SEGMENTATION AND CLASSIFICATION OF HAND-DRAWN PICTOGRAMS IN CLUTTERED SCENES – AN INTEGRATED APPROACH

Stefan Müller, Stefan Eickeler, Christoph Neukirchen, Bernd Winterstein

Department of Computer Science
Faculty of Electrical Engineering
Gerhard-Mercator-University Duisburg, Germany
e-mail: {stm,eickeler,chn,winter}@fb9-ti.uni-duisburg.de

ABSTRACT

In this paper, a new approach to identification of handwritten symbols in arbitrary complex environments is presented. 20 different pictograms drawn in different backgrounds can be identified with a recognition accuracy of 90%. In order to perform this challenging task, we use pattern spotting techniques based on pseudo 2-D Hidden Markov Models (P2DHMMs). Practical applications of our approach can be found in many typical mulitmedia document processing tasks, such as localization and recognition of non-rigid objects in image databases, detection of objects in complex scenes, finding trademarks in presence of clutter within videos, processing distorted document images in digital libraries, or content-based image retrieval based on handwritten query symbols.

1. INTRODUCTION

Segmentation is the first essential and important step of low level vision and can be described as a process of partitioning an image into some non-intersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous [6]. Applications which involve segmentation are e.g. the identification of people in natural images or the spotting of a set of keywords in noise contaminated images of scanned text. Besides the low-level processing, the previously mentioned tasks also require the involvement of a recognition step. In many conventional computer vision approaches the segmentation and recognition steps are developed independently, thus breaking down the task into sub-problems which are pipelined together. This division into independent *black boxes*, according to the popular *top-down* approach, may in many cases lead to suboptimal solutions. The reason for this is the existence of an information theory theorem, which states that during the communication of pipelined stages important information may be lost. Formally stated, the information loss $\Delta I_\Omega(X,Y) = I(X;\Omega) - I(Y;\Omega)$ about some feature variable Ω across a chain of N–1 stages can be expressed by

$$\Delta I_\Omega(X_1, X_N) = \sum_{i=1}^{N-1} I(X_i;\Omega) - I(X_{i+1};\Omega)$$

Due to the fact that $\Delta I_\Omega(X,Y)$ is always non-negative for any transform $Y = f(X)$, the information lost across any transform in this series can never be regained. Hence, a standard approach is to minimize the ΔI_Ω for the individual transforms, bearing the risk that an early stage of the chain transforms the data into a form which makes it difficult for the following stage to minimize its own information loss [7].

In order to overcome these limitations of concatenated *black boxes*, integrated approaches for e.g. the delineation and classification of closed contour shapes [1] or the simultaneous rotation angle estimation and recognition of pictograms [5] have been proposed. 1D-Markov Random Fields (MRFs) have been used in [1] to extract contours and simultaneously classify objects in images. Due to the high computational burden usually associated with MRFs and the simulated annealing optimization procedure, Hidden Markov Models (HMMs) have been chosen in [5] for the joint alignment and recognition of arbitrarily rotated, hand-drawn pictograms. In this publication, 20 different isolated symbols with different characteristic shapes are recognized in a rotation invariant mode, with a recognition accuracy of 99.5%. This has been achieved by using concatenated HMMs with modified *filler*-models in conjunction with a rotating feature extraction. The joint classification and segmentation abilities of the HMM-framework (a property continuous speech heavily relies on) can be utilized to simultaneously recognize shapes and give an estimate of the rotation angle of that pictogram. This method is much more elegant than the complicated preprocessing steps suggested in [2, 4], where shapes are rotated to the same orientation prior to the feature extraction and classification steps.

Although dealing with planar objects and shapes, the previously mentioned publications are using one-dimensional models, namely 1D-MRF and linear HMMs, respectively. In the present paper, an integrated approach to segmentation and classification of hand-drawn pictograms in cluttered scenes is presented, using a two-dimensional modeling technique. Fig. 1 shows the 20 different handwritten symbols to be spotted and recognized embedded in different backgrounds. The previously mentioned publications [2, 4] introduced the classes 1–8 of these symbols. Fig. 2 presents six images, taken from our database, where hand-drawn pictograms of *Class 16* are surrounded by three different kinds of clutter. It is noteworthy, that the pictograms itself, although taken from the same class, as well as the background show deformations, different scaling and slanting effects due to the well-known varieties in handwriting of even one single person. As one can see in Fig. 2, the background consists of strokes, which are similar to the lines that compose the object to be recognized.

