$c(x_g) = c_a(d_m(x_g, c_g)) \cdot c_r(x_g) \quad (6)$$

## 6.2. Specialised Analysers

Three specialised analysers are defined: the *confidence-based classifier*, *binary order classifier* and *binary direction classifier* that are used respectively for *shape*, *order* and *direction* criteria.

### 6.2.1. Confidence-Based Classifier

The confidence-based classifier is based on the *fusion* model. It introduces a method to normalise the confidence score to be meaningful for teachers and children.

Firstly, the *absolute confidence* is computed with the *generative* model and a set of positive samples to define a confidence function using a Log-Cauchy distribution. This confidence function defined in equation (7) corresponds to its reliability function where the median and the median absolute deviation of  $\ln(x)$  are robust estimators of parameters  $\mu$  and  $\sigma$  [23].

$$c_a(x) = c_{\mu,\sigma}(x) = \frac{1}{2} - \frac{1}{\pi} \arctan \left[ \frac{\ln(x) - \mu}{\sigma} \right] \quad (7)$$

However, using this estimator does not correspond to the confidence of positive samples as the computed confidence varies from 0 to 1 for positive samples. To give a useful feedback to children, it is expected that the confidence of negative and positive samples correspond respectively to ranges 0–0.5 and 0.5–1.0. Consequently,  $\mu$  must be estimated as the rejection threshold between positive and negative samples, and  $\sigma$  as the median absolute deviation.

The estimation process illustrated in Figure 11 works as follows. First, for each gesture  $g$ , positive and negative samples are generated with artificial data synthesis by geometrical distortion [24] (rotation, inclination, curvature changes). Then, the rejection threshold relatively to the Mahalanobis distance  $\mu_g$  is learnt by using a greedy algorithm that minimises a function depending on the ratio of false positives  $r_{FP}$  and false negatives  $r_{FN}$  defined in equation (8). The parameter  $\beta_g$  fixes the importance of having less false negatives than false positives that depends on the application.

$$e^{\mu_g} = \arg \min_{\theta \in [0, +\infty[} [\beta_g \cdot r_{FN}(\theta) + (1 - \beta_g) \cdot r_{FP}(\theta)] \quad (8)$$

Secondly, the *relative confidence* defined in equation (10) is based on the output of the *discriminative* classifier and extends the definition of [5] corresponding to equation (9) by penalising continuously the score when there is a confusion ( $s_g$  corresponds to the membership degree relatively to the class  $g$ ).

$$c_r^o(x_g) = \frac{s_g - \max_{p \in \mathbf{G} \setminus \{s_g\}} s_p}{\max_{p \in \mathbf{G}} s_p} \quad (9)$$

$$c_r(x_g) = \begin{cases} 1 & \text{if } c_r^o(x_g) \geq 0 \\ 1 + c_r^o(x_g) & \text{if } c_r^o(x_g) < 0 \end{cases} \quad (10)$$

### 6.2.2. Binary Order Classifier

*Order* is a binary criterion that is considered correct when all the *elementary strokes* have been drawn in the correct order. Each *elementary stroke* is represented as its median stroke of size  $s_m$  that is invariant to drawing directions.

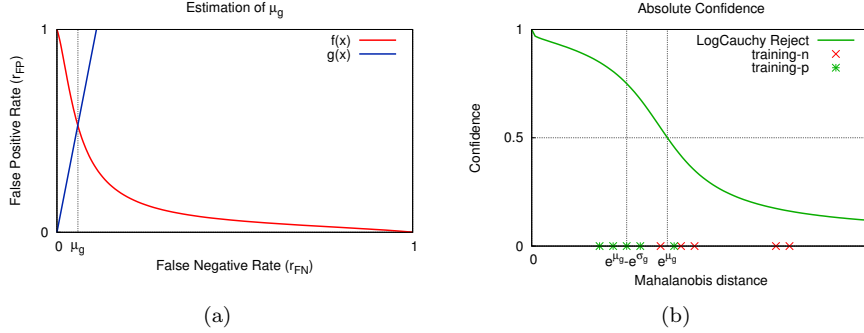


Figure 11: Building of the LogCauchy rejection confidence from positive (training-p) and negative (training-n) samples.  $\mu_g$  is computed with equation (8) and corresponds graphically (a) to the intersection of the FN/FP curve  $f(x) = r_{FP}(x)$  and  $g(x) = \frac{\beta_g \cdot x}{1 - \beta_g}$ . (b) illustrates the building of the LogCauchy reject confidence with regards to positive and negative training samples. It is noted that the confidence is 0.5 for the rejection value:  $e^{\mu_g}$ .

A *discriminative* model is used to discriminate them efficiently. Therefore, the order model  $\mathbf{MD}_o$  is defined by equation (11) and depicted in Figure 12 where  $N_e^g$  is the number of *elementary strokes* associated to a gesture  $g$  and  $E_c^g$  are the rules conclusion.

$$\mathbf{MD}_o = \left[ (\mu_{c_g^k}, \Phi_{c_g^k})_{k \in \llbracket 1, N_e^g \rrbracket}, E_c^g \right]_{g \in \mathbf{G}} \quad (11)$$

During the analysis, a spatio-temporal decomposition resulting from a spatial sampling applied to the original gesture is used to build the gesture sequence  $(g_t)_{t \in \llbracket 1, N_p^s - s_m \rrbracket}$  where

$$g_t = \cup_{i=t}^{t+s_m} p_i^s \quad (12)$$

and  $g^s = \cup_{i=1}^{N_p^s} p_i^s$  is the gesture after applying the spatial sampling. Then, given a gesture  $g$ , elements of the gesture sequence are compared to median stroke models to identify the  $e_g$  elements that have the best matching based on *discriminative* distances of the class recognised by the *generative* model (see the sliding window in Figure 13). More precisely, for each median stroke  $k$ , a set of stroke candidates  $\mathbf{G}_c$  is built with strokes having the best *generative* scores among the gesture sequence, and the best matching is found as follows:

$$\arg \min_{g \in \mathbf{G}_c} d_m(g, c_g^k) \quad (13)$$

where  $c_g^k$  represents the cluster of the  $k$ -th median stroke. Finally, the order is correct if all elements matched are in the same order than the median-stroke model.



