Interactive Off-line Handwritten Text Transcription Using On-line Handwritten Text as Feedback

Daniel Martín-Albo, Verónica Romero and Enrique Vidal
Universitat Politècnica de València, ITI
{dmartinalbo, vromero, evidal}@dsic.upv.es

***Abstract*—Handwritten Text Recognition has gained attention in the last years mainly due to the interest in the transcription of historical documents. However, automatic transcription is ineffectual in unconstrained handwritten documents, thus human intervention is typically needed to correct the results, even though post-editing is generally inefficient and uncomfortable. To alleviate these problems, multimodal interactive approaches have begun to emerge in the last years. In this scheme, the user interacts with the system by means of an e-pen. This multimodal feedback not only allows us to improve the accuracy of the system but also increases user acceptability. In this work, we present a new approach for interaction based on character sequences. We present developments that allow taking advantage of interaction-derived context to significantly improve feedback decoding accuracy. Empirical tests suggest that, despite the loss of the deterministic accuracy of traditional peripherals, this approach can save significant amounts of user effort with respect to non-interactive post-editing correction.**

I. Introduction

In general, Handwritten Text Recognition (HTR) is difficult because of the inherently variable and noisy nature of the objects (character and word images) to be recognized, among many other adversities. Therefore, as in many other Pattern Recognition problems, recognition results are not (and probably will never be) directly usable in many applications. To cope with this situation, approaches have begun to emerge in the last years in which the user and the system interact hand-in-hand to improve the accuracy of the system. For example, Civera *et al.* [1] proposed an interactive approach to language translation and Vidal *et al.* [2] applied a similar interactive approach to speech transcription.

Following similar ideas, Toselli *et al.* [3] proposed an interactive approach to transcription of handwritten text called Computer Assisted Transcription of Text Images (CATTI). Experiments have proven that these kind of systems can save significant amounts of overall human effort. In [3], an interactive handwritten transcription (IHT) system was proposed where user interaction was always performed at whole word level; that is, the user must detect and correct complete word errors. Continuing the previous work, Romero *et al.* [4] introduced a new form of interaction based only in character-level corrections.

Furthering the goal of making the interaction with the system more comfortable to the user, Toselli *et al.* [5] presented a multimodal IHT system using e-pen corrections at whole-word level. The underlying idea behind this work is that the use of more ergonomic interfaces should result in an easier and more comfortable computer-human interaction. However, this kind of feedback leads to the loss of determinism in the interaction feedback. That is, if we use a deterministic interface as the keyboard, the system knows what we have typed with 100% accuracy. The use of a non-deterministic peripheral, such as an e-pen, implies that the possible feedback decoding errors may increase the overall interaction cost. Nevertheless, by using contextual information derived from the interaction, the feedback decoding accuracy can be significantly improved over that of an out-of-the-box recognizer, which can not take advantage of the interaction context. Following this *multimodal* approach, we presented presented a work which focused on e-pen interactions at character-level [6]. The results of both works suggest that multimodal interaction, despite the loss of the deterministic accuracy of traditional peripherals, can save significant amounts of user effort with respect to fully manual transcription as well as to non-interactive post-editing correction. However, despite the good results, these works fail to provide a sufficiently comfortable approach since they restrict user corrections to whole words or isolated characters. The ideal system would be one that allows the user to correct any subset of characters of the first mistranscribed word.

The aim of this work is to present a first approximation to the ideal system discussed above. Here we present a new approach on interaction using pen-strokes based on character sequences. In this approach, the user can produce corrections, formed by pen-strokes, to fix a set of characters of the first incorrect word in the transcription. These pen-strokes interactions may represent an isolated character, a word substring or an entire word. However, this new, hopefully more ergonomic and friendlier, interaction level entails important difficulties and challenges which have been addressed in this work. In particular, to cope with word substring feedback, different character language models have been studied. In addition, we have also studied how to take advantage of the contextual information derived from the interaction process to improve the accuracy of the feedback decoding system.

To validate this approach we present a comprehensive set of experiments. Clearly, in these interactive scenarios, assessing system performance should require human work. However, the difficulty and high cost of these tests make more feasible to use the classic pattern recognition assessment paradigm based on labelled corpora to obtain adequate estimates of human effort required to achieve the goals of the considered tasks. In fact, evaluating this interactive system requires on-line data, thus we properly simulated user interactions using an on-line handwritten corpus as in [5]. In this way, the experiments in this paper are fully reproducible.

