A Statistical Approach for Phrase Location and Recognition within a Text Line: An Application to Street Name Recognition

ARTICLE *in* IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE · MARCH 2002

Impact Factor: 5.78 · DOI: 10.1109/34.982898 · Source: IEEE Xplore

Mounim A. El Yacoubi
Institut Mines-Télécom

A Statistical Approach for Phrase Location and Recognition within a Text Line: An Application to Street Name Recognition

Mounim A. El-Yacoubi, Michel Gilloux, and Jean-Michel Bertille

Abstract—In this paper, we describe a new approach to conjointly locate and recognize a street name within a street line. The system developed is based on a probabilistic framework that naturally integrates various knowledge sources to emit a final decision. At the handwriting signal level, hidden Markov models are extensively used to provide the needed matching scores. Several optimization techniques are employed to speed up the processing time. Experiments carried out on large data sets of street line images, automatically extracted from real French mail envelope images, show very promising results.

Index Terms—Phrase detection and recognition, handwriting recognition, statistical modeling, hidden Markov models.

1 Introduction

Handwriting recognition has been the focus of much attention during the last years. The interest devoted to this field is not explained only by the exciting challenges involved, but also by the huge benefits that a system, designed in the context of a commercial application, could bring. Two classes of recognition systems are usually distinguished: online systems [31] for which handwriting data are captured during the writing process, which makes available the information on the ordering of the strokes, and offline systems [30] for which recognition takes place on a static image captured once the writing process is over. Standard recognition systems usually work under the hypothesis that the processed image contains only the unknown text transcription (word, word string, etc.) to be recognized. In real-world tasks, however, the recognizer task is sometimes more complicated and consists of simultaneously detecting and recognizing a word or a word string (phrase) within a text line or a text page. Examples of such applications could be the search of a keyword in a text document, the identification of the recipient name in a document fax, or the recognition of a street name on mail envelopes. The approach presented here concerns the latter example and, therefore, the system to be built is classified into the offline category. Street name identification is an important subtask of address recognition because in a majority of postal organizations, street names, and street numbers determine the carrier walk concerned by the delivery of the mail item bearing the address. This information is used in automatic sorting machines to sort the mail with respect to carrier walks and also to sequence the mail in the order of delivery. Also, street name and street number identification may help cross-check the recognition of ZIP codes in order to reduce the error rate for ZIP code recognition.

Street names usually consist of a sequence of words, each pertaining to a given semantic or grammatical class. The most encountered classes are (Fig. 1): *Street Types*, *Articles*, *First Names*, *Titles*, and *Key Names*. Throughout this paper, *key name* will designate the last and main name in a street name, i.e., excluding prefixes, suffixes, etc. (e.g., `FOCH` in `RUE DU MARECHAL FOCH`).

Street name recognition is a very challenging problem because it requires tackling many levels of complexity, thus making the approach be developed generic enough to be adapted to other similar real-life tasks. It involves dealing with many sources of variability that make the recognizer task particularly difficult. The most significant are:

- Street line location: The hypothesis that a street line is immediately above the postal code/city name line is not always true. Thus, automatic location of street lines is subject to failures.
- The huge diversity of writing styles, inherent to any omni-writer handwriting recognition task.
- The typographical variability of words inside a street line. The fact that there are many words that do not bear the same semantic (or grammatical) information makes the writers subjectively choose the writing type of each word, thus explaining the various forms encountered (Fig. 2).

- *M.A. El-Yacoubi is with Parascript, LLC, 7105 La Vista Place, Niwot, CO 80503. E-mail: yacoubi@parascript.com.*
- *M. Gilloux and J.-M. Bertille are with La Poste, Service de Recherche Technique de la Poste (SRTP), 10, Rue de l'Ile-Mabon, BP 86334, FR-44263 NANTES Cedex, France. E-mail: {michel-gilloux, jean-michel.bertille}@laposte.fr.*

Manuscript received 11 Jan. 2001; revised 24 May 2001; accepted 17 July 2001.
Recommended for acceptance by L. Vincent
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113450.

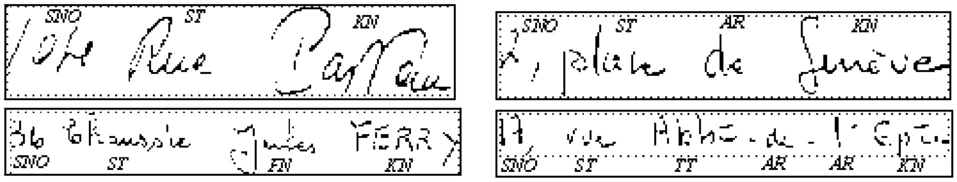

Fig. 1. Examples of street names with the notations: (*SNO*): street number, (*ST*): street type, (*AR*) article, (*FN*) first name, (*TT*) title, and (*KN*): key name.

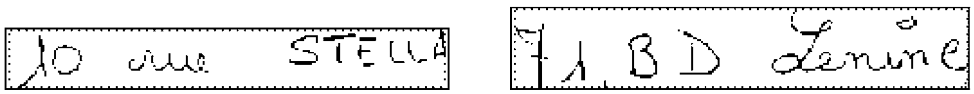

Fig. 2. Variability of the writing type in street names.

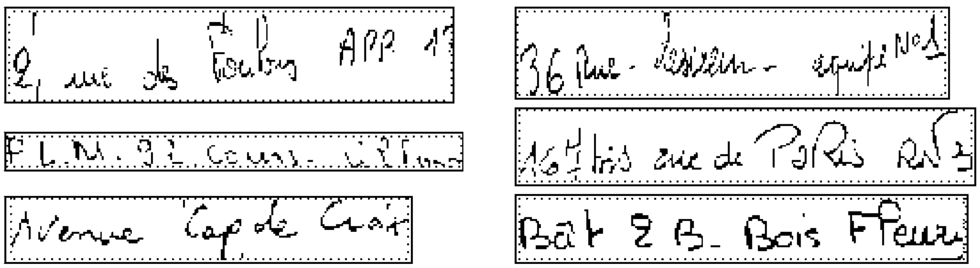

Fig. 3. Structure variability in street names.

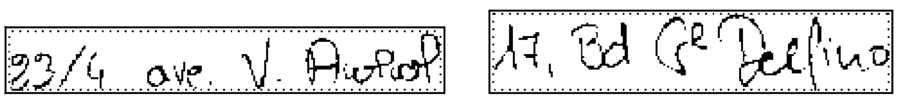

Fig. 4. Variant phenomena in street names.

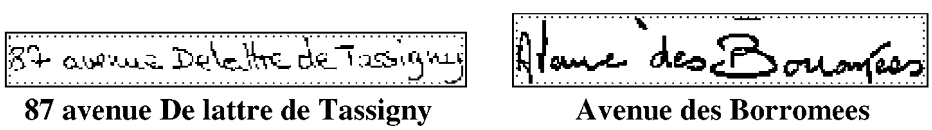

Fig. 5. Difficulty of street line segmentation into words.

- The lack of a standard structure for street lines. A simple hypothesis would be to assume that a street line is composed of a street number followed by a street name (Fig. 2). Unfortunately, this configuration, albeit being the most frequent, is not standard (Fig. 3). Sometimes the street number is absent or is written after the street name. Furthermore, some extra information, such as a comma, the apartment number, the P.O. box, etc. may surround the street name, either on the left or on the right side. This structure variability explains the need of the supplementary task of locating the street name in addition to recognizing it.
- The phenomenon of variants or abbreviations (Fig. 4). The variants are numerous and depend on each grammatical word category. As a matter of fact, the number of variants at the street name level grows exponentially with respect to the number of words.
- Street line segmentation: Segmenting a street line into its component words without resorting to recognition is a hard task. The main reason for this difficulty is that common used criteria for word segmentation, such as space sizes, punctuation marks, presence of uppercase letters, etc. [19], [28], turn out to not be robust enough to face all the variability encountered (Fig. 5).

On the other hand, context is rich for street names since they generally consist of a relatively large number of words. Thus, recognition could be facilitated if this context is used the right way, for instance, if the street line is correctly preprocessed[1] and if the variants are correctly taken into account, so as to avoid a mismatch in recognition.

In light of these points, our objective is to design a recognition system having, among others, two desirable capabilities:

- The system has to take full advantage of the rich context of street names. By correctly exploiting the available redundancy, recognition could be much easier.

1. For instance, if a street line with a highly slanted baseline is not correctly normalized, some words may be wrongly considered as components belonging to neighboring lines.

- The system has to avoid the a priori segmentation of street lines into words in order to prevent the occurrence of premature errors.

In this prospect, we propose, in this paper, an original approach, which allows conjointly locating and recognizing a street name within a handwritten street line. In our system, each street name is viewed as a *keyword* to be detected in the street line. The two fields of handwritten information possibly surrounding the street name are viewed as *nonkeywords* and are not modeled explicitly.[2] The designed system privileges soft decisions over hard ones to a large extent and permits an interaction between various sources of knowledge—character level, word level, phrase level, variant generation, street name information, street name extra-information—which simultaneously participate to the elaboration of the system decision. At the handwriting signal level, hidden Markov models are extensively used to provide the needed matching scores. Several optimization techniques are used to reduce the street name lexicon size, so as to speed up the processing time. Experiments carried out on large data sets of street line images, automatically extracted from real French mail envelope images, show very promising results. The technique described in this paper consists of several components, one of which is an already developed system, designed for city name recognition [12]. This is the reason why here we will emphasize mainly on the novelty aspects of our approach.

This paper is divided into seven sections in addition to the present one. Section 2 sketches previous work regarding street name recognition and other concepts related to the approach we present. Section 3 gives an overview of the proposed approach. Section 4 presents a theoretical formulation of the street name recognition problem and Section 5 describes the various components of the recognition system. Section 6 is devoted to the reduction of the street name lexicon. Section 7 details the experiments carried out to validate the approach. Finally, Section 8 draws some conclusions and sketches future directions of our work.

