Statistical Pattern Recognition Techniques for Multimodal Human Computer Interaction and Multimedia Information Processing

G. Rigoll, S. Müller
Department of Computer Science, Faculty of Electrical Engineering
Gerhard-Mercator-University Duisburg, Germany
Bismarckstr. 90, 47057 Duisburg, Germany

Abstract: This paper presents an extensive overview on statistical pattern recognition methods for a variety of different tasks, related to multimodal human-computer interaction and multimedia information processing. Typical tasks in the area of human-computer interaction include handwriting and gesture recognition, as well as pen-based retrieval of image databases. Multimedia information processing includes algorithms for document processing, video indexing or face recognition. The aim of the paper is to demonstrate to the speech community the usability of classical speech recognition algorithms, such as Hidden Markov Models and related statistical pattern recognition techniques, for a much larger variety of related problems in man-machine-communication and the efficient processing and retrieval of multimedia information.

1. INTRODUCTION

Hidden Markov Models (HMMs) represent a well-known statistical pattern recognition technique and can be considered as the most powerful tool in speech recognition [1]. The efficiency of HMMs basically has several reasons: One is the fact that HMMs are perfectly suited for warping of patterns of almost arbitrary origin, and another major reason are the effective self-organizing learning capabilities of HMMs using large databases. Other advantages include the decoding capabilities of HMMs and their ability to perform recognition and segmentation in one single step. It turns out, that exactly those capabilities can be also useful for a variety of other challenging pattern recognition problems. One could expect that such problems include especially the processing of time-sequential data, such as speech signals. This is indeed true for applications as e.g. handwriting recognition and gesture recognition, because it is obvious, that such applications directly imply the generation of feature sequences, derived from the pen input in case of handwriting or from a video sequence in case of gesture recognition. These applications happen to be also very popular modalities for human-computer interaction. However, as will be shown later, the use of HMMs is not only restricted to time-sequential data, but also "static" data, such as still images, can be succesfully modelled and processed using HMMs. In this case, the image space is used to subdivide the image into segments, and the spatial order of these segments serves as substitute in order to artificially generate sequences that can be processed by HMMs. In this way, images can be warped in vertical and horizontal direction, leading to elastic matching capabilities for images using HMMs.

In this paper, we will not explicitly explain the fundamentals of HMMs, since it is assumed that these are well-known for most of the people active in speech recognition. Instead, we will concentrate on presenting some examples for the use of HMMs in the application areas mentioned above.

2. HMM-BASED HANDWRITING RECOGNITION

It is obvious that the most popular application area of HMMs has been speech recognition, but recently, HMMs have also been more and more used for handwriting recognition. This is especially true for on-line handwriting recognition, since the pen trajectory captured in this case can be interpreted as one-dimensional, non-stationary pattern sequence. It has turned out that the superior learning abilities of HMMs, their warping capabilities and the existence of very fast and effective decoding procedures for connected patterns are also the major advantages of HMMs over other traditional pattern processing techniques for handwriting recognition. Therefore, some of today's most powerful large vocabulary handwriting recognition systems are based on HMM-technology. Usually, in HMM-based handwriting recognition, each letter of the alphabet is represented by one HMM. The various HMM states for one letter represent different sections of this letter, and the topology has an important influence on the way the production of that letter is modeled. It is also important for the connection of various letters in order to build HMMs for entire words. As in speech recognition, also in handwriting, the major HMM component is the output observation component, consisting of the probability density functions that model the generation of the various strokes for each state. For on-line handwriting recognition, features are extracted from the pen trajectory that contain some dynamic information about the handwriting production process. It has turned out that further improvements can be obtained if additionally some static features are incorporated into the recognition process. The large vocabulary handwriting recognition system developed by the authors evaluates the following features:

- angle α of the spatially resampled strokes (coded as sinα and cosα)
- difference of consecutive angles (sin$\Delta\alpha$, cos$\Delta\alpha$)
- pen pressure during writing (p)
- 9-dimensional vector (x) representing a bitmap slided along the pen trajectory in order to incorporate a fraction of the actual image of the currently written letter

Fig. 1 shows a few of the above described features along with an HMM with left-to-right topology typically used for handwriting recognition.

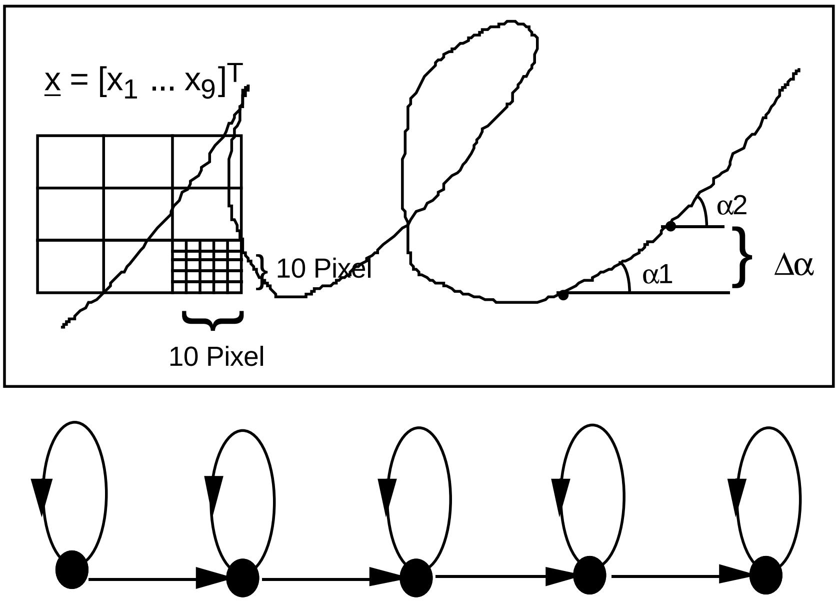

Figure 1: HMM for modeling the handwriting production process and typical features for recognition

