A contribution to automatic recognition of handwritten Arabic scripts

Hanene Boukerma

April 25, 2019

Acknowledgments

I extend my deepest thanks to all those who contributed directly or indirectly to complete this dissertation.

I would like to thank Nadir Farah, my dissertation advisor, for his help, encouragement, and kindness. He was also my magister's thesis advisor; he was the first professor who introduced me to the world of research.

I would like to express my sincere gratitude to Christophe Choisy. This work would have been very different without his great help and guidance throughout my research in general, and the NSHP-HMM in particular. I am very grateful for his encouragement, continuous support and presence during the last six years even remotely through email.

I am also extremely grateful to Mohamed Cheriet for his invaluable help and kindness. It was a real pleasure to work with him in the Synchromedia laboratory: he gave me the opportunity of running my experiments on the two servers Galar and Guillimin that really helped me in optimizing my time and energy. He also provided English proofreading service for my articles.

My deep appreciation also goes to Hakima Ould-Slimane, Ghania Atek, and Yosra Njah for their great help, kindness, and support with their presence as a family during my internships in Montreal.

I also wish to thank the chair and members of my dissertation defense committee: Labiba Souici-Meslati, Smaine Maazouzi and Hamid Seridi for accepting the responsibility of reviewing my dissertation and for their valuable feedbacks. I am especially grateful to Labiba Souici-Meslati for her generosity, kindness and, availability to help me any time with suggestions and clarifications.

I am also grateful to Samira Hazmoune and Abdallah Benouareth for sharing with me their knowledge on the HMM model.

Words could never be enough to express my love and appreciation to my dear mother and father. I will forever be grateful for all that you have done, doing, and will do for me. May Allah be pleased with you!

Words are also powerless to express my gratitude to my sisters, brothers and all my family members, especially Mohamed !, for their help, support, and patience.

I am all gratitude to my friends Gehan, Zhour, Meriem, Soumaya, Rafika and Hanifa for their help, encouragement, and kindness.

I am indebted to all my teachers, especially Ali Bouachari, for their help, patience, and generosity. Finally, I would like to extend my sincere gratitude to all those who have taught me new ideas, knowledge, and skills through my life.

Abstract

The performance of unconstrained handwritten words remains a challenging pattern recognition problem. In recent years, stochastic two-dimensional (2-D) models, especially Non-Symmetric Half-Plane Hidden Markov Model (HSHP-HMM), have been successfully applied to the area of off-line handwriting recognition. Contrary to all existing HMM-based 2-D model, the NSHP-HMM does not suffer from the problem of exponential complexity in terms of transitions. It solves the approximations by using a real 2-D context estimated in term of pixels. Consequently, the NSHP-HMM brings the efficient training and recognition algorithms of 1-D HMM to the 2-D modeling of spatial data. This characteristic makes the NSHP-HMM one of the most efficient optimal 2-D recognizers. However, the main drawback of this model is the short 2-D context measured in terms of only a few pixel context, and hard decision on its value: only black and white pixels are allowed, needing a binary transformation that breaks down a part of its advantages.

In this dissertation, we demonstrate that the use of the NSHP Markov random field to describe the contextual information at the *'zone'* level rather than the *'pixel'* level gives an efficient solution to the reference model's limits. Using the proposed model, called the NSHP$^{\mathrm{Z}}$-HMM (Z for 'zone'), the 2-D context is extended allowing a better modeling of the spatial property of an image. Therefore, the use of high-level features extracted directly on the gray-level or color zones is possible, unlike what is done in a recognition based on classical NSHP-HMM, where the model, mandatorily, operates at a pixel level on normalized binary images; consequently, the applicability of our model is more general compared to the classical NSHP-HMM.

Throughout this dissertation, we demonstrate the efficiency of the proposed approach at two stages. Firstly, in the theoretical study, we show the advantage of our model over other HMM-based 2-D classifiers. In this part, we present to our

best knowledge, the first complete overview of 2-D recognition approaches; we also explain why the NSHP^{Z} is a natural extension of the classical NSHP. Secondly, the experimental evaluations performed on recognition of handwritten digits/words provides the effectiveness of the NSHP^{Z}-HMM against all other HMM-based 2-D recognizers.

To improve the NSHP^{Z}-HMM performance on Arabic script recognition, we introduce a new zoning design approach based on baseline localization. The effectiveness of the proposed approach is tested separately and according to recognition accuracy. The key point is to adequately divide the image of Arabic word into zones considering an efficient baseline estimation method.

Finally, we propose a hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities. Analyzing the effect of the hybrid model parameters showed that extending the 2-D context at the zone-level is more effective than its enlargement at the pixel-level, which justifies once more the superiority of the NSHP^{Z}HMM over the classical NSHP-HMM.

keywords Two-dimensional Hidden Markov Models, Non-Symmetric Half-Plane Markov chain, zoning, handwriting recognition, Arabic sub-words, baseline.

Résumé

Le développement d'un system de reconnaissance de mots manuscrits sans contrainte reste un des défis dans le domaine de la reconnaissance de formes. Dans ces dernières années, des modèles stochastiques bidimensionnels (2-D), en particulier le modèle de Markov caché à demi-plan non-symétriques (HSHP-HMM), ont été appliqués avec succès dans le domaine de la reconnaissance hors ligne de l'écriture manuscrite. Contrairement à tous les modèles 2-D à base de HMM, le NSHP-HMM ne souffre pas du problème de la complexité exponentielle en termes de transition entre états. Il résout les approximations par l'utilisation d'un contexte 2-D réel en terme de pixels. Par conséquent, le NSHP-HMM hérite les algorithmes d'apprentissage et de reconnaissance initialement conçus pour les HMM 1-D à la modélisation 2-D des données spatiales. Cette caractéristique rend le NSHP-HMM un des classificateurs 2-D les plus optimaux. Cependant, le principal inconvénient de ce modèle est le contexte 2-D limité à quelques pixels seulement, et une forte limite quand à leurs valeurs : seuls le noir et le blanc sont considérés, ce qui impose une binarisation qui pénalise de fait les avantages de ce modèle.

Dans cette thèse, nous démontrons que l'utilisation des chaînes de Markov à demi-plan non-symétriques (NSHP) pour modéliser les informations contextuelles au niveau *'zone'* plutôt qu'au niveau *'pixel'* offre une solution efficace aux limites du modèle classique. En utilisant le modèle proposé, appelé NSHP$^{\mathrm{Z}}$-HMM (Z pour 'zone'), le contexte 2-D est étendu permettant une meilleure modélisation de la propriété spatiale d'une image. Par conséquent, l'utilisation de caractéristiques de haut niveau extraites directement sur des zones en niveau de gris ou des zones en couleurs est possible; contrairement à ce qui se fait dans une reconnaissance basée sur le NSHP-HMM classique, où le modèle s'applique au niveau des pixels sur des images binaires normalisées. Par conséquent, l'applicabilité de notre modèle est plus

générale que celle du NSHP-HMM classique.

Tout au long de cette thèse, nous démontrons l'efficacité de l'approche proposée en deux étapes. Premièrement, dans l'étude théorique, nous montrons l'avantage de notre modèle par rapport aux autres classificateurs 2-D à base de HMM. Dans cette partie, nous présentons, à notre connaissance, la première étude complète des approches de reconnaissance bidimensionnels ; nous expliquons également pourquoi le NSHP$^{\mathrm{Z}}$ est une extension naturelle du NSHP classique. Deuxièmement, l'évaluation expérimentale effectuée sur la reconnaissance des chiffres/mots manuscrits montre l'efficacité du NSHP$^{\mathrm{Z}}$-HMM par rapport à tous les autres systèmes de reconnaissance 2-D à base de HMM.

Afin d'améliorer les performances du NSHP$^{\mathrm{Z}}$-HMM sur la reconnaissance de l'écriture Arabe, nous introduisons une nouvelle approche de division en zones basée sur la localisation de la ligne de base. L'efficacité de l'approche proposée a été testée individuellement et en fonction du taux de reconnaissance. Le point clé réside dans la division adéquate des images de mots arabes en s'appuyant sur une méthode efficace d'extraction de la ligne de base.

Finalement, nous proposons un NSHP-HMM hybride combinant les probabilités d'observation au niveau pixels avec celles au niveau zones. En analysons l'influence des paramètres du modèle hybride sur ses capacités de reconnaissance, nous concluons que l'extension du contexte 2-D au niveau de la zone est plus efficace que son extension au niveau du pixel, ce qui justifie une fois de plus la supériorité du NSHP$^{\mathrm{Z}}$HMM sur le NSHP-HMM classique.

Mots-clés modèle de Markov caché bidimensionnels, chaînes de Markov à demi-plan non-symétriques, zonage, reconnaissance de l'écriture manuscrite, pseudo-mots Arabe, ligne de base.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The recognition of unconstrained handwritten words is an area of pattern recognition that continues to pose a major challenges to researchers. The current research is addressed towards some difficult handwritten scripts such as Chinese, Indian and Arabic.

In this dissertation, we introduce a recognition system of handwritten Arabic words. The novelty of our work consists in two key points. Firstly, we propose a new HMM-based two-dimensional (2-D) recognizer, called the NSHP$^{\mathrm{Z}}$-HMM (for Non-Symmetric Half-Plane Hidden Markov Model based on conditional Zone observation probabilities). Secondly, we introduce a zoning design approach based on baseline to adapt the NSHP$^{\mathrm{Z}}$-HMM to better model the particularities of Arabic writing.

The proposed recognizer is an HMM-based model. In the machine learning area, HMM has emerged as a powerful technique [Rabiner, 1989]. The strength of this technique lies in its consistent statistical framework that offers computationally efficient algorithms to solve the three fundamental problems of HMM: probability evaluation, finding the optimal state sequence and model parameter estimation.

Using an adequate modeling of observation sequences is an important factor in determining the success of any pattern recognition system based on HMMs. For one-dimensional (1-D) speech signals, temporal information can be accurately modeled by 1-D HMMs. This can justify the success of 1-D HMMs in the field of the automatic speech recognition. In the image recognition area, 1-D HMMs are widely used

by making a simplifying assumption that consists in modeling the 2-D properties of the image through a sequence of feature vectors extracted by a sliding window method. The reported results from this approach are generally not as good as for speech recognition; obviously, the penalty comes from losing information about planes that presents a fundamental difference from speech.

It is why during the last few years, considerable research effort has been devoted to how to extend 1-D HMM to plane in order to properly describe the 2-D features of spatial data. Planar-HMM (PHMM), also called "pseudo" 2-D HMM in common sense that it is not fully connected 2-D HMM, presents the first attempt in this direction.

The causal Markov Random Fields (MRF) is the 2-D counterpart of the 1-D Markov chain in which the natural ordering of past, present, and future is replaced by the spatial concept of neighborhood [Perronnin, 2004]. Two types of causal MRFs have been extensively used in image processing: the Markov Mesh Random Fields (MMRFs) and the unilateral Markov random fields which is also called the Non-Symmetric Half-Plane (NSHP) Markov chain [Jeng and Woods, 1987].

Signal modeling by HMM consists of properly setting the state transition probability matrix A and the observation probability matrix B. If the signal is 2-D in nature, modeling it can be done via MRF at *A-level* or *B-level*. This gives as two principal types of "truly" 2-D HMM: i) Hidden Markov Mesh Random Field (HMMRF) with a 2-D state transition matrix modeled by an MMRF [Devijver and Dekesel, 1987] and ii) Non-Symmetric Half-Plane Hidden Markov Model (NSHP-HMM) in which the NSHP is used at the B-level to model the 2-D property of an image [Saon, 1999].

The main drawback of the models of the first category of 2-D HMM (MRF for 2-D modeling at the *A-level*) is their exponential complexity for parameter estimation and image recognition. To deal with this problem, all the proposed approaches were developed on the basis of restrictive assumptions regarding models, such as approximating a 2-D HMM with many 1-D HMMs [Likforman-Sulem and Sigelle, 2008], and using sub-optimal training and decoding algorithms (for example, the decision-directed algorithm [Devijver and Dekesel, 1987] and the one-row-one-column look-ahead technique [Park and Lee, 1998]). In practice, the used approximations and assumptions reduce the modeling accuracy and limit the exploitation of the full 2-D structure of spatial data.

On the other hand, without suffering from exponential complexity in terms of transitions, as do other HMM-based 2-D models, the NSHP-HMM brings the efficient training and recognition algorithms of 1-D HMM to the 2-D modeling of spatial data. This characteristic makes the NSHP-HMM one of the most efficient optimal 2-D recognizers. However, the main drawback of this model is the short 2-D context measured in terms of only a few pixels. For example, a maximum number of the neighborhood (V parameter) equaling 4 'pixels' was used in [Saon, 1999] and [Choisy and Belaïd, 2002] for cursive word recognition; the reason was to maintain a suitable compromise between accuracy and NSHP-HMM model complexity in terms of parameter numbers.

In this dissertation, we have demonstrated that the use of the NSHP Markov random field to describe the contextual information at the *'zone'* level rather than the *'pixel'* level gives an efficient solution to the reference model's limits. The goal is to extend the context in order to produce a better modeling of the spatial properties of an image. This new model, called NSHP$^{\mathrm{Z}}$-HMM, provides an optimal solution that combines the effectiveness of 2-D modeling by NSHP-HMM with an appropriate pattern representation by zoning, which allows also bypassing any vertical normalization by the adjustment of vertical zone sizes. In this work, we also introduce a new zoning approach based on baseline localization to overcome the vertical normalization constraint inherent to the classical NSHP-HMM. Therefore, the use of low- to high-level features extracted directly on the gray-level or color zones is permitted, unlike what is done in a recognition based on classical NSHP-HMM, where the model, mandatorily, operates at a pixel level on normalized binary images.

Note that the proposal is actually a generalization of the former NSHP-HMM; as if we reduce zones to individual pixels and symbols as {black, white}, we exactly reproduce the former approach.

1.2 Contributions and outline

The outline of this dissertation is presented below. We note that the content of each chapter corresponds to original contribution.

In chapter 2, we will first review and classify the two-dimensional-based recognizers. Our goal is to clearly show the value of our contribution and to make

the comparison with other related approaches easier for the reader to understand. We emphasize that the contents of this chapter offer, to our knowledge, the first complete overview of 2-D based recognition approaches [Boukerma et al., 2018b]. Therefore, the review part of this dissertation is far to be just a list of references. Below are more specific arguments:

(i) We propose a classification of 2-D recognition approaches that have not been published in literature before. The proposed taxonomy involves two levels: Non-HMM and HMM-based approaches. For the second class, we have two sub-categories: MRF for 2-D modeling at the A-level and MRF for 2-D modeling at the B-level.

(ii) In analyzing the existing works of HMM-based 2-D recognizers, we define a general criterion that considers the optimality issue of the training and decoding algorithms. We discuss the advantages and the weakness of the existing approaches in the category "MRF for 2-D modeling at the A-level" and come to the conclusion that efficient algorithms to solve the three problems of HMM do not exist for the proposed 2-D models of this category.

(iii) We introduce then the existing systems of the second category "MRF for 2-D modeling at the B-level". At this stage, we define another criterion to evaluate the existing works. This criterion considers the four drawbacks of the classical NSHP-HMM. We evaluate the existing systems in terms of their ability to overcome these four drawbacks.

(iv) At this point of discussion, we introduce our model and demonstrate how it advances the state of the art by resolving the four mentioned drawbacks.

Through chapter 3, we define in more detail the proposed model and derive their re-estimation formulas. We then provide a complete description of the principal steps involved in the image recognition system based on the NSHP$^{\mathrm{Z}}$-HMM. In order to evaluate our proposed model in ascending order of difficulty, the first series of experiments are performed on handwritten digits recognition [Boukerma et al., 2014] [Boukerma et al., 2018b].

In chapter 4, we consider the problem of Arabic handwritten word recognition [Boukerma et al., 2015]. We introduce a new zoning design approach based on base-

line localization [Hanene et al., 2018]. Our goal is to appropriately partition the image of Arabic word into zones that facilitates its modelization using the NSHP$^{\mathrm{Z}}$-HMM.

Through chapter 5, we introduce a hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities [Boukerma et al., 2018a]. Despite the promising results of the hybrid model, further investigations are required. This is why we have decided to present the preliminary results of the hybrid model in the appendix part of our dissertation.

Finally, in chapter 6, we conclude this dissertation.

Chapter 2

Two-dimensional based recognition approaches: the State of the Art

2.1 Introduction

Before presenting our model, namely the NSHP^{Z}-HMM (see chapter 3), we are aware that it is necessary to clearly show the value of our contribution regarding other existing works. For this reason, we present in this chapter a methodology overview of 2-D recognition engines. To the best of our knowledge, this overview presents the first complete survey of 2-D recognition approaches. Our goal is to clearly show the value and the novelty of our research work by defining the drawbacks of the existing methodologies and discussing how the proposed method solves the issue.

2.2 Classification of 2-D based recognition approaches

Our own interest is the 2-D HMM models; but in order to give a complete review, we also discuss non-HMM based 2-D recognizers. Thus, the first level of the proposed classification is HMM-based and non-HMM-based 2-D recognizers. The taxonomy of 2-D based recognition approaches is shown in Fig. 2.1. In order to present our contribution methodically, we will first start with the non-HMM based 2-D recognition approaches.

2.2.1 Non-HMM based 2-D recognition approaches

In [Uchida and Sakoe, 1999] [Uchida and Sakoe, 2005] [Ronee et al., 2001], a pixel-to-pixel mapping technique using dynamic programming based Piecewise Linear two-Dimensional Warping (PL2DW) is presented. In PL2DW, the mapping of each column of one image into another image is given by the linear interpolation of the mapping of some specific points, called pivots, on that column. Thus, the mapping is controlled by K pivot-points with K less than or equal to the image size. The computational complexity of this technique is exponential order of the number of pivots K and polynomial order of the image size. The PL2DW was tested on English handwritten character recognition. In order to keep computations tractable, small K is used with additional constraints for individual categories of similar characters like 'H' and 'M'.

In [Chevalier et al., 2003], the authors proposed a 2-D approach for handwriting recognition based on Markov Random Fields models and 2-D dynamic programming (DP). The 2-D DP is used to compute the optimal configuration of an MRF model. However, to keep the model computationally tractable, the authors used a pruning strategy in order to retain, at each step of the computation, only the most probable configuration. Tested on the MNIST database, the error rate of the proposed system was 5.4%. An extension of this technique to handwritten words recognition was presented in [Chevalier et al., 2005].

A fuzzy approach to 2-D shape recognition was introduced in [Lazzerini and Marcelloni, 2001]. This approach is based on the fuzzy description of shapes in 2-D space. The horizontal and vertical coordinates of shape points and the triangular membership functions in horizontal and vertical spaces are the two basic elements for modeling shape instances. The work was tested on two applications: recognition of olfactory signals and recognition of Latin handwritten characters, with recognition rate equals to 87% and 75.86%, respectively.

Graves et al. [Graves et al., 2007] introduced Multi-Dimensional Recurrent Neural Networks (MDRNN) as an efficient extension of the Recurrent Neural Network (RNN) to multi-dimensional data. In MDRNN, the single recurrent connection found in standard RNNs is replaced by as many recurrent connections as there are dimensions in the data. The overall complexity of MDRNN training is linear in the number of data points and the number of network weights. In this work,

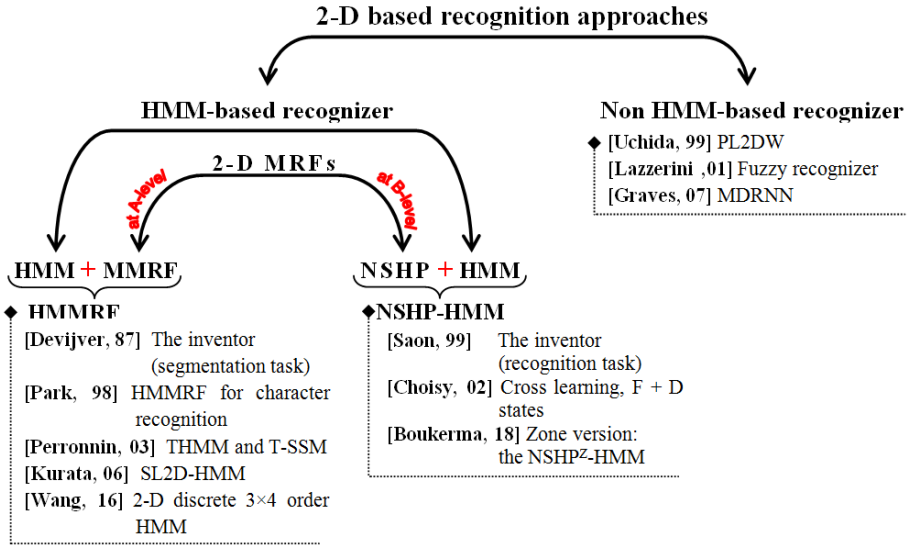


Figure 2.1: Two-dimensional based recognition approaches: a proposed classification and important advancements

the proposed MDRNN is, specifically, Multi-Dimensional Long Short-Term Memory (MDLSTM). Applied to the image segmentation problem (recognition of the MNIST digit is regarded as a segmentation task), the MDLSTM outperforms the Convolution Neural Network (CNN); its recognition error rate was 0.9%. However, the CNN outperformed the MDRNN in terms of training time. On the MNIST database, the MDRNN training time was over two weeks. In [Graves and Schmidhuber, 2009], the MDLSTM was combined with Connectionist Temporal Classification (CTC) and a hierarchical layer structure to create an efficient 2-D recognizer. More details can be found in [Graves, 2008].

The work in [Graves and Schmidhuber, 2009] and [Graves, 2008] was applied to Arabic script; a sort discription of this work with their obtained results will be presented in Section 4.3.1.

Brand et al. [Brand et al., 1997] proposed an algorithm for coupling and training HMMs. The proposed technique consists of introducing "tables conditional probabilities" between the state variables of the two coupled HMMs; this is done by taking the Cartesian product of their states and transition parameters. The proposed coupled HMM (CHMM) was used to classify two-hand gestures from Chinese martial art and meditative exercise. On this action recognition task, the CHMM outperforms the 1D HMM and linked HMMs and offers superior training speeds,

model likelihoods, and robustness to initial conditions.

In [Likforman-Sulem and Sigelle, 2007] and [Likforman-Sulem and Sigelle, 2008], the authors tried to capture the 2-D nature of character images by coupling 1-D separate HMMs. These latter are a vertical HMM and horizontal HMM whose observable outputs are the image columns and image rows, respectively. The coupling method was done according to Dynamic Bayesian Networks (DBNs) formalism. Two coupled architectures were proposed: state-coupled between the vertical and the horizontal HMMs and the Auto-Regressive (AR) model, which coupled vertical and horizontal AR models. Size normalization of images is mandatory in this system. Experimented tests conducted on the MNIST database show that the AR-coupled models perform better than independent ones and outperform the SVM classifier on degraded characters.

