Coupling of a local vision by Markov field and a global vision by Neural Network for the recognition of handwritten words

Christophe Choisy and Abdel Belaïd
LORIA, Campus Sciences
Boulevard des Aiguillettes
54500 Vandoeuvre-lès-Nancy, France
choisy@loria.fr, abelaid@loria.fr

Abstract

In this paper, an idea for the combination of global and local view models is presented. These two type of models have proved their capabilities independantly. Some combination were proposed, using global view models for local analysis, and local view models to synthetize local results. An opposite approach is proposed here : local view models are used as a normalization tool, while global view models are used for the recognition of the normalized image. The use of local view models for normalization is justified by their capability to perform a non-linear normalization according to the image information. We propose Markov models as local view models, and Neural Netwok as global view models. Using Markov models for the normalization increases results up to 3% better than a classical linear normalization. Global results are improved, performing 2% better than the Markov model itself. The extension of the system to an analytic approach is discussed.

1. Introduction

Elastic models such as HMM (Hidden Markov Model), PHMM (Planar Hidden Markov Model) and Markov random fields have proven their efficiency in word modeling through many works these last decades [1, 2, 9, 11, 9]. Their power lies in the combination of local observations and dynamic programming, that allow them to focus on interesting local features by absorbing distortions. Because of the local view of these models, we introduced the concept of "Local View Models" (**MVL**) for such models [4]. The drawback of these models is the observation independence hypothesis that limits their global representativeness. This point was reflected in some works in the fact that "HMM correctly distribute their state on a sample but they fail to give an appropriate estimation of the observation probability" [6]. In other words, HMMlike models are efficient in modeling but are less adapted in discrimination tasks.

To discriminate forms, models such as neural networks are more efficient. That can be explained by their global vision allowing correlation on the whole form; we classified them as "Global View Models" (**MVG**) [4]. Their drawbacks are an huge sensibility to important distortions and a fixed input size. Because of these limits they are usually used to estimate local observation probabilities [7, 8, 12], reducing their global efficiency.

This analysis of MVL and MVG shows that they are complementary. The aim of this work is to combine efficiently the power of these two kind of models. We propose to use MVL to focus on important local features and absorb distortions; according to the localization of these features a form can be normalized to a standard size. Then it can be analyzed by a MVG to estimate global correlation between these features.

To validate our approach, we choose the NSHP-HMM (Non-symmetric Half-Plane Hidden Markov Model) [10] as MVL and a Multi-Layers Perceptron (MLP) as MVG.

The NSHP-HMM is an HMM model where the observation probability in the states is estimated by a random field. This $2D$ Markov model is acting directly on binary images analyzed column by column. Based on an HMM, it can deal with various word lengths. Thus this model seems particularly suited for image normalization.

The MLP is a classical and efficient neural network. Its analysis can be guided using a specific topology of hidden layers to better take into account the $2D$ nature of writing. This model was successfully applied on digit image classification thus it seems adapted to analyze normalized images. We preferred MLP to SVM (Support Vector Machine) previously used [4] because of their output that approximate a posterior probability : SVM

Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
0-7695-1960-1/03 $17.00 © 2003 **IEEE**

output is difficult to interpret.

Section 2 describes the training and testing schemes of this approach. Section 3 gives some results and discuss them. Finally section 4 shows proposes an analytic scheme for such an approach. The objective of this paper is not to compare different methods or models, but to validate the proposed approach and to suggest an analytic extension.

2. General system description

Let N be the word class number in the learning and testing bases. figure 1 illustrate the learning scheme of our approach.

The system training needs 4 stages :

1. **Local Vision Model training :** each word class is modeled by a NSHP-HMM, giving a total of N NSHP-HMM.
2. **Normalization :** each NSHP-HMM normalize all the learning database. Thus we obtain N normalized databases.
3. **Global Vision Model training (first-level MLP) :** a MLP is associated to each normalized database, leading to N NN for the first level. The MLP are trained to separate the different classes in a normalized database.
4. **Second-level MLP training :** each first-level MLP gives as many values as classes, leading to a total of $N * N$ values. The second-level MLP is used to synthesize all these values giving one final value for each class.