The paper is organized as follows. Section 2 gives a brief introduction to the modeling of two-dimensional data based on P2DHMMs. Section 3 describes the integrated approach to segmentation and classification of pictograms in cluttered scenes. Section 4 presents experimental results. Conclusions are given in Section 5.

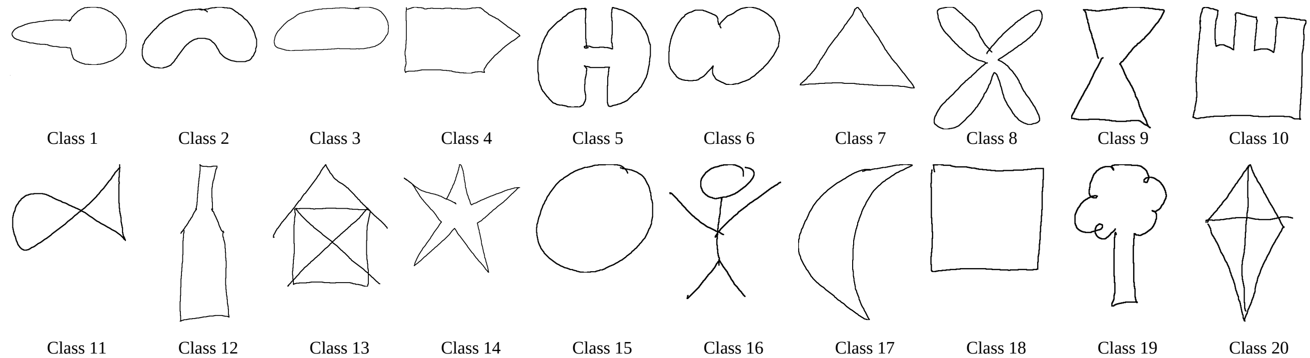

Figure 1: List of 20 different handwritten symbols to be recognized. The classes 1–8 have been introduced in [2, 4].

2. P2DHMMS FOR STOCHASTIC MODELING OF TWO-DIMENSIONAL DATA

Hidden Markov Models are finite stochastic automata and represent one of the most powerful tools for modeling dynamic, time-varying patterns. It is therefore not amazing that they became popular in speech recognition [8]. Their major advantage in time series classification results from their capability to align a pattern along their states using a probability density function for each state in order to compute the probability that a certain part of the pattern belongs to that state (see [8] for details).

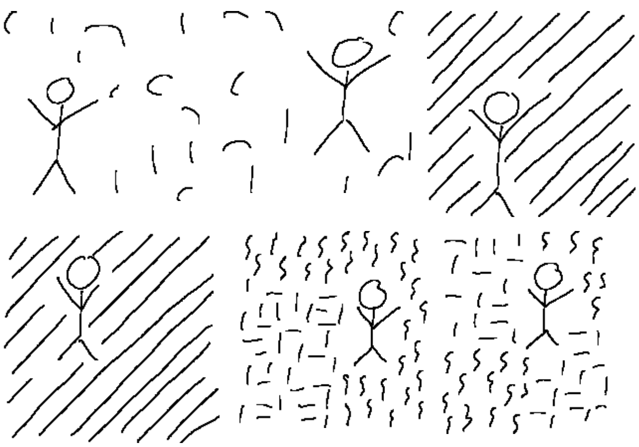

Figure 2: Handwritten symbols of *Class 16* embedded in three different backgrounds, taken from our database.

In recent years, HMMs have been more and more applied to other pattern recognition problems, especially in on-line handwriting recognition [10], where the recognition procedure is similar to speech recognition. It has also been shown that HMMs can be not only applied successfully to time series problems, but also to pattern recognition problems where the pattern varies in space rather than in time. The problem most closely related to the above mentioned examples is off-line handwriting recognition, where the characters represent a pattern changing in space while moving horizontally from left to right [10]. Therefore, HMMs have been recently also applied to image recognition problems with promising results [9] and the symbol recognition problems already mentioned in [2, 5] have been tackled using HMMs, too. For our integrated segmentation and recognition approach, we propose the use of pseudo 2-D HMMs (P2DHMM), which are also known as planar HMMs. A P2DHMM is a stochastic automata with a two-dimensional arrangement of the states, as outlined in Fig. 3[1].