Even though the HMM is used in the two last mentioned systems [Brand et al., 1997] and [Likforman-Sulem and Sigelle, 2008] , we decided to cite these works in the non-HMM-based category because the 2-D property in these models is not modeled at either the A-level or the B-level.

2.2.2 HMM-based 2-D recognition approaches

We point out at the beginning of this section that our key vision in reviewing the proposed works for HMM-based 2-D recognizers is formulated as follows: extending 1-D HMM to truly 2-D HMM requires an efficient extension of the two algorithms, namely, the Viterbi and the Baum-Welch algorithms to the 2-D case. In other words, the optimality issue of the training and decoding algorithms in the 2-D case is our main objective. According to this vision, we evaluate older and more recent published literature in HMM-based 2-D recognition approaches.

2.2.2.1 Pseudo 2-D HMM

Talking about the research effort in 2-D HMM traditionally requires starting with the Pseudo 2-D HMM (PHMM), also sometimes referred to as Planar or embedded HMM [Agazzi et al., 1993]. This model presents the first attempt to extend 1-D HMM to 2-D. Theoretically speaking, PHMM is treated as nested one-dimensional models, rather than being truly two dimensional. However, the advantage of PHMM

compared to HMM lies in its nice elastic matching property in both the horizontal and vertical directions, which makes this recognizer less sensitive to size normalization and slant correction. In practice, such as keyword spotting in 'printed' documents [Kuo and Agazzi, 1994] and 'printed' word/character recognition [Agazzi et al., 1993], PHMM performs much better than HMM. However, the drawback of PHMM lies in the hypothesis of lines (or columns) independency which constitutes the main limitation of this model for non-regular images such as 'handwritten' text recognition.

To overcome this drawback, Gilloux [Gilloux, 1995] proposed a combination of PHMM with Markov meshes for handwritten character recognition. Gilloux's model is based on the assumption that the assignment of hidden states to image pixels is properly performed by PHMM, which cannot be guaranteed. Markov meshes are then used to estimate the generation probability of an image and its associated states.

In [Li et al., 2017], the likelihood probability of twofold HMM was used as the health index for the rolling element bearings. The proposed model is composed of a supper 1-D HMM with a simple 1-D HMM embedded. The simple HMM was used to deal with the internal data property among the multiple features and the supper HMM was used to deal with the integral property of the features.

2.2.2.2 Truly 2-D HMM

Contrary to PHMM, causal Markov random fields (MRFs) are actually 2-D statistical models. Two types of causal MRFs have been extensively used in image processing: the Markov mesh random fields (MMRFs) (see *Definition 2.1*) and the unilateral Markov random fields which is also called non-symmetric half-plane (NSHP) Markov chain (see *Definition 2.2*). Jeng and Woods make in [Jeng and Woods, 1987] a comparative study between both models: MMRFs and NSHP Markov model differ in their choice of past and local state (see Fig. 2.2), they are equivalent when each has a quarter-plane local state. However, the MMRFs are conditionally independent on 40° diagonal which reduces their capacity to capture strokes having these orientations. The authors concluded that the NSHP Markov chain is the better model to use when an accurate model of the spatial data is required.

In image recognition, truly 2-D HMM can be built by integrating causal MRFs at two distinguish modeling outlooks: - a MMRF characterizing the 2-D state-

transition probability distribution (MRF for 2-D modeling at the A-level), and – a NSHP Markov random field realization (pattern image) that is considered as an observation sequence of columns (NSHP for 2-D modeling at the B-level).

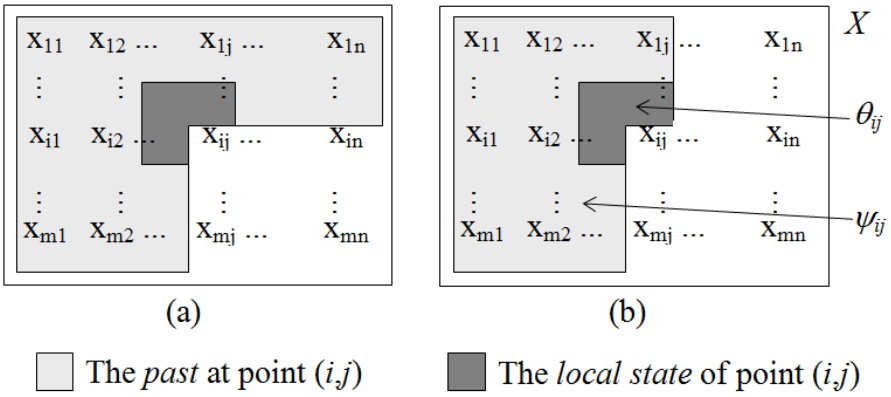


Figure 2.2: Past and local state regions of : (a) Markov mesh random field, (b) NSHP Markov chain.

2.2.2.2.1 MRF for 2-D modeling at the A-level

Before describing the HMMRF (Hidden Markov Mesh Random Field), also referred in the literature as 2-D HMM, we first define the MMRF.

Definition 2.1 Lets us consider a random field $X = \{X_{ij}\}_{(i,j)\in L}$ defined over an $m \times n$ integer lattice L. Let $\psi_{ij} = \{(k,l) \in L | 1 < k < i \; or \; 1 < l < j\}$ the past at the site[1] (i,j). $\theta_{ij} \subset \psi_{ij}$ is the support of the site (i,j), also called the local state of (i,j) (see Fig 2.2.a). Then X is a causal bidimensional MMRF if and only if for all $(i,j) \in L$:

$$P(X_{ij}|X_{\psi_{ij}}) = P(X_{ij}|X_{\theta_{ij}}) \tag{2.1}$$

The 2-D HMMRF assumes that there is set of observations $X = \{x_{ij},\ i = 1..m,\ j = 1..n\}$, which is a probabilistic function of an MMRF generated by a set of states $Q = \{q_{ij},\ i = 1..m,\ j = 1..n\}$. Let λ be the set of all HMMRF parameters, the likeli-

[1]We note that the 'site' can corresponds to one pixel or a bloc of pixels.

hood of observation X given the model λ is defined as:

$$P(X|\lambda) = \sum_{Q} (X, Q|\lambda) = \sum_{Q} (X|Q, \lambda) \cdot P(Q|\lambda) \tag{2.2}$$

$$P(X|\lambda) = \sum_{Q} \prod_{i=1}^{m} \prod_{j=1}^{n} (x_{ij}|q_{ij}, \lambda) \cdot P(q_{ij}|q_{kl}, (k,l) \in \theta_{ij}, \lambda) \tag{2.3}$$

Consider a second-order HMMRF with the support set $\theta_{ij} = \{(i, j-1), (i-1, j)\}$. The order[2] of the HMMRF corresponds here to the number of the site in θ_{ij}. Now equation (2.3) taking into account a second-order HMMRF becomes:

$$P(X|\lambda) = \sum_{Q} \prod_{i=1}^{m} \prod_{j=1}^{n} (x_{ij}|q_{ij}, \lambda) \cdot P(q_{ij}|q_{i,j-1}, q_{i-1,j},\ \lambda) \tag{2.4}$$

The interpretation of the computation in equation (2.4) clearly shows that even for the simple second-order HMMRF the direct extension of the Viterbi and the Baum-Welch algorithms to the 2-D case is exponential to the size of the data $m \times n$.

In order to avoid the exponential complexity inherent to a complete decoding of 2-D state transition matrix, all proposed works, to our knowledge, without exception are based on a set of approximations and assumptions regarding their models and their proposed algorithms. In the following, we will discuss the important efforts made in this direction. Our discussion concerns principally the algorithms and assumptions used to make computationally feasible methods. Another important element is the application of the discussed works and their obtained results.

Devijver et al. were the first to propose HMMRF to image restoration and segmentation. They also proposed the Pickard MRF for the same goal [Devijver and Dekesel, 1988]. The proposed learning technique is based on a simplified version of the Expectation-Maximisation (EM) algorithm called Decision-Directed (DD) [Devijver and Dekesel, 1987]. This algorithm makes the hypothesis that the lines and the columns are mutually independent, which may decrease the modeling accuracy of the model.

Devijver's work was further developed by Park et al. [Park and Lee, 1998]. Their work presents the first attempt to apply HMMRF to off-line handwritten character recognition. The model is a third-order HMMRF that was proposed for digit

[2]Other studies such as [Perronnin, 2004] consider the order as the distance between sites. Thus, according to the latter view, the HMMRF defined here is a first-order Markovian model.

recognition [Park and Lee, 1998] and handwritten Korean Hangul character recognition [Park et al., 2001]; the obtained results were 90.80% and 87.20%, respectively. As noted by the authors, the computational concerns in their HMMRF model necessitated certain simplifying assumptions on the model and approximations on the implementation of the estimation algorithm, such as the one-row-one-column look-ahead technique for the decoding problem and approximations for the implementation of the estimation algorithm. One significant drawback of the look-ahead technique is its computational complexity: for a third-order HMMRF model, this technique requires a total complexity of $O(H^4)$ operations per pixel where H is the number of states. For model parameters estimation, the authors evaluated two methods: the DD algorithm and an extension of the look-ahead technique. As noted by the authors, the convergence of both methods is not guaranteed.

Li et al. [Li et al., 2000] proposed the first analytic solution using the Expectation-Maximization (EM) algorithm. In this work, the re-estimation formulas for the means, the covariance and the transition probabilities are approximated by assuming that the single most likely state sequence accounts for virtually all the likelihood of the observations (suboptimal version of the Viterbi training algorithm). An enhancement of computational difficulty is proposed by the same authors in [Joshi et al., 2006]. The proposed parameter estimation algorithm is a polynomial in time considering the number of states and linear in time considering the number of pixels of an image, and was applied for both 2-D HMM and 3-D HMM. Furthermore, another extension of the work in [Li et al., 2000] was proposed by [Ma et al., 2007], in which the model allows state dependency in a diagonal direction. Its applicability is also for image segmentation.

In order to reduce the complexity of models proposed by [Park and Lee, 1998] and [Li et al., 2000], Othman et al. [Othman and Aboulnasr, 2003] make the assumption of conditional independence among the neighboring feature blocks. This assumption allows the separation of the transition matrix into vertical and horizontal state transition matrices. As in [Devijver and Dekesel, 1987] and [Park and Lee, 1998], the DD learning algorithm is used as an approximate solution to model learning. The proposed model was tested on the face recognition problem.

In [Wang et al., 2000], the second-order HMMRF model was used for offline handwritten Chinese character recognition. In this model, each character lattice

is regarded as a random field consisting of three HMMs, namely, the horizontal HMM, the vertical HMM, and 2-D inner HMMRF. The incomplete parameter estimation algorithm proposed by [Devijver and Dekesel, 1987] was used for model training. For state sequence decoding three algorithms were evaluated: maximum a posteriori (MAP), the Viterbi algorithm and iterated conditional mode (ICM). The ICM algorithm performs better than the other two methods and achieves a recognition rate of 88.23%.

In [Merialdo et al., 2000], an approximate Viterbi decoding algorithm was proposed. The used assumption and approximation lead to tractable computation, at price of a loss in full optimality. The proposed algorithm was tested on the segmentation and recognition tasks of the NIST database.

In [Feng et al., 2000], a bloc-based ICM was used as an approximation solution for the decoding problem of the Hidden NSHP Markov Chain Model (HNSHPMCM). In this model, the state transition matrix is modeled by an NSHP (see *Definition 2.2*) instead of MMRF. The model's accuracy on handwritten Chinese characters was 72.36%.

Furthermore, The NSHP Markov model with 4 neighborhoods was used for modeling the state transition matrix of dimension $H^4 \times H$ with H the number of model states [Baggenstoss, 2011]. In this work, the author proposed an approximate version of the Baum-Welch algorithm for the model parameter estimation and the joint probability density function. This model was used in [Madhogaria et al., 2015] for cars detection in aerial images.

A novel two-dimensional distributed hidden Markov model (2D-DHMM) was introduced in [Ma et al., 2008].The main objective of the authors is to define a general non-causal HMM that allow state dependencies from neighbors in all directions and all its neighbors. To overcome the effect that there is no analytic solution to the non-causal problem, the authors proposed an efficient way to break the non-causality by distributing the non-causal model into multiple distributed causal HMMs. They extended the training and classification algorithms presented in [Li et al., 2000] to a general causal model. Applied to aerial image segmentation, an interesting idea was proposed. This idea consists in defining 16 basic mage block patterns that cover all possible mixture of "man-made" and "natural" regions. Consequently, the variability of states was enriched, which in return improved the accuracy of state

estimations and segmentation performance.

A Separable Lattice 2-D HMM (SL2D-HMM) for face recognition was introduced in [Kurata et al., 2006]. This model has the composite structure of multiple hidden state sequences that interact to model the observation on a two-dimensional lattice. In other words, the SL2D-HMM structure consists of two independent 1-D Markov chains. Recently, an enhancement of the SL2D-HMM model was proposed by [Tamamori et al., 2014] in order to capture dependencies between adjacent observations by imposing explicit relationships between static and dynamic features. The training algorithm for both models is computationally intractable, and to make this problem tractable, the authors used the single-path Viterbi approximation algorithm.

In [Perronnin et al., 2003], an interesting work was proposed by Perronnin et al.; it presents a novel algorithm for decoding 2-D HMM. Its basic idea is to approximate a 2-D HMM with a Turbo-HMM (THMM), which consists of horizontal and vertical 1-D HMMs that communicate and allow iterated decoding of rows and columns by a modified version of the forward-backward algorithm. The authors also proposed the Turbo State-Space Model (T-SSM), which is an extension of the THMM to continuous states. For more details, the reader is referred to [Perronnin, 2004].

Later, the THMM was used in [Shenoy et al., 2016] to capture the deformation between pairs of images for the registration of unimodal and multimodal biomedical images.

A theoretical study of 2-D HMM was presented in [Yujian, 2007]. The key idea of this study is to consider the sequences of states on columns or rows of a 2-D HMM as states of a 1-D HMM, which reduces the context dependency as a 1D dependency, that cannot cover a real 2-D dependency.

A two-dimensional discrete 3×4 order HMM was proposed by Wang et al. [Wang et al., 2016a]. In this model, the state transition probability depends on 3 states: immediate horizontal, vertical and diagonal states; while the observation symbol probability depends on 4 states: the current state as well as the immediate horizontal, vertical and diagonal states. All algorithms presented in this work to solve the three basic problems of HMM were derived using the restrictive assumption that the sequences of states on rows or columns of the model can be seen as states of a 1-D discrete 1×2 order HMM. As in [Yujian, 2007], this work introduces

the theoretical aspects of the proposed model without any experimental results. By adapting the restrictive assumption in [Wang et al., 2016a] to the continuous case, the same authors proposed a two-dimensional continuous 3×3 order HMM [Wang et al., 2016b]. In this model, the probability density of the observation relies on 3 states: the current state, the immediate horizontal and vertical states.

In [Sargin et al., 2008], the authors presented a conditional iterative decoding (CID) algorithm for the approximate decoding of 2-D HMMs. In this algorithm, the posteriors from the previous row and column are used to calculate the next row and column. Tested on synthetic data, CID algorithm outperformed the THMM [Perronnin, 2004] and the decoding algorithm presented by Li et al. [Li et al., 2000].

Baumgartner et al. [Baumgartner et al., 2016] introduced an approximation of the second-order 2-D HMM for image segmentation. For parameter estimation problem, instead of summing over all possible state maps, the authors suggest estimating model parameters (transition probabilities, mean, and standard deviation) by using only the state map of the current iteration. Consequently, there is no theoretical guarantee that their 2-D HMM converges. For decoding problem, a new algorithm, called complete enumeration iteration, was proposed. This algorithm is based on a restrictive assumption that two diagonal pixels are independent.

In [Bobulski, 2017], the author tried to introduce an ergodic 2-D HMM for 2D and 3D face images recognition. The computational problem of parameter estimation and model decoding was neither mentioned or treated in this paper.

Recently, [Wan et al., 2018] introduced a saliency detection system based on 2-D HMM. The model exploits the hidden semantic information of an image to detect its salient regions. The authors defined an observed discrete variable map and a hidden state map and they proposed a 2D-Viterbi algorithm to infer the hidden state map. Their algorithm takes into account three directions for a given discrete variable: left, top-left, and top. This algorithm suffers from an ambiguity problem in which the state is determined by different regions. To solve this problem, the authors proposed to tack into account only one state with the maximum probability.

From the above studies, we can conclude that efficient algorithms to solve the three problems of HMM do not exist for proposed 2-D models of this category «MRF for 2-D modeling at the A-level». In fact, all proposed approaches were developed based on some restrictive assumptions regarding the models (such as approximating

a 2-D HMM with many 1-D HMMs) and using sub-optimal training and decoding algorithms (e.g. the Decision-Directed algorithm [Devijver and Dekesel, 1987], and the one-row-one-column look-ahead technique [Park and Lee, 1998], etc.). *In practice, the used approximations and assumptions reduce the modeling accuracy and limit the exploitation of the full two-dimensional structure of spatial data.*

2.2.2.2.2 NSHP for 2-D modeling at the B-level

As has already been mentioned, the second approach of a 2-D HMM based on MRF consists of using NSHP Markov random fields at the observation probability level. This model, called NSHP-HMM was, first proposed for handwritten word recognition [Saon, 1999] [Saon and Belaïd, 1997]. To the best of our knowledge, the NSHP-HMM is the only 2-D model in which the 2-D property is modeled at the B-level.

In this section, we will first define the NSHP Markov chain. We will then introduce the NSHP-HMM model and discuss its strengths and weaknesses. We will also review the published NSHP-HMM based research works. Finally, we will describe our proposed model, which presents an effective solution for the weaknesses of the reference model.

Definition 2.2 Let $X = \{X_{ij}\}_{(i,j)\in L}$ be a random field defined over an $m \times n$ integer lattice L. Let $\psi_{ij} = \{(k,l) \in L \,|l < j \text{ or } (l = j,\ k < i\}$ be the non-symmetric half-plane and $\theta_{ij} \subset \psi_{ij}$ the support of the 'pixel' $(i,j) \in L$ (also called the neighborhood set of (i,j)) (see Fig 2.2.b). X is called a non-symmetric half-plane Markov chain if for all $(i,j) \in L$

$$P\left(X_{ij}|X_{\psi_{ij}}\right) = P\left(X_{ij}|X_{\theta_{ij}}\right) \tag{2.5}$$

It is clear from *Definition 2.1*, *Definition 2.2* and Fig. 2.2 that the only difference between the MMRF and the NSHP Markov models is the choice of the structure of their past and local state regions.

According to *Definition 2.2*, we can calculate the joint field mass probability by using the chain decomposition rule of conditional probabilities as follows:

$$P\left(X\right) = \prod_{j=1}^{n}\prod_{i=1}^{m} P\left(X_{ij}|X_{\psi_{ij}}\right) = \prod_{j=1}^{n}\prod_{i=1}^{m} P\left(X_{ij}|X_{\theta_{ij}}\right) \tag{2.6}$$

Following the above equation, $P(X)$ can be computed row by row or column by column in a sequential manner. The original idea proposed by Saon [Saon, 1999] consists of associating a stochastic state process (a first-order Markov chain) to the field columns in order to tie the conditional probabilities relative to a column to a specific state. In other words, the observation probabilities in the NSHP-HMM states are estimated by an MRF. This probability is performed as the product of elementary probabilities performed on each pixel in the observed column (see Eq. 2.8). The elementary probability is determined by the NSHP according to a 2-D neighborhood fixed in the half-plane previously analyzed. The proposed model operates in a holistic manner at the 'pixel' level.

As in our discussion of HMMRF-based approaches (see section 2.2.2.2.1), it is very important to check the possibility of extending the two algorithms, namely, the Viterbi and the Baum-Welch algorithms, to the NSHP-HMM model. To show that, let us calculate $P\left(X|\lambda^{NSHP-HMM}\right)$ the pattern likelihood given the model $\lambda^{NSHP-HMM}$.

Let $Q = \{q_1, q_2, ..., q_n\}$, with n the number of columns of X. Q is a stochastic state process associated to the columns of X, where the random variable q_j takes values in a finite set of states $S = \{s_1, s_2, ..., s_H\}$, with H the number of states of the HMM, here $H \leq n$.

Given the calculation of $P(X)$ in Eq. (2.6), $P\left(X|\lambda^{NSHP-HMM}\right)$ can be computed as follows:

$$\begin{aligned} P\left(X|\lambda\right) &= \sum_{Q} \left(X,\ Q|\lambda\right) = \sum_{Q} \left(X|Q,\lambda\right) P\left(Q|\lambda\right) \\ &= \sum_{Q} \prod_{j=1}^{n} \left(q_j|q_{j-1}\right) P\left(X_j|X_{j-1}...X_1, q_j\lambda\right) \\ &= \sum_{Q} \prod_{j=1}^{n} \left(q_j|q_{j-1}\right) \prod_{i=1}^{m} P\left(X_{ij}|X_{\psi_{ij}}, q_j, \lambda\right) \\ &= \sum_{Q} \prod_{j=1}^{n} \left(q_j|q_{j-1}\right) \prod_{i=1}^{m} P\left(X_{ij}|X_{\theta_{ij}}, q_j, \lambda\right) \end{aligned} \tag{2.7}$$

Equation (2.7) is quite similar to the case of 1-D HMM, the only difference being

the calculation of $b_k(O_t)$:

$$b_k\left(O_t\right) = \prod_{i=1}^{m} b_k\left(i,\ c,\ \theta_{it}\right) = \prod_{i=1}^{m} P\left(X_{it} | X_{\theta_{it}},\ q_k\right) \tag{2.8}$$

The elementary probability of the conditional pixel observation matrix $B = \left\{b_k\left(i,\ c,\ \theta_{it}\right), 1 \leq k \leq H\right\}$ is the probability of observing in state k a pixel (i,t) of color c at height i knowing the neighborhood set θ_{it}. We note here that $T = n$.

Clearly, equations (2.7) and (2.8) show a straightforward extension of the Viterbi and the Baum-Welch algorithms to the case of the NSHP-HMM model, which makes this model one of the most optimal 2-D HMM recognizers. Detailed equations of the Baum-Welch algorithm are presented in section 3.3.2.

In contrast to the proposed models of category «2-D property at the A-level» (section 2.2.2.2.1), where the complexity is exponential in the size of the data, the complexity here is exponential in the model order V and it is treated at $b_k(O_t)$ count (see Eq. (2.8)). V is the number of pixels in the neighborhood set θ.

In addition to the optimality of the NSHP-HMM, since it allows an efficient applicability of the two algorithms (the Viterbi and the Baum-Welch algorithms), two other strong points of the HSHP-HMM are as follows: i) the conditional pixel observation probabilities $b_k\left(i,\ c,\ \theta_{it}\right)$ depend simultaneously upon the line index i and the observation state k. (i,k) presents the 2-D location of the pixel in horizontal and vertical spaces respectively. ii) $b_k\left(i,\ c,\ \theta_{it}\right)$ also depend on the color c (black or white) of the 'pixel' x_{it} and its neighborhood 'pixel' configuration θ_{it} which provides useful 2-D contextual information.