2 Previous Work and Motivation

2.1 Street Name Recognition

Few studies have been conducted for street name recognition. Srihari et al. [29] use the latter to determine the delivery point code for US handwritten addresses. After locating and recognizing the street number, a dictionary of possible street names, represented each by the expected sequence of words, is generated. The matching between each dictionary candidate and the unknown street name image, considered as everything at the right of the recognized street number, is carried out by combining an Hypothesis Generate and Reduce paradigm and an HMM recognizer. This approach does not take into account variations in suffixes, which are frequent in street names (`street-str-st, avenue-ave-av,` etc.). Chen et al. [7] directly applied their HMM-based word recognition system to street name recognition. An adaptation of the system consisted of expanding the lexicon by duplicating the prefix and/or suffix to take into account abbreviations. More recently, Kim and Govindaraju have presented a different approach in [19]. First, a neural network is used to segment the street name into its constituent words. The inputs to the neural net correspond to features extracted from the bounding boxes of presegmented characters. After word segmentation, a lexicon is generated so that it consists only of the main parts of street names (key names). A dynamic programming-based word recognizer performs the matching between the lexicon candidates and the input image. The reason why only the main part of the street name was considered for matching is to avoid facing the several variations of suffixes and the presence of extraneous words appearing at the end of street names. This approach permits overcoming these problems, but has two limitations: first, the recognition module cannot correct premature word segmentation errors and, second, basing recognition on key names only does not allow for taking full advantage of the rich context available in whole street names.

It is worth observing that, in the approaches mentioned above, street name recognition takes place only if the preceding recognition of the street number is reliable. In this case, using postal directory look-up, the street number and the 5-digit zip code, a very restricted lexicon of possible street names is created. In our approach, as we do not make the recognition of the street name depend on the street number information, the lexicon turns out to be of a much higher size since its creation relies only on the postal code information. This offers the possibility of combining, at a later stage, the hypotheses provided by street name recognition with those provided by a street number recognizer in an independent way.

2.2 HMMs

Hidden Markov Models (HMMs) [27] are finite stochastic processes that have been proven to be one of the most powerful tools for modeling speech and later on a wide variety of other real-world signals. HMMs are today widely used for text recognition in various contexts: word recognition, whether online [2], [17] or offline [3], [7], [8], [15], [16], machine printed [23], [32], and handwritten [4], [26] character recognition. These probabilistic models offer many desirable properties for modeling character or word strings. One of the most important is the existence of efficient algorithms to automatically train the models without any need of labeling presegmented data. As will be seen in the remainder of this paper, this constitutes a key feature of the approach we developed for street name recognition.

2.3 Keyword Spotting

In document processing, most of the previous work on the concept of keyword and nonkeyword models has been restricted so far to spotting keywords embedded in printed documents. In [21], an HMM (bidimensional) is built for the expected keyword and another HMM is considered for all extraneous (nonkeyword) words. After the presegmentation of the printed text into words, each input word is matched against the keyword and nonkeyword models and classified into one of these two categories according to the

2. The recognition of the street number is not addressed in this paper.

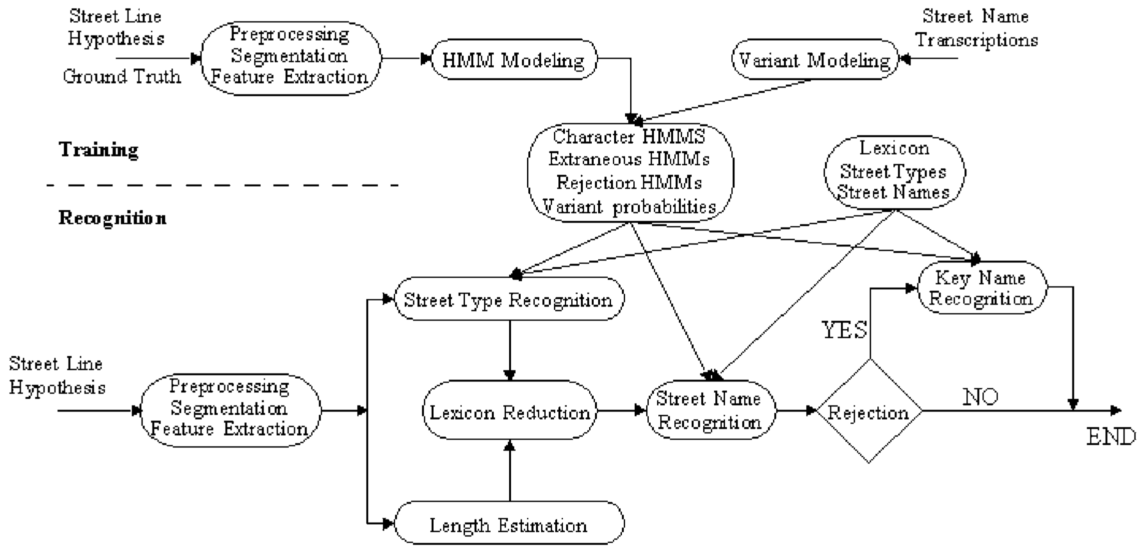

Fig. 6. Schematic diagram of the street name recognition system.

model providing the highest matching score. In [6], a similar approach, based on HMMs, is described for detecting and locating specified keywords in images of printed text. Here, no presegmentation of the text into words is performed. A spotting network, in which keyword and nonkeyword models are connected in parallel by a space model, is constructed. This network takes as input the whole text and a keyword is detected if the optimal path in the network lies through the keyword model.

Even though our approach also uses the concept of keyword and nonkeyword models, it is different in at least two fundamental points from the approaches mentioned above. First, we consider, in our work, unconstrained handwritten word strings instead of printed word strings. Hence, recognition does not start from presegmented words but rather determines the frontiers of words as a by-product of recognition itself. Second, here, as we will see in the next sections, keyword and nonkeyword models are not competing against each other. Rather, their scores simultaneously participate to provide the matching score for each lexicon candidate (street name).

3 Overview of the Street Name Recognition System

Basically, the street name recognition system consists of two phases: training and recognition (see Fig. 6 for a schematic diagram). In the first phase, each street line image of the training set is preprocessed and presegmented into graphemes and transformed into a sequence of observations. The output of the training module is a set of models associated with three categories: a set of elementary letter HMMs, and two other models, modeling extraneous handwriting information possibly surrounding the street name on the left and on the right side, respectively. These two models, which we call left extraneous (LE) and right extraneous (RE) models are represented by two ergodic (fully connected) HMMs. The use of ergodic models is motivated by the fact that we do not attempt to explicitly recognize these extraneous data, but merely to detect their presence and locate them.

In recognition, the unknown street line is automatically located and extracted from the envelope image and then passed through the same steps of preprocessing, segmentation, and feature extraction. The obtained observation sequence is then fed to the recognition module, along with a dynamic lexicon of street names corresponding to the postal code already identified on the envelope image. This module consists of a network made up of street name models aligned in parallel and connected each via simple space HMMs to the LE model on the left side and to the RE model on the right side. In this network, each street name, the model of which is built by an appropriate concatenation of letter HMMs, is viewed as a *keyword* to be detected in the street line, while the information possibly surrounding the street name is viewed as *nonkeywords* and is dealt with by extraneous models (LE and RE). The keyword HMM, whose combination with the two nonkeyword HMMs provides the best matching with the unknown image, gives the identity of the recognized street name. As a by-product of recognition, the location of the street name within the street line can also be recovered.

The lexicon mentioned above does not consist only of the standard form of street names, but also of their variants. As the standard lexicon in France has an average size of 1,000 and as the number of variants grows exponentially with respect to the number of words in the standard street name, the dynamic lexicon actually consists of thousands of entries in most cases, making recognition intractable. To solve this problem, we implemented two efficient methods to reduce the lexicon size. The first is carried out by recognizing the street type and filtering the lexicon accordingly. The second computes an a posteriori probability for each street name candidate, by combining the probability of its variant form given the standard form with the probability of the length in graphemes of the unknown street name image. The street name in question is then removed from the lexicon if its a posteriori probability is below a fixed threshold. Our system also contains an HMM-based rejection module, which aims at detecting street line mislocation cases and rejecting street names recognized with low confidence or not belonging to the lexicon.

The strategy of conjoined location and recognition of street names summarized above works well in general since it is based on a solid probabilistic framework. However, it may lack optimality when dealing with some few street lines having configurations not taken into account by the variant generation module. To properly deal with such configurations, we added a new module that is activated each time a street line is rejected. This module performs a similar recognition strategy to the previous one; the only difference is that keywords now correspond to *key names* only instead of whole street names. With this modification, the left extra-information now contains anything at the left of the key name, such as the street number, the street type, articles, first names, etc. The LE HMM is of course trained on similar data extracted from street line images. The RE HMM remains the same since no change is considered here. The rationale behind this new module is to recover rejection cases caused by the presence of street name variants, which are not modeled by the system due to their low a priori probability. In this case, the new recognition relies only on key names, which are not generally subject to variation (abbreviation, etc.), hence avoiding the mismatch occurring in the previous recognition step.

