Combination of Dynamic Bayesian Network classifiers for the recognition of degraded characters

Laurence Likforman-Sulem and Marc Sigelle

TELECOM ParisTech/TSI and CNRS LTCI, 46 rue Barrault F-75634, Paris cedex 13, France.

ABSTRACT

We investigate in this paper the combination of DBN (Dynamic Bayesian Network) classifiers, either independent or coupled, for the recognition of degraded characters. The independent classifiers are a *vertical* HMM and a *horizontal* HMM whose observable outputs are the image columns and the image rows respectively. The coupled classifiers, presented in a previous study,[1] associate the vertical and horizontal observation streams into single DBNs. The scores of the independent and coupled classifiers are then combined linearly at the decision level. We compare the different classifiers -independent, coupled or linearly combined- on two tasks: the recognition of artificially degraded handwritten digits and the recognition of real degraded old printed characters. Our results show that coupled DBNs perform better on degraded characters than the linear combination of independent HMM scores. Our results also show that the best classifier is obtained by linearly combining the scores of the best coupled DBN and the best independent HMM.

Keywords: Graphical models, Hiddden Markov models, Dynamic Bayesian networks, Handwritten digit recognition, Degraded characters

1. INTRODUCTION

Noisy and degraded text recognition is still a challenging task for a classifier.[2] In the field of historical document analysis, old printed documents have a high occurrence of degraded characters, especially broken characters due to ink fading. When dealing with broken characters, several options are generally considered: restoring and enhancing characters[3–5] or recovering characters through sub-graphs within a global word graph optimization scheme.[6] Another solution is to combine classifiers or to combine data. Several methods can be used for combining classifiers,[7] one of them consists of multiplying or summing the output scores of each classifier. In the works of,[8, 9] two HMMs are combined to recognize words. A first HMM, modeling pixel columns, proposes word hypotheses and the corresponding word segmentation into characters. The hypothesized characters or sub segments are then given to a second HMM modeling pixel rows. This second HMM normalizes and classifies single characters. The results of both HMMs are combined by a weighted voting approach or by multiplying scores.

In a previous study,[1] we have built independent and coupled models using Dynamic Bayesian Networks. The independent models are HMM-based architectures which observe either the sequence of pixel columns (the vertical-HMM) or the sequence of pixel rows (horizontal-HMM). The coupled models combine these two HMM architectures into single DBN classifiers: the two HMMs, each including an observation stream associated with state variables, are linked in a graphics-based representation. In coupled models, the two streams, vertical and horizontal, are jointly observed and model parameters

Further author information: (Send correspondence to Laurence Likforman-Sulem)
Laurence Likforman-Sulem: E-mail: likforman@telecom-paristech.fr, Telephone: +33 (0)1 45 81 73 28

(state transition matrices) reflect the spatial correlations between these observations.

Coupled DBN classifiers are single classifiers which can be combined with other classifiers. In this study, we combine classifiers at the decision level from their output scores. We first combine independent HMMs, basic and auto-regressive, and compare this combination with single coupled DBNs. We also compare several combinations at the decision level, by associating pairs of DBN models, independent or coupled. Our results show that the best combination is obtained by combining the best coupled DBN with the best independent HMM.

The paper is organized as follows. In Section 2, we introduce the independent HMM and coupled DBN classifiers, and the combination of their scores. We apply this approach for the recognition of artificially degraded handwritten numerals and old printed characters in Section 3. We draw some conclusions in Section 4.

2. COMBINATION OF DBN CLASSIFIERS

We introduce in this section the independent HMM-based architectures and their combination into single coupled DBNs. Then, we present the linear combination of the scores of these classifiers, independent or coupled.

2.1. Independent HMMs

The independent HMM-based classifiers are constructed using the DBN formalism.[1] There is one observation variable and one state variable for each discrete time $t \geq 1$. We consider four independent classifiers: two basic and two autoregressive HMM classifiers. The vertical HMM is constructed using the vertical writing stream (observation variables Y_t^1), as depicted in Fig. 1-a. Observations consist of columns of pixels (normalized values) obtained from scanning the character image from left to right. Similarly, observations $\{Y_t^2\}$ for the horizontal-HMM consist of rows of pixels obtained from scanning the character image from top to bottom (Fig-1-b). In the following, $i \in \{1, 2\}$ denotes the stream, vertical ($i = 1$) or horizontal ($i = 2$), and $\{X_t^i\}$ are the hidden state variables in stream i.