The key NSHP-HMM parameters are image height m, HMM state number H and model order V.

On the other hand, the drawbacks of the NSHP-HMM are:

(a) Pixel-based analysis and the number of properties which grows exponentially with the neighborhood size (V) lead practically to considering very short 2-D contexts measured in terms of a few pixels. In all NSHP-HMM-based systems [Saon, 1999] [Choisy and Belaïd, 2002] [Cecotti et al., 2005] [Vajda and Belaïd, 2005], a maximum value of V equaling 4 neighborhood 'pixels' was used; the reason is to maintain a suitable compromise between accuracy and model complexity.

(b) Applying on each column conduces to the use of a significant number of states. In practice, the number of states is equal to half the average image length. This point increases the model's complexity.

(c) From Eq. (2.8), the number of lines m must be the same for all images. Consequently, the NSHP-HMM requires a height normalization procedure prior to training and recognition. The normalization itself is not a problem, but in practice m must be small to keep the complexity fairly low ($m = 20$ in [Saon, 1999][Choisy and Belaïd, 2002] for handwritten word images). In fact, normalizing an image in a small number of lines introduces considerable deformation and loss of useful details on pattern shape.

(d) The NSHP-HMM deals with binarized images in order to reduce the model's complexity. c in Eq. (2.8) takes two values: black or white; extending c to G gray level values increases the number of parameters in the B matrix by a factor of $\frac{G^{V+1}}{2^V}$. For example, for an NSHP-HMM that deals with 4 gray-level values ($G = 4$) and $V = 1$ neighborhood pixel, we have an increase of a factor 8 for the number of parameters compared with an NSHP-HMM that deals with binarized images, which is very large compared to a modest modeling enhancement (an image with 4 gray levels is not as informative as a binary image). This point has a negative effect since it limits the model's applicability, i.e. the NSHP-HMM cannot be applied to color-image recognition, such as the face recognition problem.

In the rest of this section, we evaluate the proposed NSHP-HMM based systems in terms of their ability to overcome the above four drawbacks.[3]

In Choisy's work [Choisy and Belaïd, 2002], an analytical recognition approach without segmentation was proposed in which each letter is modeled by an NSHP-HMM. This latter is the same model proposed by Saon except the addition of two specific states, D and F, which model the probability of beginning and ending in each state, respectively. Obviously, the analytical approach has its known advantages to the global approach. Furthermore, according to our evaluation criteria, the analytical approach gives a solution to the second NSHP-HMM's drawback (b')

[3]The corresponding solution for each drawback is noted by ('), i.e. (a') is the proposed solution of drawback (a).

: in the global approach of Saon, the number of states of each class is based on the mean length of image word; however, in the analytical approach of Choisy, this number is reduced to the mean length of the image letter. In practice, for French bank check word recognition, the total number of states in the global system was 615 states [Saon, 1999] compared to 87 states in the analytical approach [Choisy and Belaïd, 2002] (reduction of a factor of 7).

The NSHP-HMM can be also used for image normalization [Choisy et al., 2003]. For this purpose, the Viterbi algorithm is applied to find the best repartition of image columns. In [Choisy, 2007], The NSHP-HMM was used for dynamic handwritten keyword spotting. For this task, the Viterbi calculation was applied rather than the Baum-Welch algorithm to reduce the calculation time, and coupled to a fix-point representation of probabilities to deal with integer values rather than real values.

In [Vajda and Belaïd, 2005], an improvement of the NSHP-HMM model is proposed by inserting high-level information coming from the structural feature of the pixel. The used features were the ascenders and descenders: a pixel was considered to be a structural pixel if it belongs to an ascender or a descender. The obtained results of this model on SRTP and Bangla city name databases were 87.52% and 86.80%, respectively. We can consider this work an attempt to overcome the fourth drawback (d') of the reference model given that it combines the color of the pixel (black/white) with its structural nature. However, like the baseline approach, the proposed model operates at the pixel level on the normalized binarized image; it suffers, consequently, from the same four weaknesses as the baseline model.

In [Boudaren and Belaïd, 2009], the authors try to adapt the NSHP-HMM to aerial image mapping. For this reason, they first introduce the NSHP-2D-HMM. The differences between this model and the NSHP-HMM are i) the 2-D property is modeled at the A-level and the B-level. ii) The proposed model operates on gray level images. Obviously, image mapping according to MAP probability is an NP-hard problem. To solve this problem, the authors approximate their model by DT-NSHP-HMM, which extends the principle of the Dependency Tree Hidden Markov Model (DT-HMM). The principle is that each state depends on only one neighboring state at a time. The 2-D state transition matrix of the NSHP-2D-HMM is consequently replaced in the DT-NSHP-HMM by 1-D horizontal and 1-D vertical state transitions that will be estimated separately. The DT-HMM Viterbi algorithm

is then used in the labeling procedure. The NSHP-2D-HMM and its approximator DT-NSHP-HMM were presented in a conceptual framework without experimental study. In addition to the approximation of the A matrix by two 1-D horizontal and vertical matrices, the remained question for us is how the authors' model can deal with gray level images and with four neighborhoods pixels. In this case, the number of parameters in the B matrix is $H \times m \times 256 \times 256^4$, with H and m the number of model states and image height, respectively. How can the authors' system deal with this large number of parameters? We do not know since no experimental study was given.

From the above study, it is clear that none of the proposed NSHP-HMM based approaches overcomes the four drawbacks of the reference model.

Now, we are ready to explain our contribution, which seems to be quite simple but keeps the efficient 2-D modeling of the NSHP-HMM without suffering from its four limitations.

In our opinion, the NSHP-HMM is one of the most optimal 2-D recognizers since it perfectly brings the efficient algorithms of the 1-D HMM to the 2-D modeling of spatial data. However, the main drawback of this model is the short 2-D context. So far, our main motivation is to extend the two-dimensional context in the NSHP-HMM to get better modeling of the spatial property of an image. For this reason, we replace the *pixel-level* in the reference model by the *zone-level*; the 2-D property is still modeled at the B matrix according to the MRF constraints. Hence, this version of the NSHP-HMM model is referred to as an NSHP-HMM based on *ZONE* observation probabilities (NSHP$^{\mathrm{Z}}$-HMM)[Boukerma et al., 2014][Boukerma et al., 2015][Boukerma et al., 20
First, the pattern image is divided into $M \times N$ sub-images, called zones. Each zone is then encoded into discrete symbols resulting from a vector quantization (VQ) technique applied to the local features of this zone. The emission probability of each zone is computed using state-related NSHP-like conditional zone distributions. The product of the M zones vertically superposed presents the observation probabilities of the NSHP$^{\mathrm{Z}}$-HMM model (see Eq. 3.1).

Even though our principal goal was to overcome the short context limitation of the reference model (a'), our proposal seems to be a promising solution for the three remaining drawbacks of the reference model. Therefore, let us evaluate the proposed solution on this prospect:

(b') The proposed model operates on a set of vertically stacked zones. Consequently, the reduction in the number of states compared to the reference model for both global [Saon, 1999] and analytical [Choisy and Belaïd, 2002] approaches is a factor W. W is zone width.

(c') Height normalization is not mandatory for our model; the only constraint is the vertical division of the image into the same number of horizontal zones (M in Eq. 3.1 is constant). The zones can have a different height, which can be advantageous for some applications like handwriting recognition where the word image can be divided into 3 or 5 bands with different heights since the middle zone is the most important part of the word (an illustrating example is in Sec. 4.5.2).

(d') After dividing the image into zones, features are extracted from each zone followed by vector quantization. Thus, the input of our model is a matrix of symbols. As a result, the proposed model can operate on binary, gray level or color images. It can even be applied to 3-D data such as MRI images. In our opinion, this point is the most important advantage of our contribution since the applicability is more general compared to the reference model.

In addition, modeling the contextual information at the zone level allows extracting useful and distinctive information on the local characteristics of patterns. In contrast to the reference model, which mandatorily operates on binary images using pixel color (black or white) as a feature, our model enables the use of a large choice of methods for relevant feature extraction. The only remaining challenges are the choice of (1) the zoning method adapted to the application domain, (2) an efficient feature extraction method, and (3) a vector quantization algorithm and the number of symbols.

In the next chapter (Sec. 3.5.2), we will discuss the potential dependency between these three points and give some practical rules to guide the development of an efficient recognition system based on the NSHP$^{\mathrm{Z}}$-HMM.

2.3 Conclution

The overview and the taxonomy of 2-D based recognition approaches present an important contribution of this dissertation. We hope that the contents of this chapter will be helpful to researchers interested in this field. By figure 2.3, we try to graphically summarize this contents.

From the study presented in the section 2.2.2.2.2 (NSHP for 2-D modeling at the B-level), it is clear that none of the proposed NSHP-HMM based approaches overcomes the four drawbacks of the reference model. The NSHP^{Z}-HMM aims to correct these limitations by replacing the *pixel-level* in the reference model by the *zone-level*, which explains the letter Z in the name NSHP^{Z}-HMM. In the next chapter, we will define in more detail the proposed model and derive their re-estimation formulas.

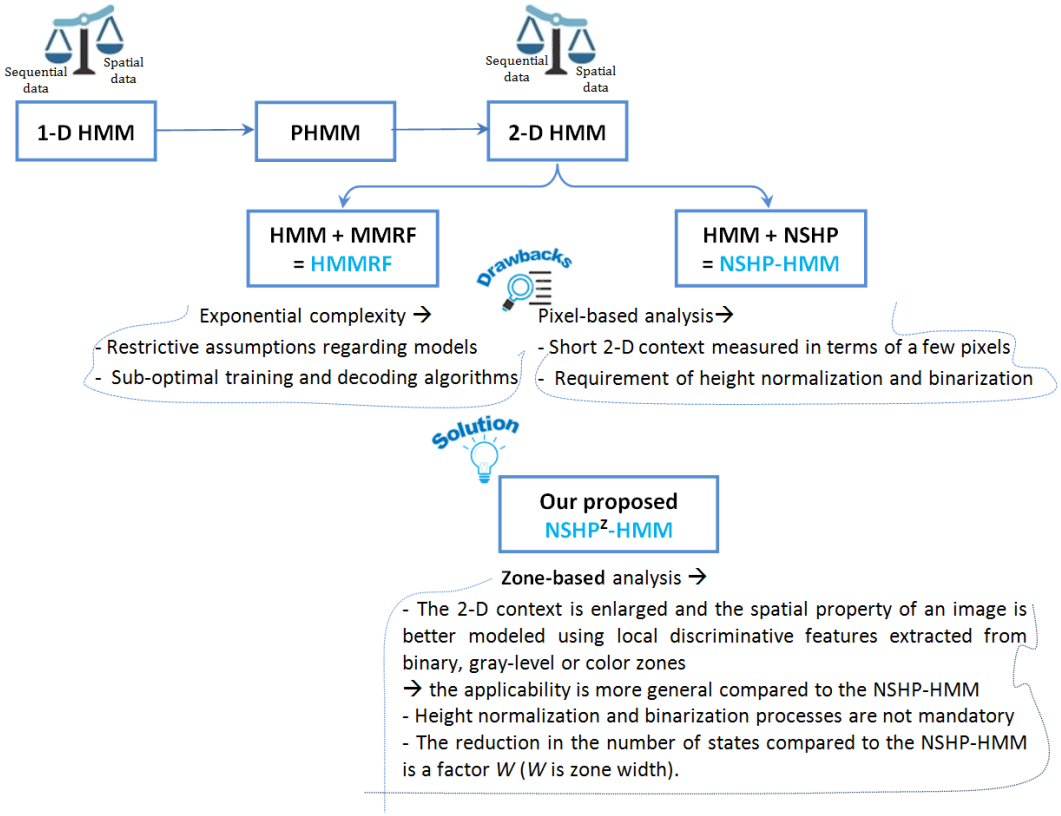


Figure 2.3: The NSHP^{Z}-HMM vs the related HMM-based 2-D models

Chapter 3

NSHP-HMM based on conditional zone observation probabilities (the NSHP^{Z}-HMM)

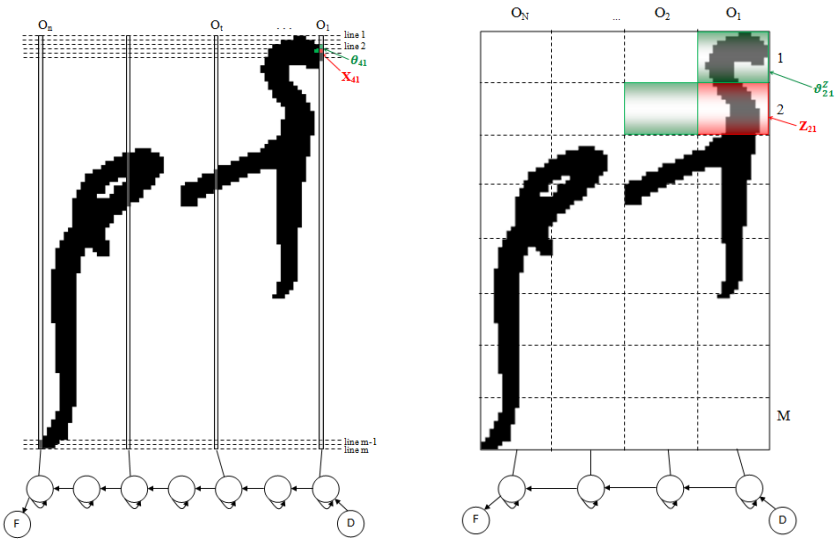
3.1 Introduction

In this chapter, we first define the proposed model and derive their re-estimation formulas. Then, we introduce the principal steps involved in the image recognition system based on the NSHP^{Z}-HMM. In order to evaluate our proposed model in ascending order of difficulty, the first series of experiments, presented in this chapter, are performed on handwritten digits recognition. In this chapter, we also give some practical rules to guide the development of an efficient recognition system based on the NSHP^{Z}-HMM.

3.2 The NSHP^{Z}-HMM: formal description

The NSHP^{Z} Markov Model is generalization of the NSHP-HMM, because the zone-level can be reduced to a pixel in order to obtain a classical NSHP-HMM. For the NSHP^{Z}, θ and ψ, in *Definition 2.2*, are replaced by θ^{Z} and ψ^{Z}, which, respectively, present the past and the local state at the zone Z_{ij} (see Fig. 3.1).

Let X be a pattern image and a zoning $Z = \{z_{11},\ z_{12}\ ,\ ..,z_{1N},\ ...,z_{M1},\ z_{M2},\ ..z_{MN}\}$ of X is a partition of X into $M \times N$ sub-images, called zones (see Sec. 3.4.2). Sup-



(a) (b)

Figure 3.1: Handwritten word recognition based on (a) the NSHP-HMM, (b) the $NSHP^Z - HMM$

posing that the interaction between zones is described by an NSHP^Z Markov chain. On the vertical axis, we define the observation probabilities of the NSHP^Z-HMM as:

$$\begin{aligned} b_k^Z\left(O_t\right) &= \prod_{i=1}^{M} B_k^Z\left(i,\ u^{Z_{it}},\ \theta_{ij}^Z\right) \\ &= \prod_{i=1}^{M} P\left(Z_{it} = u^{Z_{it}} \mid,\ Z_{\theta_{it}^Z} = \theta_{it}^Z,\ q = s_k\right) \end{aligned} \tag{3.1}$$

Equation (3.1) replaces Eq. (2.8) for the case of the NSHP^Z-HMM, where $b_k^Z\left(O_t\right)$ represents the probability of observing inside the analyzed image M vertically stacked zones $Z_t = \{z_{1t},\ z_{2t},\ ..., z_{Mt}\}$ at state k.

The complete parameter set of the NSHP^Z-HMM are $\lambda = \left(A,\ B^Z,\ \theta^Z,\ V\right)$, where:

- $\theta^Z = \left\{\theta_{ij}^Z\right\}_{(i,j)\in L}$ the set of V neighborhoods of zone Z_{ij} fixed in the half plane. We note here that the number of neighborhood zones V (also called model order) and their shapes can vary from one zone to another (see Fig. 3.2).

– $U = \{u_1,\ ...,\ u_P\}$ the set of P discrete symbols, i.e. the NSHP realization at zone level. We note by $u^{Z_{it}}$ the corresponding symbol u of the zone Z_{it}.

– $S = \{s_1,\ ...,\ s_H,\ D,\ F\}$ the set of H normal states with two specific states D and F that model the probability of beginning and ending in each normal state respectively.

– $A = \{a_{kh} \cup \{a_{Dk},\ a_{kF}\}\}_{1 \leq k,\, h \leq H}$, the state transition probability matrix, where $a_{kh} = P\left(q_{t+1} = s_h \mid q_t = s_k\right)$, $a_{Dk} = P\left(q_1 = s_k \mid D\right)$, $a_{kF} = P\left(F \mid q_T = s_k\right)$. T denotes the number of observations (here, $T = N$ the number of zones according to the vertical partitioning of the image).

– $B^Z = \left\{b_k^Z\left(i,\ u^{Z_{it}},\ \theta_{ij}^Z\right)\right\}$, where $s_k \in S,\ s_k \neq D,\ F$. B^Z is the probability of observing in state k a discrete symbol u corresponding to the zone Z_{ij} given the neighborhood set of zone θ_{ij}^Z. We note by θ_{ij}^Z the set of symbols $\{u_1,\ ...,\ u_V\}$ corresponding to the given configuration of V neighboring zones.

The key NSHP$^{\mathrm{Z}}$-HMM parameters are the number of HMM states H, model order V, codebook size P, number of horizontal zones M, zone width W and vertical overlap rate between zones Rz.

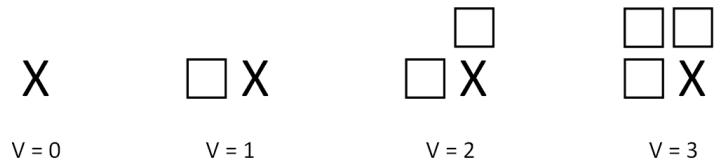


Figure 3.2: Different model orders (V) and neighborhood configurations

3.3 Model parameter estimation

3.3.1 The modified forward-backward algorithm

The forward-backward procedure defined in [Rabiner, 1989] can be easily adapted for the NSHP$^{\mathrm{Z}}$-HMM. The modification consists in the count of $b^Z(O_t)$ according to Eq. 3.1.

3.3.1.0.1 The forward variables

$$\begin{gathered}\alpha_1(k) = a_{Dk} \cdot b_k^Z(O_1) = a_{Dk} \cdot \prod_{i=1}^{M} B_k^Z\left(i,\ u^{Z_{i1}},\ \theta_{i1}^Z\right), \\ 1 \le k \le H \\ \alpha_t(k) = \left[\sum_{h=1}^{H} \alpha_{t-1}(h) \cdot a_{hk}\right] \cdot \prod_{i=1}^{M} B_k^Z\left(i,\ u^{Z_{it}},\ \theta_{it}^Z\right), \\ 1 \le k \le H;\ 2 \le t \le N\end{gathered} \tag{3.2}$$

$$P(X|\ \lambda) = \sum_{k=1}^{H} \alpha_N(k) \tag{3.3}$$

3.3.1.0.2 The backward variables

$$\begin{gathered}\beta_N(k) = a_{kF},\ 1 \le k \le H \\ \beta_t(k) = \sum_{h=1}^{H} a_{kh} \cdot \beta_{t+1}(h) \cdot \prod_{i=1}^{M} B_k^Z\left(i,\ u^{Z_{i,t+1}},\ \theta_{i,t+1}^Z\right), \\ 1 \le k \le H;\ t = N-1,\ ..,\ 1\end{gathered} \tag{3.4}$$

3.3.2 Parameter estimation

Training of the NSHP^{Z}-HMM is based on the Baum-Welch algorithm. The state transition matrix (A) is re-estimated similarly as in a classical NSHP-HMM [Choisy and Belaïd, 2002] with little modification. The re-estimation formula of the conditional zone observation probabilities (B) is performed by applying the maximum likelihood (ML) count of the number of times that a given zone configuration is encountered in R images of the training set. P_r is the emission probability of the sample X^r calculated using

Eq. (3.3).

$$\overline{a}_{Dk} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{P_r}\alpha_1^r(k)\,\beta_1^r(k)$$

$$\overline{a}_{kF} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\alpha_{N_r}^r(k)\,a_{kF}}{\sum_{r=1}^{R}\frac{1}{P_r}\left[\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,\beta_t^r(k)+\alpha_{N_r}^r(k)\,a_{kF}\right]} \tag{3.5}$$

$$\overline{a}_{kh} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,a_{kh}\beta_{t+1}^r(h)\prod_{i=1}^{M}B_h^Z\left(i,u^{Z_{i,t+1}},\theta_{i,t+1}^Z\right)}{\sum_{r=1}^{R}\frac{1}{P_r}\left[\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,\beta_t^r(k)+\alpha_{N_r}^r(k)\,a_{kF}\right]}$$

$$\overline{B}_k^Z\left(i,u^Z,\theta^Z\right) = \begin{cases} \dfrac{\displaystyle\sum_{r=1}^{R}\frac{1}{P_r}\sum_{\substack{j=1\\U_{ij}^Z=u^Z\\\theta_{ij}^Z=\theta^Z}}^{N_r}\alpha_j^r(k)\,\beta_j^r(k)}{\displaystyle\sum_{r=1}^{R}\frac{1}{P_r}\sum_{\substack{j=1\\\theta_{ij}^Z=\theta^Z}}^{N_r}\alpha_j^r(k)\,\beta_j^r(k)} & ,denominator\neq 0 \\ b_k^Z\left(i,u^Z,\theta^Z\right) & ,otherwise \end{cases} \tag{3.6}$$

3.3.3 NSHP$^{\mathrm{Z}}$-HMM model complexity

In this section, we compare the complexity of the proposed NSHP$^{\mathrm{Z}}$-HMM with that of the reference model. First, let us denote by I the number of re-estimation iterations. Let $\overline{n}$ and $\overline{N}$ be the average image length in columns for the NSHP-HMM and in number of vertical zones (N) for the NSHP$^{\mathrm{Z}}$-HMM. During training, B re-estimation, in case of NSHP-HMM, requires $\mathcal{O}(RIHm\overline{n}2^V)$ operations of addition and $\mathcal{O}(RIHm\overline{n})$ multiplications. For the NSHP$^{\mathrm{Z}}$-HMM, B^Z re-estimation (Eq. (3.6)) requires $\mathcal{O}(RIHM\overline{N}P^V)$ additions and $\mathcal{O}(RIHM\overline{N})$ multiplications. During recognition, the count of the pattern likelihood using Eq. (3.3) requires $\mathcal{O}(H\overline{n})$ additions and $\mathcal{O}(H\overline{n}m)$ multiplications for the NSHP-HMM; and it requires $\mathcal{O}(H\overline{N})$ additions and $\mathcal{O}(H\overline{N}M)$ multiplication for the NSHP$^{\mathrm{Z}}$-HMM.