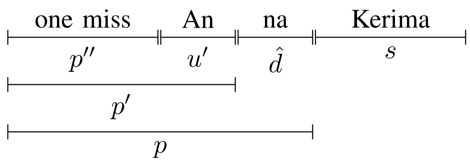

Fig. 1. A simple example of notation for the sentence *"one miss Anna Kerima"*. The system has incorrectly recognized *"one miss And Bring"*. After this, the user corrects the erroneous character *"d"* with *"na"* writing some pen-strokes (t) (see Fig.2). By making this action, the user implicitly validates the prefix (p') *"one miss An"* . In this validated prefix, the fragment of the prefix formed by complete words (p'') is *"one miss"* and the last incomplete word of the prefix (u') is "*An*". Using the available contextual information, the interaction decoding system decodes the user pen-strokes (t) resulting in the best possible decoding ($\hat{d}$) "*na*". Once the prefix is available the interactive transcription system generates *"Kerima"* as a new suffix (s).

II. Interactive Off-line Handwritten Text Transcription

As explained in [3], in the CATTI framework the human transcriber is involved in the transcription process. Formally speaking, let x be a feature vector sequence extracted from a handwritten text line. The IHT system starts proposing a transcription of x. Then, the user validates an initial part of this transcription (p'), which is error-free, and introduces a correct word (w) thereby producing a correct transcription *prefix* ($p = p'w$). After that, the system must take into account this information to suggest a new suitable continuation *suffix* (s). This process continues until the user accepts the transcription as correct. At each step of this interaction process, the system must take into account the image representation (x) and the prefix (p) to search the most likely suffix ($\hat{s}$):

$$\hat{s} = \arg\max_s P(s \mid x, p) = \arg\max_s P(x \mid p, s) \cdot P(s \mid p) \quad (1)$$

Since the concatenation of p and s constitutes a full transcription hypothesis, $P(x \mid p, s)$ can be approximated by concatenated character Hidden Markov Models (HMMs) [7], [8]. On the other hand, $P(s \mid p)$ is approximated by a dynamically modified n-gram to cope with the increasingly consolidated prefixes [3].

III. Using on-line handwritten text as feedback

The idea of using on-line handwritten text as feedback has pros and cons. The use of more ergonomic interfaces should result in an easier and more comfortable computer-human interaction. However, this implies the introduction of an on-line HTR decoding system in order to deal with this nondeterministic information.

Continuing with the above notation, let x be the representation of the input image. Let t be the on-line pen-strokes provided by the user, they are intended to correct errors in the previously suggested suffix (s'). The position of these corrections allows us to define the user-validated-error-free prefix (p'). This prefix p' consists of two parts: whole words (p'') and the last incomplete word (u'). Finally, the system has to find a new suffix ($\hat{s}$) as a valid continuation of the prefix (p'), considering all possible decodings (d) of the on-line data (t). An example of this notation is shown in Fig. 1. Following [5], $\hat{s}$ can be formulated as:

$$\hat{s} \approx \arg\max_s \max_d P(t \mid d) \cdot P(d \mid p'', u', s') \cdot P(x \mid s, p', d) \cdot P(s \mid p', d) \quad (2)$$

Eq. (2) can be approximately solved using a two-step approach as seen in [5]. In the first step, the on-line subsystem must decode the on-line data (t) into the most probable sequence of characters ($\hat{d}$), knowing that this decoding must be a valid continuation of the prefix (p'). Once $\hat{d}$ is available, a new consolidated prefix p is produced joining the previous prefix (p') and the most probable decoding ($\hat{d}$). Then, the second step searches, in a similar fashion to (1), for the most probable suffix using the new consolidated prefix. These two steps are repeated until p is accepted by the user as a full correct transcription of x. An example of this process is shown in Fig. 2.

In this work we focus on the first step of (2); i.e., the restricted decoding of the user interactions, which can be isolated in a single expression as:

$$\hat{d} = \arg\max_d P(t \mid d) \cdot P(d \mid p'', u', s') \quad (3)$$

where, $P(t \mid d)$ is the morphological model provided by HMMs and $P(d \mid p'', u', s')$ can be approached by a language model dynamically constrained by information derived from the interaction process. Different scenarios arise depending on the assumptions and constraints adopted for $P(d \mid p'', u', s')$.

The first and simplest scenario corresponds to a naive approach where any kind of interaction-derived information is ignored; that is, $P(d \mid p'', u', s') \equiv P(d)$. This scenario will be used as *baseline* result in our experiments.