4 Theoretic Formulation of the Street Name Recognition Problem

Owing to the fact that postal code recognition and street line location are prone to errors, street name detection and recognition involve dealing with not only street lines, in which the street name is guaranteed to belong to the lexicon, but also other kinds of lines. To correctly address this problem, all configurations the system may face, when taking as input an unknown line, must be taken into account:

Configuration 1 (***cfg*^1^**). The line extracted from the envelope image corresponds to a street line and the street name it contains belongs to the dynamic lexicon. This is the normal and expected configuration, hence:

$$\Pr(O \Big| cfg^1) = \sum_{sn} \Pr(O|sn) \cdot \Pr(sn), \tag{1}$$

where the summation is over all street names sn belonging to the dynamic lexicon. Assuming that the likeliest street name has a matching score much larger than the remaining scores, this summation can be replaced by a maximization over sn:

$$\Pr(O \big| cfg^1) = \max_{sn} \Pr(O|sn) \cdot \Pr(sn). \tag{2}$$

As explained earlier, we assume that the sequence O results, in this case, from the concatenation of three subsequences:

$$O = O^l + O^{sn} + O^r, \tag{3}$$

where + stands for the string concatenation operator, O^{sn} is the street name observation subsequence and O^l and O^r are the observation subsequences associated, respectively, with the left and with the right side of the street name. These sides are modeled by two ergodic HMMs, hmm_{SN}^{LE} and hmm_{SN}^{RE}, while the street name is modeled by one composite HMM, hmm^{sn}. Assuming that the three fields are mutually independent, $Pr(O|sn)$ can be expressed as:

$$\begin{aligned}\Pr(O \Big| sn) = \sum_{O^l, O^{sn}, O^r; (O^l + O^{sn} + O^r = O)} & \Pr(O^l \big| hmm_{SN}^{LE}) \\ & \cdot \Pr(O^{sn} \big| hmm^{sn}) \cdot \Pr(O^r \big| hmm_{SN}^{RE}),\end{aligned} \tag{4}$$

where the summation is over all possible segmentations of O into three subsequences. Generally, the matching score associated with the optimal segmentation is much larger than the remaining ones; therefore, this summation can be replaced by a maximization:

$$\begin{aligned}\Pr(O \Big| sn) = \max_{O^l, O^{sn}, O^r; (O^l + O^{sn} + O^r = O)} & \Pr(O^l \big| hmm_{SN}^{LE}) \\ & \cdot \Pr(O^{sn} \big| hmm^{sn}) \cdot \Pr(O^r \big| hmm_{SN}^{RE}).\end{aligned} \tag{5}$$

Equation (5) is the basis for computing the score associated with each street name sn. Substituting, in (2), $\Pr(O|sn)$ by the second term of (5) leads to:

$$\begin{aligned}\Pr(O \Big| cfg^1) = \max_{sn} \max_{O^l, O^{sn}, O^r; O^l + O^{sn} + O^r = O} & \Pr(O^l \big| hmm_{SN}^{LE}) \\ & \cdot \Pr(O^{sn} \big| hmm^{sn}) \cdot \Pr(O^r \big| hmm_{SN}^{RE}) \cdot \Pr(sn).\end{aligned} \tag{6}$$

Due to the huge size of street name vocabularies, it becomes hard in practice to collect sufficient training samples to reliably estimate the a priori probability of street names, $\Pr(sn)$, within a street line. This is the reason why we consider all street names, for a given postal code, as equiprobable:

Configuration 2 (***cfg*^2^**). The extracted line does not correspond to a street line (mislocation). Such lines ($\overline{SL}$) are assumed to be produced by an ergodic HMM, $hmm^{\overline{SL}}$, trained on a set of line images extracted from envelopes, but which are different from street lines:

$$\Pr(O \Big| cfg^2) = \Pr(O \big| hmm^{\overline{SL}}). \tag{7}$$

Configuration 3 (***cfg*^3^**). The extracted line does correspond to a street line, but the associated street name does not belong to the lexicon. This may occur when the postal code does not match the street name because of a postal code misrecognition or because of a writer error. We model such lines by the street type (instead of street name) detector and recognizer.

$$\begin{aligned}\Pr(O \Big| cfg^3) = \max_{st \in ST} \max_{O^l, O^{st}, O^r; O^l + O^{st} + O^r = O} & \Pr(O^l \big| hmm_{ST}^{LE}) \\ & \cdot \Pr(O^{st} \big| hmm^{st}) \cdot \Pr(O^r \big| hmm_{ST}^{RE}) \cdot \Pr(st),\end{aligned} \tag{8}$$

where hmm_{ST}^{LE} is identical to hmm_{SN}^{LE} and hmm_{ST}^{RE} is an ergodic HMM modeling anything on the right side of the

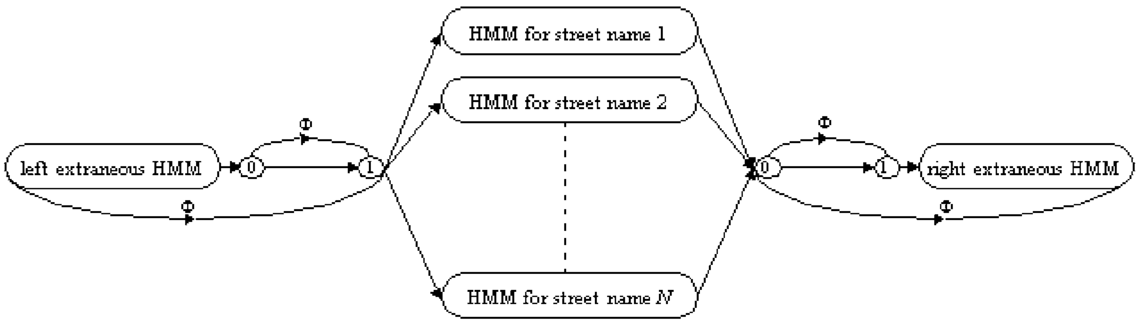

Fig. 7. Global HMM network for street name recognition.

street type. $\Pr(st)$, the a priori probability of street type st, can be easily estimated on the training set.

Now, the problem of street name recognition can be formulated by applying the standard statistical approach to pattern recognition, which consists of maximizing the a posteriori probability $\Pr(cfg^i|O)$ to find the optimal configuration of the unknown line. Using the Bayes' rule, this amounts to maximizing:

$$\begin{aligned}\Pr(cfg^i\Big|O) &= \frac{\Pr(O\big|cfg^i)\cdot\Pr(cfg^i)}{\Pr(O)} \\ &= \frac{\Pr(O\big|cfg^i)\cdot\Pr(cfg^i)}{\sum_{i=1}^{3}\Pr(O\big|cfg^i)\cdot\Pr(cfg^i)},\end{aligned} \tag{9}$$

where $\Pr(cfg^i)$ is the a priori probability of configuration i, estimated by its empirical frequency on the training set. System decision is then set using the following rule:

set

$P_{\max} = \max_i \ \Pr(cfg^i|O)$

$c = \arg\max_i \ \Pr(cfg^i|O)$

$\Delta =$ global decision threshold

if $c = 1$ (first configuration) **then** {
 if $(P_{max} > \Delta)$ **then**
  the street name with the highest matching score is accepted
 else
  the street name is rejected
}
else
 the street line is directly rejected.

This decision rule is applied for both cases when the keyword is considered either as the whole street name or only the key name (when the previous step of whole street name recognition yields a rejection decision by the system).

5 Description of the System Components

To conjointly locate and recognize street names within street lines, the system takes as input an automatically extracted street line image and a dynamic lexicon of possible street names derived from the recognized postal code/city name. For each candidate, a global model (Fig. 7) is built by connecting, via space HMMs, the LE HMM, the street name HMM, and the RE HMM. Adopting a model discriminant recognition strategy and Viterbi as the matching algorithm, the optimal street name is recovered. The latter is the one whose corresponding HMM model, when connected to the LE and RE models, provides the highest probability of generating the street line observation sequence. As a by-product of the Viterbi alignment, the location of the street name within the street line can also be retrieved. In the following sections, we give a more in-depth description of the various stages composing our global recognition system.

5.1 Street Line Location

While street line location is manually done for training purposes, we need, in recognition, an automatic technique to perform this operation. To this end, we use the technique developed in [25], which takes as input an envelope binary image and provides as output a list of address bloc lines. This technique uses, first, a bottom-up data driven strategy, with the aim at merging the connected components on the address image into words, lines, and address blocs successively. A filtering phase is applied to retain, among the address bloc hypotheses, only the most plausible one. Then, the technique uses a top-down strategy, which carries out, when necessary, merging and/or splitting operations on the lines of the retained address bloc. Finally, we consider as a single street line hypothesis the line located immediately above the postal code/city name line (the latter being already recognized in a previous stage).

5.2 Preprocessing, Segmentation, and Feature Extraction

Given the global character of our approach, each street line is transformed as a whole into an observation sequence by the successive application of preprocessing, segmentation, and feature extraction. A detailed description of these stages is given in [9], [12]. Preprocessing consists of normalizing the lower and upper baselines, correcting the average character slant, and smoothing the normalized image. The segmentation algorithm uses the upper contour minima and some heuristics to split the word string image into a sequence of *segments* or *graphemes*, each of which consists of a correctly segmented, an undersegmented, or an oversegmented letter. The discussion on the rationale behind using a grapheme-based segmentation is out of the scope of this paper, but is addressed in [5], [12], [14] for

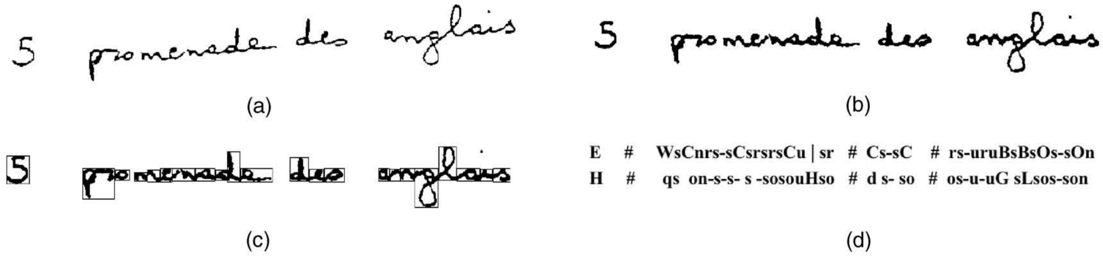

Fig. 8. (a) Input image, (b) preprocessing, (c) segmentation, and (d) observation sequence.