As already mentioned (see pp. 39), the reduction in H (number of states), for the NSHP$^{\mathrm{Z}}$-HMM, is a factor W compared to the NSHP-HMM. Furthermore,

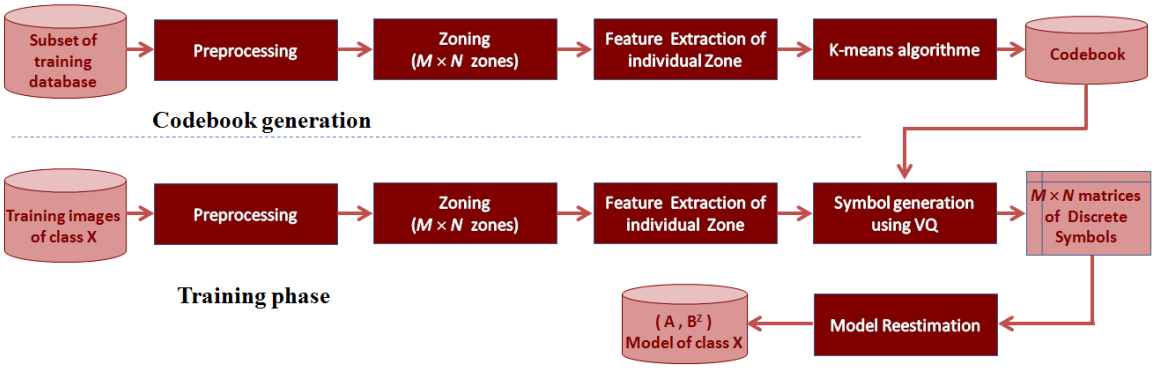


Figure 3.3: Image recognition system using the NSHP^{Z}-HMM

M (the number of horizontal zones) is sgnificantly smaller than m (image lines). Consequently, the NSHP^{Z}-HMM has a fewer cost in term of complexity during training and recognition, even if theoretically speaking its complexity estimation is in the same order of the classical NSHP-HMM (reducing zones to pixels will lead at the same cost).

3.4 2-D recognition based on the NSHP^{Z}-HMM: a general framework

In this section, we present an outline of the recognition system based on the NSHP^{Z}-HMM using the zoning principle. First, we start by presenting the main steps in case of any 2-D spatial data. Then, in chapter 4, we present our vision of how to adapt the NSHP^{Z}-HMM to properly model the images of Arabic script.

Our 2-D recognition system based on the NSHP^{Z}-HMM includes five steps : preprocessing, zoning, feature extraction, vector quantization and training/recognition (see Fig. 3.3).

3.4.1 Preprocessing

The performance of any text recognizer depends on the quality of the input text and the lack of noise, even more so than with human analysts. Document images of poor quality and high intra-class variation pose greater difficulty for recognizers. In [Likforman-Sulem et al., 2009] and [El Abed and Margner, 2007], the authors ex-

perimentally analyze the effects of distortion, noise and preprocessing stages on the recognition rate of their systems. In this regard, the input text image could be preprocessed to simplify recognition by removing all distortion and uninteresting variations in writing styles. Preprocessing operations are usually specialized image processing operations that transform the image into another one with reduced noise and variation. Those operations include binarization, noise removal, smoothing, thinning, contour analysis, baseline estimation, and text normalization such as skew correction and character normalization. The application of all these operations is not imperative in each recognition system.

In case of the reference model [Saon, 1999], image binarization and height normalization are two mandatory pre-processing steps. The number of image lines m is one of the NSHP-HMM parameters; consequently, m must be constant. Furthermore, the pixel's color (black/white) presents the model vocabulary. The potential disadvantage of these two processes on system performance and application generality is studied in section 3.5. Theoretically, the normalization itself should not be a problem, but in practice m must be small to keep the complexity fairly low. For example, for handwritten word images, Ref. [Saon, 1999] and [Choisy and Belaïd, 2002] used $m = 20$. Consequently, normalizing an image in a small number of lines introduces considerable deformation and loss of useful details on pattern shape, which are mostly accented considering the binarization level, even for the differential normalization proposed by [Choisy and Belaïd, 2002].

The NSHP^{Z}-HMM aims to correct these limitations by replacing the *pixel-level* in the reference model by the *zone-level*, which explains the letter Z in the name NSHP^{Z}-HMM. As a result, image binarization and height normalization are not mandatory for our model; the only constraint is the division of the image into the same number of horizontal zones (M in Eq. (3.2), (3.4) and (3.6) is constant).

3.4.2 Zoning design

A possible way to extract local distinctive characteristics from patterns is the use of a zoning technique adapted to an application domain. In the literature, a large number of zoning methods have been proposed; for a good survey see [Impedovo and Pirlo, 2014].

The parameters of our zoning method are the number of horizontal zones M, the zone width W and the vertical overlap rate between zones Rz. The N parameter is

computed as follows: $N = round(\frac{n}{W})$, where n is the number of image columns and Rz is supposed to be zero.

Zoning is a crucial stage in our approach for two principal reasons:

- First, zoning must be done in such a way as to increase the 2-D correlation between the samples of the same class. Let us explain this point in more detail: zoning can be seen as a segmentation, and correct segmentation helps to correctly classify patterns. Similarly, adequate division in zones helps our NSHP^{Z}-HMM to efficiently model the spatial data. In fact, the content of a given zone Z_{ij} (represented by the symbol u_{ij}^{Z}) and its horizontal position i present a part of the NSHP^{Z}-HMM parameters.

- Secondly, zoning has an important impact on the subsequent steps. As an example, according to W and Rz parameter values, we set the number of states H and we chose the number of clusters P and the kinds of features to use.

3.4.3 Feature extraction and vector quantization

After dividing the image in zones, a feature vector is calculated from each zone; it can be composed of low-level and high-level features. Then, the extracted feature vectors are encoded through cluster prototypes using a vector quantization algorithm.

K-means is one of the most widely used algorithms for data clustering (see Fig 3.4). A good reference to this field is presented in [Jain, 2010]. The K-means algorithm requires three user-specified parameters : number of clusters P, cluster initialization and distance metric [Jain, 2010].

Increasing P always decreases the squared error of the vector quantization phase. However, for the NSHP^{Z}-HMM, the number of parameters increases linearly with P leading, in the case of a sufficient number of training samples, to improve the modeling power of the model (a sufficient number of samples is needed to train the model, which grows consequently with P). It, therefore, seems logical to choose P value as a trade-off between the complexity of the NSHP^{Z}-HMM and the number of samples available to train the model. For the distance metric, Euclidean distance is used, assuming the normal distribution of features extracted from the zones. For

cluster initialization, we apply the K-means algorithm for different initial partitions and we choose the result with the smallest squared error.

The obtained codebook is then used to generate a matrix of discrete observations from zones. This matrix presents the input of the NSHP^Z-HMM model. The codebook size P is determined empirically on the validation dataset.

Algorithm. K-means algorithm
Input: a sample set $\omega = \{x_1, x_2, \ldots x_S\}$ of example vectors and the desired codebook size P.

1. Initialization
Choose the first P vectors of the sample set as initial codebook Y^0
$Y^0 = \{x_1, x_2, \ldots x_P\}$
initialize iteration count $m \leftarrow 0$

2. Iteration
for all vectors not yet processed x_t, $P < t \leq S$

a) Classification
for x_t determine the optimal reproduction vector y_i^m in the current codebook Y^m
$y_i^m = argmin\, d(x_t, y), y \in Y^m$

b) Partition Update
determine the new partition by updating the cell of the codebook vector selected

$$R_j^{m+1} = \begin{cases} R_j^m \cup \{x_t\}\ if\ j = i \\ R_j^m\ otherwise \end{cases}$$

c) Codebook Update
determine a new codebook by updating the prototype of the cell modified in the previous step

$$y_j^{m+1} = \begin{cases} cent(R_j^{m+1})\ if\ j = i \\ y_j^m\ otherwise \end{cases}$$

Output: codebook

Figure 3.4: Vector quantization using K-means clustering algorithm.

3.4.4 Training

Training of the NSHP^Z-HMM consists in estimating its parameters that optimally fit a set of training data. The training algorithm involves the following steps:

1 Initialize parameters A and B^Z.

2 Extract the bounding box of the pattern image and divide it into $M \times N$ zones.

3 From each zone, a feature vector is calculated; it can be composed of low-level and high-level features.

4 Encode the feature vector of each zone into a discrete symbol using a vector quantization algorithm.

5 Repeat steps 2-4 for each image from R images in the training set.

6 For each image, use the generated $M \times N$ matrix of discrete symbols as an input to calculate $b_k^Z(O_t), 1 \leq k \leq H, 1 \leq t \leq N$ by Eq. (3.1).

7 Estimate the NSHP^{Z}-HMM parameters using Eq. (3.6) for the conditional zone observation probabilities and Eq. (3.5) for the state transition matrix.

8 Repeat steps 6-7 I times. I denotes the number of training iterations.

3.4.5 Recognition

1 Apply steps 2-4 and 6 on the test image to generate the matrix of discrete symbols.

2 Use Eq. (3.3) to compute the pattern likelihood for all models and assign the test image according to the model of maximum likelihood.

3.5 Experimental results on handwritten digit recognition

In order to evaluate our proposed model in ascending order of difficulty, the first series of experiments are performed on handwritten digits recognition. Here, the MNIST database [LeCun et al., 1998] is used; this database contains gray-level images of handwritten digits divided into a training set of 60,000 examples and a test set of 10,000 examples.

Experiments on cursive handwritten word recognition will be the subject of the next chapter.

3.5.1 Influence of image height normalization

As already mentioned, one of the reference model's drawbacks is that it mandatorily operates on normalized images. In this experiment, we study the influence of this step on the recognition rate of both models, namely, the reference model and our proposed model.

To gain a better understanding of whether the superiority of the NSHP^{Z}-HMM on the classical NSHP-HMM is

> due to i) extending 2-D context: *zone vs pixel*,
>
> or due to ii) the use of local discriminative features of the zone: *Features vs binary pixel*,
>
> or due to iii) avoiding the negative effect of the normalization and binarization steps: *normalized vs non-normalized images* and *binarized vs gray-level or color images*,

we first compare the NSHP^{Z}-HMM to the classical NSHP-HMM on normalized binarized images. In this experiment, three models are compared: NSHP-HMM, NSHP^{Z}-HMM$_1$ and NSHP^{Z}-HMM$_2$. The first two models were trained on normalized images of 14 lines and the third was trained on original images. The state number of three models is proportional to the average image length in columns for NSHP-HMM and in a number of vertical zones (N) for NSHP^{Z}-HMM$_1$ and NSHP^{Z}-HMM$_2$. As for the discrete HMM, the quantization of a continuous signal into discrete symbols can introduce degradation and loss of useful information for classification. In order to examine this effect on the proposed approach, NSHP^{Z}-HMM$_1$ and NSHP^{Z}-HMM$_2$ worked without codebook and used five symbols extracted directly from the 2×2 binary zone. These discrete symbols represent the number of foreground pixels within the zone (0, 1, 2, 3 or 4 black pixels), which is in some sense equivalent to the NSHP-HMM ability.

Developing our approach under poor circumstances by working only on extending the 2-D context thanks to the zone-level and using simple techniques for zoning design and feature extraction provides us with important knowledge on the capability of our approach.

Fig. 3.5 plots the top 1 to 10 recognition rates of the three models for different values of model order V = (0, 1, 2, or 3). V equals zero means modeling the site[1](i, j) independently of its neighborhood. In general, the NSHP^{Z}-HMM$_1$ achieves results similar to the NSHP-HMM : the NSHP^{Z}-HMM$_1$ outperforms the NSHP-HMM for $V = 1$ and $V = 2$. On the other hand, the NSHP-HMM gives better results than the NSHP^{Z}-HMM$_1$ for $V = 0$ and $V = 3$. We note here that the results obtained for the NSHP-HMM as re-implemented by ourselves (93.62% for $V = 3$, top 1) is quite close to the published result in [Cecotti et al., 2005] for this model (93.44% for $V = 3$, top1); see Table 3.1. Working on original images (without height normalization) significantly improves performance: the NSHP^{Z}-HMM$_2$ performs better than the NSHP^{Z}-HMM$_1$ and the NSHP-HMM, with an improvement of 2.47% and 1.79% respectively at $V = 3$, top1. *These results are relatively good when one remembers that simple methods for zoning design and feature extraction were used.* Further investigations of these two steps are the subject of the following experiments.

3.5.2 Effect of system parameters

From the above experiments and our previous tests presented in [Boukerma et al., 2014] and [Boukerma et al., 2015], we can bring three main remarks:

- Increasing model order (V) and working on non-normalized gray-level images significantly improve the recognition rate.
- Parameters of zoning design (W, M and Rz) present the most important parameters of our model because of their influence on the others as follows:
  - The number of states H is equal to half of the average observation sequence length (N). H consequently depends on W and Rz parameters.
  - The choice of used features and the number of symbols P depend on the zone size.
- In the case of a sufficient amount of training data, using a large number for V, P and H improves the modeling power and increases the system's performance.

[1]The site (i, j) corresponds to the pixel (i, j) for the NSHP-HMM model and to the zone Z_{ij} for the NSHP^{Z}-HMM.

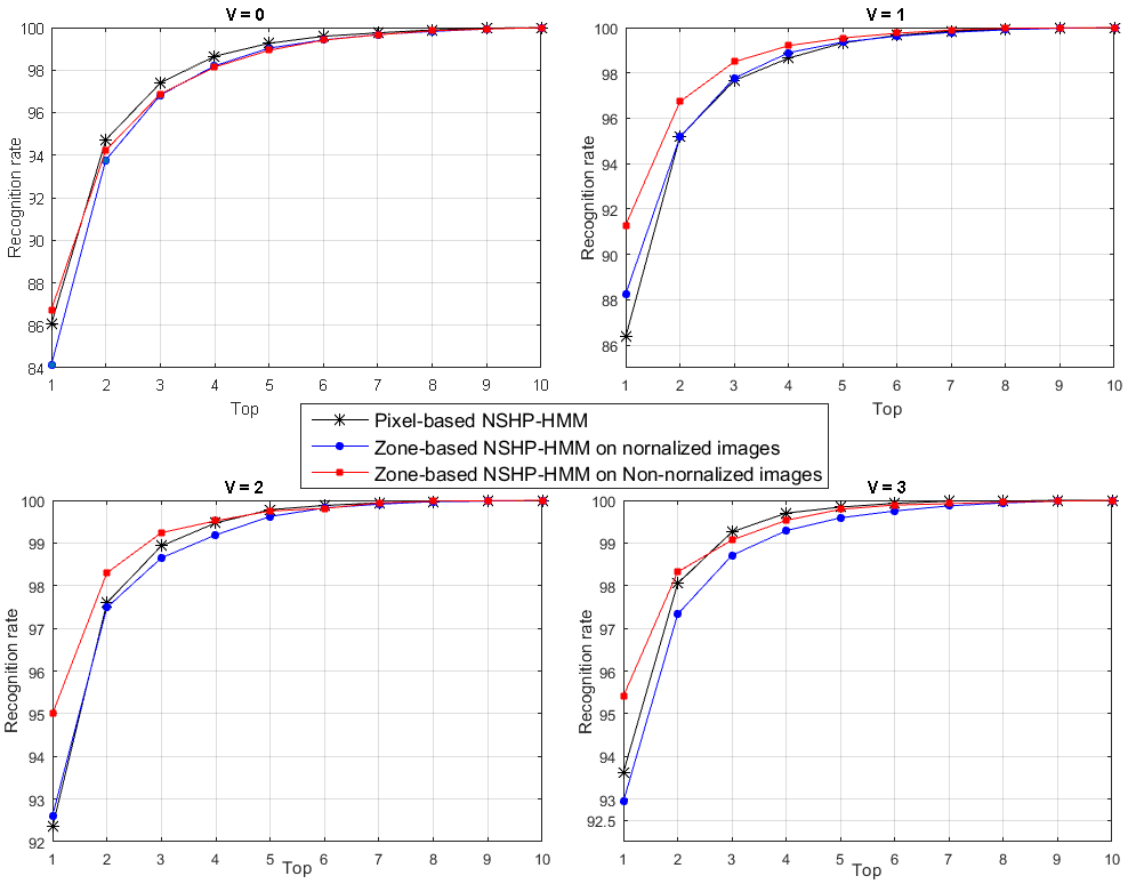


Figure 3.5: Influence of image normalization on recognition rate for different model orders (the x-axis represents the top 1 to 10 recognition rates).

Like other HMM-based models, the values of these parameters must be a trade-off between the complexity of the NSHP^{Z}-HMM and the number of samples available to train the model.

Considering the above remarks, several experiments were done using different feature extraction methods, such as pixel values of zone, pixel densities of zone and of each column of zone, background/foreground transitions between the columns of the zone and between the rows of the zone, density from the first line/column of image up to the zone, and the upper, lower, left and right profiles.

Our best results were obtained using $V = 2$, $M = 7$, $Rz = 0.6$, $P = 15$ and the pixel values of the zones as features. The used zoning technique was a simple uniform partition using regular grid, with $M = 7$ horizontal zones and zone width $W = 3$ pixels.

Using these optimal parameters, four NSHP^{Z}-HMMs were developed that correspond to four images: original image, mirrored horizontally, mirrored vertically, mirrored horizontally and vertically. These four models are referred to as NSHP^{Z}-HMM$_{(img)}$, NSHP^{Z}-HMM$_{(img^{lr})}$, NSHP^{Z}-HMM$_{(img^{ud})}$ and NSHP^{Z}-HMM$_{(img^{lr\text{-}ud})}$, respectively , and will be combined in the next section.

Results are given in Table 3.1 for individual models compared to the four NSHP-HMMs in [Cecotti et al., 2005] also developed on four mirrored images. Table 3.1 confirms that the NSHP^{Z}-HMM outperforms the NSHP-HMM for any given version of the input images. When each pair of models that correspond to the same input image version is compared, an improvement of 3.11%, 1.78%, 2.44% and 1.01% on the test set is obtained.

	2-D models	Train	Test
Classical NSHP-HMMs in [CEC,05]	NSHP-HMM$_{(img)}$	93.69	93.44
	NSHP-HMM$_{(img^{lr})}$	95.00	94.91
	NSHP-HMM$_{(img^{ud})}$	94.42	94.00
	NSHP-HMM$_{(img^{lr\text{-}ud})}$	95.22	95.25
Proposed NSHP^{Z}-HMMs	NSHP^{Z}-HMM$_{(img)}$	99.16	96.55
	NSHP^{Z}-HMM$_{(img^{lr})}$	99.32	96.69
	NSHP^{Z}-HMM$_{(img^{ud})}$	98.87	96.44
	NSHP^{Z}-HMM$_{(img^{lr\text{-}ud})}$	98.95	96.62
	NSHP^{Z}-HMM$_{(img^{90^{\circ}})}$	98.97	96.03
	NSHP^{Z}-HMM$_{(img^{270^{\circ}})}$	99.19	96.28

Table 3.1: Performance of the proposed model against the classical NSHP-HMM for different versions of the input images

3.5.3 Performance improvement using models combination

Combining the four models trained separately on four mirrored images is a traditional and successful way to improve the performance of NSHP-HMM-based systems [Saon, 1999][Choisy and Belaïd, 2002][Cecotti et al., 2005]. The reason behind this process is the causal nature of the NSHP and the HMM according to which the scanning order of the image is left-to-right and top-to-bottom. Fig 3.6.a shows the obtained top1-5 recognition rates of three different combination rules: sum, product,

and majority vote rules. These results demonstrate that a combination of models permits a significant enhancement of recognition rate for any given combination rule, which is, of course, the expected result. The best result (97.74%, top1) is achieved with the product combination rule.

Because of the structure of the neighborhoods set (θ) in the NSHP Markov chain (see *Definition 2.2*), the conditional observation probability of site (i, j) is calculated independently of the sites located to the right and below it. By considering this crucial characteristic of the NSHP Markov chain, a rational design for a multiple NSHP$^{\mathrm{Z}}$-HMMs combination method is the creation of 4 images: original image, rotated by 90°, rotated by 180° and rotated by 270°, which will be processed by 4 NSHP$^{\mathrm{Z}}$-HMM. We create here two new models NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{90^\circ})}$ and NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{270^\circ})}$ and combine them with the two models NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img})}$ and NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{\text{lr-ud}})}$. The latter is the same model that corresponds to 180° rotated images.

The results of all six models on the training and test sets are given in Table 3.1 and Fig. 3.7. Both the NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img})}$ and the NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{\text{lr-ud}})}$ models perform slightly better than the NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{90^\circ})}$ and the NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{270^\circ})}$. These results confirm the discriminatory power of the columns over the rows of handwritten digits, as already pointed out in [Likforman-Sulem and Sigelle, 2007] for the same database.

Combining the four models NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img})}$, NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{90^\circ})}$, NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{\text{lr-ud}})}$ and NSHP$^{\mathrm{Z}}$-HMM$_{(\mathrm{img}^{270^\circ})}$ gives, as expected, better results than the traditional combination method. The results in Fig. 3.6.b show that the proposed combination method always outperforms the traditional combination method for any given combination rule. The best result (98.22%, top1) is achieved by product rule. When compared to the traditional combination of the four reference models presented in [Cecotti et al., 2005] (see Table 3.2), an improvement of 1.78% is obtained; and when compared to the traditional combination method of four NSHP$^{\mathrm{Z}}$-HMMs, an improvement of 0.48% is obtained.