A more restrictive scenario arises when we regard the portion of word already validated (u'). In this case the decoding can be more accurate, since we know beforehand that the sequence of pen-strokes to decode must be a valid continuation of the part of word accepted so far. This scenario can be written as $P(d \mid p'', u', s') \equiv P(d \mid u')$.

The last scenario emerges if we add, to the previous one, the set of complete words p''. In this case, the possible decodings are constrained to be a suitable continuation of the whole prefix accepted so far. This scenario can be written as $P(d \mid p'', u', s') \equiv P(d \mid p'', u')$.

IV. Dynamic language modeling

Language model restrictions are implemented on the base of n-grams. As we mentioned earlier, the *baseline* scenario given by $P(d)$ (being $d = \{c_1, c_2, ..., c_l\}$) does not take into account any information derived from the interaction. Here, character n-grams have been used for modelling $P(d)$.

The next scenario, given by $P(d \mid u')$, is approached also using a character n-gram language model, but it is conditioned by the fragment of word (u'). Following similar discussion as in [5] this can be written as:

$$P(d \mid u') = \prod_{i=1}^{n-1} p(c_i \mid c_1^{i-1}, u'^{k}_{k-n+i+1}) \cdot \prod_{i=n}^{l} p(c_i \mid c_{i-n+1}^{i-1}) \quad (4)$$

		the audience at the awards was particularly enthusiastic when one Miss Anna Kerima
	$\hat{s}$	the audience cit then ackerd was particularly enthusiastic when one miss And bring
Interaction 1	t	at
	$p = p'\hat{d}$	the audience at
	$\hat{s}$	the awards was particularly enthusiastic when one miss And bring
Interaction 2	t	na
	$p = p'\hat{d}$	the audience at the awards was particularly enthusiastic when one miss Anna
	$\hat{s}$	Kerima
	p≡T	the audience at the awards was particularly enthusiastic when one miss Anna Kerima

Fig. 2. Example of interactive transcription using e-pen strokes as feedback to transcribe an image containing the sentence *"the audience at the awards was particularly enthusiastic when one miss Anna Kerima"*. Each interaction consists of two steps. In the first step, the user writes some pen-strokes (t) to amend the suffix ($\hat{s}$) proposed in the previous step. This defines a correct prefix p', which can be used by the on-line HTR subsystem to obtain a more accurate decoding of t. In the second step, a new prefix (p) is built from the previous correct prefix p' concatenated with the decoded on-line handwritten text $\hat{d}$. Using this information, the system proposes the most probable suffix. The process ends when the user accepts the suffix as a full correct transcription.

where $u' = \{u'_1, u'_2, ..., u'_k\}$. The first term of (4) accounts for the probability of the $n-1$ character of the suffix, whose probability is conditioned by known characters from the validated prefix, and the second one is the usual n-gram probability for the rest of the unknown characters.

The scenario defined by $P(d \mid p'', u')$ uses the last incomplete word of the prefix (u') and the complete words of the prefix (p''). This scenario has been approached in two different ways. The first one uses a character n-gram model conditioned by $p' = p''␣u'$ (where '␣' is a white space). We can restrict this model in a similar manner to (4):

$$P(d \,|p') = \prod_{i=1}^{n-1} p(c_i \,|\, c_1^{i-1}, p'^{z}_{z-n+i+1}) \cdot \prod_{i=n}^{l} p(c_i \,| c_{i-n+1}^{i-1}) \quad (5)$$

where $p = \{p_1, p_2, ..., p_z\}$.

The second approach for $p(d \mid p'', u')$ is modeled using a special kind of n-gram that combines words and characters. We called this $n+m$-grams, where n accounts for the maximum number of complete words in each $n+m$-gram and m is the number of characters after this n words. Given that we want to recognize a sequence of characters (d) and part of our prefix (p') is composed by complete words (p''), makes sense to use models that combine two levels of representation (words and characters). This *two-level-model* approach is not new; this idea has been previously used successfully in speech recognition. For example, Bazzi *et al.* [9] combined lexical entries with sub-lexical units to generate a hybrid language model.

The main advantage of using n+m-grams, instead of normal n-grams, is that the former model can be more informed, being smaller. For example, if we model the words `sleep furiously` with the first approximation of $P(d \mid p')$, we would need a total of seven 9-grams to represent the string. In contrast, using, for example, an 1-8-grams we would need only two elemets; i.e., "`sleep f u r i o u s l`" and "`f u r i o u s l y`".