The recognition scheme is similar to the learning scheme. 3 stages are applied for the recognition of a word image :

1. **Normalization :** the input image is normalized by each NSHP-HMM. Thus we obtain N normalized images.
2. **Global analysis of normalized images :** each first-level MLP analyses the image normalized by the corresponding NSHP-HMM. At the end of this stage, we have $N * N$ values corresponding to the N output of the N MLP.
3. **First-level output synthesis :** the $N * N$ values are given in input of the second-level MLP. Its output give the likelihood of the analyzed image to belong to each word class.

2.1. Model training

NSHP-HMM are trained in an embedded analytic approach based on *Baum-Welch* formulas [5]. The basic idea of this method is to gather the letter information of different word models in the corresponding letter model. Training is made at the word level and allows an automatic learning of letters models without word segmentation. Letter NSHP-HMM are then used to build word NSHP-HMM.

After training, the word NSHP-HMM are used to normalize the whole learning database, leading to N different normalized databases.

MLP are trained with a classical gradient retro-propagation method. Each of the N first-level MLP is trained on the database normalized by the corresponding NSHP-HMM. After training these MLP are used to generate a new database containing their results : for each image the $N * N$ of the N first-level MLP form a corresponding pattern.

The second-level MLP is trained on this new database. It learns to synthesize all the values to produce the good classification decision.

2.2. NSHP-HMM normalization

The normalization of an image by a NSHP-HMM is based on the *Viterbi* algorithm. This algorithm allows to find the best repartition of image columns in the NSHP-HMM states. At each NSHP-HMM state is associated a column in the normalized image. This column is calculated as the mean value of all the columns gathered in the corresponding NSHP-HMM state [4]. The figure 2 illustrate this method and its effect on a sample of the word “et”.

The figure 3 compares this normalization and a classical linear normalization on several samples of the words “et” and “cent”. First column shows the original image, second one let see the NSHP-HMM normalization result, as last one shows the linear normalization result. We observe a better synchronization of the images in the NSHP-HMM normalization case. “t” bar are reduced in only one state except for the samples containing a “ts”; in this case the normalized “s” contains the overlapping part of the “t” bar.

3. Application and discussion

We applied our concept for the recognition of french bank check words. The global database contains 25249 images from an industrial application. This base is divided in a $2/3 - 1/3$ manner giving a training database of 16650 images and a learning database of 8599 images. Due to the lack of samples the whole training base is used at each part of the learning scheme.