The states in horizontal direction are denoted as *superstates*, and each superstate consists of a one-dimensional HMM in vertical direction. P2DHMMs have been already used for character recognition in [3]. If one considers one of the symbols in Fig. 1 subdivided into e.g. vertical stripes, it is possible to use P2DHMMs for modeling a two-dimensional object in the following manner: Each stripe is aligned to one of the superstates of the P2DHMM, resulting in a horizontal warping of the pattern. Furthermore, within the superstate, the pattern representing the stripe is aligned to the one-dimensional HMM states, resulting in a vertical alignment of the stripe. In a similar way, it is also possible to model data, which is considered as consisting of horizontal stripes.

The P2DHMM shown in Fig. 3 can be trained from data, after features have been extracted, using the segmental k-means algorithm as outlined in Fig. 4. The feature extraction used throughout the paper is a simple subsampling technique. Once the models have been trained for each class, the recognition procedure is accomplished by calculating the class-dependent probability that the (unclassified) data has been generated by the corresponding HMM. For this procedure, the doubly embedded Viterbi algorithm can be utilized, which has been proposed by Kuo and Agazzi in [3]. Alternatively, Samaria shows in [9], that a P2DHMM can be transformed into an equivalent one-dimensional HMM by the insertion of special *end-of-line* states. Thus, these equivalent HMMs can be

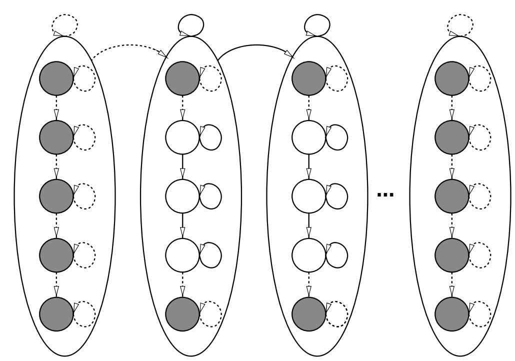

Figure 3: Pseudo 2-D Hidden Markov Model[1]

[1]Up to this point, the *shaded states* as well as the *dotted transitions* can be treated as normal states and transitions, respectively. Their *special* meaning will be explained in Sec. 3.

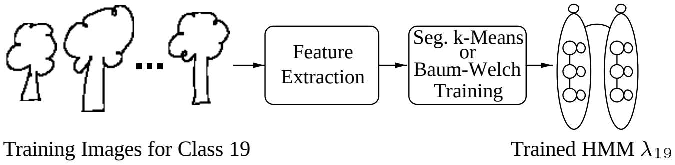

Figure 4: Training procedure for a pictogram HMM

trained by the standard Baum-Welch algorithm and the recognition step can be carried out using the standard Viterbi algorithm.

3. PATTERN SPOTTING USING PSEUDO 2-D HMMS WITH FILLER STATES

So far, we have not yet answered the question how one can identify symbols within complex backgrounds as shown in Fig. 2. In Section 1, it has been argued that due to the information loss in a system of pipelined stages, a classical two-step approach based on separation of the symbol and a subsequent classification should lead to a less successful solution compared to an integrated approach. Fortunately, the P2DHMM technique allows the integration of the segmentation and classification process in one single step. The basic idea of our integrated approach is to surround the already trained P2DHMMs for each symbol with additional so-called *filler*-states, whose Gaussian distribution model the generation of the strokes in the background. The structure of the augmented model is illustrated in Fig. 3, where the white circles represent the states of the original pseudo 2-D HMM (see also Fig. 4), whereas the shaded circles denote the additional filler-states. The dotted transitions in Fig. 3 have not been trained and have to be chosen arbitrarily. By setting the transition it is possible to integrate an expected background to object ratio. The filler-states surrounding the original HMM all share the same output probability function (pdf), a strategy which is commonly referred to as tying [11]. We developed two different strategies for the estimation of the tied pdf of the filler states, which will be explained in the next two sections.

3.1. Filler State Parameter Estimation Based on Prior Knowledge about the Background

The first method is based on prior knowledge about the background to be expected. A large number of images showing symbols embedded in background-strokes which are similar to the expected background of the test image are used in order to estimate the parameters of the filler-state pdf. Of course, one could have also used images showing only background strokes, however using the proposed technique, we can reuse the images from our database. Fig. 5 gives a schematic overview about the proposed technique and shows also the recognition procedure. A test image is presented to every augmented model and the probability $Pr(\vec{O}|\lambda)$ is used to classify the unknown pattern according to the maximum-likelihood(ML)-decision. These probabilities are jointly calculated on the samples representing the background as well as the pictogram samples which have been aligned by the Viterbi-algorithm to the symbol states and the filler states, respectively. Note that the filler states' pdfs in Fig. 5 are also tied across the models of different classes.