The parameters of these basic HMMs are CPTs (Conditional Probability Tables) A, CPDs (Conditional Probability Distributions) B and the initial distribution Π. CPTs A are the state-transition matrices associated to nodes X_t^i, $t \geq 2$, CPDs B to observed nodes Y_t^i (because of the stationarity assumption, all nodes X_t^i share the same CPT A and all nodes Y_t share the same matrix B, and the initial state distribution Π is associated to node X_1. State-transition matrices are left-right. Each observation variable Y_t^i is assumed to follow a *single* Gaussian probability density function (pdf) with full covariance matrix. Estimating a mixture of Gaussian pdfs rather than a single one may be more appropriate but would require more training data (approximately twice data for a mixture of two Gaussian pdfs).

The other two independent architectures are auto-regressive HMMs: the observations of the *vertical-AR* and the *horizontal-AR* are dynamically linked in time. The AR models proposed here are switching Markov models [10] and the model order is one. An observed node Y_t^i depends on both the current state X_t^i and the previous observed node Y_{t-1}^i. The two resulting vertical-AR and horizontal-AR single stream architectures remain however independent (Fig. 1-c and d). The only parameters which differ from basic HMMs are the Conditional Probability Distributions CPDs B. The mean μ_k^i of the Gaussian probability density function associated to the current state k is shifted by $W_k^i\ y_{t-1}^i$ according to the previous observation y_{t-1}^i and the regression matrix W. Matrix W_k^i is of size dxd, with d being the

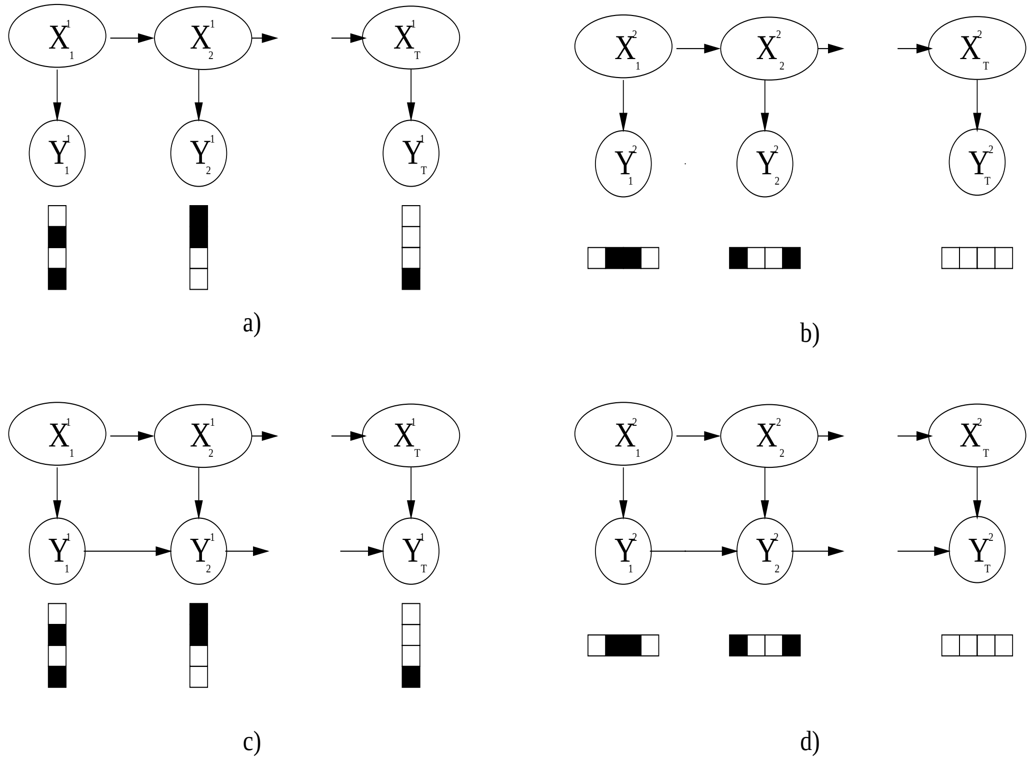

Figure 1. Independent HMM-based architectures represented as DBNs . a) vertical-HMM b) horizontal-HMM. c)vertical-AR d) horizontal-AR.