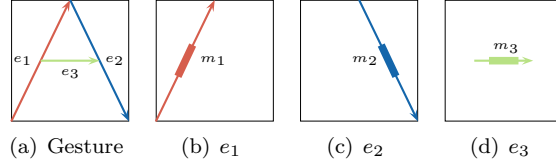


Figure 12: During the learning stage of the *order* classifier, each gesture (a) is decomposed in *elementary strokes*. Then, for each *elementary stroke* ( $e_k$ ), features are extracted from the *median strokes* (b) (c) (d) represented as thicker lines, and used to learn a classifier.

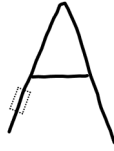


Figure 13: During the *order* analysis, an approach based on a sliding windows (represented with the dotted line rectangle) is used to search the best substrokes corresponding to the *elementary strokes* modelled in Figure 12.

### 6.2.3. Binary Direction Classifier

Similarly to *order*, *direction* is a binary criterion that is considered correct when all the *elementary strokes* are in the correct direction independently of the order. For each gesture  $g$ , a *discriminative* model  $\mathbf{MD}_d$  is learnt to discriminate the direction correctness. A class is created for each combination of *elementary strokes* inversion direction. This model is defined in equation (14) where  $I^g$  represents the number of inversion of direction (*i.e.*  $2^{N^g}$ ). Therefore, the binary decision follows the best membership degrees  $s_{best}$ , *i.e.* correct if the correct direction class is recognised and incorrect if one of the incorrect direction class is recognised.

$$\mathbf{MD}_d = \left[ (\mu_{c_g^k}, \Phi_{c_g^k})_{k \in [1, I^g]}, E_c^g \right]_{g \in \mathbf{G}} \quad (14)$$

### 6.3. Multi-Criteria Analyser

Finally, a refined *multi-criteria* confidence gives to children a feedback related to the legibility and the ductus. It is defined as the combination of the previous criteria by equation (15) where  $c_s$ ,  $c_o$  and  $c_d$  are the confidence associated to *shape*, *order* and *direction*, and  $(\alpha_k)_{k \in \{1, 0\}}$  a maximum score when there is one or two errors, and  $\alpha_s$  a weight associated to the shape.

$$c = \min(\alpha_s c_s, 1.0) \cdot \mathcal{P}(c_o, c_d) \text{ where } \mathcal{P}(c_o, c_d) = \begin{cases} 1 & \text{if } c_o + c_d = 2 \\ \alpha_1 & \text{if } c_o + c_d = 1 \\ \alpha_0 & \text{if } c_o + c_d = 0 \end{cases} \quad (15)$$

This confidence is obtained by combining the continuous confidence of the *shape* classifier with a penalisation when the *order* and/or the *direction* are incorrect.

305 The weights  $(\alpha_k)_{k \in \{0,1,s\}}$  are fixed by teachers and depends on the children skills and teacher expectations.

## 7. Experiments and Results

This section presents experimental results of the *multi-criteria* analyser against the *recognition-based* approach of [5] that uses only recognition features. First, datasets are introduced followed by the approach to generate the ground truth and the evaluation protocol. Finally, experiments and results are explained.

### 7.1. Datasets

*Training* and *testing* datasets are extracted from a dataset collected from 315 952 children from four to five years old in 20 French preschools. Children were practising block letters with exercises mainly related to week words (*e.g.* *lundi*, *mardi* meaning respectively Monday and Tuesday in French). 20 correct samples<sup>2</sup> for each letter (*i.e.* A, C, D, E, H, I, L, M, N, O, P, R and U) with various number of strokes are used in the *training* dataset, and the rest constitute the *Original Testing* (OT) dataset described by Table 4(a). However, 320 the OT dataset contains few samples of incorrect *order* (4%) and *direction* (3%) probably because the input was done at the end of the school year when children produce less mistakes. Therefore, to fully validate *order* and *direction* classifiers, negative samples described by Table 4(b) are generated from positive samples by synthetically changing order and direction (*Negative Order/Direction Testing* (NODT) dataset). Finally, three groups of similar letters relatively to the *shape* (*i.e.* A,H,R,P / C,O,U,D / E,I,L / M,N) are used to build the *Negative Shape Testing* (NST) dataset presented in Table 4(c): 50 samples for each letter are selected from the testing dataset and negative samples are chosen within a group (*e.g.* E and I for L).

330 This dataset containing symbols drawn by 952 children is publicly available on a dedicated web page<sup>3</sup> with a ground truth for each criterion, and results of the presented approach to be used in the future as a baseline for the *analysis* task.

### 335 7.2. Generation of a Ground-Truth with an Expert

In the OT dataset, it is often difficult to estimate the correctness of a sample as children are learning to write. Therefore an expert has manually created the ground truth with the following rules. First, a sample is correct with regards to *shape* if it is similar to the expected letter, it cannot be confused with another letter and if there are no strokes modifying the letter aspect. Moreover, 340 as it exists different levels of deformation for the shape, the expert can qualify

---

<sup>2</sup>This 20 samples are used to generate 580 synthetic training samples with geometrical distortion [24].

<sup>3</sup> <https://www-intuidoc.irisa.fr/partial-block-letters/>

Table 4: Number of gestures ( $N_g$ ) per letter in the *Original Testing* (OT) (a), *Negative Order/Direction Testing* (NODT) (b) and *Negative Shape Testing* (NST) (c) datasets with the ratio of positive and negative samples per criterion. Order and direction as a sample is only valid when shape is correct. The shape is also characterise with a *strong* and *weak* confidence. Letters have been drawn by 952 children from preschools.