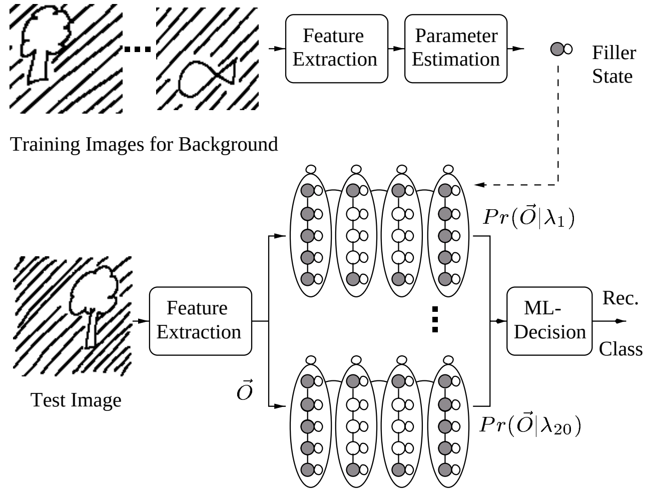

Figure 5: Schematic overview of the filler state parameter estimation based on prior knowledge about the background. All the shaded states share a single pdf.

3.2. Adaptive Parameter Estimation

The experiences gained in the experiments using the filler state estimation technique presented in Sec. 3.1 lead to the development of an improved estimation procedure which does not depend on any prior knowledge about the expected background and thus works in an adaptive way. The estimation of the background state is directly carried out on the (unsegmented) test image (see also Fig. 6). Since the background fills out most of the space in the images of our database, the filler state is adapted mainly to the generation of background strokes, rather than symbol strokes.

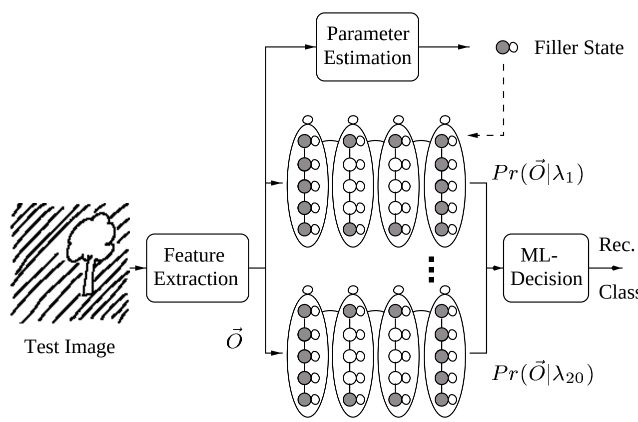

Figure 6: Adaptive parameter estimation

4. EXPERIMENTS AND RESULTS

The techniques presented in the previous sections have been evaluated on a database consisting of 200 isolated pictograms (10 different for each class in Fig. 1) and 300 images of pictograms in cluttered scenes (5 different for each class and for each background in Fig. 2). As one can see in Fig. 2, not only the sizes of these complex scenes and the positions of the pictograms vary, but also the ratio of pixels belonging to the pictogram and the background varies between 11–41%. The database has been build by scanning the images, which have simply been drawn by hand on a sheet of paper by one of the authors. The feature extraction consists of a subdivision of the drawings into blocks of 30×30 pixels, using an overlap of 75%. For each square, a feature vector is extracted by a further subdivision into nine small squares for which the mean grey value is calculated. Following the feature extraction, an individual pseudo 2-D Hidden Markov Model has been trained for each class on 10 isolated pictograms utilizing the Baum-Welch reestimation formula (see also Fig. 4). These HMMs consist of 5 superstates with 5 states each. After the filler state parameters (each pdf is modeled by a mixture of six Gaussian densities) have been estimated using one of the techniques described in Sec. 3, the augmented HMM of size 7×7 states is constructed. The Viterbi-algorithm then computes the most likely HMM sequence that might have generated the unknown pattern sequence, resulting in an alignment of the entire image to the white and shaded states in Fig. 3. Fig. 7 shows the state alignment for one of the examples in Fig. 2. In Fig. 7, all white areas have been aligned to the white HMM states, while all shaded areas have been assigned to states modeling the background. The result of the Viterbi-algorithm is not only this *optimal* state alignment, but also a probability that the presented pattern has been generated by this P2DHMM.