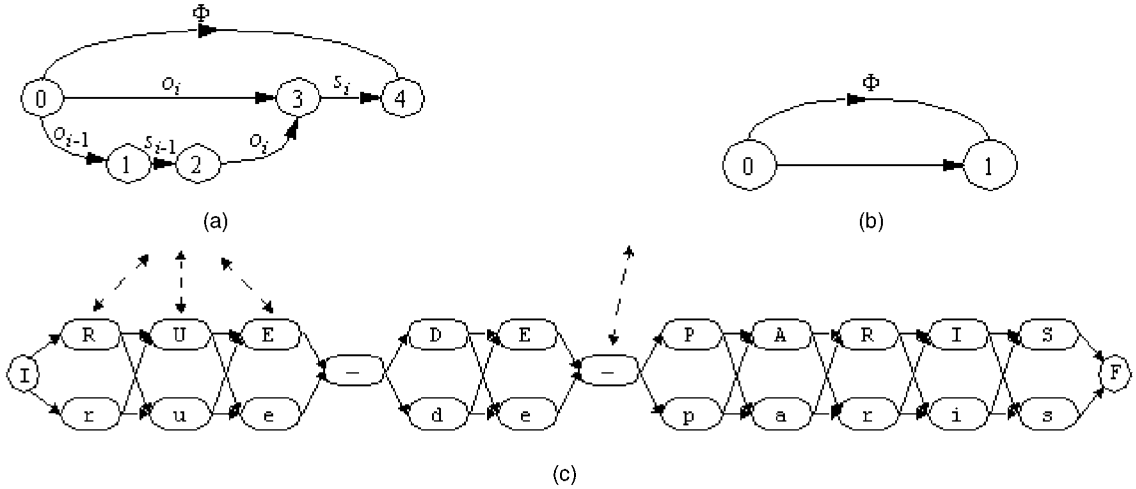

Fig. 9. (a) Elementary letter HMM. (b) Space HMM. (c) Street name HMM for recognition.

instance. The sequence of graphemes is transformed into a string of observations formed by pairs of symbols taken in two different sets. The first set, targeted to cursive handwriting, is based on global features such as loops, ascenders, and descenders, while the second one, targeted to handprinted handwriting, is based on the analysis of the bidimensional contour transition histogram of each segment in horizontal and vertical directions. From each of these two sets, a discrete alphabet is derived by considering all possible and relevant combinations of the basic features within a grapheme. An objective evaluation of these two feature sets based on entropy and perplexity is described in [11]. We also consider segmentation features whose aim is to characterize physical segmentation points or spaces between letters (intergrapheme segments). Given that the two sets of shape-features are independently extracted from the image, a word string is represented by two observation sequences of equal length, each consisting of an alternating sequence of shape symbols and segmentation symbols. Observations of odd rank (starting at rank 1) represent graphemes, whereas observations of even rank represent intergrapheme segments. Fig. 8 shows the transformation steps of a street line image into a pair of discrete observation (symbol) sequences.

5.3 Markovian Modeling of Street Lines

As discussed earlier, we assume that an observation sequence extracted from a street line image results from the concatenation of three subsequences associated with the street name and the left and right extra-information parts possibly surrounding the street name.

5.3.1 Street Name Modeling

A street name usually consists of a word string (street type, articles, first names, titles, key name). Due to the huge size of the vocabulary of street names, our Markovian modeling is based on the character level. The street name model is similar to the model we already developed for city names [10], [12] and is created by appropriately concatenating elementary letter HMMs and by considering the interword spaces as special characters, modeled by simple HMMs. All letters are modeled in the same way (same architecture and number of states). The architecture of the letter HMMs (Fig. 9a), emitting observations along transitions and allowing null transitions (represented by transitions emitting the null symbol Φ), has been chosen to model insertion, deletion, and substitution errors, in order to account for the imperfections of the segmentation and feature extraction process [1]. In Fig. 9a, o_i represents a grapheme observation of rank i while s_i represents an intergrapheme observation of rank i. Null Transition t_{04} models deletion corresponding to undersegmented letters. The path along transitions t_{03} and t_{34} models substitution errors corresponding to correctly segmented letters. Last, the path along transitions t_{01}, t_{12}, t_{23}, and t_{34} models letter insertion (oversegmentation). We must add that transitions associated with shapes actually emit two discrete symbols encoding the associated grapheme, as described in Section 5.2. The details of the Baum-Welch and Viterbi algorithms for this particular HMM modeling are given in [13]. Space HMMs (Fig. 9b) consist of two transitions modeling, respectively, the presence or the absence (null symbol Φ) of a space between two contiguous words. In recognition, each letter model actually consists of two parallel letter HMMs modeling the uppercase and lowercase alternatives. Fig. 9c shows the composite HMM corresponding to street name `rue de Paris`.

5.3.2 Modeling Left and Right Extra-Information in Street Lines

To conjointly locate and recognize a street name without its a priori segmentation, the extra-information surrounding it has to be modeled. The left extra-information usually corresponds to a street number possibly followed by a comma. Sometimes, however, additional information could be present (Fig. 3). To model the left extra-information, an *ergodic* (fully connected) HMM with single initial and final states is built. A null transition connecting the initial state of the LE model and the final state of the following space model is added to model the situation when no extra-information precedes the street name. The objective of this model is not to explicitly recognize the extra-information, but merely to detect and locate it. We have found that an ergodic model is more flexible in modeling this kind of data and also is more powerful in taking into account the context than a *trigram* model which we considered first for this purpose [10].

Since we intend to model the right extra-information with the same goal in mind, it is not surprising that we assume that these data are generated by a similar right extraneous (RE) ergodic HMM. The major difference with the LE model lies in the associated null transition which has here a much higher probability since the a priori probability that extra-information follows a street name is very low.

To train the LE and RE HMMs, only the street name transcription along with the whole observation sequence extracted from the street line are needed. First, an embedded HMM is built by connecting, via space HMMs, the LE HMM, the street name HMM, and the RE HMM. The Baum-Welch algorithm then conjointly optimizes (in the maximum likelihood sense) the parameters of these three HMMs and the segmentation of each training street line image into left extra-information, street name, and right extra-information subimages. Since we are considering character HMMs to build street name models, the training set for the latter needs not be composed necessarily of street name images, but of any data collected in the same context and conditions. In our approach, we directly use the character HMMs obtained in the system described in [12] and trained on city name images and initialize, without re-estimation, the parameters of character HMMs of the street name model by those used in our city name recognition system. This means that the street name model acts as a segmentation engine of the street line into the three kinds of information described above.

5.3.3 Variant Generation

Compared to city names in handwritten addresses, street names are far more subject to the variant phenomenon. The reason explaining this difference is twofold. First, street names usually consist of a larger number of words. Second, these words do not bear the same semantic (grammatical) meaning. For instance, key names, which are commonly considered as the main identifiers of street names, are generally not subject to variations. Conversely, street types, articles, first names, or titles bear less information and their abbreviation, omission, or insertion is widely used in postal addresses as well as in other contexts. Another peculiarity of the variant phenomenon is that the number and the type of variants are closely related to the word grammatical category. Therefore, in our approach, the variant corresponding to each word category is properly dealt with:

- Street type variants: Street types are often abbreviated or substituted by other street types. Typically, numerous variants are observed for each street type. For practical reasons, however, only those with significant frequency are retained.
- Article variants: Articles such as `DE`, `DU`, `DES`, `LA`, etc. are sometimes omitted (e.g., `RUE` **DU** `CAPITAL GALINAT` becoming `RUE CAPITAL GALINAT`) or inserted (e.g., `RUE PROFESSEUR ROLLET` becoming `RUE` **DU** `PROFESSEUR ROLLET`).
- First name variants: Variants involving first names are numerous. However, the most common consist of abbreviating a first name by its initial or completely omitting it (e.g., `RUE` **ANDRE** `MAGINOT` becoming `RUE` **A** `MAGINOT` or `RUE MAGINOT`, respectively).
- Title variants: The variants involving titles are similar to first name variants. The main difference is that their most observed abbreviation is not the use of the initial (e.g., `DOCTEUR ==> DR`; `PRESIDENT ==> PDT`, `PT`; `MARECHAL ==> MAL`).

Since the vocabulary of first names and titles is open, it is almost impossible to list and store all the corresponding variants. To overcome this problem, we established variant generation rules which take into account only the grammatical category of these entities. The rules retained are given in Table 1.

TABLE 1
Rules for Generating First Name and Title Variants

	FIRST NAME	TITLE
1st Variant	Φ (omission)	Φ (omission)
2nd Variant	*F* (first letter)	*TE* (first and last letters)

Although the variant *TITLE* ==> *TE* is the most common (`DOCTEUR==>DR`, `SAINT==>ST`), it does not cover all the cases (e.g., `MARECHAL==>MAL`, `SAINTE==>STE`). This lack of optimality, however, is balanced by the fact that we get in this way a unique rule to generate all the variants independently of the lexical information of each title. The rule, which generates variants, is applied offline, so that the dictionary of street names used during recognition includes all the variants in an extensive way. The rationale for using rules rather than full forms is that the number of variants may be large and that most of the variants are formed in a regular way. The number of variants, generated in this way, gets significantly reduced, thus allowing a more robust estimation of the associated probabilities, given our limited amount of variant training data. Besides, the imperfections resulting from using a unique rule are dealt with to a certain extent by our HMM modeling, the structure of which is well suited to the phenomena of character substitution, deletion, and insertion. Finally, all variants that do not result from a regular production rule (e.g., `RUE JEAN-JACQUES` for `RUE JEAN-JACQUES ROUSSEAU`) are lexicalized.

To estimate the probability of the variant associated with each grammatical word category, we use a lexicon that assigns to each street name field its grammatical category (street type: *st*, article: *ar*, first name: *fn*, title: *tt*, and key name: *kn*). A training set consisting of a set of pairs (standard street name, variant actually observed on the envelope[3]) allows us to estimate the probabilities of the variants of each field by their frequencies. In the recognition phase, we generate, for each street name (in standard form) belonging to the lexicon, all its variants by considering all possible combinations of the variants of its constituent fields. The number of these combinations varies exponentially with respect to the number of fields in the street name. Assuming, for simplification, that the variants corresponding to different fields are independent, the probability of a variant at the street name level is given by:

$$\Pr(sn^j \Big| sn^{std}) = \prod_i \Pr(sn_i^j \big| sn_i^{std}), \tag{10}$$

where sn^{std} is the standard street name form, sn^j is the *j*th variant of sn^{std}, and sn_i^j is the elementary variant associated with sn_i^{std}, the *i*th field of sn^{std}.

6 Lexicon Reduction

Although dynamically derived from the postal code identity, street name vocabularies consist of a large size with an average around 1,000 candidates. Moreover, the addition of variants typically increases the vocabulary size of one order of magnitude. In address recognition applications, images must be processed at a high throughput of typically 20 images/second and within a range of time of typically a second. These constraints are directly related, respectively, to the throughput of address sorting machines and to the maximal time available to take a decision to divert the letter to the appropriate sorting bin. This explains why lexicons of size 1,000 appear to be large although much larger lexicons of size typically of the order of 60,000 are dealt with in speech recognition applications. In addition, contrary to speech recognition where syntactic and semantic constraints help reduce the language perplexity to a few hundreds, street name recognition has a perplexity of the same order of magnitude as the size of the lexicon. To speed up the recognition process time, we implemented two lexicon reduction strategies to reduce the street name vocabulary size: street type recognition and estimation of the street name length.