More theoretically speaking, let $w = \{w_1, w_2, ..., w_l\}$ be a sequence of words. To compute the probability of this sequence with a $(n+m)$ language model, w can be considered split into two fragments, the first part represented as a sequence of complete words and the second part as a sequence of characters $w = \{w_1, w_2, ..., w_b, c_1, c_2, ..., c_s\}$, where $\mathrm{length}(w_{b+1}, w_l)$ is the length in characters from word w_{b+1} to w_l. Considering the boundary point b as a hidden variable, we can write:

$$p(w) = \sum_{1 \le b \le l} \prod_{i=1}^{b} p(w_i \mid w_{i-n+1}^{i-1}) \quad (6)$$
$$\cdot \prod_{i=1}^{n+m-1} p(c_i \mid c_{i-n+1}^{i-1}, w_{b-n+1}^{b}) \cdot \prod_{i=n+m}^{s} p(c_i \mid c_{i-n-m-1}^{i-1})$$

where the first term of (6) accounts for the probability of the b complete words, the second one accounts for the probability of the characters influenced by characters and words and the last one is the usual n-gram.

$p(d \mid p'', u')$ can be approached by adapting this $n+m$-gram language model so as to cope with the consolidated prefix. The previous equation provides a model for the probability of $p(w)$, where w can be seen as the concatenation of $p''u'd$. But now both p'' and u' are fixed. Therefore some changes are needed. Following similar discussion to that presented in [5] and considering only the last complete word of the prefix ($p'' \equiv w_p$) we can write $p(d \mid p'', u')$ as:

$$P(d \mid u', w_p) = \prod_{i=1}^{n-k-1} p(c_i \mid c_1^{i-1}, u', w_p)$$
$$\cdot \prod_{i=n-k}^{n-1} p(c_i \mid c_1^{i-1}, u'^{k}_{k-n+i+1}) \cdot \prod_{n}^{l} p(c_i \mid c_{i-n+1}^{i-1}) \quad (7)$$

where the first term of (7) accounts for the probability of the $n-k-1$ characters of the suffix, whose probability is conditioned by known characters from the validated prefix and the previous complete word, the second one accounts for the probability of the $n-1$ characters of the suffix, whose probability is conditioned by characters of the last incomplete word from the validated prefix, and the last one is the usual n-gram probability for the rest of the unknown characters.

V. Experimental details

The details of the HTR systems, the corpora and the assessment measures used in the experiments are given below.

A. Off- and On-line HTR Baseline Systems

Besides the main off-line recognition system, an additional subsystem is needed here in order to cope with the e-pen interactions. In this section a general overview of both HTR systems is presented.

The two systems follow the classic scheme of pattern recognition: *preprocessing*, *feature extraction* and *recognition*. The first two blocks of each system entail different techniques, since they dealing with different kind of information. However, the last one is the same for both subsystems.

Off-line HTR preprocessing is aimed at correcting image degradations and geometry distortions: skew and slant corrections and size normalization [10]. On the other hand, on-line handwriting preprocessing involves, in this case, only two steps: duplicated point removal and noise reduction [11].

Feature extraction in the off-line case transforms a preprocessed image into a sequence of feature vectors representing grey levels and gradients [10]. The on-line feature extraction module transforms the preprocessed trajectories into a new temporal sequence of seven-dimensional feature vectors [12].

The recognition process is based on HMMs. Characters are modeled by a continuous density left-to-right HMMs. A Gaussian mixture is used to model the emission of each HMM state. For each system, the number of Gaussians was 8 for the on-line system and 128 for the off-line.

B. Corpora

Three corpora have been used in the experiments. First, we used the IAMDB [13] corpus to evaluate the off-line system. The IAMDB is a publicly accessible corpus composed of 1,539 scanned text pages, handwritten by 657 different writers. The database is provided at different segmentation levels, here we use sentence-segmented images. To better focus on the essential issues of the considered problems, no punctuation marks, diacritics, or different word capitalizations are included in the transcriptions. From 2,324 sentences that forms the corpus, 200 were used as test, leaving the rest as training.

IAMDB consists of hand-copied sentences from the much larger electronic text LOB corpus [14], which contains about one million running words. This corpus was used here (after removing the test sentences) to train the language models based on n-grams.

Finally, the on-line UNIPEN [15] corpus was employed to simulate the user e-pen interactions. The UNIPEN corpus comes organized in several categories: lower and upper-case letters, digits, symbols, isolated words and full sentences. Here, three categories were used: digits, lowercase letters and symbols. Three arbitrary writers were chosen as test partition and 17 as training data [5].