(a) Original Testing (OT)

|     | $N_g$ | Shape Positive<br>(Strong/Weak) | Shape Negative<br>(Weak/Strong) | Order<br>(Positive/Negative) | Direction<br>(Positive/Negative) |
|-----|-------|---------------------------------|---------------------------------|------------------------------|----------------------------------|
| A   | 737   | 0.10 / 0.49                     | 0.19 / 0.22                     | 0.56 / 0.00                  | 0.40 / 0.00                      |
| C   | 192   | 0.21 / 0.18                     | 0.35 / 0.26                     | 0.39 / 0.00                  | 0.39 / 0.00                      |
| D   | 634   | 0.02 / 0.67                     | 0.11 / 0.2                      | 0.62 / 0.00                  | 0.62 / 0.00                      |
| E   | 243   | 0.01 / 0.37                     | 0.29 / 0.33                     | 0.13 / 0.23                  | 0.32 / 0.00                      |
| H   | 141   | 0.35 / 0.33                     | 0.08 / 0.24                     | 0.01 / 0.52                  | 0.51 / 0.11                      |
| I   | 908   | 0.00 / 0.75                     | 0.22 / 0.03                     | 0.75 / 0.00                  | 0.75 / 0.00                      |
| L   | 607   | 0.02 / 0.75                     | 0.11 / 0.12                     | 0.77 / 0.00                  | 0.77 / 0.00                      |
| M   | 538   | 0.05 / 0.38                     | 0.27 / 0.30                     | 0.29 / 0.00                  | 0.29 / 0.01                      |
| N   | 768   | 0.00 / 0.27                     | 0.48 / 0.25                     | 0.21 / 0.00                  | 0.21 / 0.00                      |
| O   | 132   | 0.02 / 0.57                     | 0.35 / 0.07                     | 0.52 / 0.00                  | 0.45 / 0.06                      |
| P   | 137   | 0.22 / 0.36                     | 0.16 / 0.26                     | 0.54 / 0.00                  | 0.55 / 0.00                      |
| R   | 315   | 0.00 / 0.34                     | 0.26 / 0.40                     | 0.29 / 0.00                  | 0.28 / 0.01                      |
| U   | 315   | 0.02 / 0.57                     | 0.26 / 0.16                     | 0.57 / 0.00                  | 0.57 / 0.00                      |
| All | 5667  | 0.08 / 0.46                     | 0.24 / 0.22                     | 0.5 / 0.04                   | 0.49 / 0.03                      |

(b) Negative Order/Direction Testing (NODT)

|     | $N_g$ | Shape<br>(Positive) | Order<br>(Positive/Negative) | Direction<br>(Positive/Negative) |
|-----|-------|---------------------|------------------------------|----------------------------------|
| A   | 312   | 1.00                | 0.58 / 0.41                  | 0.31 / 0.69                      |
| C   | 61    | 1.00                | 1.00 / 0.00                  | 0.00 / 1.00                      |
| D   | 1456  | 1.00                | 0.75 / 0.25                  | 0.25 / 0.75                      |
| E   | 342   | 1.00                | 0.39 / 0.61                  | 0.61 / 0.39                      |
| H   | 72    | 1.00                | 0.58 / 0.42                  | 0.42 / 0.58                      |
| L   | 112   | 1.00                | 0.75 / 0.25                  | 0.25 / 0.75                      |
| M   | 608   | 1.00                | 0.39 / 0.61                  | 0.61 / 0.39                      |
| N   | 372   | 1.00                | 0.58 / 0.42                  | 0.42 / 0.58                      |
| O   | 34    | 1.00                | 1.00 / 0.00                  | 0.00 / 1.00                      |
| P   | 224   | 1.00                | 0.75 / 0.25                  | 0.25 / 0.75                      |
| R   | 240   | 1.00                | 0.58 / 0.42                  | 0.42 / 0.58                      |
| U   | 105   | 1.00                | 1.00 / 0.00                  | 0.00 / 1.00                      |
| All | 3938  | 1.00                | 0.63 / 0.37                  | 0.37 / 0.63                      |

(c) Negative Shape Testing (NST)

|       | A   | C   | D   | E   | H   | I   | L   | M  | N  | O   | P   | R   | U   |
|-------|-----|-----|-----|-----|-----|-----|-----|----|----|-----|-----|-----|-----|
| $N_g$ | 150 | 150 | 150 | 100 | 150 | 100 | 150 | 50 | 50 | 150 | 150 | 150 | 150 |

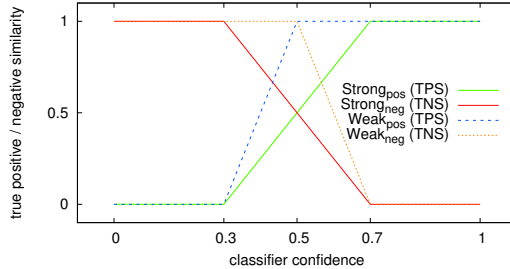


Figure 14: The *true positive similarity* (TPS) and *true negative similarity* (TNS) values measure the similarity of a confidence-based system with an expert. Thus, a sample with a *strong* confidence will be more penalised than a sample with a *weak* confidence when the result is different from the expert. These curves show respectively the *true positive similarity* for *strong* and *weak* positive samples, and the *true negative similarity* for the *strong* and *weak* negative samples with regards to the classifier confidence.

positive/negative samples with a *strong* or *weak* confidence. Then, a sample is correct relatively to *order* and *direction* respectively if the order and direction of *elementary strokes* are correct. Ground truth of NODT and NST datasets are automatically deduced from the generation process.

Two metrics called *true positive similarity* (TPS) and *true negative similarity* (TNS) are introduced to measure the similarity between the output of the system (*i.e.* a confidence value) with the expert decision. These metrics illustrated in Figure 14 penalise more errors of the classifier when a sample is ground-truthed with a *strong* confidence.

### 7.3. Evaluation Protocol

The evaluation protocol is based on two types of evaluation metrics reflecting performances of confidence-based and binary classifiers. The former uses a ROC curve created by changing the confidence threshold to show system performances at different operating points. The latter computes the analysis rate of positive and negative samples. Finally, to have a uniform comparison of criteria, analysis rates of positive and negative samples are also computed for a confidence-based classifier by using a confidence threshold of 0.5.

Each criterion is tested on specific testing datasets: *shape* (OT and NST), *order* (OT and NODT) and *direction* (OT and NODT) to avoid data duplication, *e.g.* the evaluation of *shape* does not need to be performed on NODT as all shapes are in OT.