3. Note that, here, training images are not required but only the associated ASCII transcriptions.

6.1 Street Type Recognition

The first strategy initially recognizes the street type and filters the street name lexicon accordingly. Such a strategy is motivated by two reasons:

- The street type lexicon is relatively small and consists of 50 entries at most, including common variants.
- The intrinsic difficulty of the street type recognition task can be evaluated by the perplexity measure P, which is given by:

$$P = 2^H = 2^{-\sum_i p_i \cdot \log p_i}, \tag{11}$$

  where p_i is the a priori probability of street type i and H is the entropy of the source assumed to produce street types. Using the estimated probabilities, the perplexity was found to be equal to 5.6. This result can be interpreted by saying that recognizing a street type image actually involves the competition between 5.6 equiprobable street type candidates in average instead of 50. Hence, this task is relatively easy, justifying its prior use to the recognition of street names.

Street types are recognized in the same way whole street names are. Therefore, the LE HMM remains the same, while the RE HMM now models anything on the right side of the street type and is trained accordingly on appropriate street line subimages. However, no null transition is added to this model since a street type is always followed by at least a key name. Finally, to avoid premature errors, the constraint of first recognizing the street type is relaxed by retaining the three best recognized street types to filter the lexicon if their scores are significant.

6.2 Estimation of the Street Name Length

Lexicon reduction based on the length of the entity to be recognized is a widely used technique in handwriting recognition [20], [22]. In [22], it was used in the context of street name verification. It consists of retaining, during recognition, only those lexicon entries having an expected length close to that of the unknown entity. Estimating the expected length of a word string requires sufficient training samples of this string. Unfortunately, in large vocabulary applications, such training data are not available. To overcome this problem, we adopted a probabilistic scheme, which estimates the expected length (in graphemes) of a street name from the expected lengths of its constituent characters. This is done in the following way. Let

$$sn = c_1 c_2 \dots c_n \tag{12}$$

be a street name composed of characters c_i. Due to the imperfection of the segmentation process, each c_i is converted into a number of graphemes ranging from 0 (undersegmentation) to 3, which is supposed to be the maximum number of segments per character. Thus, we consider l_i, the length in graphemes of c_i as a random variable. The length of the whole street name sn is given by:

$$L = l_1 + l_2 + \dots + l_n. \tag{13}$$

TABLE 2
Possible Configurations in the Decision Process

TYPE OF INPUT IMAGES	REJECTION	ERROR	CORRECT RECOGNITION
Invalid	1	2	3
Valid	4	5	6

If all lengths l_i have the same mean μ and variance σ^2,

$$Y = \frac{L - n\mu}{\sigma\sqrt{n}} \tag{14}$$

tends, as n becomes large, to a standard normal *pdf* (central limit theorem). It is clear that these two conditions are not strictly valid here. However, this assumption is reasonable since n is relatively large (street names usually consist of several words) and the means and variances of the lengths of characters are of the same order of magnitude. As the letters l_i do not have the same mean and variance, we can replace

$$n\mu \text{ by } \sum_{\mathrm{i}=1}^{\mathrm{n}} \mu_{\mathrm{i}} \text{ and } \sigma^2 \text{ by } \sum_{i=1}^{n} \sigma_i^2, \tag{15}$$

where μ_i and σ_i are the mean and the variance of letter l_i. In this way, for each street name sn, we dynamically derive a probability function, depending on the transcription of sn, which estimates the probability that sn produces an observation sequence of any length L. The parameters μ_i and σ_i are easily estimated using the trained character HMMs, thanks to their particularity in our approach of modeling character deletion, insertion, and substitution. Now, to filter the lexicon, two inputs are considered:

- The street line observation sequence from which we derive the length L of the unknown street name after the street type is recognized.
- The dynamic lexicon of street names which is appropriately prefiltered using the result of street type recognition. This lexicon is expanded by generating, for each street name in standard form, its most common variants with associated probabilities.

For each entry sn (standard or variant form), we compute the a posteriori probability of sn given the length of the street name in the input image, $\Pr(sn|L)$. By the Bayes' rule:

$$\Pr(sn\Big|L) = \frac{\Pr(L\big|sn) \cdot \Pr(sn)}{\Pr(L)}. \tag{16}$$

As $\Pr(L)$ does not depend on sn,

$$\begin{aligned}\Pr(sn\big|L) \propto P(L, sn) &= \Pr(L\big|sn) \cdot \Pr(sn) \\ &= \Pr(L\big|sn) \cdot \Pr(sn\big|sn^{std}) \cdot \Pr(sn^{std}),\end{aligned} \tag{17}$$

where, applying the central limit theorem,

$$\Pr(L\big|sn) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{\left(L - \sum_i \mu_i\right)^2}{2\sum_i \sigma_i^2}\right) \tag{18}$$

and $\Pr(sn|sn^{std})$, the probability that sn^{std} generates variant sn is given by (10). Here, we have assumed that a variant sn may be the variant of only one standard form, which is almost always the case in practice. $\Pr(sn^{std})$, the a priori probability of sn^{std} is dropped due to the lack of sufficient training data.

To retain a street name sn in the lexicon, we require that $P(L, sn) > P_0$, where P_0 is a system threshold depending on the desired rate of lexicon reduction (typically, $P_0 = 10^{-3}$, 10^{-4}, etc.). This strategy has the advantage of using only one control parameter (the threshold P_0) and has been found to be very efficient in reducing the street name lexicon size.

7 Experiments

This section is devoted to the experiments carried out to evaluate our approach. We begin, first, by describing the databases used and by defining the evaluation process. After presenting the results obtained, we analyze, with some examples, the behavior of the recognition system. Finally, we give an analysis of rejection and error cases.

7.1 Databases and Performance Evaluation

Our street name recognition system was tested on two databases corresponding to real French mail envelopes. These sets consist of 1,287 and 2,856 envelope images corresponding to the cities of `ARRAS` and `TOULOUSE`, respectively, and were collected following the departmental sorting mode. In this mode, 95 percent of the envelopes have the expected postal code (here either the `ARRAS` or the `TOULOUSE` postal code) as the destination point, while the remaining 5 percent correspond to preparation errors and, thus, cover randomly the other cities of the country. It is important to note that each of the two sets described above contains a nonnegligible quantity of invalid addresses. An address is considered to be valid if five human operators (five out of any number) from `La Poste` (`French Post Office`) can, without ambiguity, visually determine its destination point (postal code, city name, street name). The existence of valid and invalid envelopes gives rise to various decision configurations street name recognition could be involved with. These configurations are given in Table 2.

Referring to Table 2, we define the recognition rate RR and the error rate ER as follows:

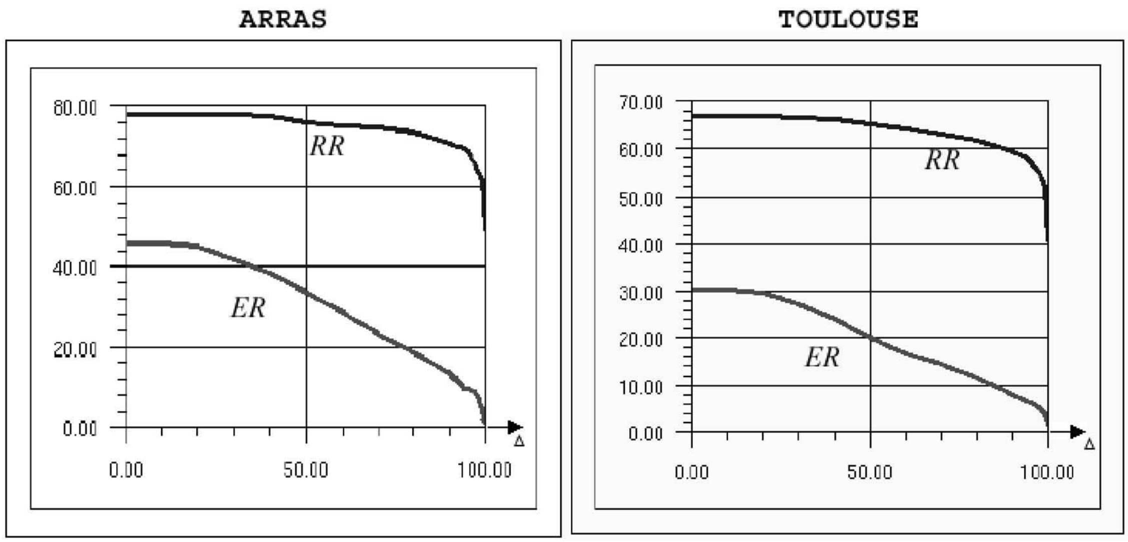

Fig. 10. Recognition Rates (RR) and Error Rates (ER) achieved by the system on the ARRAS and TOULOUSE databases.

$$RR = \frac{6}{4+5+6} \text{ and } ER = \frac{2+5}{2+5+6}, \qquad (19)$$

the rejection rate being the complement of their sum and equal to $1 - (RR + ER)$. It is important to note that the error rate is computed as the percentage of misinterpreted images among the accepted ones only (excluding rejection cases).

7.2 Training Phase

Our modeling of street lines involved considering various ergodic HMMs to represent the rejection, left and right extra-information models (see Section 4). The training of these models was carried out on a training set of approximately 2,000 street line images, using the cross-validation principle. Initially, the training set is split into two sets: the first is used to actually train the model parameters and the other is used for validation purposes. The optimal number of states is then determined as the one yielding the highest likelihood of generating the validation set. The use of the latter allows the trained models to acquire better generalization capabilities over unseen data. Using this strategy, the number of states for the left ergodic HMM, the right ergodic HMM used in street type recognition, the right ergodic HMM used in street name recognition, and the rejection ergodic model were determined:

$$N(hmm^{LE}_{SN}) = N(hmm^{LE}_{ST}) = 3$$
$$N(hmm^{RE}_{ST}) = 16$$
$$N(hmm^{RE}_{SN}) = 10$$
$$N(hmm^{\overline{SL}}) = 20.$$

As pointed out earlier, the character HMMs used to build up street type or whole street name models were those obtained by training the city name recognition system described in [12]. The a priori probabilities related to the possible street line configurations discussed in Section 4 were estimated by their frequencies on our training sets:

$$\Pr(cfg^1) = 0.8 \;\; \Pr(cfg^2) = 0.15 \;\; \Pr(cfg^3) = 0.05.$$

7.3 Results

Once all the parameters required by the decision system were appropriately and automatically set, we ran our recognition module to compute the recognition and error rates as defined by (19). Fig. 10 shows the recognition and the error rates as a function of the global decision threshold $\Delta(\times 100\%)$ $(0 \leq \Delta < 1)$, obtained on the ARRAS and TOULOUSE test databases. The value $\Delta = 0$ corresponds to the case when rejection is considered only when the highest matching score does not correspond to a street name (case $c \# 1$, see Section 4).