C. Assessment Measures

Two kinds of measures have been adopted to assess the systems. On the one hand, the quality of non-interactive transcriptions can be properly assessed with the well known word error rate (WER). It is defined as the minimum number of words that need to be substituted, deleted or inserted to convert a sentence recognized by the system into the corresponding reference transcription, divided by the total number of reference words. On the other hand, the effort needed by a human transcriber to produce correct transcriptions using an IHT system is estimated by the word stroke ratio (WSR), which can be also computed using the reference transcriptions. After each system hypothesis, the longest common prefix between the hypothesis and the reference is obtained and the first unmatching word from the hypothesis is replaced by the corresponding reference word. This process is iterated until a full match with the reference is achieved.

Therefore, the WSR can be defined as the number of word level[1] user interactions that are necessary to achieve the reference transcription of the text image considered, divided by the total number of reference words. This definition makes WER and WSR comparable. Moreover, the relative difference between them gives us a good estimate of the reduction in human effort that can be achieved by using CATTI with respect to using a conventional HTR system followed by human post-editing. This estimated effort reduction will be denoted as EFR.

Finally, the conventional classification error rate (ER) was used to assess the accuracy of the on-line HTR subsystem.

VI. Experimental Results

The aim of the developed experiments were twofold: 1) check which degree of synergy can actually be expected by taking into account information derived from the interactions; and, 2) study how much effort is increased using an e-pen compared to using a deterministic device, such as the keyboard, during the interaction process.

To address these questions we simulated the process of a user performing a transcription using our system. We employed the IAMDB corpus as a document to be transcribed. For each sentence, the system proposes a new potential transcription. If the answer contains any mistakes, some amendments must be introduced. To simulate this user interactions we used the UNIPEN corpus. After correcting the error and having a new consolidated prefix, the system generates a new suffix. This process runs until the transcription is equal to the reference.

TABLE I. Statistics of the test set

	Word Fragments	Complete words	Total
IAMDB	414	1848	2262

In order to increase the reproducibility and simplify the experiments, we assumed that the rectifications correct from the beginning of the erroneous fragment until the end of the first mistranscribed word; i.e., if the system recognizes the word `january`, but the reference is `janitor`, the correction used here is `itor`. Moreover, samples used as corrections were generated as follows: we created three samples for every erroneous sequence to correct, one for each test writer. Each of these samples was generated by binding isolated characters from each writer defined as test from the UNIPEN corpus. Continuing with the previous example, if the amendment

[1]Although in this paper we have focused on correcting word segments, we believe that correcting a word segment costs as much as correct the whole word.

needed is the sequence `itor`, we need to paste together an `i`, a `t`, an `o` and a `r` of each user. These characters are chosen at random. Table I shows the statistics of the data employed in the experiment.

TABLE II. FEEDBACK DECODING ERROR RATES

	CN	CN_p	W-CN_p	M-CN_p	Comb.
Word Fragments	11.4	3.2	8.5	8.4	3.2
Complete Words	12.7	12.7	10.9	8.5	8.5
Average	12.5	11.0	10.5	8.5	7.5

Table II reports the average feedback decoding error considering the different scenarios described before. The first one, called here CN, corresponds to the baseline given by $P(d)$. Here we use a 9-gram character model constructed using the characters of isolated words since in this scenario there is no need of context between words. The second scenario ($P(d \mid u')$), called here CN_p, uses the same character 9-gram language model as above, but in this case is prefix-conditioned. The third one, called W-CN_p, is a a whole-prefix-conditioned character 9-gram ($P(d \mid p'', u')$). This language model has been constructed using separated characters of words grouped in pairs, thus this language model contains information about how words are connected. The fourth one, named by M-CN_p, is a whole-prefix-conditioned 1-8-gram. Finally, the last column represents our best system, created by combining the best results (CN_p for word fragments and M-CN_p for complete words).

As expected, the more information available, the highest feedback decoding accuracy. The excellent result achieved by CN_p recognizing word fragments may be due to the way that language model is created, since it only uses isolated words. Honestly speaking, it is also possible that the use of the prefix beyond the *intra-word* context when recognizing word fragments can mean more hassle than help.