The baseline of this work is the approach presented in [5] with *recognition-based* features - summarised in Table 2 - combined with the *confidence-based* classifier. The binary decision of *order* and *direction* classifiers are set up empirically in [5] by fixing a confidence threshold (0.2 and 0.25 respectively). It is

important to note that the presented approach uses three distinct analysers for the criteria *shape*, *order* and *direction* corresponding to the *confidence-based*, *binary order* and *binary direction* classifiers as opposition of the baseline that is only based on the *confidence-based* classifier.

In the presented approach, the importance factor  $\beta_g$  in equation (8) is equal to 0.9 to have less false negatives during the estimation of the rejection threshold, and the size of the median stroke is  $s_m = 5$  (*i.e.* 5 points resulting from the HBF49 [7] spatial sampling) to be discriminant. The choice of a high value for the importance factor  $\beta_g$  is motivated by the fact that the system should not penalise a children if the letter can be interpreted as ‘correct’, therefore false positives are tolerated but not false negatives. The adjusting weights  $\alpha_s$  for the shape in equation (15) is fixed to 1.1. The maximum score  $\alpha_0$  and  $\alpha_1$  in equation (15) are respectively set up to 0.5 and 0.35 to be in the orange or red part of the colour scale indicator when there is an error of order and/or direction.

#### 7.4. Results

Before giving quantitative results about handwriting analysis, it is important to analyse feedback of children about this first experimentation with tactile digital devices. In-class experiments conducted with the *multi-criteria* classifier have shown that children get quickly familiar with the application and try to improve the evaluation score that is well understood. Moreover, the personalised feedback allows children to progress with autonomy at their own speed, and to stay concentrated during the whole session of 20 minutes which is difficult for young children. Some of them were so involved that they did not want to stop.

Quantitative and qualitative results of all criteria are summed up in Table 5 and Figure 16. In all qualitative figures, evaluation results are displayed with an encompassing rectangle: green (true positive), blue (true negative), orange (false negative) and red (false positive) where the decision threshold is 0.5. Moreover, confidence scores are displayed on the bottom left corner with four levels of correctness: very good (green with a dark star), good (green), average (orange) and incorrect (red) corresponding to a uniform partition of confidence ranges. However, before to describe in details these results per criterion, it is important to note that the *shape* correctness is a necessary condition to perform *order* and *direction* analysis as shown in Figure 15 and justifies the choice of the presented hierarchical classifier. In Figure 16, results on the three criteria are given for five letters ‘R’. The samples #0, #1 and #2 correspond to letters that are correct for all criteria. The samples #3 and #4 have respectively an incorrect *order* (*b*) and *direction* (*c*) although the *shape* (*a*) is correct. In #3, the order is incorrect because the first and second stroke drawn correspond to the second and first median stroke of the model, therefore there is an inversion between the two first strokes. In 4#, the direction is incorrect because the vertical line is not a descending stroke. The rest of this section presents results per criterion as follows:

Table 5: Comparison between the previous approach [5] and the multi-criteria classifier. Results are presented with *True Positive Similarity* and *True Negative Similarity* (TPS - TNS) evaluation scores on the three criteria: *Shape*, *Order* and *Direction*. Features used in [5] and introduced in this paper are respectively described in details in Table 2 and Table 3.

(a) Confidence-based classifier (with features of the previous approach [5])

| Classifier | Shape [5]                 | Order [5]          | Direction [5]      |
|------------|---------------------------|--------------------|--------------------|
|            | Confidence-Based          | Confidence-Based   | Confidence-Based   |
| Features   | A subset of HBF49         | A subset of HBF49  | A subset of HBF49  |
| A          | <b>0.99</b> - 0.77        | 0.68 - 0.55        | <b>0.97</b> - 0.34 |
| C          | <b>0.96</b> - 0.93        | -                  | 0.97 - <b>1.00</b> |
| D          | <b>0.99</b> - 0.85        | 0.24 - <b>0.99</b> | 0.88 - <b>0.99</b> |
| E          | <b>0.99</b> - <b>0.96</b> | 0.56 - 0.64        | <b>1.00</b> - 0.44 |
| H          | <b>0.96</b> - <b>0.96</b> | 0.28 - 0.45        | 0.73 - 0.79        |
| I          | <b>0.97</b> - 0.76        | -                  | 0.89 - <b>1.00</b> |
| L          | <b>0.99</b> - 0.65        | 0.86 - <b>1.00</b> | 0.73 - 0.81        |
| M          | <b>0.97</b> - 0.65        | 0.59 - 0.61        | <b>0.95</b> - 0.75 |
| N          | 0.95 - 0.86               | 0.46 - 0.81        | 0.94 - 0.71        |
| O          | 0.89 - 0.96               | -                  | <b>1.00</b> - 0.0  |
| P          | 0.94 - 0.95               | 0.27 - <b>1.00</b> | 0.85 - <b>0.99</b> |
| R          | <b>0.94</b> - 0.89        | 0.64 - 0.76        | 0.99 - 0.85        |
| U          | <b>0.99</b> - 0.95        | -                  | 0.98 - 0.00        |
| All        | <b>0.96</b> - 0.85        | 0.51 - 0.76        | 0.91 - 0.67        |

(b) Multi-Criteria classifier (new result)

