A Time–Length Constrained Level Building Algorithm for Large Vocabulary Handwritten Word Recognition

Alessandro L. Koerich[1, 2], Robert Sabourin[1, 2], and Ching Y. Suen[2]

[1]Laboratoire d'Imagerie, de Vision et d'Intelligence Artificielle
École de Technologie Supérieure
Montréal, QC, Canada

[2]Centre for Pattern Recognition and Machine Intelligence
Concordia University
Montréal, QC, Canada

Abstract

In this paper we introduce a constrained Level Building Algorithm (LBA) in order to reduce the search space of a Large Vocabulary Handwritten Word Recognition (LVHWR) system. A time and a length constraint are introduced to limit the number of frames and the number of levels of the LBA respectively. A regression model that fits the response variables, namely, accuracy and speed, to a non–linear function of the constraints is proposed and a statistical experimental design technique is employed to analyse the effects of the two constraints on the responses. Experimental results prove that the inclusion of these constraints improve the recognition speed of the LVHWR system without changing the recognition rate significantly.

1 Introduction

In spite of recent advances in the field of handwriting recognition, few early studies have addressed the problem of large vocabulary off–line handwritten word recognition [1] [2] [3]. The most frequent simplification has been a pre–selection of possible candidate words before the recognition based on other sources of knowledge [4]. The majority of works have focused on improving the accuracy of small vocabulary systems while the speed is not taken into account.

In HMM–based systems, to handle the huge search space and keep search effort as small as possible, generally beam search is used together with the Viterbi algorithm. Beam search finds locally, i.e. at the current frame, best state hypothesis and discard

all other state hypotheses that are less probable than the locally best hypothesis by a fixed threshold [5]. The conventional LBA does not incorporate any kind of time or length constraint. Rabiner and Levinson [7] introduced global duration constraints built into the algorithm to limit the duration of the models.

In this work, we introduce two constraints to the LBA, one to limit the number of frames at each level and another to limit the number of levels of the LBA. Furthermore, we characterize the performance of the system by two responses, recognition rate (*RR*) and recognition speed (*RS*), and we assume that these responses are governed by the two constraints. A statistical experimental design technique [8] is employed to better characterize the behaviour of the LVHWR system as well as to optimise its performance as a function of these two constraints.

This paper is organized as follows. Section 2 gives an overview of the LVHWR system. Section 3 introduces the two constraints to the LBA. Section 4 describes the experimental plan, the statistical analysis of the experimental data and the results of the verification experiment over another database. Finally, some conclusions are drawn in the last section.

2 The LVHWR System

This section presents a brief overview of the structure and the main components of the LVHWR system. The system is composed of several modules: pre–processing, segmentation, feature extraction, training and recognition. The pre–processing normalizes the word images in terms of slant and size. After, the images are segmented into graphemes and the sequence of segments is transformed into a sequence of symbols (or features). There is a set of 69 models among characters, digits and special characters that are modelled by a 10–state–arc–based HMM [4]. Training of the HMMs is done by using the Maximum Likelihood criterion and through the Baum–Welch algorithm. Recognition is based on a syntax–directed level building algorithm (SDLBA) using a tree–structured lexicon generated from a 36,100–word vocabulary.

The lexicon is organized as a character tree (Fig. 1). If the spelling of two or more words contains the same n initial characters, they share a single sequence of n character HMMs representing that initial portion of their spelling. The recognition engine works in such a way that for a certain lexicon size (from 10 to 30,000) made up of words randomly chosen from the global vocabulary, the corresponding word HMMs are made up by the concatenation of character HMMs. All words are matched against the sequence of observations extracted from the word image and the probability that such word HMMs have generated that sequence of observations are computed. The word candidate is that one that provides the highest likelihood.

A crucial problem of such a system is the recognition speed. Since we do not know a priori the case of the characters we need to test both uppercase and lowercase characters at each level of the LBA and that increase the size of the search space. For digits and symbols, only a single model is tested at each level of the LBA. This approach provides good recognition rates but at the cost of low speed for lexicons that contain more than 1,000 entries. Therefore, our goal is to find a best

compromise between the recognition rate and the recognition speed when considering large vocabularies.