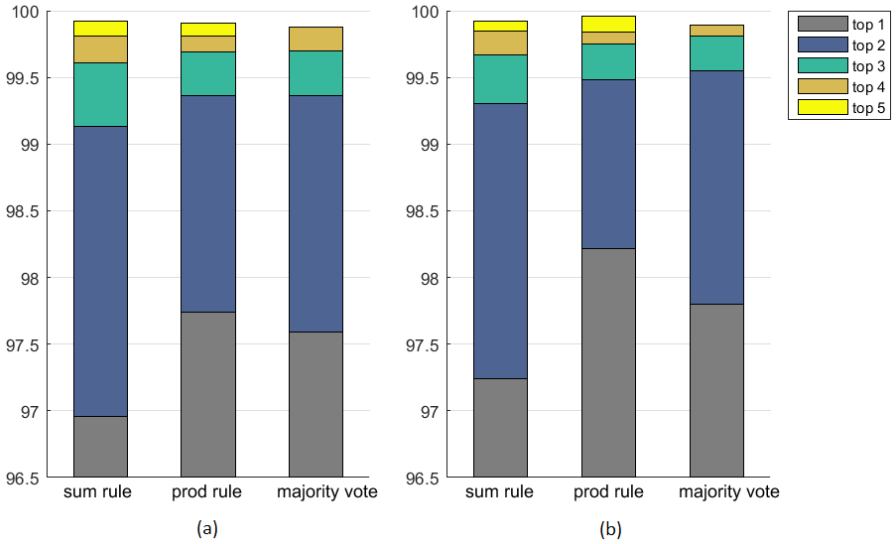


Figure 3.6: Recognition rate versus different combination rules of (a) traditional combination of four models corresponding to four mirrored images, (b) combination of four modes corresponding to four rotated images

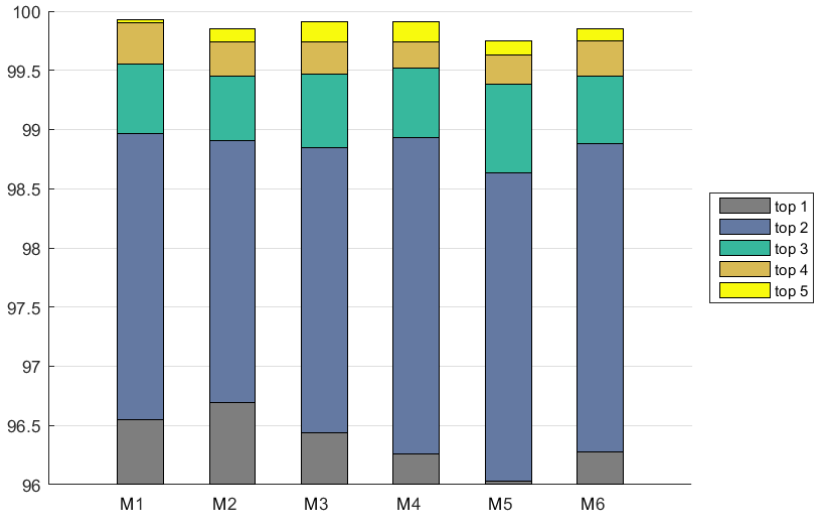


Figure 3.7: Recognition accuracy of six NSHP^{Z}-HMM models trained on different versions of the input images. M1, M2, M3, M4, M5 and M6 correspond respectively to NSHP^{Z}-HMM$_{(\text{img})}$, NSHP^{Z}-HMM$_{(\text{img}^{\text{lr}})}$, NSHP^{Z}-HMM$_{(\text{img}^{\text{ud}})}$, NSHP^{Z}-HMM$_{(\text{img}^{\text{lr-ud}})}$,NSHP^{Z}-HMM$_{(\text{img}^{90^\circ})}$ and NSHP^{Z}-HMM$_{(\text{img}^{270^\circ})}$

3.5.4 Comparison with the state-of-the-art HMM and non-HMM based 2-D recognizers

We compare here our digit recognition system presented in section 3.5.3 against the other 2-D recognizers described in section 2.2. All these recognizers were trained with the MNIST database except the HMMRF in [Park and Lee, 1998], which was trained on the database of Concordia University of Montreal, Canada.

The results in Table 3.2 show that our model outperforms all other HMM-based 2-D recognizers. The improvement in the accuracy ranges from 1.78%, when compared to the classical NSHP-HMM in [Cecotti et al., 2005], to 7.42%, when compared to the HMMRF in [Park and Lee, 1998].

2-D recognizers		Recognition rate
[Graves et al., 2007]	MDRNN	99.10%
Our model	**Combination of four 2^{nd}-order NSHP$^{\mathbf{Z}}$-HMM**	**98.22%**
[Cecotti et al., 2005]	Combination of four 3^{rd}-order NSHP-HMM	96.44%
[Likforman-Sulem and Sigelle, 2008]	Coupling of the vertical and horizontal AR-HMMs	94.90%
[Chevalier et al., 2003]	Combination of MRF and 2-D DP	94.60 %
[Park and Lee, 1998]	3^{rd}-order HMMRF	90.80%

Table 3.2: Performance of the proposed digital recognition system against the state of art HMM and non-HMM based 2-D recognizers.

3.6 Conclusion

The experimental study presented in this chapter demonstrates the effectiveness of the proposed model over other HMM-based 2-D recognizers and confirms the NSHP$^{\mathrm{Z}}$-HMM optimality issue examined in the theoretical study described in chapter 2.

Analysing the results obtained by the NSHP$^{\mathrm{Z}}$-HMM shows the major influence of height normalization, zoning parameters and models combination on system per-

formance.

Because of the importance of the zoning step in our approach, further investigations on this step may improve the system's performance. This point will present the main subject of the next chapter.

Chapter 4

2-D recognition system of Arabic handwritten words based on the NSHP^{Z}-HMM

4.1 Introduction

The contribution presented in this chapter consists in improving the NSHP^{Z}-HMM to better model the particularities of Arabic writing.

The proposed enhancement is implemented in the zoning step. It consists of dividing the Arabic word according to the position of its baseline. We call this method the "zoning approach based on baseline localization". The effectiveness of the proposed approach will be tested separately and according to recognition accuracy. In fact, the performance of this approach resides in the ability of our baseline to perfectly follow the writing curve even in case of vertical ligatures and of different slant angles within the same word.

It is important to note that the proposed method is the first truly 2-D HMM-based recognition system for handwritten Arabic words. The previous efforts made in this direction involve using the Planar-HMM [Touj et al., 2005][Touj et al., 2007]. This model is also called the "pseudo" 2-D HMM, in the sense that it is not a fully connected 2-D HMM in consideration of the independence between each column (each one is modeled by an independent HMM with no 2-D contextual information).

4.2 Particularities of Arabic writing

Arabic handwritten has its particularity in comparison with other scripts such as Latin which may pose significant challenges in employing conventional algorithms for preprocessing, feature extraction and recognition. Its main three peculiarities are the following:

- Arabic writing is semi-cursive; words in Arabic consist of a sequence of connected components called pseudo-words or PAWs (Piece of Arabic Word, see *Definition 4.1*).

- Some Arabic letters share the same main shape and differ only by the number and position of diacritical marks.

- Some couples or triples of letters can be joined vertically, forming what is known as a vertical ligature.

For more details on the specificities of the Arabic script with illustrative examples, the reader is referred to our previous works [Boukerma and Farah, 2012] and [Boukerma, 2010].

4.3 Related stat-of-art recognition systems of handwritten Arabic words

4.3.1 Arabic script recognition using Non-HMM based 2-D approaches

In [Graves, 2012] [El Abed and Märgner, 2011], three versions of MDLSTM (see Sec. 2.2.1) were proposed to recognize Arabic handwritten words of IFN/ENIT database [Pechwitz et al., 2002]. The difference between these three systems is slightly and concerns mainly with the recognizers parameters.

The proposed recognition system is multilingual. It operates directly on the word image without any preprocessing or feature extraction. Consequently, no adaptation

to Arabic specificities was done in this work. In the ICDAR 2009 Arabic Handwriting Recognition Competition [El Abed and Märgner, 2011], the MDLSTM outperformed all other recognition systems in terms of both recognition rate and speed. The obtained top1 recognition accuracy on test sets f and s were 93.37 % and 81.06 %, respectively.

The same MDLSTM network architecture of Graves [Graves, 2012] was used in [Chherawala et al., 2013] in order to evaluate the performance of automatically learned features (by the MDLSTM) compared to handcrafted features. Four sets of features were used: two distribution features, concavity features, and visual-descriptor-based features. Experiments were done on IFN/ENIT database and using the RNNLIB implementation of the recurrent neural network [Graves, 2013]. This work shows that the recognition system based on handcrafted features (precisely the distribution features) outperformed that using automatically learned features and achieved a top1 accuracy of 83.8 % compared to 76.5 % for the learned feature system.

The normalization algorithm used in [Breuel et al., 2013] before applying 1D LSTM for printed English OCR was adopted by [Yousefi et al., 2015] for Arabic handwritten words of the IFN/ENIT database. Using this normalization step allowed to simplify the LSTM architecture by applying 1D LSTM instead of its 2D architecture. In terms of training time and convergence, the proposed system was faster compared to 2D LSTM. Furthermore, in terms of recognition performance, the proposed system outperformed the 2D LSTM system and the 1D LSTM network trained with manually crafted features presented in [Chherawala et al., 2013]. The achieved top1 accuracy was 87.5 %.

In [Abandah et al., 2014], an interesting segmentation-based approach was proposed by Abandah et al. Their system, called JU-OCR2, has some common points with the MDLSTM of Graves et al. [Graves et al., 2007]. The main difference between the two systems is the feature extraction approaches. Whereas the MDLSTM extracts features directly from the raw pixels of the image (holistic approach), the JU-OCR2 extracts an efficient well selected 30 features from the segmented subwords (graphemes). The used classifier is also different: the BLSTM (bi-direction LSTM). In general, MDLSTM differs from ordinary BLSTM in that there are four distinct hidden layers instead of two, and each of these layers receives information

from two previous states instead of one. To improve the convergence and generalization ability of the BLSTM, the authors used an interesting technique named "weight noise" method. The proposed system was tested on the IFN/ENIT database. Compared with MDLSTM2 (an improved version of the MDLSRM in [Graves, 2012] that used the "weight noise" method), JU-OCR2 reduces the graphemes error, word error, and execution time by 18.5, 22.3, and 31%, respectively.

In [Khemiri et al., 2015] and [Khémiri et al., 2014], a similar work of [Likforman-Sulem and Sigelle (see section 2.2.1) was adopted for Arabic script recognition. The used classifiers were independent and coupled HMM-based models. The independent Vertical-Horizontal HMM considers features extraction from both rows and columns of the word image. However, the coupled classifier is performed through a DBN architecture which combines two basic HMM: the H-HMM (horizontal HMM) whose outputs are structural features (ascenders, descenders, loops and diacritic points) extracted from word image columns and V-HMM (vertical HMM) whose outputs are statistical features (pixel distributions and local pixel configurations) extracted from word image rows. Two DBN architectures were proposed which consist of adding connections between states of the H-HMM and V-HMM. Tested on a subset of the IFN/ENIT vocabulary, the independent Vertical-Horizontal HMM outperforms DBNs and achieves 92.19% [Khémiri et al., 2014] and 90.42% [Khemiri et al., 2015] recognition accuracy on 21 word classes and 50 word classes, respectively.

In [AlKhateeb et al., 2011], the authors presented a comparative study between 1D HMM and DBN recognizers. Like in [Likforman-Sulem and Sigelle, 2008], the used coupled architecture of the DBN framework was the autoregressive coupled model. Applied to IFN/ENIT database, the 1D HMM outperformed the DBN. The authors explained this surprising result by the fact that using sliding window method might simplify handwritten recognition to a linear case, hence 1D HMM works more effectively that DBN which is more suitable for spatial data.

Coupling vertical and horizontal HMM using the DBN formalism was also proposed in [Mahjoub et al., 2013]. The recognizer was tested on a subset of 18 classes of the IFN/ENIT database and achieved 83.7% recognition rate.

4.3.2 Arabic script recognition using HMM based 2-D approaches: PHMM

To the best of our knowledge, the truly 2-D HMM-based recognizer has never been applied to Arabic script recognition. It has been applied to Latin [Saon, 1999][Choisy and Belaïd, 200 korean [Park et al., 2001], Bangla and Chinese [Wang et al., 2000] [Feng et al., 2000] script recognition but not for Arabic. All existing works related to 2-D HMM were PHMM-based. *Our contribution presents the first truly 2-D HMM-based recognition system for Arabic script.*

The PHMM was used by Ben Amara et al. [Amara and Belaïd, 1996] to recognize 11 classes of connected printed Arabic words or subwords. The authors enhanced their system through the use of the normal density to represent the distribution of super-state duration. They report that their system achieves 100 % recognition rate.

The precedent system [Amara and Belaïd, 1996] was further developed by Miled and Ben Amara [Miled and Amara, 2001] to recognize a large vocabulary of 100 printed subwords. The recognition rate achieved was 99.84 %. A PHMM for 'handwritten' Arabic words was also presented in this paper. In the proposed model, jumps are allowed between 7 super-states; which are: beginning super-state, upper diacritics, ascenders, median, descenders, lower diacritics, and end super-state. The authors opted for an analytical modeling and they defined a new alphabet of pseudo-characters. However, no results about the performances of this work in the handwritten case were provided by the authors.

Further developments on median zone modeling were proposed in [Touj et al., 2005] and [Touj et al., 2007]. Tested on two subsets of 25 and 30 word classes of the INF/ENIT database, the obtained recognitions rates were, respectively, 88.7% [Touj et al., 2005] and 86.1% [Touj et al., 2007].

Recently, the PHMM was used by [Cheikh and Laffet, 2017] and [Cheikh and Allagui, 2015] to recognize wide then open vocabulary of Arabic decomposable words. In this approach, each PHMM was trained on a set of different words but which are derived from the same root. Since each super state of the principal HMM represents a definite morphological element (root consonant, infix enclitic, etc.) and each state of secondary HMM represents letter horizontal bands and well-defined classes of

primitives. The proposed system was tested on a vocabulary of about 7000 printed words derived from 90 roots. These words are from APTII database and they are in different fonts and sizes. In [Cheikh and Laffet, 2017], the authors discussed the ability of their system to recognize new words that have not been learned. The proposed technique consists of instant conceiving of appropriate PHMM whose pieces are intelligently collected from others PHMMs.

4.4 The proposed two-dimensional recognition system for handwritten Arabic words

4.4.1 Pre-processing

A set of pre-processing algorithms for handwritten Arabic script were presented in [Boukerma and Farah, 2012]. Here, to implement the proposed idea of NSHP^Z-HMM adaptation to Arabic script, we apply two pre-processing steps: diacritics extraction and baseline estimation. The estimated baseline is next used to find optimal zoning of Arabic word.

4.4.1.1 Diacritics extraction to PAWs localization

Diacritics such as dots and zigzags should be eliminated before baseline detection to avoid disturbance of two processes: the selection of the local minima points and the localization of PAWs.

For the diacritics extraction step, we use the algorithm presented in [Boukerma and Farah, 2012]. This algorithm is based on the area, height and relative position of the connected components. The thresholds used are determined empirically and depend on the thickness of a word (pen size). This latter is estimated by counting the run-lengths of black pixels in each column and row of the word image; the most frequent run-length is then adopted as the pen size. The use of thickness in threshold estimation makes this estimation more adaptable to the size variation of different words. A complete description of the algorithm used, with qualitative and quantitative evaluations on Arabic words from the IFN/ENIT database, is presented in [Boukerma and Farah, 2012].

4.4.1.2 Baseline estimation

The specificity of our baseline estimation algorithm (presented in Fig. 4.1) is that the extracted baseline is not a straight line. The problem in estimating the baseline as a straight line is that the PAW level creates a discontinuity of cursiveness that can cause the appearance of different slant angles in the same Arabic word. Furthermore, the existence of vertical ligatures between Arabic letters introduces another form of word slant. Fig. 4.2 shows some examples of such cases. The reader may be completely convinced from these examples that no straight line can represent the optimal baseline of these words. In these examples, the proposed baseline is marked in red. The blue straight line presents the maximum value of horizontal projection profile and serves here to give a comparison with our baseline. The reader is referred to [Boukerma and Farah, 2010] for more details on the baseline detection algorithm.

4.4.2 Optimal zoning design based on baseline localization

In the proposed system, dividing the word image in zones is relative to the baseline of this word. Fig. 4.3 describes the proposed zoning approach.

For each frame and considering the horizontal position, we first calculate the `y_means` of the baseline in this frame. Then, we divide the part of the frame above the baseline into $M1$ zones of the same height, and we also divide the part of the frame below the baseline into $M2$ zones of the same height. $M1$ and $M2$ must be the same for all images; their sum gives the parameter M of the NSHP$^{\mathrm{Z}}$-HMM model ($M = M1 + M2$).

In the other direction, a possible way to deal with changes in pattern localization relative to the vertical division is to extract the bounding box prior to zoning and to use large zone widths that overlap. For our system, we propose two forms of overlap. The first is the overlap between zone $Z_{i,t}$ and zone $Z_{i,t-1}$. The overlap rate at this level is noted by the Rz parameter. The second one is the overlap between zone Z_{it} and its neighborhood set of zones θ_{it}^{Z}; the overlap parameter is noted by Rn. A combination between these two forms is also possible. Consequently, the parameters of our zoning design approach are the number of vertical zones above and below the baseline ($M1$ and $M2$ respectively), the zone width W, the overlap rate between zones Rz and the overlap rate between the current zone and its neighborhood set of

Algorithm Baseline detection
Input: image, contour, skeleton, and thickness of word

1. Remove all diacritics and consider the remaining connected components as subwords.
2. For each subword image, estimate its appropriate horizontal band (*HB-subword*) by applying the following steps:
   a. Divide the image of the word on three equal horizontal partitions and use the second partition as the first horizontal band of whole word (*HB-word*).
   b. In *HB-word*, find the lowest points of closed loops and the branch points and cross points of word skeleton.
   c. Determine *HB-subword* as a horizontal range centered by the feature points selected in (b) and that has a total height equal to five times the approximate thickness of the word.
3. If there is subword *SWi* which do not contains any feature points of (b), so make *SWi* inherit the horizontal band of its nearest neighbor subword in the image.
4. If no feature points of (b) are extracted on all subwords of the word, then use of the horizontal projection method for baseline estimation.
5. Else, inside the *HB-subword* of each subword, detect the baseline relevant support points which are: - the local minima points of the lowest outer contour and - the lowest points of closed loops located next to lower *HB-subword*.
6. Trace the baseline of the entire word by applying the linear interpolation on each two consecutive support points selected in (5).

Output: not a straight baseline.

Figure 4.1: Baseline estimation algorithm for Arabic handwritten

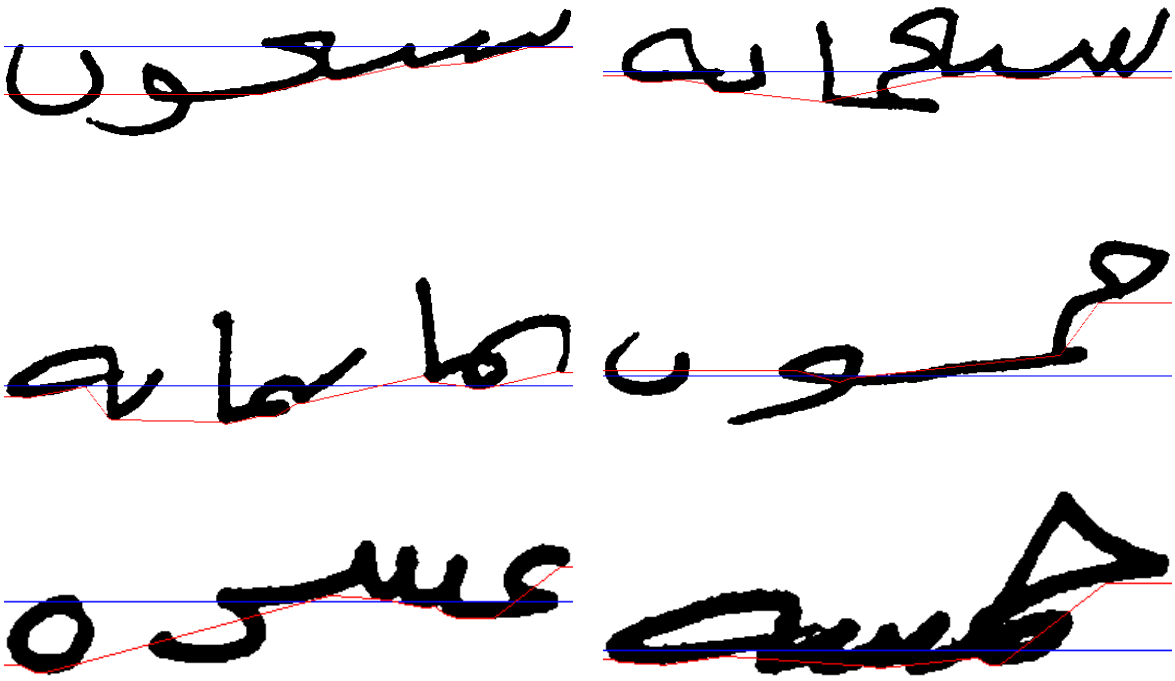


Figure 4.2: Good results of our baseline estimation algorithm (red line: our extracted baseline, blue line: baseline as the maximum value of horizontal projection profile). Left: words with different slant angles. Right: words with vertical ligatures

zones Rn.

4.4.3 Feature extraction and vector quantization

After dividing the image in zones, a feature vector is calculated from each zone; it can be composed of low-level and high-level features. Then, the extracted feature vectors are encoded through cluster prototypes using a vector quantization algorithm. In the proposed recognition system, we use the K-means clustering algorithm. The codebook size P is determined empirically on the validation dataset.

4.4.4 Training and Recognition

Training of the NSHP$^{\mathrm{Z}}$-HMM consists in estimating its parameters that optimally fit a set of training data. For this purpose, we use the ML function by performing the Baum-Welch re-estimation algorithm. The re-estimation formulas of the conditional

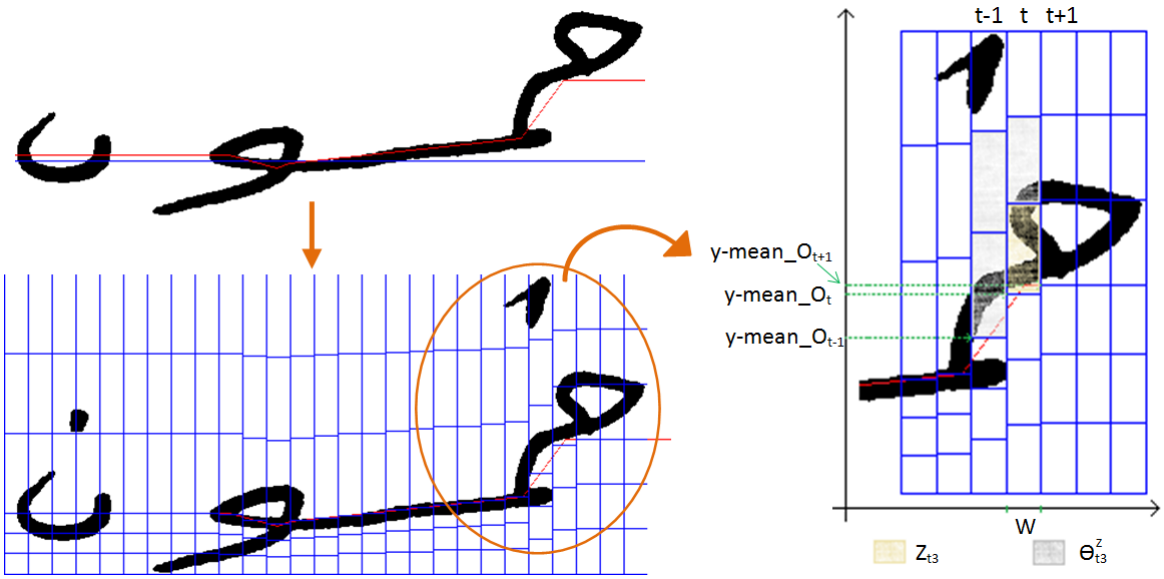


Figure 4.3: The proposed zoning approach based on baseline localization

zone observation probabilities (B) and the state transition matrix (A) are given in chapter 3.