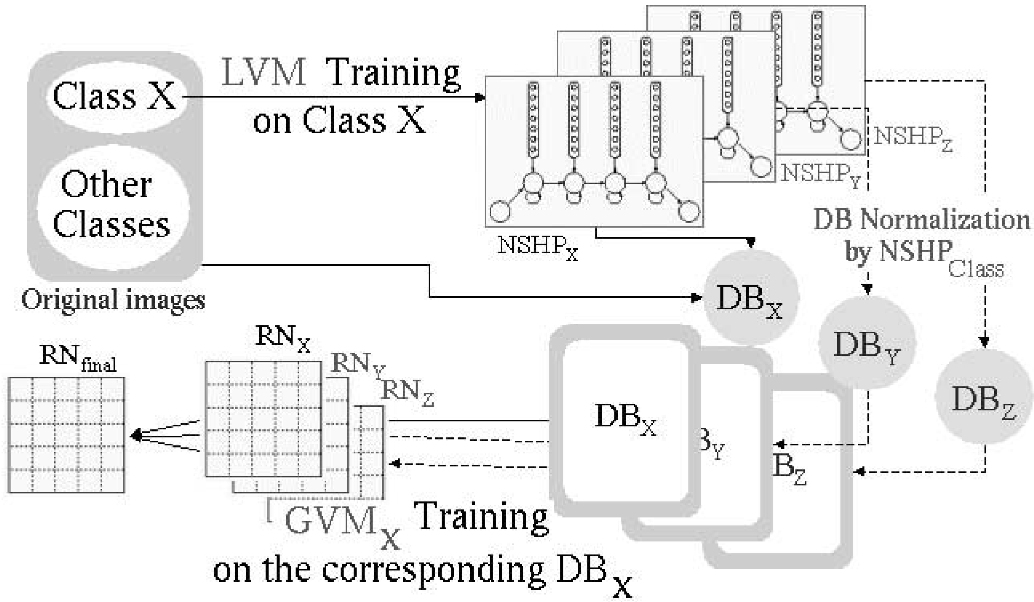

Figure 1. Learning scheme of the proposed approach : 1) MVL are trained on their class, 2) MVL are used to normalize all the database, 3) normalized database are used to train MVG, 4) the final MLP is trained on MVG results

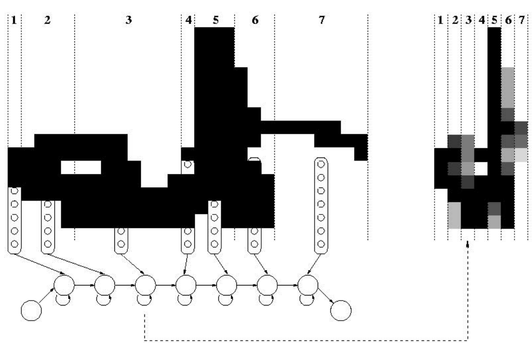

Figure 2. Normalization of the word "et" by the NSHP-HMM modeling this class. Normalized columns are given by the mean of the columns observed by the corresponding NSHP-HMM state in the input image.

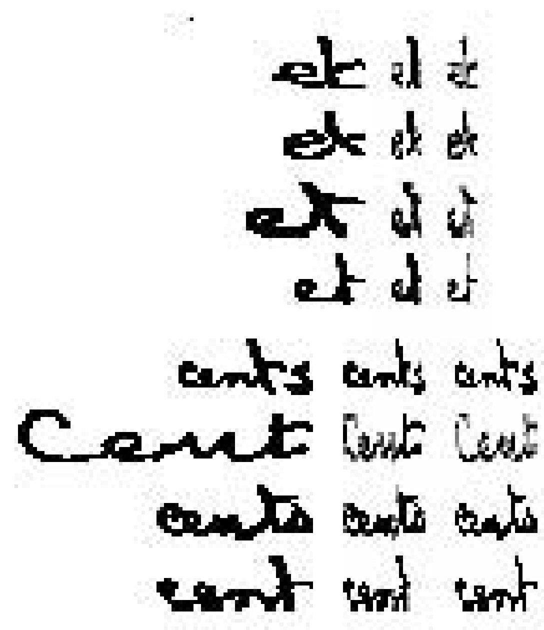

Figure 3. Examples of samples before normalization (first column), after NSHP-HMM normalization (second column) and after linear normalization (third column).

table 1 compares the results obtained with the NSHP-HMM itself, with our approach considering a classical linear normalization, and for our approach with NSHP-HMM normalization. Linear normalization gives same image size that NSHP-HMM one. Notice that NSHP-HMM results consider the *Baum-Welch* probability estimation that gives better results than *Viterbi* algorithm used for the NSHP-HMM normalization.

NSHP-HMM normalization performs 3% better than linear one, proving the efficiency of this normalization method. Despite the lack of samples and the use of the same database for each training part our approach performs 2% better than the NSHP-HMM recognition, showing the superiority of global vision models upon local vision models when data are correctly normalized.

4. Interest of this work

The first interest of the proposed approach is to combine the adaptation capabilities of the NSHP-HMM and the classification power of neural networks. Results show the interest of this idea.