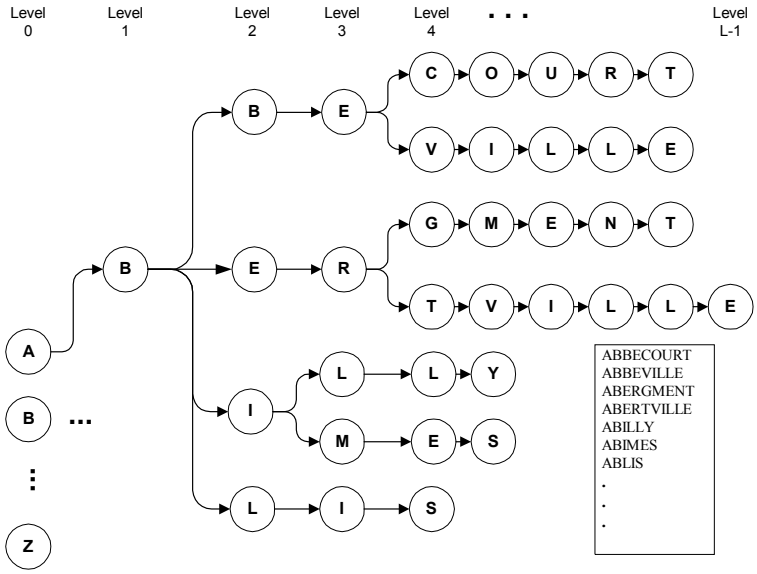

Figure 1: Tree–structured lexicon

2.1 Level Building Algorithm (LBA)

The LBA has been used for a long time in speech recognition [7], and more recently on handwriting recognition. Given a set of individual character models $c = \{c_0, c_1, c_2, ..., c_{K-1}\}$ where K denotes the number of models and a sequence of observations $O = \{o_0, o_1, ..., o_{T-1}\}$, where T denotes the length of the sequence, recognition means decoding O into the sequence of models. Namely, it is to match the observation sequence to a state sequence of models with maximum joint likelihood. The LBA jointly optimises the segmentation of the sequence into subsequences produced by different models, and the matching of the subsequences to particular models.

In the LVHWR system, we have adapted the LBA to take into account some particular characteristics of our character model since it is modelled by a 10–state–left–right–arc–based HMM, and also to take into account some contextual information. Since the lexical tree guides the recognition process, the LBA incorporates some constraints to handle the language syntax provided by the lexical tree as well as the contextual information related to the character class transition probabilities. Different from an open vocabulary problem where all character HMMs are permitted in all levels of the LBA, here the character HMMs that will be tested in each level depend on the sequence of nodes of the lexical tree. Furthermore, since only two character models compete in each level of the LBA, one corresponding to the uppercase and other corresponding to the lowercase character, it will be only necessary to compute the likelihood of two character HMMs at each level of the LBA. For digits and special characters, only one model is computed by level.

3 Incorporating Time and Length Constraints to the LBA

The SDLBA presented by Koerich *et al.* [3] is constrained only by the HMM topology and the lexical tree. The SDLBA implies the testing of the whole sequence of observations at each level. Due to the fact that our HMMs do not include self–transitions, we know that such a model can emit a limited number of observations. In other words, we have a priori knowledge of the model duration since it is implicitly modelled by the HMM topology [9]. Furthermore, it seems to be wasteful to align the whole observation sequence at all levels of the LBA, since it is expected that in average four observations be emitted at each level of the LBA. Therefore, limiting the number of observations at each level could reduce the size of the search space.

If we take into account again the topology of our HMM, it is easy to verify that short observation sequences are more likely to be generated by short words. Therefore, it seems useless to align the observation sequences with nodes of high levels if the sequence is short. Nevertheless, we know in advance the length of the sequence of features and considering that each character model can emit 0, 2, 4 or 6 observations, we can estimate from the length of the sequence of features the length of the words that could have generated such a sequence and use such information to limit the search to words with appropriate lengths. Therefore, it is expected that the performance of the system will be improved by constraining the LBA both in time and in length without changing the accuracy significantly.