length of observation vectors. Each observation variable Y_t^i follows a Gaussian probability density function $\mathcal{N}(y_t^i;\ \mu_k^i + W_k^i\ y_{t-1}^i,\ \Sigma_k^i)$. Characters are first preprocessed by a 3×3 Gaussian mask with standard deviation 0.5 . The resulting pixel values are then normalized in $[0., 1.]$. Two observation sequences of length $T = 28$ are obtained from the respective vertical and horizontal streams.
All independent character models share a single DBN architecture but their parameters differ for each class. Parameters are learnt using the EM (Expectation-Maximization) algorithm and inference. For independent HMMs, observation parameters are initialized by assigning observations to states linearly. For AR independent models, observation parameters are initialized randomly. The number of states is common for all models and set to $Q = 14$ which offers the best compromise between complexity and performance.
For training handwritten digit models, we use 5000 characters from the MNIST training database and a crossvalidation methodlogy. For training old printed character models, 50 characters per class are selected from the first three standard pages. The 16 character classes the most represented are used in this study. We use the *BayesNet* toolbox[11] which provides general MatLab source code for training and inference in static and dynamic Bayesian networks.

2.2. Coupled DBNs

The previous architectures are now combined into single classifiers, denoted by coupled-DBNs, where the vertical and horizontal observation streams are observed jointly. Several combined architectures can be obtained (see[1]) and we select in this study the two best ones. The *State-coupled* model results from the combination of the basic vertical and horizontal HMMs. The *AR-coupled* model results from

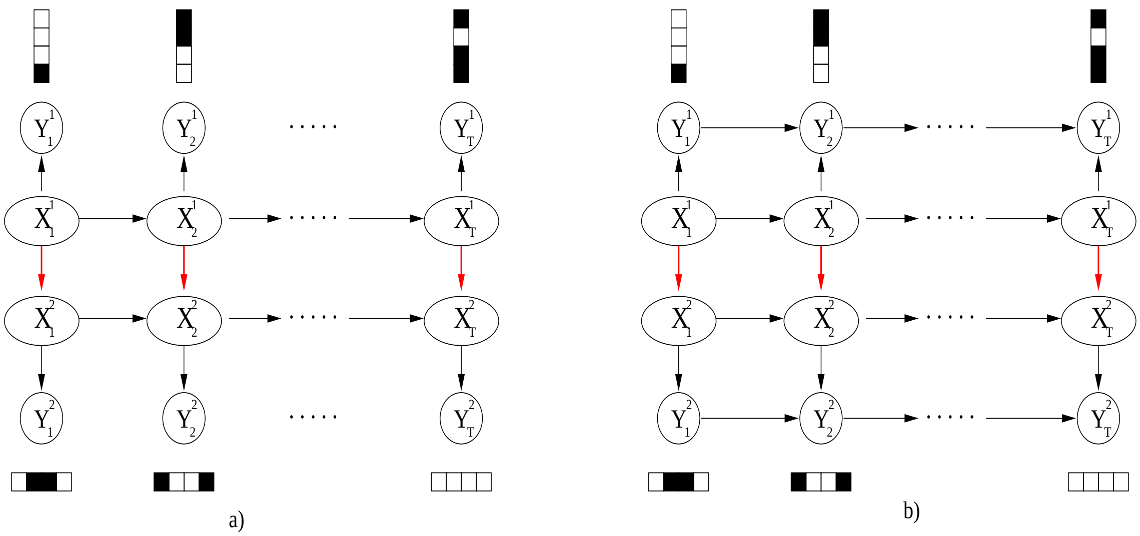

Figure 2. Coupled DBNs. The State-coupled model (a) combines the two basic HMM architectures. The AR-coupled model (b) combines the two auto-regressive HMMs.

the combination of the autoregressive vertical and horizontal AR-HMMs (Fig. 2).
These classifiers are constructed by adding edges between state variables from the vertical and the horizontal streams. Edges are directed from the vertical stream to the horizontal one in order to enhance the influence of the vertical stream. We have noticed that the vertical HMM is more reliable than the horizontal one since vertical strokes are predominant for the shapes considered.