In postal applications, error rates are required to be low enough for the system to be exploited in practice. By fixing the error rate at 1.5 percent, we obtain the results given by Table 3.

As Table 3 shows, our approach is very promising. The difference in the two recognition rates shown in Fig. 10 and Table 3 is explained by the fact that the ARRAS database comprises much more invalid addresses than the TOULOUSE database. Comparison in the same context with other approaches is not possible, our system being the first to propose a system to automatically recognize street names for the French mail. Comparison with other street name recognizers in general is very hard also because of the difference not only in the used databases, but also in the evaluation process conditions. For instance, error and rejection cases in our system are not caused only by

TABLE 3
Recognition, Error, and Rejection Rates Obtained on the TOULOUSE and ARRAS Databases

DATABASE	*RECOGNITION*	*ERROR*	*REJECTION*
ARRAS	40.1%	1.5%	58.4%
TOULOUSE	45.6%	1.5%	52.9%

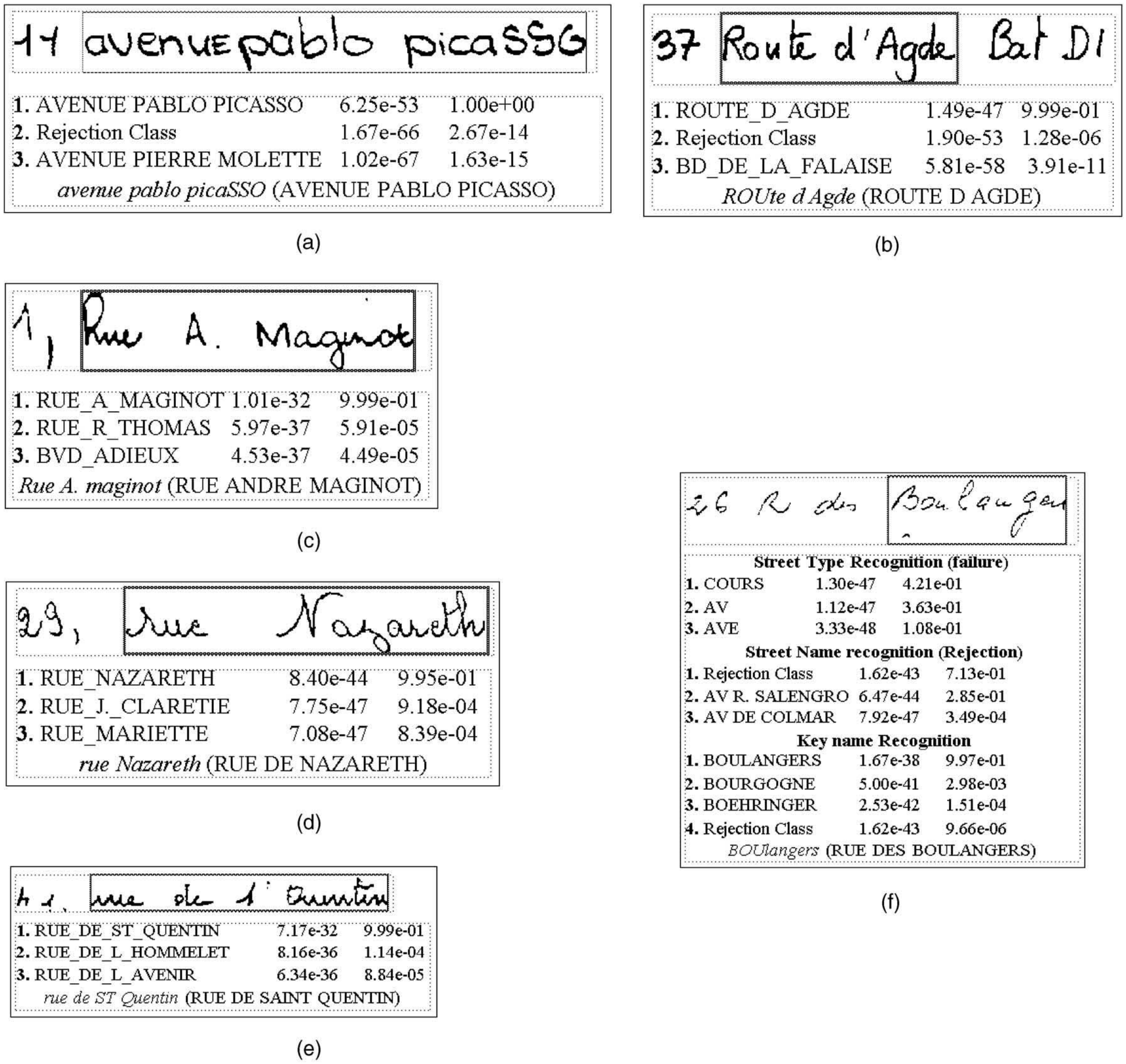

Fig. 11. Some recognized street names and their location as given by the Viterbi algorithm: The first score is $\Pr(O, sn, cfg^1)$ and the second one is $\Pr(cfg^1, sn|O)$ (similar scores are given for street type recognition or for rejection models).

classical handwriting recognition problems (segmentation, matching, etc.) but also by street line location errors, a problem which is not addressed in most systems. Also, the performance of our system is dramatically weakened by the presence of invalid inputs.

The processing time is not constant for each input image but depends on the difficulty of the recognition task and mainly on the rate of lexicon reduction and the success of the street type recognition stage. Nonetheless, the average processing time per image, from preprocessing to recognition, has an order of magnitude of one second on a Sun workstation Sparc 10.

7.3.1 Analysis of the Behavior of the Recognition System

In this section, we explain more in detail the behavior of the recognition system according to the difficulty of the image to be recognized. Fig. 11 gives some examples of correctly recognized street names. For each example, we show, in decreasing order, the best street name outputs provided by the recognition system. Associated with each street name sn are two scores: the first one is the conjoint probability $\Pr(O, sn, cfg^1)$ and the second score is the a posteriori probability $\Pr(cfg^1, sn|O)$ (similar scores are given in the case of street type recognition or as outputs for one of the rejection models). The recognized street name transcription in terms of variant and writing type, implicitly provided by the Viterbi algorithm, is given at the end in *italic*. The actual standard street name is given between parentheses. Street name location as a by-product of recognition is shown by a rectangle delimiting the frontiers of the best hypothesis.

Fig. 11a shows the benefits of first performing street type recognition. The street type is recognized with high confidence, leading to a large reduction of the lexicon. The rich context (presence of three words) then allows an easy recognition of the street name. This example also shows the ability of the system to recover the writing type in terms of the type of case letters. Here, a noncommon configuration is correctly detected (the first four letters of the word `picaSSO` are lowercase while the last three are uppercase). Fig. 11b shows the ability of the system to conjointly locate and recognize the street name. Even though roughly modeled by ergodic HMMs, left and right

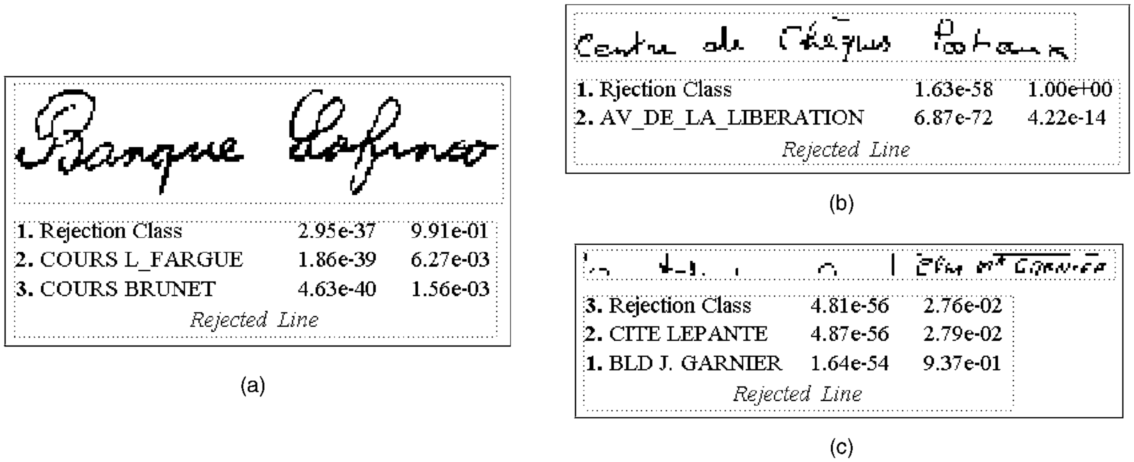

Fig. 12. Examples of rejected lines: the first score is $\Pr(O, sn, cfg^1)$ and the second one is $\Pr(cfg^1, sn|O)$ (similar scores are given for rejection models).

extra-information were almost always correctly detected and located. In fact, the ergodic models are not the actual extra-information detectors. It is their combination with the street name model, via space models, which permits an optimal (in the maximum likelihood sense) segmentation of the street line into possibly three parts. Fig. 11c and Fig. 11d show examples of correct recognition of street name variants. In the first example, the first name `ANDRE` was abbreviated as `A` while, in the second example, the article `DE` was omitted. Appropriate modeling of variants allows for integrating all the information contained in the street name to help recognition. Fig. 11e shows an example of the difficulty of performing context-free handwriting recognition. Here, the street name is hard to identify, even visually. However, the availability of the lexicon allows for correctly recognizing the ambiguous street name image and constitutes one of the strongest advantages of lexicon-driven recognition strategies. Finally, Fig. 11f shows an example when the actual street type is not among the three best hypotheses. Here, the variant `R` of `RUE`, being noncommon, is not modeled. In this case, the first recognition stage erroneously filters the actual street name. Fortunately, the first hypothesis of the second stage (whole street name recognition) corresponds to the rejection of the street line, correctly considered as not belonging to the lexicon. As explained in the end of Section 3, a third stage is then activated to perform street name recognition by considering this time key names only, which allowed the system here to recognize the correct word `BOULANGERS` and then to derive the associated whole street name `RUE DES BOULANGERS`.