3.1 Time Constraint

The time constraint concerns the limitation of the number of frames aligned at each level of LBA. We introduce two variables: $FL_{IT}(l)$ and $FL_{FT}(l)$. The first one denotes the index of the first frame while the second one denotes the index of the last observation frame that will be aligned at each level of the LBA. Both variables are functions of the level (l). Figure 2 shows how these two constraints are incorporated to the LBA.

To incorporate these two constraints into the LBA, the equations of the LBA are not modified, but just the range of the variable t that denotes the frame index. The variable t, that originally ranges from 0 to T–1, now, its range will be given by equation (1).

$$t = FL_{IT}(l), FL_{IT}(l)+1, FL_{IT}(l)+2, \ldots, FL_{FT}(l) \tag{1}$$

where $FL_{IT}(l)$ and $FL_{FT}(l)$ must be integers and they are given by equations (2) and (3) respectively. The lower and upper limits for $FL_{IT}(l)$ and $FL_{FT}(l)$ are 0 to T–1 respectively.

$$FL_{IT}(l) = \begin{cases} 0 & \text{if } l = 0 \\ l - FL_I & \text{if } l > 0 \end{cases} \tag{2}$$

$$FL_{FT}(l) = 6.(l + FL_F) \tag{3}$$

where FL_l and FL_F are the two control factors to be determined.

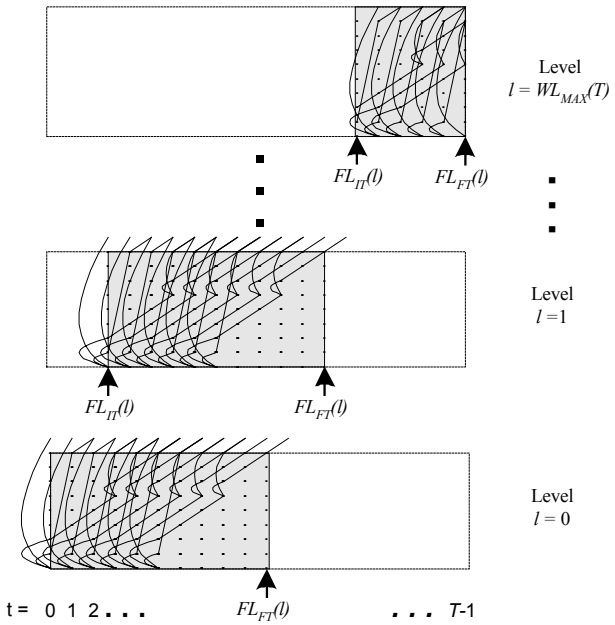

Figure 2: LBA incorporating time and length constraints

3.2 Length Constraint

As we have discussed in the first paragraphs of section 3, the length constraint concerns the limitation of the number of levels of the LBA. We introduce the variable $WL_{MAX}(T)$ which denotes the maximum number of levels of the LBA for a given observation sequence with length T. To incorporate this constraint to the LBA, the equations of the LBA are preserved, but the range of the variable l that denotes the level index, is modified slightly. Now, instead of ranging from levels 0 to L–1, the range will be given by equation (4).

$$l = 0,1,2,...,WL_{MAX}(T) \tag{4}$$

where $WL_{MAX}(T)$ is an integer given by equation (5) and its lower and upper limits are given by the shortest and the longest word in the lexicon respectively.

$$WL_{MAX}(T) = \frac{T}{WL} \tag{5}$$

where WL is the control factor to be determined.

4 Factorial Analysis

Here we introduce a formal method to determine the values of the control factors (FL_I, FL_F, and WL) based on a statistical experimental design technique that optimises the performance of the LVHWR system. In order to simplify our analysis, we take into account only two control factors, FL_I and WL. The value of the other control factor, FL_F, is fixed equal to 1.

First, we derive two regression models where the independent variables are FL_I, and WL, and the dependent variables are the responses of the system: recognition rate (RR) and recognition speed (RS). Afterwards, a complete factorial plan is employed to gain information on the control factors and to determine the coefficients of the regression models. Based on these regression models, the optimal values of the control factors that jointly optimise both RR and RS can be determined.