The coupled models have two observation variables for each discrete time t: Y_t^1 and Y_t^2 representing the pixel column and pixel row respectively. At each time, coupled models are in two states, the state corresponding to the column observation (the *vertical* state) and the state corresponding to the row observation (the *horizontal* state). A transition to the vertical state X_t^1 depends only on the value of the preceding state X_{t-1}^1 like classical left-right HMMs. But a transition to the horizontal state X_t^2 depends on both the value of the preceding state X_{t-1}^2 and the value of the current vertical state X_t^1. This dependence between the horizontal and the vertical states expresses the dependence of the observations, i.e. between row t and column t. In the CPT U - which can couple up to 3 states together and encodes $U_{j,k,l} = P(X_t^2 = l \mid X_{t-1}^2 = j, X_t^1 = k)$ - the state transitions within the same stream are left-right whereas the state transitions between different streams can be ergodic. This means that X_t^2 is equal to or greater than X_{t-1}^2 but X_t^2 is equal to, less than or greater than X_t^1.

We set no contraints on the covariance matrices for observation parameters: they may be full as our experiments show better accuracy for such matrices. However, we have noticed that the most significant terms in covariance matrices are arranged within block structures around the diagonal. This can be explained by the fact that correlations mostly occur within neighboring pixels. Thus correlation terms for distant pixels could be set to zero, avoiding learning all terms of the full matrix. The coupling proposed here requires that both observation sequences have the same length since streams are synchronized at each time slice: each image column is associated with one row. The observation length is $T = d$ as the character image is previously normalized to a square of size dxd with $d = 28$ pixels. The coupled DBN models are trained similarly as independent models. For the State-coupled model, observation parameters are initialized to a common value for all states and each stream i.e, the empirical mean and covariance matrix obtained from the sample data. The parameters of the AR-coupled model are initialized randomly.

2.3. Linear combination of DBN scores

This section describes the combination of DBN classifiers at the decision level. We first combine independent HMMs of same type (basic, or auto-regressive) from their scores. These scores are the log-likelihoods output during the decoding process and a weighted linear combination is used. As single classifiers, coupled DBNs can be combined linearly with other classifiers, coupled or independent. Second, we combine pairs of classifiers and for each pair, the optimal weight is searched on a separate validation database (see Section 3).

The combined score for a pattern given a class model results from the weighted sum of the log-likelihoods (scores) provided by each classifier. The weights $\alpha_i, i = 1, 2$ must satisfy the constraints: $\alpha_i \geq 0$ and $\alpha_2 = 1 - \alpha_1$. Thus, only the value $\alpha = \alpha_1$ dedicated to one classifier needs to be optimized. The validation set consists of 1000 MNIST digits from the MNIST training database and distinct from the digits used to train the DBN models. It consists of 1038 characters for old printed characters.

As an example, Fig. 3 shows recognition rates versus α for the combination of basic HMMs (vertical and horizontal). For digits, the maximum is reached for $\alpha = 0.5$. We have observed that log-likelihoods provided by the vertical HMM are in average higher than those provided by the horizontal HMM. Consequently, even $\alpha = 0.5$ gives more weight to the vertical HMM.

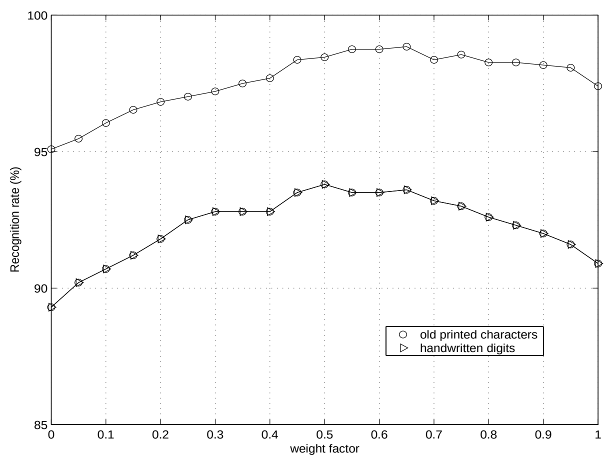

Figure 3. Linear combination of basic HMM scores: recognition rate versus weight factor α

3. EXPERIMENTAL RESULTS

We evaluate the various DBN classifiers and their combination for the problem of the recognition of degraded characters. Two off-line recognition tasks are considered : the recognition of artificially broken handwritten digits and the recognition of real degraded printed characters. The robustness of the combination to degradation is evaluated against different degradation parameters.