7.3.2 Rejection Analysis

As discussed in the previous sections, system rejection corresponds to several configurations, which may be encountered in a real-world street name recognition task. Fig. 12a and Fig. 12b show examples where the street line is mislocated, which means that the street line is not immediately located above the postal code/city name line. In both cases, however, the rejection ergodic HMM provided the highest matching score, leading to the rejection of the processed line. The example of Fig. 12c is more difficult. Indeed, even if the processed line is not a street line, the last word coincidentally corresponds to a key name belonging to the lexicon. However, even if this key name is proposed as the best hypothesis, it is eventually rejected owing to its insufficient score confidence.

7.3.3 Error Analysis

As for rejection, system errors are caused by several sources. Table 4 shows the weight in percentage of the most significant error sources observed in our experiments when the global system threshold is fixed at 0.99, which roughly corresponds to the error rates reported in Table 3. Note that the sum of these weights exceeds 100 percent since they are not necessarily mutually exclusive, which means that an error could be provoked by more than one source.

- Street line location errors stem from the fact that the hypothesis that the street line is located immediately above the postal code/city name line is true in practice only for approximately 80 percent of the addresses. Our global ergodic HMM, designed to

TABLE 4
Error Classification in the Street Name Recognition System

ERROR SOURCE	LOCATION	WORD RECOGNITION	STREET TYPE	VARIANT	POSTAL CODE
Weight	34.6%	32.7%	25.0%	07.7%	05.8%

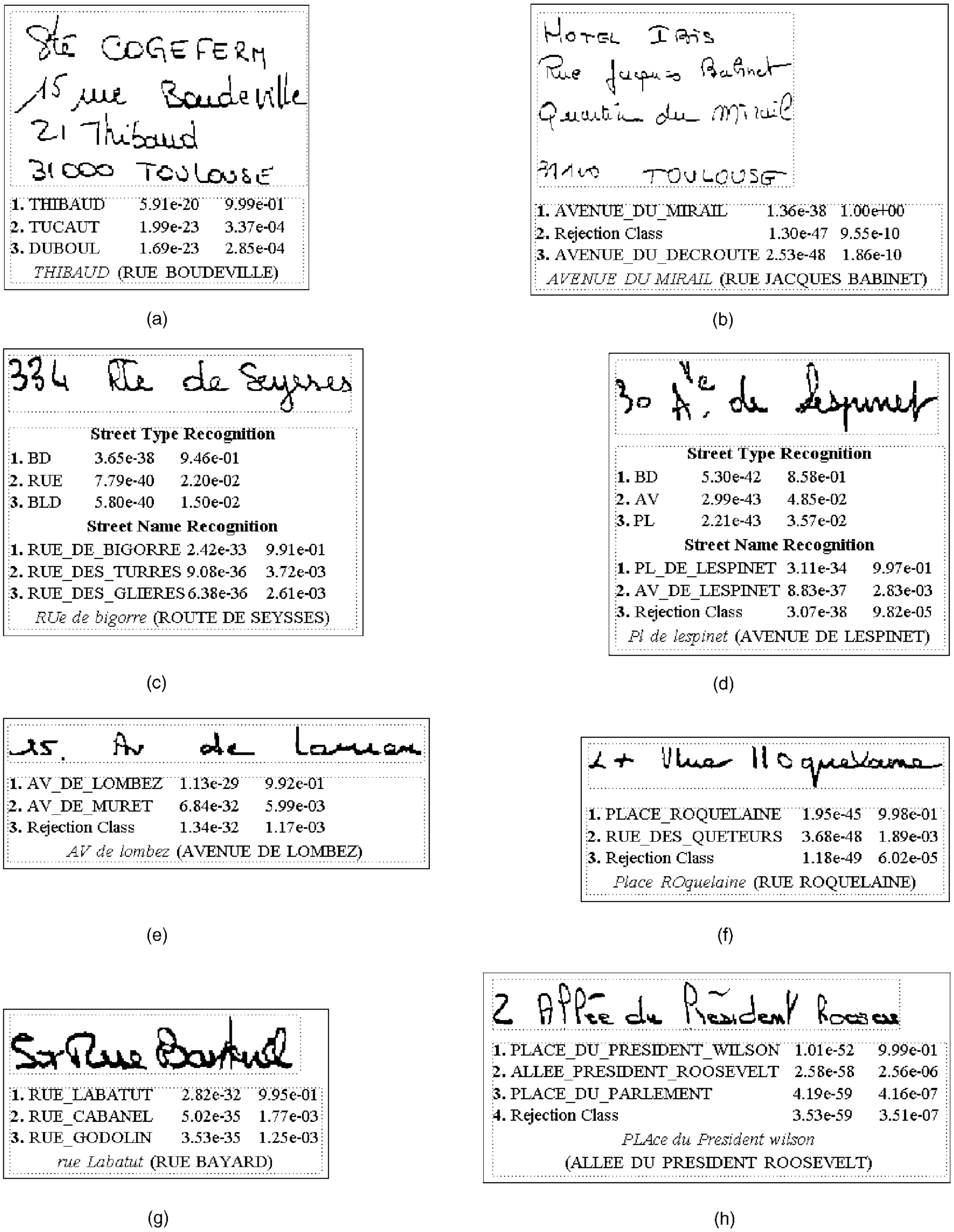

Fig. 13. Examples of misrecognized street names: the first score is $\Pr(O, sn, cfg^1)$ and the second one is $\Pr(cfg^1, sn|O)$ (similar scores are given for street type recognition or for rejection models).

handle such cases, works quite well in general. However, it becomes less robust when the wrongly located line contains a word that coincidentally exists in the dynamic street name lexicon, in the form of a key name usually. Fig. 13a gives an example of this case. Here, the second line from bottom is not a street line. However, `THIBAUD` is recognized as the key name of street name `RUE THIBAUD`. A similar example is given in Fig. 13b, where `QUARTIER DU MIRAIL` matches an existing street name (`AVENUE DU MIRAIL`).

- Street type errors have two main causes: errors due to high street type distortions (Fig. 13c) and errors generated by variants not modeled owing to their low frequency, especially when the street type is as relevant to recognition as the key name. This case is shown in Fig. 13d where both `AVENUE LESPINET` and `PLACE LESPINET` are two different street names sharing the same key name and belonging to the same lexicon.

- The postal code error type corresponds to errors occurring when the located line does correspond to a street line, but the associated street name does not belong to the lexicon, due to a postal code misrecognition. We do use, to reject this kind of line, an appropriate rejection model (Section 4). However, this model may lack robustness when there is a high ambiguity between the actual street name and an existing street name in the lexicon, which is the case in Fig. 13e for instance (`Av de Larieu` vs. `Av de Lombez`).
- Finally, word type errors are common errors inherent to handwriting recognition and caused by high image distortions, ambiguity, truncations, and overall the imperfection of our recognizer. Fig. 13f, Fig. 13g, and Fig. 13h give some examples of these errors.

8 Conclusion and Future Directions

In this paper, we presented a full system to automatically recognize a street name within a street line. This system operates in adverse conditions, such as the difficulty of street line location, the high level of distortion and variability in handwritten data, the high frequency of variant phenomena, and the fact that the input image usually contains extra-information in addition to the word string of interest (street name).

The main strength of our approach lies in its modeling phase. By building an embedded HMM network connecting street name HMMs with extraneous HMMs, our system is able to conjointly locate and recognize a street name without the need of an a priori segmentation of the street line, hence avoiding the occurrence of premature errors. A key element of our system is the optimal and extensive use of the rich context available for street names by considering all the fields a street name is composed of (street types, articles, etc.). This was possible only thanks to our variant generation tool, which permitted us to take into account the most common abbreviation, deletion, and insertion variations consistent with these fields. The consideration of robust rejection mechanisms makes our system able to detect some failures inherent to early stages of the recognition phase, such as street line mislocation or postal code misrecognition. Thus, our approach actually performs a conjoint *detection*, *location*, and *recognition* of a street name within a street line.

The developed system shows very promising results, especially if we bear in mind the real character of our application. After analyzing system errors and rejection, straightforward and significant improvements could be achieved in the following way:

- Street Line Location: The recognition performance of our system is limited by the correctness rate of the street line location stage, which is only about 80 percent. We could overcome this limitation by processing in parallel the two lines above the *postal code/city name* line and retain the one yielding the best matching score in recognition. Alternatively, we could call for the same concepts used in the current approach to detect the street line in the envelope image prior to the recognition phase. This could be carried out by first extracting from each envelope line the associated observation sequence according to our feature extraction phase. Then, our module of conjoint *street type* detection and recognition could be applied to detect the line with the maximum a posteriori probability of containing a street type. In this case, special care should be considered in comparing the matching scores associated with different lines.
- Street Type Recognition: This nontime consuming step is the basis of the reduction of the street name lexicon and, thus, is crucial to the optimal behavior of the overall recognition system. It would be very beneficial to make it as accurate as possible by looking for more robust features, more detailed models for street types, etc.
- Another improvement of the current approach could be the use of context-dependent HMMs to accommodate the writing styles particularly inherent to handwritten street lines. For instance, many characters appearing at the end of abbreviations are written as superscript (S^{te}, D^{r}, etc.) and, hence, would be more accurately modeled by specific HMMs.
- In order to further speed up the recognition time, we could adopt some successful speech recognition search techniques by considering dynamic tree lexicons and efficient search techniques like Viterbi beam search, stack decoding, N-Best, and multipass search strategies [18], [24].