4.1 Multiple Regression Model

We need to establish a multiple regression model before carrying out any experiment. We assume that the responses RR and RS are approximated by the mean (M), the two control factors (FL_I and WL), the square of the control factors (FL_I^2 and WL^2), and the interaction between them ($FL_I.WL$). We assume that equation (6) gives the regression model for the recognition rate (RR) while equation (7) gives the regression model for the recognition speed (RS).

$$RR \cong M_{RR} + a_1 FL_I + a_2 WL + a_3 FL_I^2 + a_4 WL^2 + a_5 FL_I.WL \tag{6}$$

$$RS \cong M_{RS} + b_1 FL_I + b_2 WL + b_3 FL_I^2 + b_4 WL^2 + b_5 FL_I.WL \tag{7}$$

By analysing RS and RR for different values of WL and FL_I, it is possible to determine the means M_{RR} and M_{RS} and estimate the coefficients a_1, …, a_5 and b_1, …, b_5 for the control factors and the interactions.

4.2 Experimental Design

Since we have only two control factors, we can use a complete factorial plan, assigning three levels to each factor to capture the linear and the quadratic effects of both constraints over the responses. For this plan we have only 9 treatment combinations and 8 degrees of freedom (*df's*) to estimate the effects in the process we are investigating. However, to accommodate the non–linear effects and the interactions, we replicate the experiments by using a different random lexicon. Therefore, we will have 18 treatment combinations from which we lose 1 *df* due to finding the mean of the data and other 17 *df's* to estimate the effects.

Eighteen experimental runs were conducted, corresponding to the 18 combinations of the two control factors (9 for each random lexicon) and both RR and RS were measured. In these experiments, we have used a validation set that has been taken

from the Service de Recherche Technique de la Poste (*SRTP*) database. The *SRTP* database is composed of digitised images of French postal envelopes. The information written on the envelopes is labelled and segmented. This dataset is composed of 3,475 images of French city names. The experiments were carried out for lexicons with 10, 100, 1,000, 5,000, 10,000, 20,000 and 30,000 entries.

4.3 Analysis of Results

In order to perform a multifactor analysis of variance for *RR* and *RS*, we have constructed various tests and graphs to determine which factors have a statistically significant effect on both responses for different lexicon sizes. Figure 3 shows an example of the effects of the linear and quadratic terms of both *WL* and FL_I in both responses for a 100–entry lexicon. The control factor *WL* has the most pronounced effect on *RR*. The effect of this control factor is approximately quadratic. The other control factor has less effect and it seems to be approximately linear. On the other hand, the control factor *WL* has the most pronounced effect on *RS*, but the effect due to the control factor FL_I is also pronounced. The effect of both factors is approximately quadratic.

For each lexicon size, the sum of squares, the mean of squares, and the Fischer coefficients were computed. All Fischer coefficients were based on the mean square error. For the different lexicon sizes, both *WL* and FL_I have a statistically significant effect on *RR*. For the control factor FL_I, the quadratic effect can be neglected. On the other hand, both the linear and the quadratic effects of the factor *WL* are significant. These results confirm what we have seen in Figure 3. Both constraints have a statistically significant effect on *RS* and the linear and quadratic effects of both constraints are significant. The effects of the interaction of the two control factors are not significant and they can be neglected. The same behaviour was observed for the different lexicon sizes.

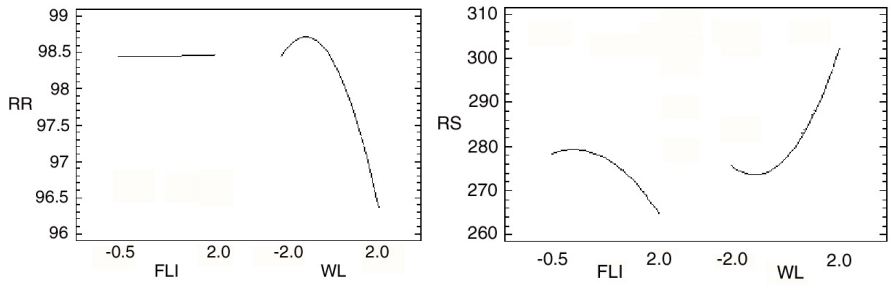

Figure 3: Main effects of the control factors on *RR* and *RS* for a 100–entry lexicon

The coefficients of the regression models can be determined by using a least–square procedure [11]. For each lexicon size, we will have different multiple regression equations that have been fitted to the experimental data.