3.1. Handwritten digits

The original digit set is the MNIST test database[12] which includes 10,000 digits. This test set is degraded artificially by creating breaks within characters as depicted in.[1] We evaluate the classifiers under parameter w which denotes the number of windows applied to characters: within each window, writing pixels are changed to background pixels. We consider three levels of degradation: no additional degradation using the original MNIST test set ($w = 0$), one break created within digits ($w = 1$) and two breaks created ($w = 2$).

3.1.1. Comparison with the combination of independent HMM scores

Independent HMMs	$w = 0$	$w = 1$	$w = 2$
vertical-HMM	90.2	86.9	83.8
horizontal-HMM	87.4	82.8	75.3
vertical-AR	93.2	89.8	85.3
horizontal-AR	87.7	81.6	75.6
Coupled DBNs			
State-coupled	92.4	90.8	87.4
AR-coupled	**94.9**	**93.4**	**90.9**
Linear Combination			
Combination of HMM scores	93.1	90.6	87
Combination of AR-HMM scores	94.7	91.9	89

Table 1. Recognition of handwritten digits, under different levels of degradation ($w = 0$: no additional degradation, $w = 1$: one break, $w = 2$: two breaks). Comparison of coupled DBNs with the linear combination of independent HMM scores. Recognition rates are given in %.

The linear combination of independent HMM scores is compared with the coupled DBNs, the State-coupled or Auto-regressive coupled models. Results for the degraded test sets are given in Table 1. When $w = 0$ (no degradation), the combination of HMM scores performs better than the State-coupled model and worse than the AR-coupled model. When the level of degradation increases ($w = 1, w = 2$) , both the State-coupled and the AR-coupled models perform better than the combination of HMM scores. The performance of individual HMMs as well as the linear combination of them, deteriorates more rapidly as degradation increases than when they are combined in a coupled DBN model.

We also compare the AR-coupled model with the weighted combination of auto-regressive HMMs. As previously, the combined score for a pattern given a class model results from the weighted sum of the scores provided by each AR-HMM, vertical and horizontal. The optimal weight α, searched on the validation set, is equal to $\alpha = 0.45$ for the AR-HMM combination. The AR-coupled model performs bettter than the combination of AR-HMM scores whatever the level of degradation. The improvement brought by the AR-coupled model is enhanced for degraded characters ($w > 0$).
Missing observations, such as in broken characters, may be predicted through auto-regressive model. Coupled architectures may also include at least one uncorrupted stream, horizontal or vertical, within each time slice and thus better cope with missing information.

3.1.2. Comparison with the linear combination of DBN scores

The coupled DBNs can be combined, as single classifiers, with other independent or coupled classifiers. We combine pairs of DBN classifiers, either independent or coupled. As previously, for each pair of classifiers, the optimal weight is searched on the validation database. Results for the weighted linear combination are given in Table 2, as well as the optimal weight α for each experiment. The best combination over single classifiers is obtained for pairs including the vertical auto-regressive HMM and a coupled DBN classifier (State-coupled or AR-coupled). The combination with the horizontal AR-HMM instead of the vertical one also improves performance but slightly. The other combinations do not improve performance over single classifiers. This shows the complementarity of the independent vertical-AR with the AR-coupled model.