The recognition phase is done according to the model discriminant approach. Given a test image, we calculate the pattern likelihood for all models, the model with maximum likelihood is the winner.

4.5 Experimental results

The holistic approach that we adopt in this work explains our orientation toward the small vocabulary size of Arabic literal amounts. To the best of our knowledge, the AHDB [Al-Ma'adeed et al., 2002] is the only available and free standard database in the field of Arabic check processing; other databases are either private [Farah et al., 2006] [Maddouri and Amiri, 2002] or not free (payment required) available databases [Al-Ohali et al., 2003]. The AHDB database contains Arabic words and texts written by a hundred different writers. For this database, we concern ourselves solely with one subdirectory which contains the most frequently used 33 words used in filling out checks; the words of this vocabulary are presented in Fig. 4.4. In this subdirectory, we have in total 3465 binary images (105 images per class); of them, 80% are used for training and 20% for testing.

Figure 4.4: Subset of the Arabic literal amounts vocabulary of AHDB database

Before presenting the recognition performances of the proposed system, we first present the results of the two used algorithms for baseline estimation and zoning design.

4.5.1 Results of the proposed algorithm of baseline estimation

An automatic quantitative evaluation of our algorithm of baseline estimation requires the availability of the baseline ground-truth with all y-positions in the word image. To the best of our knowledge, there is no Arabic handwritten text database that includes the ground-truth of a non-straight baseline. For this reason, we performed a subjective evaluation of the proposed algorithm. The adopted evaluation protocol is as follows:

First, let L^c be the number of letters in word class c. For each word image X of class c, we count e, the error of the baseline estimation algorithm, as the ratio between the number of letters with false baseline Fl and the total number of letters L^c in word class c, that is, $e = \frac{Fl}{L^c}$; the baseline error rate of word class c is the summation over all samples normalized by the number of samples in this class.

In the evaluation process, we used 3465 images of the AHDB database of Arabic literal amounts [Al-Ma'adeed et al., 2002]. Fig 4.2 illustrates some of the efficient results of our algorithm even in difficult cases of words with vertical ligatures and words with different slant angles. On the other hand, the algorithm partially fails

to correctly estimate the baseline of some parts of word; see Fig 4.5 for an example. Often, this error is caused by the erroneous selection of local minima points located at the bottom curve of descenders written near the baseline. We show at the bottom of the images in Fig. 4.5 the value of the error e.

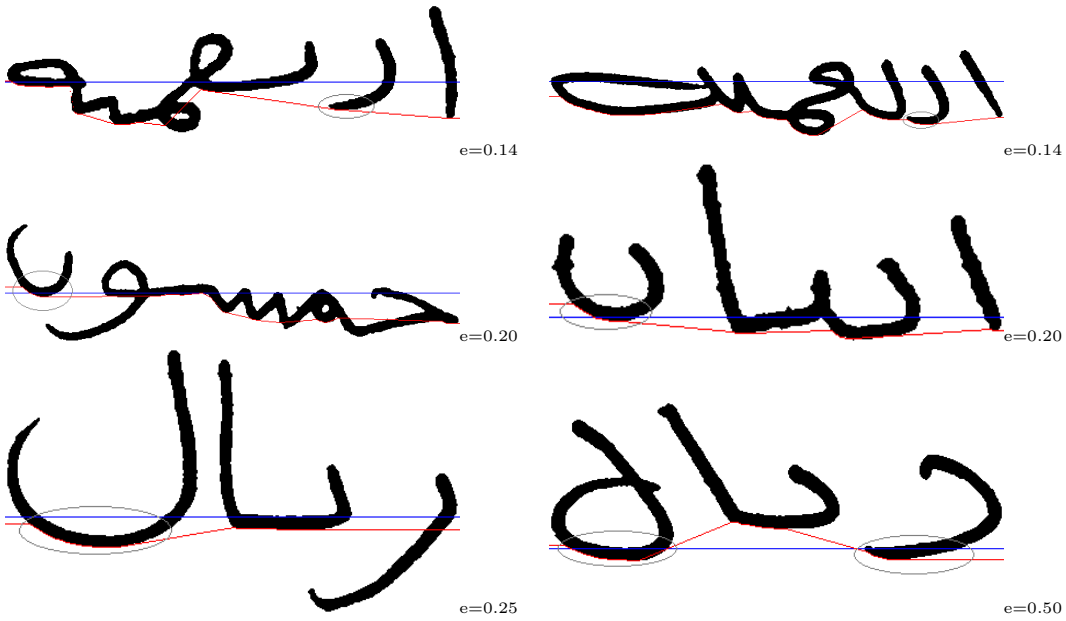


Figure 4.5: Partially failed cases of baseline estimation algorithm. Firs line: the descender isolated letter 'raa'. Second line: the descender isolated letter 'noon'. Third line: the word class 'ryal'

Following the evaluation protocol described above, the obtained error rate of baseline estimation algorithm is 10.47%. This result must not be interpreted as implying that the baselines of 10.47% of 3465 images are falsely extracted, but rather that, given a word image, the algorithm partially fails to correctly estimate the baseline with an average of 10.47% of the number of letters that constitute that word. The baseline of the remaining parts of the word is correctly estimated.

The challenging cases of our method are generally isolated descender letters written on the line (see the surrounding region in Fig. 4.5).

We note that the failure cases of our algorithm are similar. For example, for the word إثنان, the isolated descender letter noun ن is either well estimated (below the line) or when it is falsely estimated it is above the line. This regularity in baseline

error is very beneficial for our zoning design approach based on baseline because it increases the possibility of having the same division in zones for different samples of the same word class.

4.5.2 Results of the proposed zoning design approach

In this section, we evaluate our zoning design approach. This evaluation is quantitative and made against the uniform partition using a regular grid with $M = 6$. $M1$ and $M2$ in our zoning approach are equal and set to 3. To simplify the illustration, the Rz and the Rn parameters are set to zero. See Fig. 4.6.

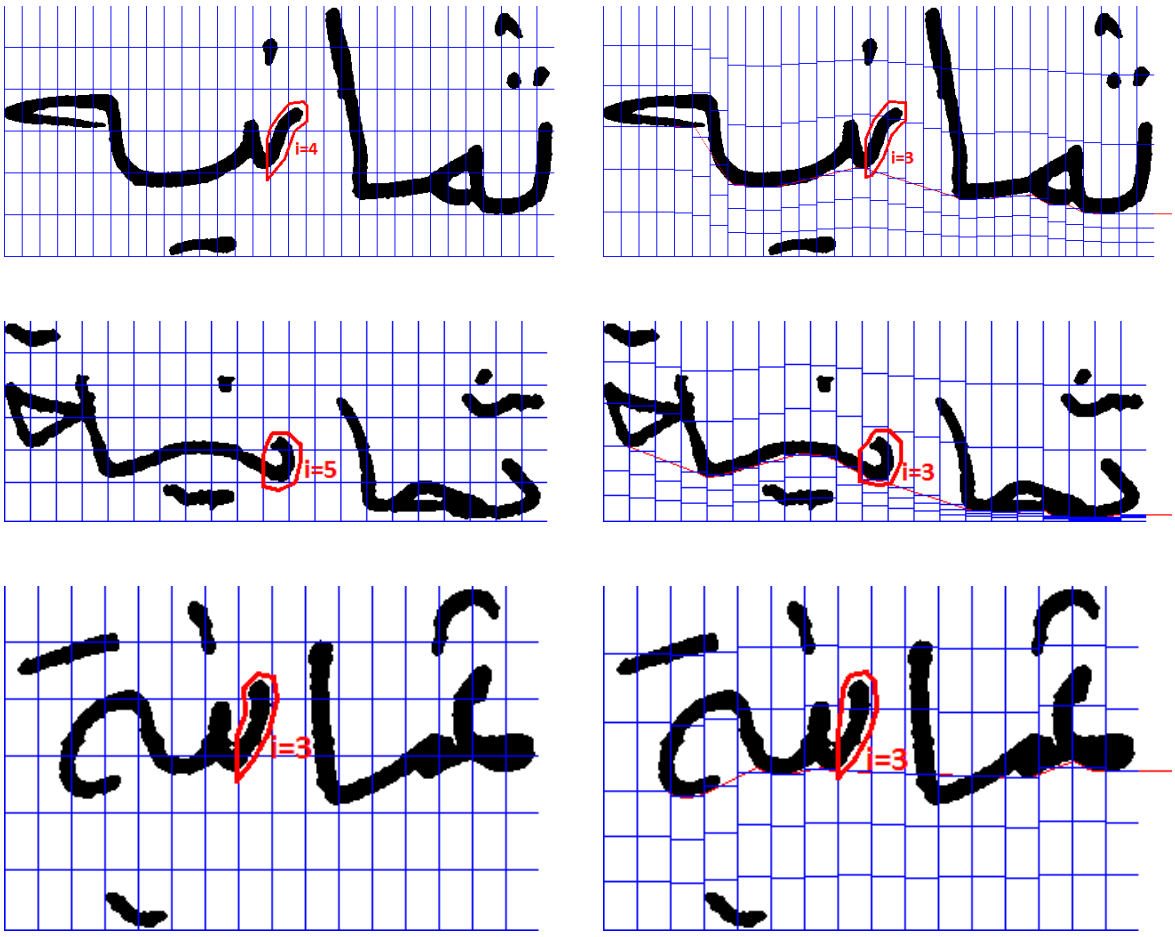


Figure 4.6: Comparison between the two zoning design methods. Right: results of our approach based on baseline localization. Left: results of the uniform partition using the regular grid

The proposed zoning design approach is clearly more efficient than the uniform partition. For example, by applying the proposed zoning approach to the class word ثمانية (see Fig. 4.6-right), the main parts of this word are located generally in the first three horizontal positions (i = 1,2 and 3) because this word does not have any descender letters. However, using the regular grid (see Fig. 4.6-left), the letters of this word change position from one sample to another. To give an example, we surrounded the letter نـ in Fig. 4.6. This letter changes positions (i = 3, 4 or 5) in the case of the zoning independent baseline (regular grid), but its position is the same (i = 3) when the zoning dependent baseline is used.

A quantitative comparison between these two methods of zoning design will be presented in Sec. 4.5.4. This comparison is made according to the recognition rate obtained by two NSHP^{Z}-HMM models one uses the zoning based baseline approach and the other uses the uniform partition (zoning-independent baseline).

4.5.3 Recognition system performances

The key parameters of NSHP^{Z}-HMM are the number of HMM states H, model order V, codebook size P, and the number of vertical zones above and below the baseline, $M1$ and $M2$ respectively. In addition, three system parameters are related to the zoning step, namely, the zone width W, the overlap rate between zones Rz and the overlap rate between the current zone and its neighborhood set of zones Rn.

To set these parameters, several experiments were done using 1050 images of the validation data set. The tested values of P were {10, 15, 20, 25, 30}. For model order V, we tested small values {0, 1, 2} because of the limited amount of training data. V equals zero means modeling zone Z_{it} independently of its neighborhood set of zones θ_{it}^{Z}. The tested values of ($M1$,$M2$) were {(2,2), (3,3), (4,2), (5,3), (6,3), (6,4)}. For zone width W, values were tested between 5 and 20 pixels. However, the tested values for the two overlap rates Rz and Rn were {0, 0.3, 0.5} and {0, 0.1, 0.2, 0.3, 0.5}, respectively.

Our best results were obtained using $V = 1$, $P = 10$, $M1 = 4$, $M2 = 2$, $W = 20$ pixels, $Rz = 0.5$ and $Rn = 0$. The number of states H was proportional to the average image length in the number of horizontal zones (N). H ranges from 29 for word class ألف to 60 for خمسمائة. We considered a right-to-left HMM for our NSHP^{Z}-HMM model. As features, we used the pixel density in 3×4 blocks in each

zone.

Using the founded optimal value, four NSHP^{Z}-HMMs were developed that correspond to four images: original image, mirrored vertically, mirrored horizontally, mirrored horizontally and vertically. Considering our preceding experiments on digital recognition, we chose the product rule to combine the outputs of this four models.

Recognition accuracies of each model and of their combination are given in Table 4.1. It is important to note that the proposed work is the first truly 2-D recognition system for handwritten Arabic words.

Table 4.1: Performance of the NSHP^{Z}-HMM with baseline-based zoning design on AHDB database

Models	**Recognition accuracy(%)**				
	Top 1	Top 2	Top 3	Top 5	Top 10
Model img	92.42	98.18	98.94	99.55	99.55
Model img^{V}	91.97	97.12	98.64	98.94	99.70
Model img^{H}	90.00	95.76	98.18	98.94	99.39
Model img^{HV}	92.12	97.12	98.48	99.39	99.55
Combination	95.45	98.48	99.09	99.55	99.70

4.5.4 NSHP^{Z}-HMM accuracy versus zoning-dependent and zoning-independent baseline

As the goal here is to explore how much an appropriate division in zones improves the performance of the NSHP^{Z}-HMM, we performed another experiment using a zoning method that is independent of baseline position. This method consists of dividing the image according to an uniform regular grid of $6 \times N$ zones, where $N = round(\frac{Nb\_column^{img}}{W})$. Therefore, the number of parameters of the two models, namely, the NSHP^{Z}-HMM using zoning dependent baseline (Sec. 4.5.3) and the NSHP^{Z}-HMM using zoning independent baseline (developed here), is equal. As in the preceding experiment, four NSHP^{Z}-HMMs were created that correspond to four mirrored images. The final system is the combination of these four models using product rule. The results are shown in Table 4.2. As can be seen, using the proposed zoning dependent baseline improves the accuracy of the NSHP^{Z}-HMM: the

increase in top1 accuracy reaches 2.87% when compared at the combination level (95.45 − 92.58); and it reaches a maximum of 4.7% when a comparison is made between each pair of individual models (92.12 − 87.42).

Table 4.2: Performance of the NSHP^{Z}-HMM using zoning-independent baseline

Models	**Recognition accuracy(%)**				
	Top 1	Top 2	Top 3	Top 5	Top 10
Model img	89.39	96.36	97.88	99.24	99.70
Model img^{V}	89.39	96.21	97.58	99.09	99.55
Model img^{H}	88.48	95.00	97.42	98.79	99.39
Model img^{HV}	87.42	95.15	97.12	98.94	99.39
Combination	92.58	97.12	98.48	99.10	99.85

4.5.5 Comparison to the state of the art

Table 4.3 provides a comparison of our results with other recognition systems developed for the same database (AHDB). Here we give a short description of these systems. In [Alma'adeed et al., 2004] a combination of rule-based classifier with a discrete 1-D HMM was presented. The authors of Ref. [Al-Nuzaili et al., 2017] proposed an enhanced quadratic angular feature model as a new statistical feature method. The classifier used is the Extreme Learning Machine (ELM). In [El-Melegy and Abdelbaset, 2007], a set of structural features was used to train four classifiers: neural network, K-NN, decision tree and Bayesian classifier. The obtained results of these four classifiers were, respectively, 86.5%, 83.1%, 80% and 79.5%. Table 4.3 shows also the results of a recognition system using the pixel-based NSHP-HMM re-implemented by ourselves. We observe that our proposed 2-D recognizer significantly outperforms the state-of-art recognition systems of AHDB Arabic literal amounts. The performance improvement in terms of top1 accuracy ranges from 3.78%, when compared to the pixel-based NSHP-HMM, to 44.45%, when compared to the 1-D HMM in Ref. [Alma'adeed et al., 2004].

Table 4.3: Comparison with the state-of-art recognition systems of AHDB Arabic literal amounts. Here, ZBL means zoning based baseline and ZRG zoning based regular grid

Systems	**Recognition accuracy(%)**			
	Top 1	Top 3	Top 5	Top 10
Rule-based classifier with 1-D HMM	51	69	77	-
ELM classifier	83.06	-	-	-
Neural network	86.5	-	-	-
Pixel-based NSHP-HMM	91.67	97.27	98.64	99.85
NSHP^{Z}-HMM + ZRG	92.58	98.48	99.10	99.85
NSHP^{Z}-HMM +ZBL	95.45	99.09	99.55	99.70

4.6 Conclution

We have proposed a new truly 2-D recognition system of Arabic handwritten words. The key point is the use of an efficient zoning design approach that exploits the two important information aspects of baseline and PAW. The result is an appropriate division in zones of an image word that facilitates its modelization through the use of the NSHP^{Z}-HMM model.

Experiments on Arabic literal amounts recognition have shown that the proposed zoning design method improves the performance of the NSHP^{Z}-HMM, and that this latter outperforms the reference NSHP-HMM model even when a simple regular grid for zoning design is used.

Chapter 5

Hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities

5.1 Introduction

In studying handwriting recognition using the classical NSHP-HMM [Choisy and Belaïd, 2002] and the NSHP$^{\mathrm{Z}}$-HMM [Boukerma et al., 2015], we find that it is useful to combine local analysis by pixel-based observation with global analysis by zoning. In fact, a possible strategy to improve the accuracy of pattern classifiers is the combination of global and local interpretations.

The contribution presented here is the introduction of a hybrid NSHP-HMM combining *pixel-based* and *zone-based* observation probabilities. The hybrid model will be theoretically presented in this chapter. Preliminary results on the MNIST database demonstrating the effectiveness of the hybrid approach over the baseline models will be presented in the appendix A.

The remaining part of this chapter is organized as follows: section 5.2 explains our motivation behind this work. Section 5.3 presents the hybrid NSHP-HMM combining *pixel-based* and *zone-based* observation probabilities and explains the mathematical reasoning behind the proposed idea. Section 5.4 provides the re-estimation formulas of the proposed model according to the expectation-maximization (EM) algorithm. Some remarks are presented in conclusion section 5.5.

5.2 The Hybrid model: motivation and advantages

The classical NSHP-HMM was successfully used in [Saon, 1999] and [Choisy and Belaïd, 2002] to efficiently model the words at a pixel level. However, the main three drawbacks of this model are the short 2D context, the requirement of the height normalization and binarization processes. Actually, pixel-based analysis and the number of properties grows exponentially with the neighborhood size (V), leading practically to considering very short 2D contexts measured in terms of a few pixels. In all NSHP-HMM-based systems [Saon, 1999] [Choisy and Belaïd, 2002] [Cecotti et al., 2005], a maximum value of V equaling 4 neighborhood "pixel" was used; the reason is to maintain a suitable compromise between accuracy and model complexity in term of number of features. For the NSHP$^{\mathrm{Z}}$-HMM, the 2D contextual information is modeled at a "zone" level, which results in extending the context and better modeling of the spatial property of an image. In addition, the use of high-level features extracted directly on the gray-level or color zones is possible. Merely, height normalization is not mandatory for the NSHP$^{\mathrm{Z}}$-HMM, as the vertical size of zones can be adapted to fit the vertical image size.

However, the main challenge with the NSHP$^{\mathrm{Z}}$-HMM is the necessity of an efficient zoning method adapted to the image content. In Chapter 4, this issue was resolved by dividing the image of Arabic word according to the position of its baseline. The baseline serves here as an efficient reference to zoning design.

In this work, we propose another way to escape the above challenge by introducing a hybrid NSHP-HMM combining *pixel-based* and *zone-based* observation probabilities. We point out at the beginning of this chapter that these two types of observations are incorporated into the same recognizer and modeled using the NSHP Markov random field. What we propose here is an hybrid model and not simple combinations between the two reference models (in appendix A, we compare our model with an external combination between the two reference models using different combination rules). The proposed model inherits the advantages of the two reference models: the NSHP-HMM with its local interpretation at pixel level without the need of dividing the image into zones; and the NSHP$^{\mathrm{Z}}$-HMM with its large context and using high-level features extracted on gray-level or color zones. In fact, a

possible strategy to improve the accuracy of pattern classifiers is the combination of global and local interpretations. The key idea is that the analysis must be global for a good synthesis of the information, while being based on local information suitable to create this synthesis.

5.3 Hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities

The proposed framework hybridizes a classical NSHP-HMM and an NSHP^{Z}-HMM, thus allowing a multi-level analysis of features (low-level and high-level information) and interpretation (pixel-based and zone-based views). We will refer to this framework as a hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities. Fig. 5.1 illustrates an image recognition system using the proposed hybrid model. In this example, $V_p = 3$ and $V_z = 1$. V_p and V_z represent, respectively, the order of pixel-based model and zone-based model (see Sec. 5.3.1).

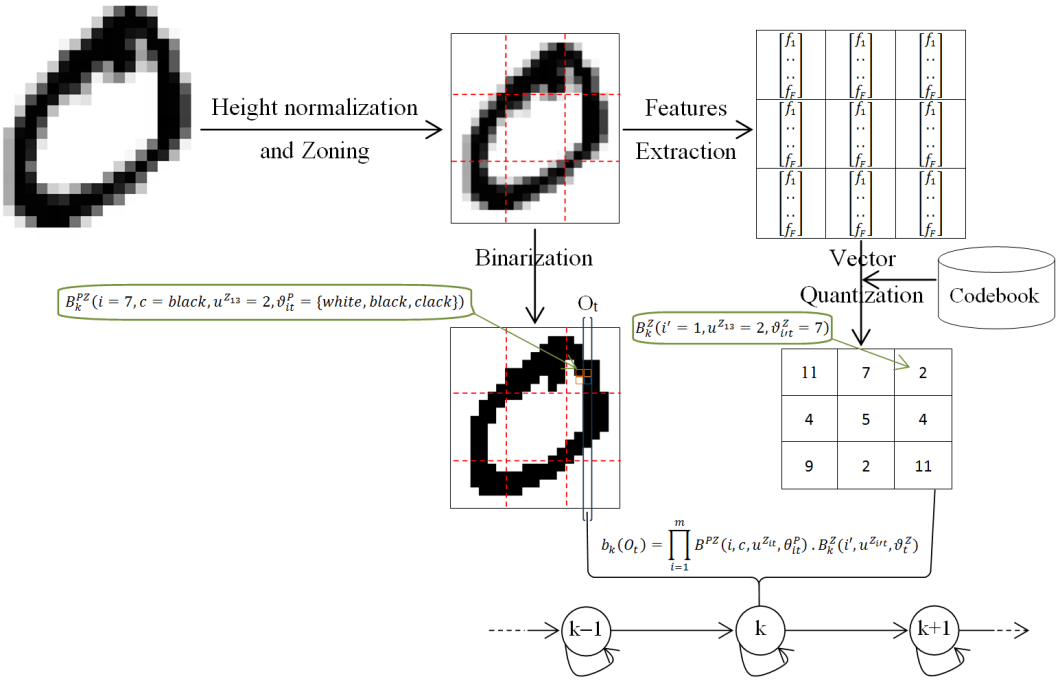


Figure 5.1: Image recognition using the proposed hybrid model.