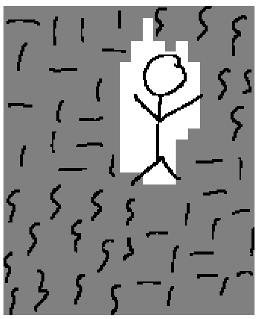

Figure 7: Segmentation result after running the Viterbi-algorithm

Table 1 summarizes the recognition accuracies that have been achieved in our experiments, separated for a horizontal and vertical orientation of the superstates. The columns denoted *backgr. 1–3* show the recognition results for the different backgrounds in Fig. 2 using the modeling technique proposed in Sec. 3.1. The rates for the adaptive modeling technique (section 3.2) are given in the final column.

direction	backgr. 1	backgr. 2	backgr. 3	adaptive
	91.0%	90.0%	90.0%	90.3%
	91.0%	89.0%	91.0%	90.0%

Table 1: Recognition accuracy achieved in the experiments.

5. SUMMARY AND OUTLOOK

In this paper, we proposed a novel integrated approach to the segmentation and classification of hand-drawn pictograms in cluttered scenes. This technique utilizes pseudo 2-D Hidden Markov Models which are augmented by so-called filler states. These states model the generation of background strokes and two different methods for the parameter estimation of the filler states' pdf have been presented. Recognition results are slightly dependent on the choice of the background, because certain backgrounds (e.g. hatching) can cause more confusion with some symbols than other backgrounds. The recognition rates are generally between 90 and 91%, which can be considered as a high recognition accuracy for this task. While most of our research has been carried out in order to solve the basic problem of spotting patterns in arbitrary environments, we believe that our algorithm has a potential for many practical applications. Such applications include localization of objects in complex scenes, finding trademarks in advertisements or videos, or locating text in images or document images. An example for this ongoing activity is shown in Fig. 8. A P2DHMM with the same topology as used for the pictograms in Fig. 3 has been constructed from the single image on the left. The right part shows the segmentation result obtained by the Viterbi-algorithm.

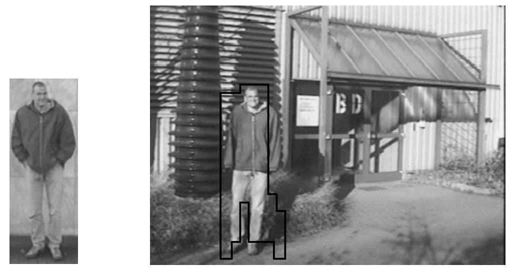

Figure 8: Example for the detection of a person using P2DHMMs.

6. REFERENCES

[1] N.S. Friedland and A. Rosenfeld, An Integrated Approach to 2D Object Recognition, *Pattern Recognition*, Vol. 30, No. 3, 1997, pp. 525–535.

[2] Y. He and A. Kundu, 2-D Shape Classification Using Hidden Markov Model, *IEEE Trans. on PAMI*, Vol. 13, No. 11, 1991, pp. 1172–1184.

[3] S. Kuo and O. Agazzi, Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models, *IEEE Trans. on PAMI*, Vol. 16, No. 8, 1994, pp. 842–848.

[4] S. Lee and B. Lovell, Modelling and Classification of Shapes in Two-Dimensions Using Vector Quantization, In *Proc. IEEE-ICASSP*, Vol. V, 1994, pp. 141-144.

[5] S. Müller, G. Rigoll, A. Kosmala, and D. Mazurenok, Recognition of Hand-Drawn Pictograms Using HMMs with a Rotating Feature Extraction, In *Proc. 6th IWFHR*, Taejon, Korea, 1998, pp. 25–34.

[6] N.R. Pal and S.K. Pal, A Review on Image Segmentation Techniques, *Pattern Recognition*, Vol. 26, No. 9, 1993, pp. 1277–1294.

[7] M.D. Plumbley, An Information-Theoretic Approach to Unsupervised Connectionist Models, Ph. D. Thesis, Cambridge University, 1991.

[8] L.R. Rabiner and B.H. Juang, An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, 1986, pp. 4–16.

[9] F.S. Samaria, Face Recognition Using Hidden Markov Models, Ph. D. Thesis, Cambridge University, 1994.

[10] R. Seiler, M. Schenkel and F. Eggimann, “Off-Line Cursive Handwriting Recognition Compared with On-Line Recognition”, In *Proc. IEEE-ICPR*, Vol. IV, 1996, pp. 505–509.

[11] S.J. Young, The General Use of Tying in Phoneme-Based HMM Speech Recognisers, In *Proc. IEEE-ICASSP*, 1992, pp. 569–572.