classifier 1	**classifier 2**	**weight** α	$w = 0$	$w = 1$	$w = 2$
vertical-HMM	State-coupled	0.55	92.4	90.4	87.5
horizontal-HMM	State-coupled	0.25	92.5	90.8	87
vertical-AR	State-coupled	0.55	94.6	92.4	89.6
horizontal-AR	State-coupled	0.4	93.1	90.9	87.4
vertical-HMM	AR-coupled	0.2	95.2	93.5	**91.7**
horizontal-HMM	AR-coupled	0.25	95.2	93.5	91.3
vertical-AR	AR-coupled	0.45	**95.6**	**93.7**	91.2
horizontal-AR	AR-coupled	0.25	95	93.3	91.2
State-coupled	AR-coupled	0.45	95	93	90.6

Table 2. Recognition of handwritten digits under different levels of degradation. The scores of DBN classifiers, independent or coupled, are linearly combined. Recognition rates are given in %. The weight corresponding to classifier 1 is α.

3.2. Old Printed Characters

3.2.1. Comparison with the combination of independent HMM scores

Character models are tested on two test sets: a standard test set (test-s) from two standard pages and a degraded test set (test-d) from two degraded pages including fainting characters (Fig. 4). The standard and degraded test sets include 1009 and 1079 characters respectively from 16 classes. The character classes are the most frequent ones: a, b, c, d, e, i, l, m, n, o, p, r, s, long s, t and u. Confusions may mostly occurr between class e and class c and between class n and class u because of fainting, and between class r and class t because of shape similarity and normalization process. Recognition accuracies for old printed characters are given in Table 3. The horizontal stream is less reliable but its coupling with the vertical stream at the decision level or through coupled DBNs, increases performance. The auto-regressive coupled model (AR-coupled) always performs better than independent models, the State-coupled model and the combination of HMM scores for which the optimal value of $\alpha = 0.65$ was found on a validation set of 1038 digits (Fig. 3). The State-coupled and AR-coupled coupled models better cope with degraded characters (test-d) than independent HMMs (basic and AR) and the combination of HMM scores. For the combination of auto-regressive HMMs (AR-HMMs), the optimal weight α searched on the validation set is equal to $\alpha = 0.4$. The AR-coupled model performs better than the combination of AR-HMMs whatever the level of degradation. However, the improvement brought by the AR-coupled model is enhanced when characters are broken due to natural fading process or low binarization threshold.

Independent HMMs	standard (test-s)	degraded (test-d)
vertical-HMM	98.3	93.8
horizontal-HMM	93.7	88.1
vertical-AR	97.9	94.5
horizontal-AR	96.2	91.2
Coupled DBNs		
State-coupled	98.7	95.5
AR-coupled	**98.8**	**96**
Linear Combination		
Comb. of HMM scores	98.4	95.4
Comb. of AR-HMM scores	98.7	95.5

Table 3. Recognition rates (in %) for standard and degraded old printed characters.

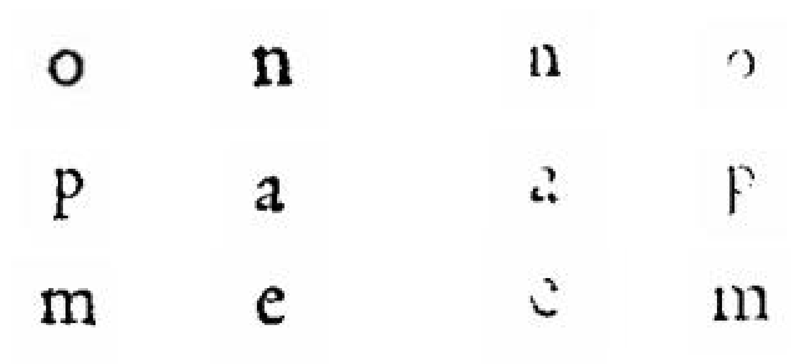

Figure 4. Sample old printed characters from the standard (left) and the degraded (right) sets.

3.2.2. Comparison with the linear combination of DBN scores

Results for the linear combination of DBN classifiers, independent or coupled, are given in Table 4. The weight α corresponding to the classifier in the first column (classifier1) is given for each experiment. Increased performance is obtained when combining a coupled DBN (State-coupled or Auto-regressive coupled) with an independent vertical classifier (basic or auto-regressive). There is a slight improvement when combining a coupled DBN with the horizontal AR-HMM.
As previously, the best combination is obtained with the pair including the AR-coupled and the vertical AR classifier.