To explain the reasoning behind the proposed idea, we first recall the count of

$b_k(O_t)$ in the classical NSHP-HMM :

$$b_k(O_t) = \prod_{i=1}^{m} P\left(X_{it} = c \,|X_{\theta_{it}},\ s_k\right) \tag{5.1}$$

Inspired by Xue's work [Xue and Govindaraju, 2006], the pixel X_{it} can be given as $\left(c, u^{Z_{it}}\right)$, where c is the color of pixel X_{it} (white or black) and u the corresponding symbol of the zone Z_{it} to which the pixel belongs. Thus, Equation 5.1 becomes:

$$\begin{aligned} b_k(O_t) &= \prod_{i=1}^{m} P\left(c,\ u^{Z_{it}}|\ X_{\theta_{it}},\ s_k\right) \\ &= \prod_{i=1}^{m} \left[P\left(c \mid u^{Z_{it}},\ X_{\theta_{it}},\ s_k\right) \cdot P\left(u^{Z_{it}}|\ X_{\theta_{it}},\ s_k\right)\right] \end{aligned} \tag{5.2}$$

In this initial version of the hybrid NSHP-HMM, we consider the zone Z_{it} as the column O_t. Therefore, $u^{Z_{it}}$ is replaced by u^{O_t}, which makes it independent of row i. Equation 5.2 becomes:

$$b_k(O_t) = \prod_{i=1}^{m} \left[P\left(c \mid u^{O_t},\ X_{\theta_{it}},\ s_k\right)\right] \cdot P\left(u^{O_t}|\ s_k\right) \tag{5.3}$$

where $P\left(c \mid u^{O_t},\ X_{\theta_{it}},\ s_k\right) = B_k^{PZ}\left(i,\ c,\ u^{O_t},\ \theta_{it}^{P}\right)$, which presents the probability to observe in state k a pixel (i,t) of color c at row i and column t when one know the neighborhood set of pixel θ_{it}^{P}, the pixel (i,t) is defined by discrete symbol u which corresponds to column O_t.

The second part of Equation 5.3, $P\left(u^{O_t}|\ s_k\right)$, presents the probability to observe in state k a discrete symbol u corresponding to column O_t. This probability can be calculated using the NSHP$^{\mathrm{Z}}$-HMM model or using a 1-D HMM. In this work, the 2-D NSHP$^{\mathrm{Z}}$-HMM model is used. This probability is then calculated as $P\left(u^{O_t}|\ s_k\right) = B_k^{Z}\left(u^{O_t},\ \theta_t^{Z}\right)$, which presents the probability to observe in state k a discrete symbol u corresponding to column O_t given the neighborhood set of column θ_t^{Z}.

The probability $b_k(O_t)$ for the hybrid model is consequently computed as

$$b_k(O_t) = \left[\prod_{i=1}^{m} B_k^{PZ}\left(i,\ c,\ u^{O_t},\ \theta_{it}^{P}\right)\right] \cdot B_k^{Z}\left(u^{O_t},\ \theta_t^{Z}\right) \tag{5.4}$$

5.3.1 Formal description

The complete parameter set of the hybrid NSHP-HMM is $\lambda = \left(A,\ B^{PZ},\ B^{Z}\right)$, where

- $U^P = \{0, 1\}$, the vocabulary of the local view model (0: white pixel, 1: black pixel).
- $U^Z = \{u_1,\ ...,\ u_P\}$, the set of P discrete symbols composing the vocabulary of the global view model.
- $\theta^P = \left\{\theta_{ij}^P\right\}$, the set of V_p neighborhoods of the pixel (i, j) fixed in the half plane. V_p is designated the order of the pixel-based model.
- $\theta^Z = \left\{\theta_{ij}^Z\right\}$, the set of V_z neighborhoods of zone Z_{ij}. V_z is designated the order of the zone-based model.
- $S = \{s_1,\ ...,\ s_H,\ D,\ F\}$, the set of H normal states with two specific states D and F that model the probability of beginning and ending in each normal state respectively.
- $A = \left\{a_{kh} \cup \{a_{Dk},\ a_{kF}\}\right\}_{1 \leq k,\, h \leq H}$, the state transition probability matrix, where $a_{kh} = P\left(q_{t+1} = s_h \mid q_t = s_k\right)$, $a_{Dk} = P\left(q_1 = s_k \mid D\right)$, $a_{kF} = P\left(F \mid q_T = s_k\right)$. T denotes the number of observations (here, $T = n$ the number of columns in the image)
- $B_k^{PZ} = \left\{b_k^{PZ}\left(i,\ c,\ u^{O_t},\ \theta_{ij}^P\right)\right\}$, where $s_k \in S,\ s_k \neq D,\ F$. B^{PZ} is the probability to observe in state k a pixel (i, j) of color c at height i and column j defined by the discrete symbol u when one know the neighborhood set of pixel θ_{ij}^P.
- $B^Z = \left\{b_k^Z\left(u^{O_j},\ \theta_j^Z\right)\right\}$, where $s_k \in S,\ s_k \neq D,\ F$. B^Z is the probability of observing in state k a discrete symbol u corresponding to the column O_j given the neighborhood set of column θ_j^Z. For the general case of the hybrid NSHP-HMM model, the "column" j is replaced by the "zone" to which the pixel (i, j) belongs (see Fig. 5.1).

The set of P discrete symbols composing the vocabulary of the global view model are computed empirically on the validation dataset. Three steps are applied: 1. dividing the image into $M \times N$ zones using a zoning technique adapted to an application domain. 2. From each zone, a feature vector is calculated; it can be composed of low-level and high-level features. 3. Encode the feature vector of each zone into a discrete symbol using the K-means clustering algorithm [Fink, 2008].

5.4 Parameters estimation of the hybrid model

For the hybrid NSHP-HMM, we need to estimate the following parameters: the state transition probability matrix A, the conditional pixel/zone observation probabilities B^{PZ}, and the conditional zone observation probabilities B^{Z}. These parameters are estimated using the Baum-Welch algorithm.

Recognition phase is done according to model discriminent approach. Given a testing image, Eq. 5.6 is used to compute the pattern likelihood for all models, then we select the model of maximum likelihood.

5.4.1 The modified forward variables

$$\begin{gathered}
\alpha_1(k) = a_{Dk} \cdot B_k^Z\left(u^{O_1},\ \theta_1^Z\right) \cdot \prod_{i=1}^{m} B_k^{PZ}\left(i,\ c,\ u^{O_{i1}},\ \theta_{i1}^P\right), \\
1 \le k \le H \\
\alpha_t(k) = \left[\sum_{h=1}^{H} \alpha_{t-1}(h) \cdot a_{hk}\right] \cdot B_k^Z\left(u^{O_t},\ \theta_t^Z\right) \cdot \prod_{i=1}^{m} B_k^{PZ}\left(i,\ c,\ u^{O_t},\ \theta_{it}^P\right), \\
1 \le k \le H;\ 2 \le t \le n
\end{gathered} \tag{5.5}$$

$$P(X|\ \lambda) = \sum_{k=1}^{H} \alpha_n(k) \tag{5.6}$$

where m and n are respectively the number of rows and columns of the image. Therefore, as for the classical NSHP-HMM, a height normalization procedure is required for the proposed hybrid model.

5.4.2 The modified backward variables

$$\begin{gathered}
\beta_n(k) = a_{kF},\ 1 \le k \le H \\
\beta_t(k) = \sum_{h=1}^{H} a_{kh} \cdot \beta_{t+1}(h) \cdot B_k^Z\left(u^{O_{t+1}},\ \theta_{t+1}^Z\right) \cdot \prod_{i=1}^{m} B_k^{PZ}\left(i,\ c,\ u^{O_{t+1}},\ \theta_{i,t+1}^P\right), \\
1 \le k \le H;\ t = n-1,\ ..,\ 1
\end{gathered} \tag{5.7}$$

5.4.3 Parameters estimation

$$\overline{a}_{Dk} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{P_r}\alpha_1^r(k)\,\beta_1^r(k)$$

$$\overline{a}_{kF} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\alpha_{N_r}^r(k)\,a_{kF}}{\sum_{r=1}^{R}\frac{1}{P_r}\left[\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,\beta_t^r(k)+\alpha_{N_r}^r(k)\,a_{kF}\right]}$$

$$\overline{a}_{kh} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,a_{kh}\beta_{t+1}^r(h)\left[\prod_{i=1}^{m}B_h^{PZ}\left(i,c,u^{O_{t+1}},\theta_{i,t+1}^{P}\right)\right]B_h^Z\left(u^{O_{t+1}},\theta_{t+1}^Z\right)}{\sum_{r=1}^{R}\frac{1}{P_r}\left[\sum_{t=1}^{N_r-1}\alpha_t^r(k)\,\beta_t^r(k)+\alpha_{N_r}^r(k)\,a_{kF}\right]} \tag{5.8}$$

The state transition matrix is estimated in a way similar to the estimation in a classical NSHP-HMM with little modification; the detailed formulas are given by Eq. 5.8. The estimation of the conditional zone observation probabilities B^Z is derived in Sec. 3.3.2 and rewritten in Eq. 5.9. The conditional pixel/zone observation probabilities B^{PZ} is computed by Eq. 5.10. Here, R is the number of samples in the training set, and P_r is the emission probability of the sample X^r calculated using Eq. 5.6.

$$\overline{B}_k^Z\left(i,u^Z,\theta^Z\right) = \begin{cases} \dfrac{\sum_{r=1}^{R}\frac{1}{P_r}\sum\limits_{\substack{j=1\\U_{ij}^Z=u^Z\\ \theta_{ij}^Z=\theta^Z}}^{N_r}\alpha_j^r(k)\,\beta_j^r(k)}{\sum_{r=1}^{R}\frac{1}{P_r}\sum\limits_{\substack{j=1\\ \theta_{ij}^Z=\theta^Z}}^{N_r}\alpha_j^r(k)\,\beta_j^r(k)} & ,denominator \neq 0 \\ b_k^Z\left(i,u^Z,\theta^Z\right) & ,otherwise \end{cases} \tag{5.9}$$

$$\overline{B}_k^{PZ}\left(i,c,u^Z,\theta^P\right)=\begin{cases}\dfrac{\displaystyle\sum_{r=1}^{R}\frac{1}{P_r}\sum_{\substack{j=1\\ pixel\ (i,t)\ is\ c\\ U_{it}^Z=u^Z\\ \theta_{it}^P=\theta^P}}^{n_r}\alpha_j^r(k)\,\beta_j^r(k)}{\displaystyle\sum_{r=1}^{R}\frac{1}{P_r}\sum_{\substack{j=1\\ U_{it}^Z=u^Z\\ \theta_{it}^Z=\theta^Z}}^{n_r}\alpha_j^r(k)\,\beta_j^r(k)} & ,denominator\neq 0\\ b_k^{PZ}\left(i,c,u^Z,\theta^P\right) & ,otherwise\end{cases} \tag{5.10}$$

5.5 Conclusion

In this chapter, a new HMM-based 2-D model has been introduced to combines the pixel-based and the zone-based observation probabilities in order to improve the modeling power of the classical NSHP-HMM.

The main drawback of the hybrid model is that the height normalization process is mandatory like for the classical NSHP-HMM. The reason is the definition of the hybrid model that was based on the pixel-based analysis of the classical NSHP-HMM (see Eq. 5.1 and 5.2). Note also that a large number of samples is needed to train the hybrid model because of the important number of its parameters (A, B^Z and B^{PZ} matrices). For this reason, the first series of experiments were performed on the MNIST database. The experimental study is not finished yet, this is why we have decided to present it in the appendix part of our dissertation.

Chapter 6

Conclusion

In this dissertation, a new HMM-based 2-D model has been introduced. The model, called the NSHP^{Z}-HMM, presents an efficient solution to the four drawbacks of the reference model, such as short 2-D context and the requirement of the height normalization and binarization processes. Thanks to zone-level, the 2-D context is enlarged and the spatial property of an image is better modeled using local discriminative features extracted from binary, gray-level or color zones. Furthermore, without suffering of exponential complexity, like the other HMM-based 2-D models, the NSHP^{Z}-HMM brings, like the reference model, the efficient training and decoding algorithms of 1-D HMM to the 2-D modeling of spatial data.

The overview and the taxonomy of 2-D based recognition approaches (chapter. 2) presents an important contribution of this work. We hope that the contents of this chapter will be helpful to researchers interested in this field.

The first part of the experimental section (chapter 3) has been done so that to gain a better understanding of the superiority of the NSHP^{Z}-HMM on the classical NSHP-HMM. For this reason, the MNIST database was used. Sec. 3.5 presents a comprehensive comparison of our system against the state-of-the-art HMM and non-HMM based 2-D recognizers on handwritten digit recognition.

Next, we turned in chapter 4 to the problem of handwritten Arabic word recognition. The key point of the proposed system is the use of an efficient zoning design approach that exploits the two important information aspects of baseline and PAW. The result is an appropriate division of an image word into zones that facilitates its modelization through the use of the NSHP^{Z}-HMM model. Experi-

ments on Arabic literal amounts recognition have shown that the proposed zoning design method improves the performance of the NSHP^{Z}-HMM and that this latter outperforms the reference NSHP-HMM model even when a simple regular grid for zoning design is used. Taking into account state-of-the-art works on the AHDB database, the results are improved from 8.95% when compared to the work in [El-Melegy and Abdelbaset, 2007] to 44.45% as compared to [Alma'adeed et al., 2004].

Therefore, in chapter 5, we introduced a hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities. Analyzing the effect of the hybrid model parameters showed that extending the 2-D context at the zone-level is more effective than its enlargement at the pixel-level, which justifies once more the superiority of the NSHP^{Z}HMM over the classical NSHP-HMM.

The contributions presented in this dissertation have been published in the following refereed conference and journal articles:

(1) Hanene Boukerma, Christophe Choisy, Nadir Farah, and Mohamed Cheriet. The efficiency of the NSHP^{Z}-HMM: theoretical and practical study. *Applied Intelligence*, 48(12), 4660-4677. 2018.

(2) Hanene Boukerma, Christophe Choisy, Abdallah Benouareth, Nadir Farah, and Mohamed Cheriet. Hybrid two-dimensional recognizer based on the NSHP-HMM model. In *16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*,pages 582-586, IEEE, 2018.

(3) Hanene Boukerma, Christophe Choisy, Nadir Farah, and Mohamed Cheriet. Baseline-based zoning design for the NSHP^{Z}-HMM 2-D model applied to Arabic script. In *2nd International Workshop on Arabic Script Analysis and Recognition (ASAR)*, pages 41-46, IEEE, 2018.

(4) Hanene Boukerma, Christophe Choisy, Abdallah Benouareth, and Nadir Farah. A performance evaluation of NSHP-HMM based on conditional zone observation probabilities application to offline handwriting word recognition. In *13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1091-1095. IEEE, 2015.

(5) Hanene Boukerma, Abdallah Benouareth, and Nadir Farah. NSHP-HMM based on conditional zone observation probabilities for offline handwriting recogni-

tion. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 2961-2965. IEEE, 2014.

(6) Hanene Boukerma and Nadir Farah. Preprocessing algorithms for Arabic handwriting recognition systems. In *International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Workshop on Islamic Applications in Computer Science and Technology*, 2012, pages 318-323. IEEE, 2012.

(7) Hanene Boukerma and Nadir Farah. A novel Arabic baseline estimation algorithm based on sub-words treatment. In *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2010, pages 335-338. IEEE, 2010.

Future work will focus on four main directions:

Reducing the number of parameters of the proposed model would be a welcome enhancement, especially when the training data are insufficient. This can be done at the B^Z matrix by considering the set of neighboring zones as one neighboring *'region'*. Consequently, the model order V is reduced to only 1 neighborhood without reducing the 2-D context. In this case, two codebooks must be used: one to encode the features of current zone and the other to encode the features of the neighboring region.

The second future direction will consist of improving the vertical division of word image into zones. Even if the use of the Rz parameter (set to 0.5) has reduced the weakness of this step, a further enhancement is possible by using some key points searched in the word skeleton and the vertical projection profile.

The third opportunity for future work considers the hybrid model. Because that this model was defined based on the pixel-based analysis of the classical NSHP-HMM (see Eq. 5.1 and 5.2 in chapter 5), its zone-based view is restricted to one column and the height normalization process is mandatory like for the classical NSHP-HMM. Enlargement of the zone view of the hybrid model requires the development of *states synchronization mechanism* between the pixel-based view and the zone-based view (the two parts of Eq. 5.4). This enhancement might improve the performances like it has been proved for the NSHP$^{\mathrm{Z}}$HMM, and presents the subject of a future publication on handwritten word recognition.

Other possible model improvements can be expected by moving toward the analytic approach. Our currently reported results on 33 words of AHDB database are relatively good when one remembers that the vocabulary of Arabic literal amounts has considerable similarity between classes (i.e. [تسعة, سبعة], [أألفان, اثنان], [تسعون, سبعون] and [ستمائة, تسعمائة]) and that the model order employed considers only one neighborhood zone. Applying our work to a large vocabulary size requires the move towards the analytic approach that has been developed for the classical NSHP-HMM [Choisy and Belaïd, 2002].

Appendix A

The hybrid model: Experimental results

The proposed hybrid NSHP-HMM has been tested on handwritten digit images of the MNIST database.

The key parameters of the hybrid NSHP-HMM combining pixel-based and zone-based observation probabilities are model state number H, the order of the pixel-based model V_p, the order of the zone-based model V_z and codebook size P.

Figure A.1 shows the influence of these parameters on the recognition rate of the proposed hybrid model. The tested values were $\{0, 1, 2\}$ for V_p, $\{0, 1\}$ for V_z and $\{5, 10, 15, 20, 25, 30\}$ for P. Model order (V_p or V_z) equals zero means that the site(here, pixel or column, respectively) is modeled independently of its neighborhood. For each experiment, the training and testing images are normalized in 14 lines and the used number of model states (H) is proportional to the average image length in columns.

From Fig. A.1, it is clear that the hybrid model is not very sensitive to the increase in V_p. For example, when the model with ($V_p = 2$, $V_z = 1$, $P = 30$) is compared against the model with ($V_p = 1$, $V_z = 1$, $P = 30$), an improvement of 0.27% is obtained; and when the model with ($V_p = 2$, $V_z = 0$, $P = 20$) is compared against the model with ($V_p = 1$, $V_z = 0$, $P = 20$), an improvement of 0.80% is obtained.

On the other hand, increasing V_z significantly increases the recognition rate. For example, when the model with ($V_p = 2$, $V_z = 1$, $P = 20$) is compared against the

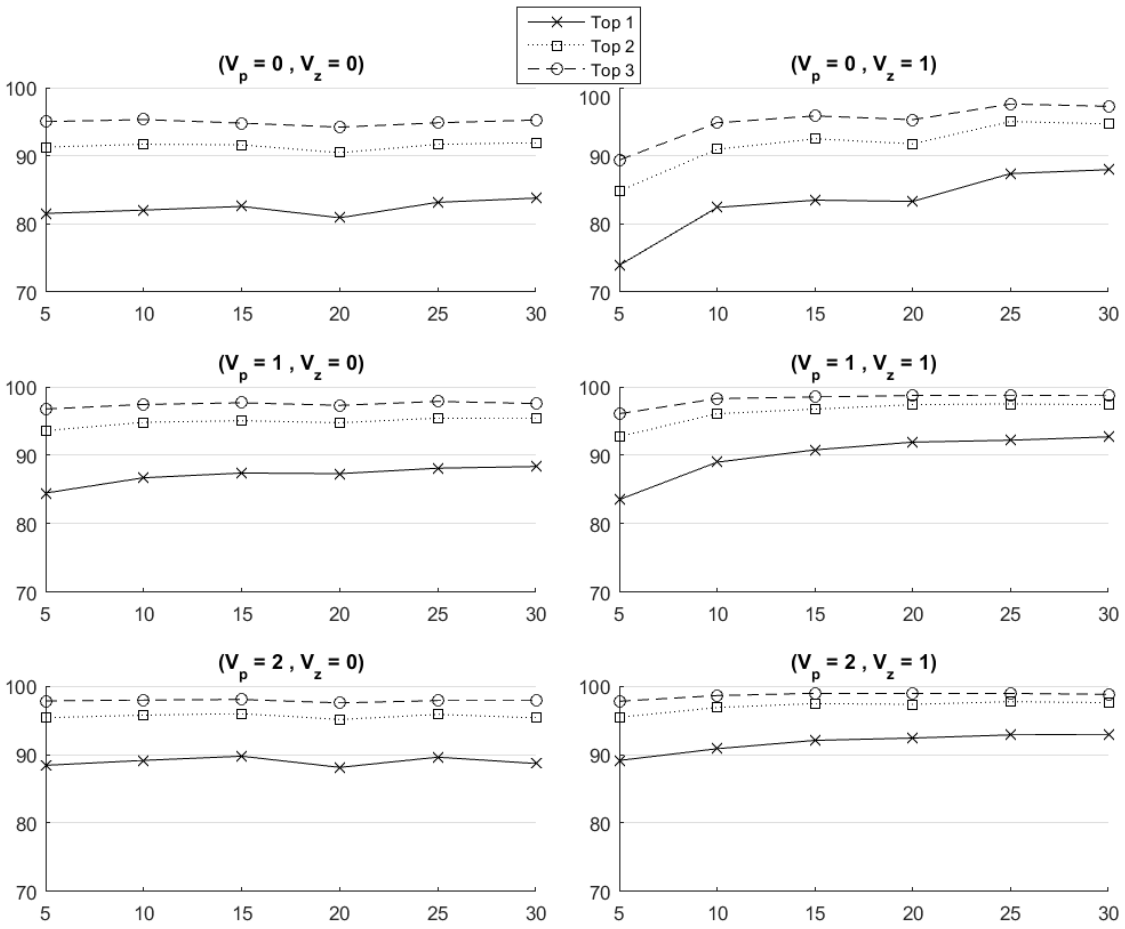


Figure A.1: Recognition rates of the hybrid model for different model orders (V_p & V_z) and different codebook sizes P (the x-axis).

model with ($V_p = 2$, $V_z = 0$, $P = 20$), an improvement of 4.28% is obtained; and when the model with ($V_p = 1$, $V_z = 1$, $P = 20$) is compared against the model with ($V_p = 1$, $V_z = 0$, $P = 20$) an improvement of 4.60% is obtained. From these experiments, we deduce that the information provided by the global view model significantly increases the performance. In other words, extending the 2-D context at the zone-level is more effective than its enlargement at the pixel-level, which justifies the superiority of the NSHP$^{\mathrm{Z}}$-HMM over the classical NSHP-HMM [Boukerma et al., 2015].