| Classifier | Shape  | Order  | Direction                    |
|------------|--|--|------------------------------|
|            | Confidence-Based                                       | Binary Order                                   | Binary Direction             |
| Features   | A subset of HBF49 and a Fuzzy Histogram of Orientation | A subset of HBF49 on <i>elementary strokes</i> | Fuzzy Histogram of Direction |
| A          | 0.91 - <b>0.92</b>                                     | <b>0.98</b> - <b>0.97</b>                      | 0.95 - <b>0.93</b>           |
| C          | 0.95 - <b>0.94</b>                                     | -  | <b>1.00</b> - <b>1.00</b>    |
| D          | 0.97 - <b>0.97</b>                                     | <b>0.99</b> - 0.98                             | <b>0.98</b> - <b>0.99</b>    |
| E          | 0.97 - <b>0.96</b>                                     | <b>0.99</b> - <b>0.98</b>                      | 0.89 - <b>1.00</b>           |
| H          | 0.94 - 0.95  | <b>0.97</b> - <b>0.97</b>                      | <b>0.97</b> - <b>1.00</b>    |
| I          | 0.92 - <b>0.84</b>                                     | -  | <b>0.98</b> - <b>1.00</b>    |
| L          | 0.97 - <b>0.85</b>                                     | <b>0.99</b> - <b>1.00</b>                      | <b>0.99</b> - <b>0.99</b>    |
| M          | 0.95 - <b>0.90</b>                                     | <b>0.99</b> - <b>1.00</b>                      | <b>0.95</b> - <b>0.92</b>    |
| N          | <b>0.96</b> - <b>0.98</b>                              | <b>1.00</b> - <b>1.00</b>                      | <b>0.96</b> - <b>0.99</b>    |
| O          | <b>0.94</b> - <b>0.99</b>                              | -  | <b>1.00</b> - <b>1.00</b>    |
| P          | <b>0.97</b> - <b>0.96</b>                              | <b>1.00</b> - <b>1.00</b>                      | <b>0.97</b> - <b>0.98</b>    |
| R          | 0.93 - <b>0.98</b>                                     | <b>1.00</b> - <b>1.00</b>                      | <b>1.00</b> - <b>0.96</b>    |
| U          | 0.97 - <b>0.96</b>                                     | -  | <b>1.00</b> - <b>1.00</b>    |
| All        | 0.95 - <b>0.94</b>                                     | <b>0.99</b> - <b>0.99</b>                      | <b>0.97</b> - <b>0.98</b>    |

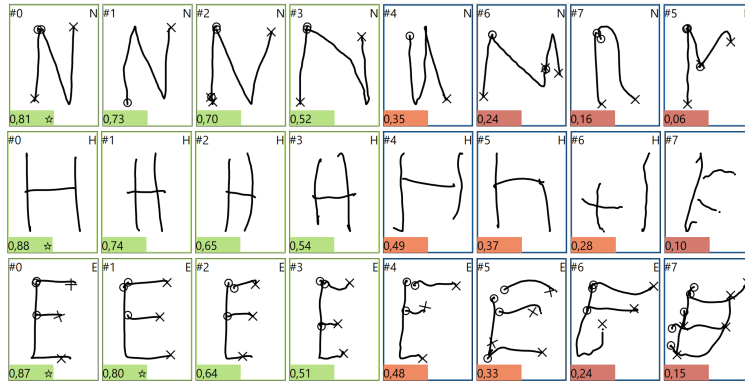


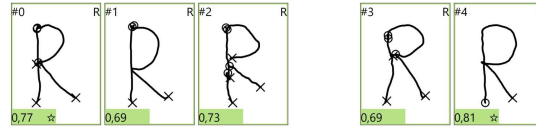
Figure 15: Results on the *shape* criteria: scores decrease coherently with the deformation of the shape.

#### 7.4.1. Shape

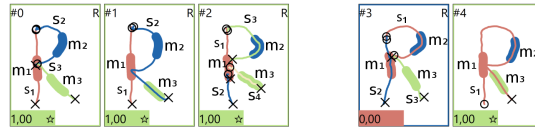
The first part of the evaluation is the operating point resulting from the Log-Cauchy distribution (*i.e.* a confidence of 0.5) of which results are presented in Table 5 and show a change in the TPS (True Positive Similarity) / TNS (True Negative Similarity) from 0.96/0.85 to 0.95/0.94. It corresponds to a better similarity of 60% with the expert for negative samples while it decreases of 20% for negatives. More precisely, ROC results presented in Figure 18 show improvements at different operating points which means *analyse-based* features (*i.e.* fuzzy histogram of orientations) have a significant impact on the results. Qualitative results in Figure 15 and Figure 16(a) show a coherent decreasing of the confidence score independently of the number of strokes and direction/order correctness. Nevertheless, some errors remains as presented in Figure 17(a):  $a\#3$ ,  $a\#4$  are borderline errors but  $a\#1$  and  $a\#2$  are shapes with missing/additional strokes and  $a\#0$  contains strokes that do not respect relative stroke proportions. Although,  $a\#0$ ,  $a\#1$  and  $a\#2$  are close to the decision border, this underlines limits of the presented model. One possible research direction to solve this problem would be firstly to represent a gesture with multiple clusters to have closest borders.

#### 7.4.2. Order

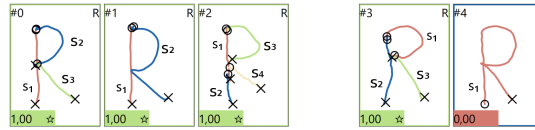
Quantitative results in Table 5 show a significant improvement of the TP-S/TNS from 0.51/0.76 to 0.99/0.99 corresponding to a decreasing of respectively 98% and 96% of the analysis of positive and negative symbols compared to the expert decision. This is mainly due to the presented approach to identify *elementary strokes* as depicted in Figure 16(b) which is more robust than a confidence-based classifier. Remaining errors cases in Figure 17(b) correspond to symbols with deformations at the limit of the shape correctness. Qualitative results in Figure 16(b) demonstrate the ability to analyse correctly the order with a various number of strokes and independently of the direction.



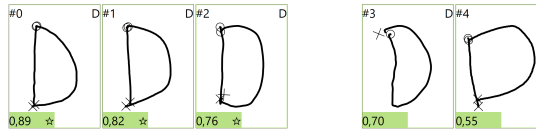
(a) Shape (R)



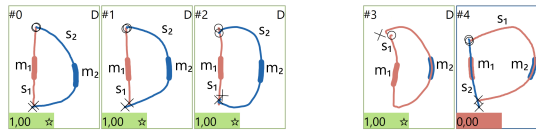
(b) Order (R)



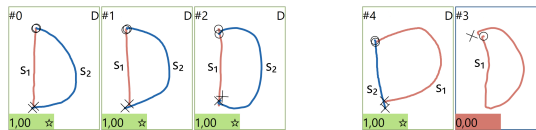
(c) Direction (R)



(d) Shape (D)



(e) Order (D)



(f) Direction (D)

Figure 16: Qualitative results on the three criteria with positive shapes and where strokes are coloured from the first to the fourth with red, blue, green and brown (the begin and end of each stroke is represented respectively with a circle and a cross). In (b) and (e) median strokes are represented with thicker lines as in Figure 10.