4. CONCLUSION

We have proposed in this paper to combine DBN architectures for the recognition of degraded characters. First, we have presented four independent classifiers, two basic HMMs and two auto-regressive HMMs represented through the DBN formalism. Each HMM observes one writing stream, either horizontal or vertical. Second, the architectures of these independent HMMs have been combined resulting in single coupled DBN classifiers observing jointly the horizontal and the vertical streams. Third, single classifiers (independent or coupled) have been combined in a decision fusion scheme by linearly combining their scores.

classifier 1	**classifier 2**	**weight** α	**test − s**	**test − d**
vertical-HMM	State-coupled	0.6	98.8	96.1
vertical-AR	State-coupled	0.4	99.4	97
horizontal-HMM	State-coupled	0	98.7	95.5
horizontal-AR	State-coupled	0.35	98.6	95.7
vertical-HMM	AR-coupled	0.7	98.9	96.8
vertical-AR	AR-coupled	0.2	**99**	**96.7**
horizontal-HMM	AR-coupled	0.2	98.7	96
horizontal-AR	AR-coupled	0.25	98.9	96
State-coupled	AR-coupled	0.35	99	96.5

Table 4. Linear combination of DBN classifiers for the recognition of old printed characters: standard characters (test-s) and degraded characters (test-d). The weight corresponding to classifier 1 (resp. classifier 2) is α (resp. $1-\alpha$). Recognition rates are given in %.

We have applied the different combination schemes to the recognition of degraded characters: artificially broken digits and naturally broken old printed characters. Accuracy is enhanced through combination. We have shown that for degraded characters, coupled DBNs perform better than the linear combination of independent HMM scores . The best performance is obtained by combining at the decision level the best independent HMM (the vertical AR-HMM model) with the best coupled DBN (the AR-coupled model).

Future directions will consist on evaluating the impact of deslanting characters previous to recognition since our models fit characters with dominant vertical and horizontal strokes. Other future directions consist of improving the combining strategy. Ensemble-based strategies such as Mixtures of Bayesian Networks or Adaboost, could be used.

REFERENCES

1. L. Likforman-Sulem and M. Sigelle, "Recognition of degraded characters using dynamic bayesian networks," *Pattern Recognition* , http://dx.doi.org/DOI10.1016/j.patcog.2008.03.022.
2. H. Baird, "The state of the art of document image degradation modeling," in *Proc. of the the 4th Workshop on Document Analysis Systems, DAS*, pp. 1–16, (Rio de Janeiro), 2000.
3. A. Whichello and H. Yan, "Linking broken character borders with variable sized masks to improve recognition," *Pattern Recognition* **29**(8), pp. 1429–1435, 1996.
4. B. Allier, N. Bali, and H. Emptoz, "Automatic accurate broken character restoration for patrimonial documents," *IJDAR* **8**(4), pp. 246–261, 2006.
5. A. Antonacopoulos and D. Karatzas, "Document image analysis for World War II personal records," in *First Int. Workshop on Document Image Analysis for Libraries, DIAL 04*, pp. 336–341, (Palo Alto), 2004.
6. M. Droettboom, "Correcting broken characters in the recognition of historical printed documents," in *Proc. of Joint Conference on Digital Libraries, JCDL'03*, 2003.
7. J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE PAMI* **3**(20), pp. 226–239, 1998.
8. W. Wang, A. Brakensiek, and G. Rigoll, "Combining HMM-based two pass classifiers for off-line word recognition," in *Proc. of ICPR*, pp. 151–154, (Quebec), 2002.
9. A. Elms, S. Procter, and J. Illingworth, "The advantage of using an HMM based approach for faxed word recognition," *IJDAR* **1**, pp. 18–36, 1998.
10. J. Hamilton, "Analysis of time series subject to changes in regime," *Journal of Econometrics* **45**, pp. 39–70, 1990.

11. K. Murphy, "BayesNet Toolbox for Matlab." Available on the web at http://www.ai.mit.edu/murphyk/Bayes/bnintro.html, 2003.
12. Y. LeCun, "The MNIST handwritten digit database." Available on the web at http://www.research.att.com/~yann/ocr/mnist/, 1998.