In the next series of experiments, we compared the hybrid NSHP-HMM with three models: the classical NSHP-HMM, the NSHP$^{\mathrm{Z}}$-HMM and an external combination between these two models (see Table A.1). The tested combination rule were : sum, product and majority vote rules. In order to make an objective comparison,

The classical NSHP-HMM

V_p	0	1	2
R.R	86.07	86.36	92.35

The NSHP^Z-HMM

V_z	0	1
R.R	81.97	89.15

Combination between the NSHP-HMM and the NSHP^Z-HMM

		(V_p, V_z)			
		(0,0)	(0,1)	(1,0)	(1,1)
Comb. rule	Sum	81.97	89.15	81.97	89.15
	Prod	87.71	90.13	90.03	91.65
	M.Vote	84.07	87.69	84.47	88.07

The proposed Hybrid model

(V_p,V_z)	(0,0)	(0,1)	(1,0)	(1,1)
R.R	83.79	88.55	88.39	93.40

Table A.1: Recognition rates (R.R) of the hybrid model against the classical NSHP-HMM, the NSHP^Z-HMM and combination between the two baseline models for different values of model orders (V_p, V_z)

the zone in the NSHP^Z-HMM is considered as one column. We should point out that the zoning is a crucial stage for the NSHP^Z-HMM, and using appropriate division into zones certainly increases the model's performance.

For these experiments, both the classical NSHP-HMM and the NSHP^Z-HMM were trained on normalized images of 14 lines with the P parameter being equal to 30 for both the NSHP^Z-HMM and the hybrid model. Owing to learning time, the maximum value of model order tested in these experiments is 2 for V_p and 1 for V_z; further experiments with higher orders must be conducted to determine an enhancement in performance.

The results in Table A.1 show that the hybrid model outperforms the baseline models for small values of model order. Furthermore, the hybrid NSHP-HMM gives better results than the combination of the baseline models for any given combination rule.

We note here that our main objective of this work is to show the basic elements of the proposed hybrid 2-D model and its superiority over the two baseline models.

Thus, the results presented here do not necessarily outperform the state-of-art approaches. However, the obtained preliminary results are comparable to results of some 2-D recognition approaches applied to the same classification problem; such as the 3^{rd}-order HMMRF based system proposed by Park et al. [Park and Lee, 1998], the coupling of the vertical and horizontal AR-HMMs (Auto-Regressive HMMs) proposed by Likforman-Sulem et al. [Likforman-Sulem and Sigelle, 2007] and the combination of MRF and 2-D dynamic programming proposed by Chevalier et al. [Chevalier et al., 2003].

Bibliography

[Abandah et al., 2014] Abandah, G. A., Jamour, F. T., and Qaralleh, E. A. (2014). Recognizing handwritten arabic words using grapheme segmentation and recurrent neural networks. *International Journal on Document Analysis and Recognition (IJDAR)*, 17(3):275–291.

[Agazzi et al., 1993] Agazzi, O. E., Kuo, S.-s., Levin, E., and Pieraccini, R. (1993). Connected and degraded text recognition using planar hidden markov models. In *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, volume 5, pages 113–116. IEEE.

[Al-Ma'adeed et al., 2002] Al-Ma'adeed, S., Elliman, D., and Higgins, C. A. (2002). A data base for arabic handwritten text recognition research. In *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*, pages 485–489. IEEE.

[Al-Nuzaili et al., 2017] Al-Nuzaili, Q., Ali, H., Siti, H., Saeed, F., and Khalil, M. S. (2017). An enhanced quadratic angular feature extraction model for arabic handwritten literal amount recognition. In *Recent Trends in Information and Communication Technology: Proceedings of the 2nd International Conference of Reliable Information and Communication Technology (IRICT 2017)*, volume 5, page 369. Springer.

[Al Ohali, 2002] Al Ohali, Y. (2002). *Handwritten word recognition: Application to Arabic cheque processing*. PhD thesis, Concordia University.

[Al-Ohali et al., 2003] Al-Ohali, Y., Cheriet, M., and Suen, C. (2003). Databases for recognition of handwritten arabic cheques. *Pattern Recognition*, 36(1):111–121.

[AlKhateeb et al., 2011] AlKhateeb, J. H., Pauplin, O., Ren, J., and Jiang, J. (2011). Performance of hidden markov model and dynamic bayesian network classifiers on handwritten arabic word recognition. *knowledge-based systems*, 24(5):680–688.

[Alma'adeed et al., 2004] Alma'adeed, S., Higgins, C., and Elliman, D. (2004). Off-line recognition of handwritten arabic words using multiple hidden markov models. *Knowledge-Based Systems*, 17(2):75–79.

[Amara and Belaïd, 1996] Amara, N. B. and Belaïd, A. (1996). Printed paw recognition based on planar hidden markov models. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 220–224. IEEE.

[Baggenstoss, 2011] Baggenstoss, P. M. (2011). Two-dimensional hidden markov model for classification of continuous-valued noisy vector fields. *Aerospace and Electronic Systems, IEEE Transactions on*, 47(2):1073–1080.

[Baumgartner et al., 2016] Baumgartner, J., Flesia, A. G., Gimenez, J., and Pucheta, J. (2016). A new image segmentation framework based on two-dimensional hidden markov models. *Integrated Computer-Aided Engineering*, 23(1):1–13.

[Belaïd and Choisy, 2008] Belaïd, A. and Choisy, C. (2008). Human reading based strategies for off-line arabic word recognition. In *Arabic and Chinese Handwriting Recognition*, pages 36–56. Springer.

[Bobulski, 2017] Bobulski, J. (2017). Multimodal face recognition method with two-dimensional hidden markov model. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 65(1):121–128.

[Boudaren and Belaïd, 2009] Boudaren, M. E. Y. and Belaïd, A. (2009). Markov models and extensions for land cover mapping in aerial imagery. In *International Conference of Signal and Image Engineering-ICSIE 2009*.

[Boukerma, 2010] Boukerma, H. (Dec, 2010). Combinaison de classifieurs flous pour la reconnaissance de l'écriture arabe manuscrite. Master's thesis, thèse Magister, Skikda university.

[Boukerma et al., 2014] Boukerma, H., Benouareth, A., and Farah, N. (2014). Nshp-hmm based on conditional zone observation probabilities for off-line handwriting recognition. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 2961–2965. IEEE.

[Boukerma et al., 2015] Boukerma, H., Choisy, C., Benouareth, A., and Farah, N. (2015). A performance evaluation of nshp-hmm based on conditional zone observation probabilities application to offline handwriting word recognition. In *13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1091–1095. IEEE.

[Boukerma et al., 2018a] Boukerma, H., Choisy, C., Benouareth, A., Farah, N., and Cheriet, M. (2018a). Hybrid two-dimensional recognizer based on the nshp-hmm model. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 582–586. IEEE.

[Boukerma et al., 2018b] Boukerma, H., Choisy, C., Farah, N., and Cheriet, M. (2018b). The efficiency of the nshp z-hmm: theoretical and practical study. *Applied Intelligence*, 48(12):4660–4677.

[Boukerma and Farah, 2010] Boukerma, H. and Farah, N. (2010). A novel arabic baseline estimation algorithm based on sub-words treatment. In *International Conference onFrontiers in Handwriting Recognition (ICFHR)*, pages 335–338. IEEE.

[Boukerma and Farah, 2012] Boukerma, H. and Farah, N. (2012). Preprocessing algorithms for arabic handwriting recognition systems. In *International Conference on Advanced Computer Science Applications and Technologies (ACSAT), Workshop on Islamic Applications in Computer Science and Technology*, pages 318–323. IEEE.

[Brand et al., 1997] Brand, M., Oliver, N., and Pentland, A. (1997). Coupled hidden markov models for complex action recognition. In *Computer vision and pattern recognition, 1997. proceedings., 1997 ieee computer society conference on*, pages 994–999. IEEE.

[Breuel et al., 2013] Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A., and Shafait, F. (2013). High-performance ocr for printed english and fraktur using lstm networks. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 683–687. IEEE.

[Cecotti et al., 2005] Cecotti, H., Vajda, S., and Belaïd, A. (2005). High performance classifiers combination for handwritten digit recognition. In *Pattern Recognition and Data Mining*, pages 619–626. Springer.

[Cheikh and Allagui, 2015] Cheikh, I. B. and Allagui, I. (2015). Planar markovian approach for the recognition of a wide vocabulary of arabic decomposable words. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*, pages 1031–1035. IEEE.

[Cheikh and Laffet, 2017] Cheikh, I. B. and Laffet, A. (2017). New morphological markovian approach for analysis and recognition of open arabic canonical vocabulary. In *Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on*, volume 1, pages 183–188. IEEE.

[Chevalier et al., 2003] Chevalier, S., Geoffrois, E., and Prêteux, F. (2003). A 2d dynamic programming approach for markov random field-based handwritten character recognition. In *IAPR International Conference on Image and Signal Processing*. Citeseer.

[Chevalier et al., 2005] Chevalier, S., Prêteux, F. J., Geoffrois, E., and Lemaitre, M. (2005). A generic 2d approach of handwriting recognition. In *Eighth International Conference on Document Analysis and Recognition (ICDAR 2005), 29 August - 1 September 2005, Seoul, Korea*, pages 489–493. IEEE.

[Chherawala et al., 2013] Chherawala, Y., Roy, P. P., and Cheriet, M. (2013). Feature design for offline arabic handwriting recognition: handcrafted vs automated? In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 290–294. IEEE.

[Choisy, 2007] Choisy, C. (2007). Dynamic handwritten keyword spotting based on the nshp-hmm. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 242–246. IEEE.

[Choisy and Belaïd, 2002] Choisy, C. and Belaïd, A. (2002). Cross-learning in analytic word recognition without segmentation. *International Journal on Document Analysis and Recognition*, 4(4):281–289.

[Choisy et al., 2003] Choisy, C. et al. (2003). Coupling of a local vision by markov field and a global vision by neural network for the recognition of handwritten words. In *null*, page 849. IEEE.

[Devijver and Dekesel, 1988] Devijver, P. A. and Dekesel, M. (1988). Champs aléatoires de pickard et modélisation d'images digitales. *Traitement du signal*, 5(5):131–150.

[Devijver and Dekesel, 1987] Devijver, P. A. and Dekesel, M. M. (1987). Learning the parameters of a hidden markov random field image model: A simple example. In *Pattern Recognition Theory and Applications*, pages 141–163. Springer.

[Diem et al., 2013] Diem, M., Fiel, S., Garz, A., Keglevic, M., Kleber, F., and Sablatnig, R. (2013). Icdar 2013 competition on handwritten digit recognition (hdrc 2013). In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1422–1427. IEEE.

[El Abed and Margner, 2007] El Abed, H. and Margner, V. (2007). Comparison of different preprocessing and feature extraction methods for offline recognition of handwritten arabicwords. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 974–978. IEEE.

[El Abed and Märgner, 2011] El Abed, H. and Märgner, V. (2011). Icdar 2009-arabic handwriting recognition competition. *International Journal on Document Analysis and Recognition (IJDAR)*, 14(1):3–13.

[El-Melegy and Abdelbaset, 2007] El-Melegy, M. T. and Abdelbaset, A. A. (2007). Global features for offline recognition of handwritten arabic literal amounts. In *Information and Communications Technology, 2007. ICICT 2007. ITI 5th International Conference on*, pages 125–129. IEEE.

[Farah et al., 2006] Farah, N., Souici, L., and Sellami, M. (2006). Classifiers combination and syntax analysis for arabic literal amount recognition. *Engineering Applications of Artificial Intelligence*, 19(1):29–39.

[Feng et al., 2000] Feng, Q., Minghua, D., Minping, Q., and Xueqing, Z. (2000). A novel algorithm for handwritten chinese character recognition. In *Advances in Multimodal Interfaces — ICMI*, pages 379–385. Springer.

[Fink, 2008] Fink, G. A. (2008). *Markov models for pattern recognition: from theory to applications*. Springer Science & Business Media.

[Gilloux, 1995] Gilloux, M. (1995). Reconnaissance de chiffres manuscrits par modèle de markov pseudo-2d. *TS. Traitement du signal*, 12(6):561–566.

[Graves, 2008] Graves, A. (2008). *Supervised Sequence Labelling with Recurrent Neural Networks*. PhD thesis, Technische Universitat Munchen, Fakultat fur Informatik.

[Graves, 2012] Graves, A. (2012). Offline arabic handwriting recognition with multidimensional recurrent neural networks. In *Guide to OCR for Arabic scripts*, pages 297–313. Springer.

[Graves, 2013] Graves, A. (2013). Rnnlib: A recurrent neural network library for sequence learning problems. *[OL][2015–07-10]*.

[Graves et al., 2007] Graves, A., Fernández, S., and Schmidhuber, J. (2007). Multidimensional recurrent neural networks. In *Artificial Neural Networks – ICANN*, pages 549–558.

[Graves and Schmidhuber, 2009] Graves, A. and Schmidhuber, J. (2009). Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in neural information processing systems*, pages 545–552.

[Hanene et al., 2018] Hanene, B., Christophe, C., Nadir, F., and Mohamed, C. (2018). Baseline-based zoning design for the nshp z-hmm 2-d model applied to arabic script. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)*, pages 41–46. IEEE.

[Impedovo and Pirlo, 2014] Impedovo, D. and Pirlo, G. (2014). Zoning methods for handwritten character recognition: A survey. *Pattern Recognition*, 47(3):969–981.

[Jain, 2010] Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666.

[Jeng and Woods, 1987] Jeng, F.-C. and Woods, J. W. (1987). On the relationship of the markov mesh to the nshp markov chain. *Pattern Recognition Letters*, 5(4):273–279.

[Joshi et al., 2006] Joshi, D., Li, J., and Wang, J. Z. (2006). A computationally efficient approach to the estimation of two-and three-dimensional hidden markov models. *Image Processing, IEEE Transactions on*, 15(7):1871–1886.

[Khemiri et al., 2015] Khemiri, A., Echi, A. K., Belaid, A., and Elloumi, M. (2015). Arabic handwritten words off-line recognition based on hmms and dbns. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 51–55. IEEE.

[Khémiri et al., 2014] Khémiri, A., Kacem, A., and Belaïd, A. (2014). Towards arabic handwritten word recognition via probabilistic graphical models. In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, pages 678–683. IEEE.

[Kuo and Agazzi, 1994] Kuo, S.-S. and Agazzi, O. E. (1994). Keyword spotting in poorly printed documents using pseudo 2-d hidden markov models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(8):842–848.

[Kurata et al., 2006] Kurata, D., Nankaku, Y., Tokuda, K., Kitamura, T., and Ghahramani, Z. (2006). Face recognition based on separable lattice hmms. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE.

[Lazzerini and Marcelloni, 2001] Lazzerini, B. and Marcelloni, F. (2001). A fuzzy approach to 2d-shape recognition. *Fuzzy Systems, IEEE Transactions on*, 9(1):5–16.

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

[Li et al., 2000] Li, J., Najmi, A., and Gray, R. M. (2000). Image classification by a two-dimensional hidden markov model. *Signal Processing, IEEE Transactions on*, 48(2):517–533.

[Li et al., 2017] Li, L., Ming, T., Liu, S., and Zhang, S. (2017). An effective health indicator based on two dimensional hidden markov model. *Journal of Mechanical Science and Technology*, 31(4):1543–1550.

[Likforman-Sulem et al., 2009] Likforman-Sulem, L., Darbon, J., and Smith, E. H. B. (2009). Pre-processing of degraded printed documents by non-local means and total variation. In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pages 758–762. IEEE.

[Likforman-Sulem and Sigelle, 2007] Likforman-Sulem, L. and Sigelle, M. (2007). Recognition of degraded handwritten digits using dynamic bayesian networks. In *Document Recognition and Retrieval XIV, DRR 2007*.

[Likforman-Sulem and Sigelle, 2008] Likforman-Sulem, L. and Sigelle, M. (2008). Recognition of degraded characters using dynamic bayesian networks. *Pattern Recognition*, 41(10):3092–3103.

[Ma et al., 2007] Ma, X., Schonfeld, D., and Khokhar, A. (2007). A general two-dimensional hidden markov model and its application in image classification. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 6, pages VI–41. IEEE.

[Ma et al., 2008] Ma, X., Schonfeld, D., and Khokhar, A. (2008). Image segmentation and classification based on a 2d distributed hidden markov model. In *Visual Communications and Image Processing 2008*, volume 6822, page 68221F. International Society for Optics and Photonics.

[Maddouri and Amiri, 2002] Maddouri, S. S. and Amiri, H. (2002). Combination of local and global vision modelling for arabic handwritten words recognition. In *Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on*, pages 128–135. IEEE.

[Madhogaria et al., 2015] Madhogaria, S., Baggenstoss, P. M., Schikora, M., Koch, W., and Cremers, D. (2015). Car detection by fusion of hog and causal mrf. *Aerospace and Electronic Systems, IEEE Transactions on*, 51(1):575–590.

[Mahjoub et al., 2013] Mahjoub, M. A., Ghanmy, N., Miled, I., et al. (2013). Multiple models of bayesian networks applied to offline recognition of arabic handwritten city names. *arXiv preprint arXiv:1301.4377*.

[Merialdo et al., 2000] Merialdo, B., Marchand-Maillet, S., and Huet, B. (2000). Approximate viterbi decoding for 2d-hidden markov models. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 4, pages 2147–2150. IEEE.

[Miled and Amara, 2001] Miled, H. and Amara, N. E. B. (2001). Planar markov modeling for arabic writing recognition: Advancement state. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*, pages 69–73. IEEE.

[Othman and Aboulnasr, 2003] Othman, H. and Aboulnasr, T. (2003). A separable low complexity 2d hmm with application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(10):1229–1238.

[Park and Lee, 1998] Park, H.-S. and Lee, S.-W. (1998). A truly 2-d hidden markov model for off-line handwritten character recognition. *Pattern Recognition*, 31(12):1849–1864.

[Park et al., 2001] Park, H.-S., Sin, B.-K., Moon, J., and Lee, S.-W. (2001). A 2-d hmm method for offline handwritten character recognition. *International journal of pattern recognition and artificial intelligence*, 15(01):91–105.

[Pechwitz et al., 2002] Pechwitz, M., Maddouri, S. S., Märgner, V., Ellouze, N., Amiri, H., et al. (2002). Ifn/enit-database of handwritten arabic words. In *Proc. of CIFED*, volume 2, pages 127–136. Citeseer.

[Perronnin, 2004] Perronnin, F. (2004). *A probabilistic model of face mapping applied to person recognition*. PhD thesis, ECOLE POLYTECHNIQUE.

[Perronnin et al., 2003] Perronnin, F., Dugelay, J.-L., and Rose, K. (2003). Iterative decoding of two-dimensional hidden markov models. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 3, pages III–329. IEEE.

[Rabiner, 1989] Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

[Ronee et al., 2001] Ronee, M. A., Uchida, S., and Sakoe, H. (2001). Handwritten character recognition using piecewise linear two-dimensional warping. In *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*, pages 39–43. IEEE.

[Samia et al., 2002] Samia, S.-M., Hamidi, A., Abdel, B., and Christophe, C. (2002). Combination of local and global vision modelling for arabic handwritten words recognition. In *Eighth International Workshop on Frontiers in Handwriting Recognition - IWFHR'02, Ontario, Canada*. IEEE.

[Saon, 1999] Saon, G. (1999). Cursive word recognition using a random field based hidden markov model. *International Journal on Document Analysis and Recognition*, 1(4):199–208.

[Saon and Belaïd, 1997] Saon, G. and Belaïd, A. (1997). High performance unconstrained word recognition system combining hmms and markov random fields. *International Journal of Pattern Recognition and Artificial Intelligence*, 11(05):771–788.

[Sargin et al., 2008] Sargin, M. E., Altinok, A., Rose, K., and Manjunath, B. (2008). Conditional iterative decoding of two dimensional hidden markov models. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 2552–2555. IEEE.

[Shenoy et al., 2016] Shenoy, R., Shih, M.-C., and Rose, K. (2016). Deformable registration of biomedical images using 2d hidden markov models. *IEEE Transactions on Image Processing*, 25(10):4631–4640.

[Tamamori et al., 2014] Tamamori, A., Nankaku, Y., and Tokuda, K. (2014). Image recognition based on separable lattice trajectory 2-d hmms. *IEICE TRANSACTIONS on Information and Systems*, 97(7):1842–1854.

[Touj et al., 2005] Touj, S. M., Amara, N. E. B., and Amiri, H. (2005). Arabic handwritten words recognition based on a planar hidden markov model. *Int. Arab J. Inf. Technol.*, 2(4):318–325.

[Touj et al., 2007] Touj, S. M., Amara, N. E. B., and Amiri, H. (2007). A hybrid approach for off-line arabic handwriting recognition based on a planar hidden markov modeling. In *null*, pages 964–968. IEEE.

[Uchida and Sakoe, 1999] Uchida, S. and Sakoe, H. (1999). An efficient two-dimensional warping algorithm. *IEICE TRANSACTIONS on Information and Systems*, 82(3):693–700.

[Uchida and Sakoe, 2005] Uchida, S. and Sakoe, H. (2005). A survey of elastic matching techniques for handwritten character recognition. *IEICE transactions on information and systems*, 88(8):1781–1790.

[Vajda and Belaïd, 2005] Vajda, S. and Belaïd, A. (2005). Structural information implant in a context based segmentation-free hmm handwritten word recognition system for latin and bangla script. In *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, pages 1126–1130. IEEE.

[Wan et al., 2018] Wan, W., Yuan, L., Zhao, Q., and Fang, T. (2018). Two-dimensional hidden semantic information model for target saliency detection and eyetracking identification. *Journal of Electronic Imaging*, 27(1):013006.

[Wang et al., 2016a] Wang, G.-g., Gan, Z.-l., Tang, G.-j., Cui, Z.-g., and Zhu, X.-c. (2016a). Basic problems solving for two-dimensional discrete 3× 4 order hidden markov model. *Chaos, Solitons & Fractals*, 89:73–82.

[Wang et al., 2016b] Wang, G.-g., Tang, G.-j., Gan, Z.-l., Cui, Z.-g., and Zhu, X.-c. (2016b). Basic problems and solution methods for two-dimensional continuous 3× 3 order hidden markov model. *Chaos, Solitons & Fractals*, 89:435–446.

[Wang et al., 2000] Wang, Q., Zhao, R., Chi, Z., and Feng, D. D. (2000). Hmmrf: a stochastic model for offline handwritten chinese character recognition. In *Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on*, volume 3, pages 1475–1478. IEEE.

[Xue and Govindaraju, 2006] Xue, H. and Govindaraju, V. (2006). Hidden markov models combining discrete symbols and continuous attributes in handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(3):458–462.

[Yousefi et al., 2015] Yousefi, M. R., Soheili, M. R., Breuel, T. M., and Stricker, D. (2015). A comparison of 1d and 2d lstm architectures for the recognition of handwritten arabic. In *Document Recognition and Retrieval XXII*, volume 9402, page 94020H. International Society for Optics and Photonics.

[Yujian, 2007] Yujian, L. (2007). An analytic solution for estimating two-dimensional hidden markov models. *Applied Mathematics and Computation*, 185(2):810–822.