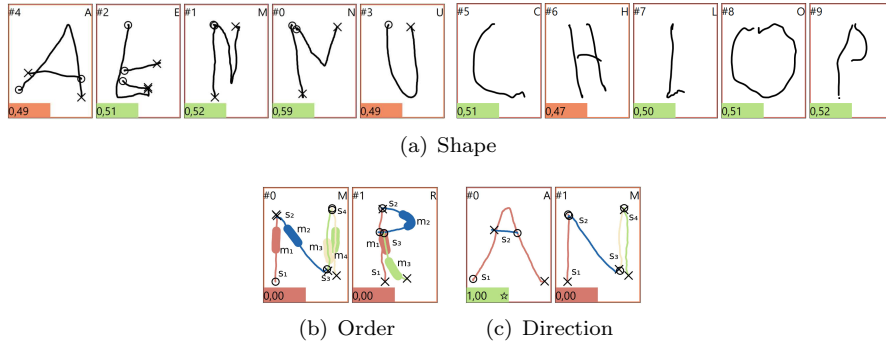


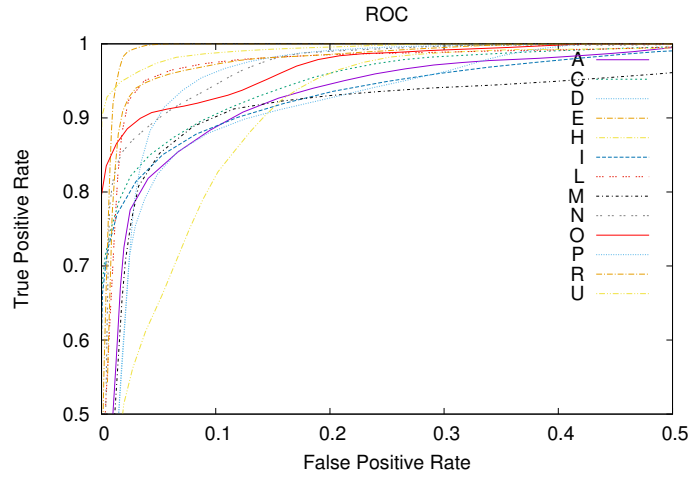
Figure 17: Qualitative results: classification errors mainly due to borderline samples (a) and shapes at the limit of the correctness to analyse order (b) and direction (c). The analysed letter is written in the top right corner.

#### 7.4.3. Direction

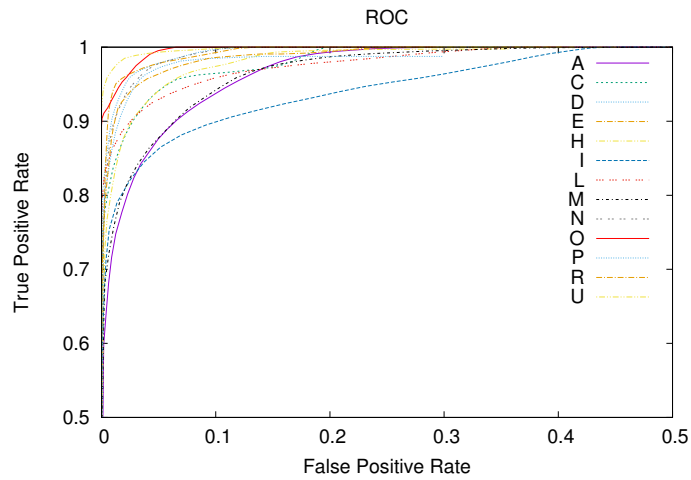
Quantitative results in Table 5 show a significant improvement of the TP-  
 440 S/TNS from 0.91/0.67 to 0.97/0.98 corresponding to a decreasing of respectively  
 67% and 94% of the analysis error of positive and negative symbols compared  
 to the expert decision. It is explained by the use of local fuzzy histograms of  
 directions that is a finer representation of local directions. It works with various  
 number of strokes and independently of the *order* as presented in Figure 16(c).  
 445 Similarly to *order*, errors cases correspond to deformations at the limit of the  
 shape correctness as depicted in Figure 17(c).

#### 7.4.4. Multi-Criteria

The *multi-criteria* confidence corresponds to the feedback given to the chil-  
 dren. Therefore, this score depends on two main aspects: the legibility and the  
 450 ductus on which the importance of each aspect is fixed by teachers based on  
 children skills and the aim of the handwriting exercise. Figure 19 illustrates  
 three possible strategies for the teacher to customise the feedback returned to  
 the children. In Figure 19(a), it corresponds to an early learning stage, *shape*,  
*order* and *direction* must correspond to the model. Therefore, errors in order  
 455 (*a*#2), direction (*a*#3, *a*#4) and shape (*a*#3) are penalised. As opposition,  
 the strategy in Figure 19(b) corresponds to an advance learning stage, where  
 children acquired their own writing style. The legibility is therefore the only  
 evaluation criterion. Finally, in Figure 19(c) and Figure 19(d) respectively,  
 there is no constraint of order and direction. These strategies are intermedi-  
 460 ate learning stages where a constraint about order or direction can be relaxed.  
 Consequently, *c*#2 and *d*#4 become correct compared to first strategy as the  
 order and direction are respectively not considered as a mistake. The sample  
 3 is incorrect in direction and slightly deformed. In *d*#3, the score is on the  
 middle as it has slightly an incorrect shape but a correct order. This illustrates  
 465 the versatility of the method to adapt to various pedagogical scenarios to satisfy



(a) Shape (previous approach [5])



(b) Shape (new result)

Figure 18: Comparison with the previous approach [5] of the *shape* criteria with the letters A, C, D, E, H, I, L, M, N, O, P, R and U on the OT and NST datasets. In both cases, a confidence-based classifier is used, only the set of feature changes.