As a final overview, Table III compares the results at the user effort level. The first column shows the WER achieved using post-editing corrections. The second one, shows the WSR achieved using an IHT system using a deterministic interface, such as the keyboard. The third column shows the WSR for our best feedback decoding approach (last column in Table II). This value is calculated under the simplification that if the system fails to recognize a sequence of characters, the user proceeds to enter it again with the keyboard (thereby combining two corrections). Finally, the last two columns show the overall estimated effort reductions (EFR) for the deterministic IHT system and our multimodal IHT approach with respect to post-edition with auto-completing.

TABLE III. EFFORT COMPARISON RESULTS

Post-editing	Word Stroke Ratio		Overall EFR	
WER	Keyboard	e-pen	Keyboard	e-pen
25.1	21.5	23.1	14.3	8.0

According to these results, the expected user effort for our approach, is only slightly higher than that of using a deterministic interaction system.

VII. CONCLUSIONS

In this work, we have studied a new way of interaction using pen-strokes as feedback. Here, this feedback is used as a part of a prefix to improve the transcription given by the computer. Thus, the system proposes a new suffix that the user can accept as a final transcription or modify in an iterative way until a full and correct transcription is finally produced.

Empirical tests presented here support the benefits of using this approach rather than traditional HTR followed by human post-editing. From the results, we observe that the use of the more ergonomic feedback modality comes at the cost of only a reasonably small number of additional interaction steps needed to correct the few feedback decoding errors. The number of these extra steps is kept very small thanks to the system ability to use interaction-derived constraints to considerably improve the on-line HTR feedback decoding accuracy.

Finally, as future work we intend to further explore the use of $n + m$-grams since its potential is clear.

ACKNOWLEDGMENT

This work was partially supported by the Spanish MEC under the STraDA research project (TIN2012-37475-C02-01), the MITTRAL (TIN2009-14633-C03-01) project, the FPU scholarship AP2010-0575, by the Generalitat Valenciana under the grant Prometeo/2009/014, and from the EU's 7th Framework Programme (FP7/2007-2013) under grant agreement n. 600707 – tranScriptorium and n. 287576 – CasMaCat.

REFERENCES

[1] J. Civera *et al.*, "A syntactic pattern recognition approach to computer assisted translation," in *Advances in Statistical, Structural and Syntactical Pattern Recognition*. Springer-Verlag, 2004.

[2] E. Vidal, L. Rodríguez, F. Casacuberta, and I. García-Varea, "Interactive pattern recognition," in *MLMI'07*, Brno, vol. 4892, pp. 60–71.

[3] A. H. Toselli *et al.*, "Computer Assisted Transcription of Handwritten Text," in *ICDAR'07*, Brazil, 2007, pp. 944–948.

[4] V. Romero, A. H. Toselli, and E. Vidal, "Character-level interaction in computer-assisted transcription of text images," in *ICFHR'10*, India, 2010, pp. 539–544.

[5] A. H. Toselli, V. Romero, M. Pastor, and E. Vidal, "Multimodal interactive transcription of text images," *Pattern Recognition*, vol. 43, no. 5, pp. 1814–1825, 2010.

[6] D. Martín-Albo *et al.*, "Multimodal computer-assisted transcription of text images at character-level interaction." *IJPRAI*, vol. 26, no. 5, 2012.

[7] F. Jelinek, *Statistical Methods for Speech Recognition*. MIT Press, 98.

[8] L. Rabiner, "A Tutorial of HMM and Selected Application in Speech Recognition," *Proc. IEEE*, vol. 77, pp. 257–286, 1989.

[9] I. Bazzi, J. Glass, and A. C. Smith, "Modeling out-of-vocabulary words for robust speech recognition," 2000.

[10] A. H. Toselli *et al.*, "Integrated Handwriting Recognition and Interpretation using FS Models," *IJPRAI*, vol. 18, no. 4, pp. 519–539, 2004.

[11] B. Q. Huang *et al.*, "Preprocessing techniques for online handwriting recognition," in *ISDA '07*, 2007, pp. 793–800.

[12] A. H. Toselli *et al.*, "On-Line Handwriting Recognition System for Tamil Handwritten Characters," in *IbPRIA'07*, vol. 4477, pp. 370–377.

[13] U.-V. Marti and H. Bunke, "A full English sentence database for off-line handwriting recognition," in *ICDAR'99*, India, 1999, pp. 705–708.

[14] S. Johansson, E. Atwell, R. Garside, and G. Leech, "The tagged lob corpus. user's manual," *Norwegian Computing Center for the Humanities*, 1996.

[15] I. Guyon *et al.*, "UNIPEN Project of On-Line Data Exchange and Recognizer Benchmarks," in *ICPR'94*, Israel, pp. 29–33.