The second interest of this approach lies in its extension to an analytical approach. Currently our system is based on a global approach of words : this choice was made to validate the concept, but it reduces the application to a small vocabulary.

As mentioned in §.2.1 the NSHP-HMM are trained in an analytical approach. Word models are built as a concatenation of letter models. By construction, it is easy to find which state of a word model correspond to which state of a letter model. Considering this point it is possible to apply the normalization process at the letter level rather than the word level, as illustrated by the figure 4. This is an important advantage of the NSHP-HMM normalization upon all normalization directly based on a transformation of the word images : these last ones have no information on letter limits, thus they cannot be applied directly to an analytic approach.

In a first step we propose to split normalized word images according to letter NSHP-HMM states. The letter NN will analyze the corresponding part. Letter results will be synthesized by a new NN that will produce the word probability. Such scheme cannot be applied on a large vocabulary because of this synthesis, but this test can validate the analytical approach.

The extension to a large vocabulary needs to solve two problems. First one is the word NN that combines letter results. This problem can be avoided by using simpler com-

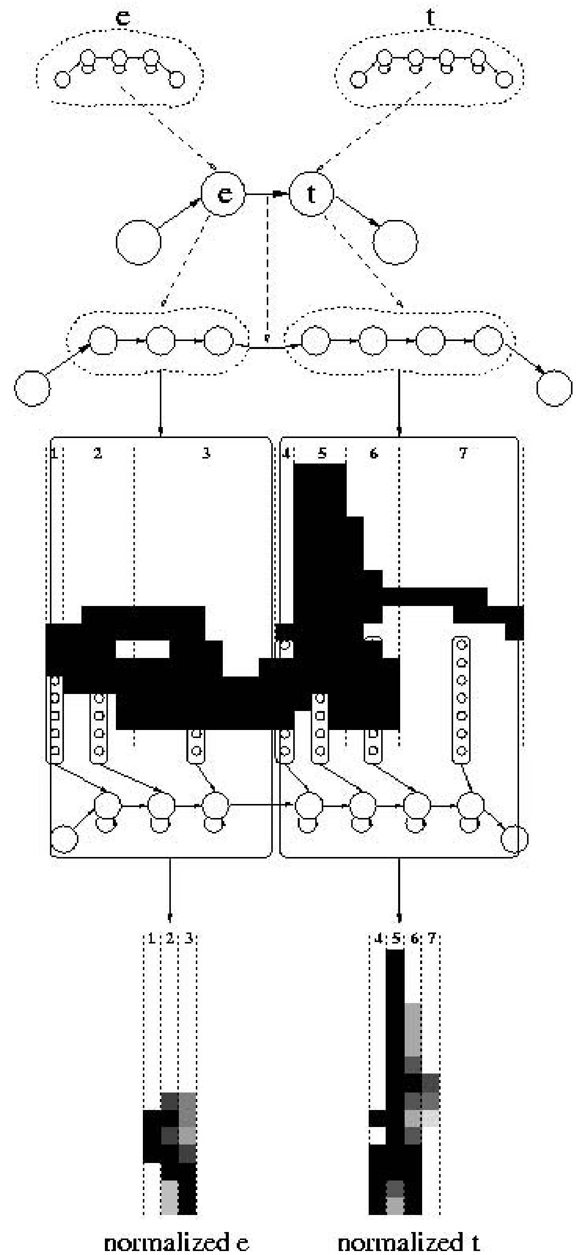

Figure 4. Analytic normalization of an image of the word "et" by the global NSHP-HMM built for this class.

Table 1. Recognition scores for several approaches

Normalization	Recognition	Top 1	Top 2	Top 3	Top 4	Top 5	Top 10
none	NSHP-HMM	85.35%	91.97%	94.39%	95.65%	96.43%	98.59%
linear	RN	84.47%	90.92%	93.43%	94.84%	95.50%	97.91%
NSHP-HMM	RN	87.41%	93.05%	95.41%	96.41%	96.86%	98.44%

bination methods, such as probability products or similarity measurement.