Table 6: Dataset and results about the cursive writing.

|     | Dataset |                                 |                                 |                        | Results            |                        |
|-----|---------|---------------------------------|---------------------------------|------------------------|--------------------|------------------------|
|     | $N_g$   | Shape Positive<br>(Strong/Weak) | Shape Negative<br>(Weak/Strong) | Direction<br>(Pos/Neg) | Shape<br>(TPS-TNS) | Direction<br>(TPS-TNS) |
| a   | 1421    | 0.15 / 0.48                     | 0.23 / 0.14                     | 0.29 / 0.23            | 0.88 - 0.93        | 1.00 - 1.00            |
| c   | 337     | 0.00 / 0.20                     | 0.73 / 0.07                     | 0.11 / 0.09            | 0.95 - 0.87        | 1.00 - 1.00            |
| d   | 1654    | 0.10 / 0.57                     | 0.25 / 0.08                     | 0.28 / 0.23            | 0.93 - 0.91        | 1.00 - 1.00            |
| e   | 974     | 0.00 / 0.69                     | 0.17 / 0.14                     | 0.48 / 0.16            | 0.89 - 0.96        | 1.00 - 0.99            |
| h   | 372     | 0.02 / 0.56                     | 0.35 / 0.07                     | 0.28 / 0.23            | 0.85 - 0.95        | 1.00 - 1.00            |
| l   | 2234    | 0.05 / 0.34                     | 0.47 / 0.15                     | 0.32 / 0.04            | 0.74 - 0.96        | 0.99 - 1.00            |
| m   | 959     | 0.20 / 0.53                     | 0.11 / 0.16                     | 0.38 / 0.32            | 0.93 - 0.83        | 0.99 - 1.00            |
| n   | 1727    | 0.17 / 0.45                     | 0.26 / 0.12                     | 0.33 / 0.27            | 0.82 - 0.96        | 1.00 - 1.00            |
| o   | 405     | 0.00 / 0.28                     | 0.64 / 0.08                     | 0.19 / 0.05            | 0.98 - 0.84        | 1.00 - 1.00            |
| p   | 386     | 0.00 / 0.61                     | 0.39 / 0.00                     | 0.21 / 0.19            | 0.94 - 0.77        | 1.00 - 0.97            |
| r   | 792     | 0.11 / 0.36                     | 0.37 / 0.17                     | 0.33 / 0.12            | 0.91 - 0.86        | 1.00 - 1.00            |
| s   | 425     | 0.00 / 0.40                     | 0.60 / 0.00                     | 0.22 / 0.17            | 0.92 - 0.94        | 1.00 - 1.00            |
| u   | 2092    | 0.00 / 0.45                     | 0.45 / 0.10                     | 0.26 / 0.14            | 0.97 - 0.88        | 1.00 - 1.00            |
| All | 13778   | 0.08 / 0.46                     | 0.35 / 0.11                     | 0.30 / 0.18            | 0.90 - 0.90        | 0.99 - 0.99            |

teacher expectations.

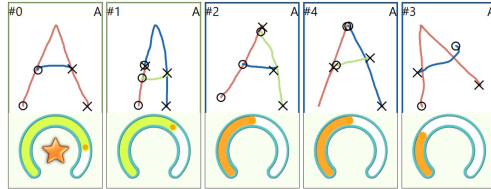
### 7.5. Experiments on Cursive Letters

In-class experiments have also been conducted with cursive writing in which the *multi-criteria* approach has been used with the *shape* and *direction* criteria as *order* is not meaningful for cursive writing. 30 samples of each letters have been used for the training. The testing dataset including the synthetic inversion of direction and results about shape and direction are presented in Table 6. In this experiment, all the parameters are the same than in block letter writing experiments.

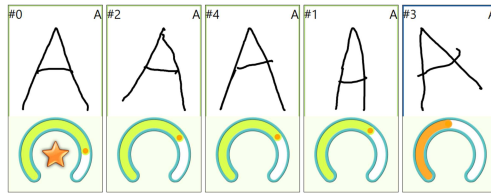
Results in terms of TPS/TNS are respectively 0.99/0.99 and 0.90/0.90 for the direction and the shape. The direction decision of the system is very close to the expert decision with 1% of dissimilarity for positive and negative samples. Compared to the expert, the system gives a shape decision with 10% of dissimilarity for positive and negative samples. 81% of samples have been annotated by the expert with weak confidence which emphasises that it is difficult to take a strict decision on this dataset. In the block writing dataset, there were 70% of samples with a weak confidence and a result of 0.95/0.94 in terms of TPS/TNS. Therefore the proposed approach gives similar results on cursive and block writing letters.

## 8. Conclusion

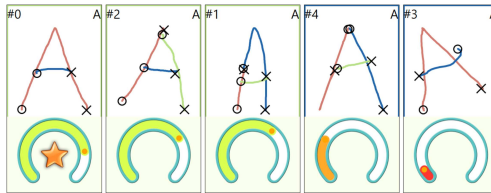
This paper has presented a *multi-criteria* analysis of block letters applied in the innovative project *IntuiScript* targeting the development of a digital workbook to help teachers and children during the handwriting learning process. A new dataset collected in 20 preschools with 952 children is made publicly available with a ground truth and results obtained in this work for comparison with future works on the handwriting *analysis*. Results on this dataset have shown



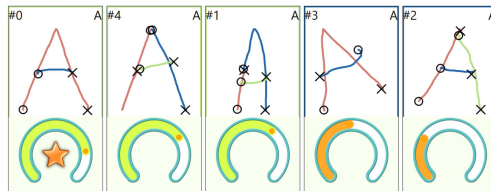
(a) Early learning stage: All Criteria.



(b) Advanced learning stage: Shape Only.



(c) Intermediate learning stage: No Order.



(d) Intermediate learning stage: No Direction.

Figure 19: Qualitative results of the *multi-criteria* classifier with the colour scale indicator returned to the children. It illustrates four possible strategies for the teacher to customise the feedback returned to the children (*i.e. multi-criteria*). In (a), it corresponds to an early learning stage, the shape, order and direction must correspond to the model. The strategy (b) only concentrates on the legibility. In (c) and (d) respectively, there is no constraint of order and direction.

significant improvements compared to [5] that are explained by the five main contributions. A *multi-criteria* approach combining the results of the three classifiers is used to give a precised feedback to children summarising kinematic  
495 and legibility aspects of the writing. The *shape* resulting from that approach is based two complementary aspects: the distance to the theoretical model and the confusion with other models (*intra* and *inter* scores). A method to identify *elementary strokes* based on *median strokes* analyses precisely the *order* independently of the drawing direction. Specific features using fuzzy histograms of  
500 orientation and direction characterise finely *shape* and *direction*. The independent analysis of each criterion is only given to teachers to help them during the remediation to identify difficulties and errors of children. Teachers can customised the feedback given to children by focusing on the legibility or/and the ductus. Finally, feedback of children and teachers about the use of tactile digital  
505 devices and the digital workbook in schools have been very positive.