The approach for street name recognition, presented here, can be adapted to the general task of detecting, locating, and recognizing a keyword or a key-phrase within a handwritten text line or a text page, a problem which is rarely addressed in current handwriting recognition systems. While the sections dealing with variant generation and lexicon reduction were specific to our application, the modeling strategy which consists of building a global network by concatenating a left ergodic HMM, a key-phrase HMM, and a right ergodic HMM can be applied to any other similar task. Our choice of considering a left and a right ergodic HMMs to, respectively, model left and right extra-information is not restrictive: One can choose different HMMs or the same HMM, depending on his or her a priori knowledge of the application at hand. In fact, when these two kinds of information are expected to be similar, it may be useful to model them with a unique HMM. In this way, the training data gets increased, allowing a more robust estimation of the HMM parameters. On the other hand, in the case of spotting tasks where the amount of extra information, on either side of the keyword, is relatively large, more attention needs to be paid to prevent false positives.

ACKNOWLEDGMENTS

This work was supported by the Service de Recherche Technique de La Poste (SRTP). The authors would like to thank the anonymous reviewers for their suggestions.

REFERENCES

[1] L.R. Bahl and F. Jelinek, "Decoding for Channels with Insertions, Deletions, and Substitutions with Applications to Speech Recognition," *IEEE Trans. Information Theory,* vol. 21, no. 4, pp. 404-411, July 1975.

[2] S. Bercu and G. Lorette, "On-Line Handwritten Word Recognition: An Approach Based on Hidden Markov Models," *Proc. Third Int'l Workshop Frontiers in Handwriting Recognition,* pp. 385-390, 1993.

[3] H. Bunke, M. Roth, and E.G. Schukat-Talamazzini, "Off-Line Cursive Handwriting Recognition Using Hidden Markov Models," *Pattern Recognition,* vol. 28, no. 9, pp. 1399-1413, Sept. 1995.

[4] J. Cai and Z.-Q. Liu, "Integration of Structural and Statistical Information for Unconstrained Handwritten Numeral Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 3, pp. 263-270, Mar. 1999.

[5] R.G. Casey and E. Lecolinet, "A Survey of Methods and Strategies in Character Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 7, pp. 690-706, July 1998.

[6] F.R. Chen, L.D. Wilcox, and D.S. Bloomberg, "Detecting and Locating Partially Specified Keywords in Scanned Images Using Hidden Markov Models," *Proc. Second Int'l Conf. Document Analysis and Recognition,* pp. 133-138, 1993.

[7] M.Y. Chen, A. Kundu, and S.N. Srihari, "Variable Duration Hidden Markov Model and Morphological Segmentation for Handwritten Word Recognition," *IEEE Trans. Image Processing,* vol. 4, no. 12, pp. 1675-1688, Dec. 1995.

[8] W. Cho, S.W. Lee, and J.H. Kim, "Modeling and Recognition of Cursive Words with Hidden Markov Models," *Pattern Recognition,* vol. 28, no. 12, pp. 1941-1953, Dec. 1995.

[9] A. El-Yacoubi, "Modélisation Markovienne de l'Ecriture Manuscrite: Application à la Reconnaissance des Adresses Postales," PhD thesis, Univ. of Rennes 1, France, Sept. 1996. (in French)

[10] A. El-Yacoubi, J.M. Bertille, and M. Gilloux, "Conjoined Location and Recognition of Street Names within a Postal Address Delivery Line," *Proc. Third Int'l Conf. Document Analysis and Recognition,* vol. 2, pp. 1024-1027, 1995.

[11] A. El-Yacoubi, M. Gilloux, R. Sabourin, and C.Y. Suen, "Objective Evaluation of the Discriminant Power of Features in an HMM-Based Word Recognition System," *Proc. Brazilian Symp. Document Image Analysis,* N.A. Murshed and F. Bortolozzi, eds., pp. 60-73, 1997.

[12] A. El-Yacoubi, M. Gilloux, R. Sabourin, and C.Y. Suen, "An HMM-Based Approach for Off-Line Unconstrained Handwritten Word Modeling and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 8, pp. 752-760, Aug. 1999.

[13] A. El-Yacoubi, R. Sabourin, M. Gilloux, and C.Y. Suen, "Off-Line Handwritten Word Recognition Using Hidden Markov Models," *Knowledge-Based Intelligent Techniques in Character Recognition,* L.C. Jain and B. Lazzerini, eds., pp. 191-229, CRC Press LLC, 1999.

[14] P.D. Gader, M.P. Whalen, M.J. Ganzberger, and D. Hepp, "Handprinted Word Recognition on a NIST Data Set," *Machine Vision and Applications,* vol. 8, no. 1, pp. 31-40, 1995.

[15] M. Gilloux, M. Leroux, and J.M. Bertille, "Strategies for Cursive Script Recognition Using Hidden Markov Models," *Machine Vision and Applications,* vol. 8, no. 4, pp. 197-205, 1995.

[16] D. Guillevic and C.Y. Suen, "HMM Word Recognition Engine," *Proc. Fourth Int'l Conf. Document Analysis and Recognition,* vol. 2, pp. 544-547, 1997.

[17] J. Hu, M.K. Brown, and W. Turin, "HMM Base On-Line Handwriting Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 10, pp. 1039-1045, Oct. 1996.

[18] F. Jelinek, *Statistical Methods for Speech Recognition,* chapter 7. Cambridge, Mass.: MIT Press, 1998.

[19] G. Kim and V. Govindaraju, "Handwritten Phrase Recognition as Applied to Street Name Images," *Pattern Recognition,* vol. 31, no. 1, pp. 41-51, 1998.

[20] F. Kimura, M. Shridhar, and Z. Chen, "Improvements of a Lexicon Directed Algorithm for Recognition of Unconstrained Handwritten Words," *Proc. Second Int'l Conf. Document Analysis and Recognition,* pp.18-22, 1993.

[21] S.S. Kuo and O. Agazzi, "Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 8, pp. 842-848, Aug. 1994.

[22] S. Madvanath, E. Kleinberg, and V. Govindaraju, "Holistic Verification of Handwritten Phrases," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 12, pp. 1344-1356, Dec. 1999.

[23] J. Makhoul, R. Schwartz, C. Lapre, and I. Bazzi, "A Script-Independent Methodology for Optical Character Recognition," *Pattern Recognition,* vol. 31, no. 9, pp. 1285-1294, 1998.

[24] H. Ney and S. Ortmanns, "Dynamic Programming Search for Continuous Speech Recognition," *IEEE Signal Processing Magazine,* pp. 64-83, 1999.

[25] J.C. Oriot, D. Barba, and J.C. Salome, "Address Bloc Locating Method Based on Transition Analysis," *Proc. First Int'l Conf. Document Analysis and Recognition,* vol. 2, pp. 665-673, 1991.

[26] H.S. Park and S.W. Lee, "A Truly 2-D Hidden Markov Model for Off-Line Handwritten Character Recognition," *Pattern Recognition,* vol. 31, no. 12, pp. 1849-1864, 1998.

[27] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE,* vol. 77, no. 2, pp. 257-286, 1989.

[28] G. Seni and E. Cohen, "External Word Segmentation of Off-Line Handwritten Text Lines," *Pattern Recognition,* vol. 27, no. 1, pp. 41-52, 1994.

[29] S.N. Srihari, E. Cohen, V. Govindaraju, and A. Shekhawat, "Determining Delivery Point Codes on Handwritten Addresses," *Proc. Fifth U.S. Postal Service Advanced Technology Conf.,* pp. 321-335, 1992.

[30] T. Steinherz, E. Rivlin, and N. Intrator, "Offline Cursive Script Word Recognition—A Survey," *Int'l J. Document Analysis and Recognition,* vol. 2, pp. 90-110, 1999.

[31] C.C. Tappert, C.Y. Suen, and T. Wakahara, "The State of the Art in On-line Handwriting Recognition—A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, no. 8, pp. 787-808, Aug. 1990.

[32] J.A. Vlontzos and S.Y. Kung, "Hidden Markov Models for Character Recognition," *IEEE Trans. Image Processing,* vol. 1, no. 4, pp. 539-543, 1989.

Mounim A. El-Yacoubi received the BS degree in physics in 1990 from the Faculty of Science I at Casablanca, Morocco, and in electrical engineering in 1991 from the University of Orleans, France. He received the MS and PhD degrees in signal processing & telecommunications from the University of Rennes I, France, in 1992 and 1996, respectively. From 1992 to 1996, he was a member of the research staff at the Service de Recherche Technique de La Poste (SRTP), the research center of the French Postal Service at Nantes, France, where he worked on address recognition tasks. From January 1997 to June 1998, he was a visiting scientist at the Center of Pattern Recognition and Machine Intelligence (CENPARMI) in Montreal, Canada. From August 1998 to December 2000, he was an associate professor of computer science at the Catholic University of Parana in Curitiba, Brazil. Recently, he joined Parascript, LLC in Boulder, Colorado, where he works on machine printed and handwriting recognition applications. His research interests include pattern recognition, hidden Markov models, machine print and handwriting recognition, document image processing, and signature verification.

Michel Gilloux graduated from the Ecole Polytechnique, Palaiseau, France (1977) and the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France (1982). He is currently the manager of research and development on postal address recognition systems and document processing at the Service de Recherche Technique de la Poste (SRTP), the research center of the French postal service La Poste, Nantes, France. Prior to this, he was with the Centre National d'Etudes des Télécommunications of France Telecom in Lannion, France, where he was in charge of research and development projects in experts systems, natural language processing, and information retrieval. His research interests are in hidden Markov models, offline handwriting recognition, character and word recognition, document image processing, and applications and systems. He has served on the program committee of several conferences and workshops. He is a member of the editorial board of the *International Journal on Document Analysis and Recognition*. He has published more than 50 referred technical papers in the areas of expert systems, natural language processing, and handwriting recognition.

Jean-Michel Bertille obtained a Diplôme d'Etudes Approfondies in artificial intelligence from the University of Montpellier, France (USTL) in 1981. He prepared a PhD concerning the possibility of automatic learning using nonlinear neural nets at the Ecole Nationale Supéreure de l'Aéronautique et de l'Espace (ENSAE), Toulouse, France. Since 1990, he has been with the French Postal Office research center (SRTP) in Nantes, France, where he manages OCR projects. His interest lies mostly in the automatic recognition of handwritten mail addresses.