The second problem concerns the word image splitting. With a small vocabulary we can split the analyzed pattern according to each word NSHP-HMM normalization. In a large vocabulary such an approach cannot be used.

A common solution to deal with large vocabularies is to keep in memory the n most probable letter sequences in order to perform a validation step. With HMM-like models it could be done using the modified *Viterbi* algorithm [3]. Assuming that the correct answer will belong to the n best answer, we can have a reduced set of splitting possibilities; thus it is realistic to apply letter NN for all these possibilities. The letter NN result synthesis will give the letter sequence likelihood, allowing to find the most probable word among the n sequences.

These ideas will be developed in future works. Our hope is to present the first analytic step in the final version of this paper.

5. Conclusion

This work describes the use of a "Local Vision Model" like the NSHP-HMM as a normalization tool. Word recognition is made by a "Global Vision Model" like NN. The proposed normalization rely on the state repartition given by the *Viterbi* algorithm. Despite of the lack of samples for some classes in the database, the obtained results are encouraging for a global approach of words.

Considering the analytic learning of the word models, we discuss how to implement an analytical approach on the same principle. It is made possible because of the knowledge provided by the word NSHP-HMM, where each state belongs to a letter by construction of the word model. Such a knowledge is not present in a topological normalization, showing the interest of the use of an elastic model as normalization tool.

References

[1] N. B. Amara, A. Belaïd, and N. Ellouze. Modélisation pseudo-bidimensionnelle pour la reconnaissance de chaînes de caractères arabes imprimées. In *Conférence Internationale Francophone sur l'Écrit et les Documents*, pages 131–140, Montréal, Canada, Aug. 1998.

[2] R. Bippus. 1-dimensonal and pseudo2-dimensional hmms for the recognition of german literal amounts. In *Fourth International Conference on Document Analysis and Recognition (ICDAR'97)*, volume 2, pages 487–490, Ulm, Germany, Aug. 1997.

[3] M. Y. Chen, A. Kundu, J. Zhou, and S. N. Srihari. Off-line handwritten word recognition using hidden markov model. In *USPS'92*, 1992.

[4] C. Choisy and A. Belaïd. Handwriting recognition using local methods for normalization and global methods for recognition. In *Sixth International Conference on Document Analysis and Recognition (ICDAR'2001)*, pages 23–27, Seattle, Washington, USA, Sept. 10-13 2001.

[5] C. Choisy and A. Belaid. Apprentissage croisé en reconnaissance analytique de l'écriture manuscrite. In *Conférence Internationale Francophone sur l'Écrit et les Documents*, pages 203–210, Lyon, juillet 2000.

[6] M. Gilloux. Reconnaissance de chiffres manuscrits par modèles de markov pseudo-2d. In *Actes du* $3^{ème}$ *Colloque National sur l'Écrit et le Document*, pages 11–17, Rouen, France, 1994.

[7] M. Gilloux, B. Lemarié, and M. Leroux. A hybrid radial basis function network/hidden markov model handwritten word recognition system. In *Third International Conference on Document Analysis and Recognition (ICDAR'95)*, pages 394–397, Montréal, 1995.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[9] H. S. Park and S. W. Lee. An hmmrf-based statistical approach for off-line handwritten character recognition. In *Proceedings of the 12th International Conference on Pattern Recognition (ICPR'96), Vienna, Austria*, volume 2, pages 320–324, 1996.

[10] G. Saon and A. Belaïd. Off-line handwritten word recognition using a mixed hmm-mrf approach. In *Fourth International Conference on Document Analysis and Recognition (ICDAR'97)*, volume 1, pages 118–122, Ulm, Germany, Aug. 1997.

[11] J. C. Simon, O. Baret, and N. Gorski. A system for the recognition of handwritten literal amounts of checks. In *Internal Association for Pattern Recognition Workshop on Document Analysis System (DAS'94), Kaiserlautern, Germany*, pages 135–155, September 1994.

[12] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.