## 9. Future Work

The next steps of the *IntuiScript* project are the analysis of the *fluidity* that is a fundamental criterion to characterise the cursive writing. To handle this problem, methods based on lognormal models [25, 26] seem promising.  
510 Finally, the *IntuiScript* project will let teachers define their own models in an *authoring* mode by firstly asking to the teacher to draw letter samples and secondly by consolidating with children samples collected during the remediation to adjust the model to teacher expectations. Consequently, the model learnt will allow several configurations of correct orders and directions. Therefore, further  
515 experiments will be conducted to validate the existing incremental *generative* and *discriminative* learning approaches presented in this research work for the *authoring* mode.

## Acknowledgements

This work takes place in the context of a three years research project founded  
520 by the French government as part of innovative projects (BPI): *IntuiScript* (<http://intuiscript.com/>). The authors are grateful to their industrial partners (Script&Go, Microsoft), educational experts of Rennes academy, ESPE, Brittany region and LOUSTIC laboratory for their collaboration.

## References

- 525 [1] C. Jolly, R. Palluel-Germain, E. Gentaz, Evaluation of a tactile training for handwriting acquisition in french kindergarten children: A pilot study, *Kindergartens: Teaching methods, expectations and current challenges* (2013) 161–176.

- 530 [2] C. W. Li-Tsang, A. S. Wong, H. W. Leung, J. S. Cheng, B. H. Chiu, F. Linda, R. C. Chung, Validation of the chinese handwriting analysis system (chas) for primary school students in hong kong, *Research in developmental disabilities* 34 (2013) 2872–2883.
- [3] T. H. Falk, C. Tam, H. Schellnus, T. Chau, On the development of a computer-based handwriting assessment tool to objectively quantify handwriting proficiency in children, *Computer Methods and Programs in Biomedicine* 103 (2011) 102–111.
- 535 [4] V. Kulesh, K. Schaffer, I. Sethi, M. Schwartz, Handwriting quality evaluation, in: *International Conference on Advances in Pattern Recognition*, 2001, pp. 157–165.
- [5] M. Bouillon, E. Anquetil, Handwriting analysis with online fuzzy models, in: *Conference of the International Graphonomics Society*, 2015, pp. 71–74.
- 540 [6] A. Almaksour, E. Anquetil, Improving premise structure in evolving takagi-sugeno neuro-fuzzy classifiers, *Evolving Systems* 2 (2011) 25–33.
- [7] A. Delaye, E. Anquetil, Hbf49 feature set: A first unified baseline for online symbol recognition, *Pattern Recognition* 46 (2013) 117–130.
- 545 [8] L. H. Dinehart, Handwriting in early childhood education: Current research and future implications, *Journal of Early Childhood Literacy* 15 (2015) 97–118.
- [9] J. A. Schickedanz, *Much More than the ABCs: The Early Stages of Reading and Writing.*, ERIC, 1999.
- 550 [10] E. Guinet, S. Kandel, Ductus: A software package for the study of handwriting production, *Behavior Research Methods* 42 (2010) 326–332.
- [11] C. Jolly, E. Gentaz, Analysis of cursive letters, syllables, and words handwriting in a french second-grade child with developmental coordination disorder and comparison with typically developing children, *Frontiers in psychology* 4 (2013) 1–10.
- 555 [12] A. P. Accardo, M. Genna, M. Borean, Development, maturation and learning influence on handwriting kinematics, *Human movement science* 32 (2013) 136–146.
- [13] Z.-H. Hu, Y. Xu, L.-S. Huang, H. Leung, A chinese handwriting education system with automatic error detection, *Journal of Software* 4 (2009) 101–107.
- 560 [14] N. Law, W. Ki, A. Chung, P. Ko, H. Lam, Childrens stroke sequence errors in writing chinese characters, *Reading and Writing: An Interdisciplinary Journal* 10 (1998) 267–292.
- 565

- [15] C. K. Tan, An algorithm for on-line strokes verification of chinese characters using discrete features, in: International Workshop on Frontiers in Handwriting Recognition, 2002.
- 570 [16] Y. Gao, L. Jin, N. Li, Chinese handwriting quality evaluation based on analysis of recognition confidence, in: International Conference on Information and Automation, 2011, pp. 211–225.
- [17] D. Impedovo, G. Pirlo, R. Plamondon, Handwritten signature verification: New advancements and open issues, in: International Conference on Frontiers in Handwriting Recognition, 2012, pp. 367–372.
- 575 [18] G. Pirlo, V. Cuccovillo, D. Impedovo, P. Mignone, On-line signature verification by multi-domain classification, in: International Conference on Frontiers in Handwriting Recognition, 2014.
- [19] A. Vinter, A. W. van Gemmert, J. G. Phillips, Special issue: Progress in graphonomics: A perceptual motor skill perspective, Human Movement Science 30 (4) (2011) 683–686.
- 580 [20] L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
- [21] A. Delaye, E. Anquetil, S. Macé, Explicit fuzzy modeling of shapes and positioning for handwritten chinese character recognition, in: International Conference on Document Analysis and Recognition, 2009, pp. 1121–1125.
- 585 [22] A. Almaksour, E. Anquetil, Ilclass: Error-driven antecedent learning for evolving takagi-sugeno classification systems, Applied Soft Computing 19 (2013) 419–429.
- [23] J. Olive David, Applied Robust Statistics, Southern Illinois University Department of Mathematics, 2008.
- 590 [24] S. Bayoudh, H. Mouchre, L. Miclet, E. Anquetil, Learning a classifier with very few examples: Analogy based and knowledge based generation of new examples for character recognition, in: European Conference on Machine Learning, 2007.
- 595 [25] R. Plamondon, M. Djioua, A multi-level representation paradigm for handwriting stroke generation, Human Movement Science 25 (2006) 586–607.
- [26] M. Djioua, R. Plamondon, A new algorithm and system for the characterization of handwriting strokes with delta-lognormal parameters, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 2060.