4.4 Optimisation of Parameters

Optimisation involves estimating the relationship between *RR* and *RS*, and the two control factors. Once the form of this relationship is known approximately, the constraints may be adjusted to jointly optimise the system performance.

In our system an optimal response means maximizing both *RR* and *RS*. Therefore, we need to determine the combination of experimental factors that simultaneously optimise both response variables. We do so by maximizing equations (6) and (7) for each lexicon size. The combination of control factor levels that achieves the overall optimum responses for each lexicon size is given in Figure 4.

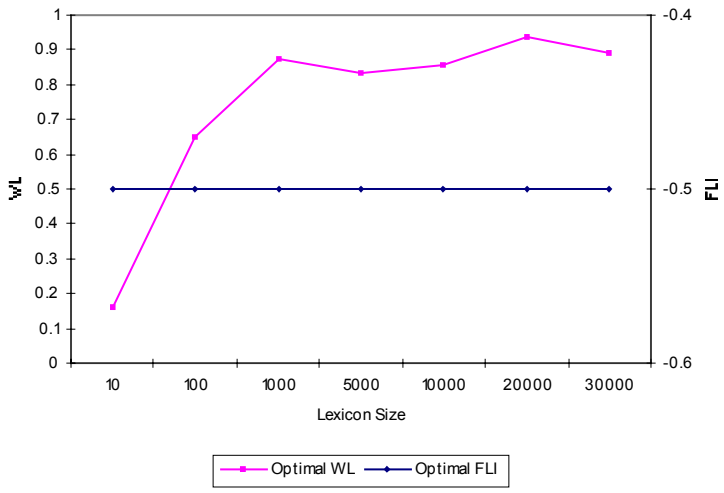

Figure 4: Optimal control factor values for each lexicon size

4.5 Experimental Results

In order to demonstrate the applicability of these constraints in the reduction of the search space and to verify the effects of the control factors in the responses of the LVHWR system, we have run a confirmation experiment. We have used a different dataset. This testing dataset contains 4,674 samples of city name images also taken from the *SRTP* database. Figure 5 shows the results obtained by the standard LBA (*STD*) and the constrained LBA (*TLC*). By comparing the results for recognition speed we can verify that by using the two constraints and setting them up to the optimal values given by the statistical experimental design technique, we improved the recognition speed significantly while keeping almost the same recognition rate.

It should be notice that, in spite of the values of the control factors are dependent on the lexicon size the improvement in speed is almost independent. Table 1 shows the approximate individual contribution of the constraints in speeding up the system. The number of character is related to *WL* while the number of frames is related to FL_I but it is also dependent on the *WL*.

5 Discussion and Conclusion

In this study, we have presented a constrained LBA where two control factors were chosen and analysed through a complete factorial plan. The effects of these two factors in the outputs of a LVHWR system were investigated. We have seen that limiting the number of observations according to the level of the LBA as well as limiting the number of levels of the LBA by taking into account the length of the observation sequences lead to an improvement of 24.4–30.3% in the recognition speed with a slight reduction of 0.28–0.77% in the recognition rate for lexicons with 10–30,000 entries respectively. If we compare with the results of a previous version of the system based on a Viterbi–flat–lexicon scheme [3] [4], the improvement in speed is more expressive (627–1,010%) with a reasonable reduction in the recognition rate (0.45–1.8%). Furthermore, the experimental design technique used for adjusting the values of the control factors provides us a robust framework where the responses of the system are non–linear functions of the control factors. Our future work will focus on the pruning the number of characters by using a beam search technique.

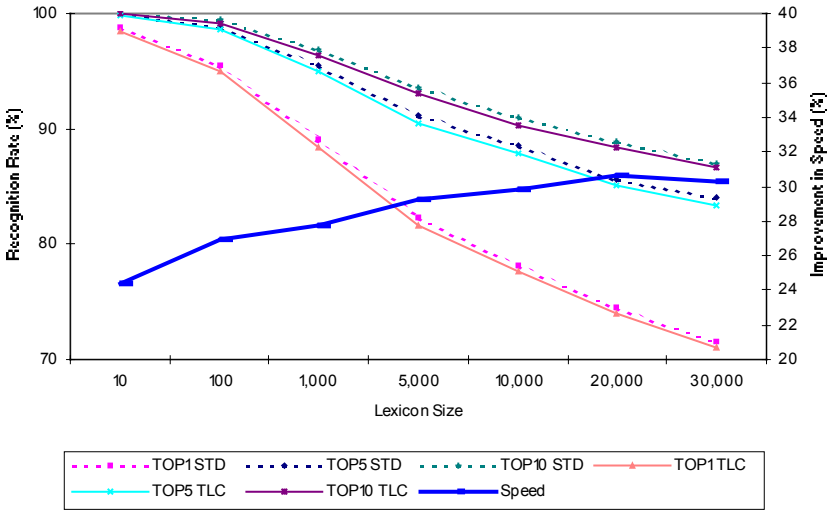

Figure 5: *RR* and *RS* for the standard LBA (*STD*) and the constrained LBA (*TLC*)

Table 1: Reduction in the number of frames and characters

Lexicon Size	Characters (%)	Frames (%)	Speed (%)
10	0	29.41	24.38
100	0.632	31.58	26.90
1,000	2.542	33.47	27.77
5,000	2.568	35.06	29.25
10,000	2.855	35.90	29.79
20,000	4.357	36.88	30.63
30,000	3.879	37.46	30.28

Acknowledgements

The authors wish to thank the CNPq–Brazil, and the MEQ–Canada, which have supported this work and the SRTP–France for supplying the baseline system and the database.

References

1. Kaltenmeier A, Caesar T, Gloger J M, Mandler E. Sophisticated topology of hidden Markov models for cursive script recognition. 2nd ICDAR, Tsukuba Science City, Japan, 1993, pp 139–142.

2. Koga M, Mine R, Sako H., Fujisawa H. Lexical search approach for character–string recognition. 3rd IWDAS, Nagano, Japan, 1998, pp 237–251.

3. Koerich A L, Sabourin R, Suen C Y, El–Yacoubi A. A syntax–directed level building algorithm for large vocabulary handwritten word recognition. 4th IWDAS, 2000, Rio de Janeiro, Brazil.

4. El–Yacoubi A, Gilloux M, Sabourin R, Suen C Y. An HMM based approach for off–line unconstrained handwritten word modelling and recognition. IEEE Trans on PAMI 1999; 21: 752–760.

5. Umbach R H, Ney H. Improvements in beam search for 10,000–word continuous–speech recognition. IEEE Trans on SAP 1984; 2: 353–356.

6. Ney H. The use of a one–stage dynamic programming algorithm for connected word recognition. IEEE Trans on ASSP 1984; 32: 263–271.

7. Rabiner L R, Levinson S E. A speaker–independent, syntax–directed, connected word recognition system based on hidden Markov models and level building. IEEE Trans on ASSP 1985; 33: 561–573.

8. Barker T B. Quality by experimental design. Marcel Dekker, NY, 1994.

9. Grandidier F, Sabourin R, El–Yacoubi A, Gilloux M, Suen C Y. Influence of word length on handwriting recognition. 5th ICDAR, 1999, Bangalore, India, pp 777–780.

10. Manke S, Finke M, Waibel A. A fast search technique for large vocabulary on–line handwriting recognition. In: Progress in handwriting recognition. World Scientific, Singapore, 1996, pp 437–444.

11. Dowdy S, Wearden S. Statistics for research. John Wiley & Sons, NY, 1991.