An interesting question also in handwriting recognition is the choice of the correct modeling technique. Our group has presented one of the first systematic investigations for that problem in [16], comparing various discrete, continuous and hybrid modeling techniques for a demanding large vocabulary handwriting recognition task. It turned out the the new hybrid technique based on information theory principles as described in [17] performed best in this case. The most advanced handwriting recognition systems are capable of handling vocabularies of up to 200.000 words with recognition accuracies over 90% for writer-dependent recognition mode, as described in [16].

3. STOCHASTIC MODELING OF IMAGE SEQUENCES

As mentioned already in the introduction, due to the fact that time series modeling has been the most successful original application of HMMs, the HMM paradigm is the obvious and most natural choice if it comes to the question of how to use the principle of statistical theories and stochastic modeling techniques for image sequences. Obviously, the most suitable applications for image sequence processing with

HMMs are problems, that involve the recognition and classification of image sequences. Thus, typical applications can be found in the area of gesture recognition, video-indexing, or tracking and surveillance. It has indeed turned out that HMMs can be used very efficiently for exactly these tasks, and our research group has performed extensive investigations on the use of HMMs in these image sequence processing areas. Because the area of image sequence processing and recognition has developed into a quite complex and demanding research area during recent years, the description of all our activities carried out in that area would be beyond the scope of this paper. Therefore, we concentrate here more on a brief description of HMM-based approaches to gesture recognition and video-indexing.

3.1 Gesture Recognition

Gesture recognition is a typical problem that requires the effective combination of two different challenges in computer vision: One is the capability to deal with image sequences rather than still images. The other is the requirement of handling flexible non-rigid objects and computing useful scores in order to classify the motion of these objects in the video sequence. HMMs have been used by a few research groups in order to deal with this problem, but many of these approaches still require the use of complicated segmentation techniques, because they are in majority based on the idea of tracking the parts of the human body that are mainly responsible for carrying out the gesture. The classical use of HMMs for gesture recognition is therefore based on tracking the hands or the arms, and evaluating the feature sequences derived from this tracking process using one-dimensional HMMs. However, this leads to problems due to the fact that the involved tracking process is relatively complicated and not at all free of errors. Therefore, in many cases, gesture recognition is only possible if the person is equipped with special sensors facilitating the tracking process, or wears special clothing or performs the gestures only in front of a specified background.

Our approach, as outlined in more detail in [15], is capable of avoiding these limitations and enables gesture recognition in a truly unconstrained way, using no additional equipment nor any background limitations. For that purpose, we calculate from each image frame of a video sequence representing a gesture some global motion features, which describe the gesture with moments derived from the difference image of the video sequence. It turns out that these features can be conveniently visualized by an ellipsis, showing the center of gravity of the motion and the motion variance as the axes of the ellipsis. Fig. 2 shows a few of the investigated gestures with the computed and overlayed motion ellipsis.

Figure 2: Typical gestures and overlayed ellipsis representing the appropriate motion features

Additionally, we use the garbage modeling techniques offered by HMMs to model additional motions that the person may perform for purposes such as relocation within the image or simply relaxing between two gestures. The result of this is an HMM-based gesture recognition system that allows the recognition of spontaneous gestures, performed at arbitrary moments, with unlimited lengths, or even performed as sequence of several adjacent gestures. Only the use of the special decoding capabilities of HMMs in the recognition process enabled the on-line recognition of spontaneously performed gestures as described in more detail in [15].

3.2 Video Segmentation and Indexing

It turned out that the global motion features used in our gesture recognition system are also very suitable for describing video sequences that do not contain gestures, but instead carry other information with a wide range of different content, as it can be found in movies or other TV material. This suggests immediately the use of 1D-HMMs for content recognition of video sequences, also known as content-based video-indexing. In [18], we present a new approach for the content-based indexing of TV news using HMMs. It turns out that in this case, the segmentation capabilities of HMMs can be used for segmenting TV-news into typical content classes, such as "news caster", "report", "weather forecast" and even such specialized items as "interview". The major innovation of this approach is the fact, that in this case HMMs allow the use of a "video-model", that indicates how different content classes can appear in typical news magazines. The result is a hierarchical HMM, that contains the stochastic model of the various content classes in its highest level, and the frame-based features indicating such elementary segmental cues as "cut" or "wipe" in its lowest level, where all levels are integrated in a large stochastic model, shown in Fig. 3. More details on this approach can be found in [18].

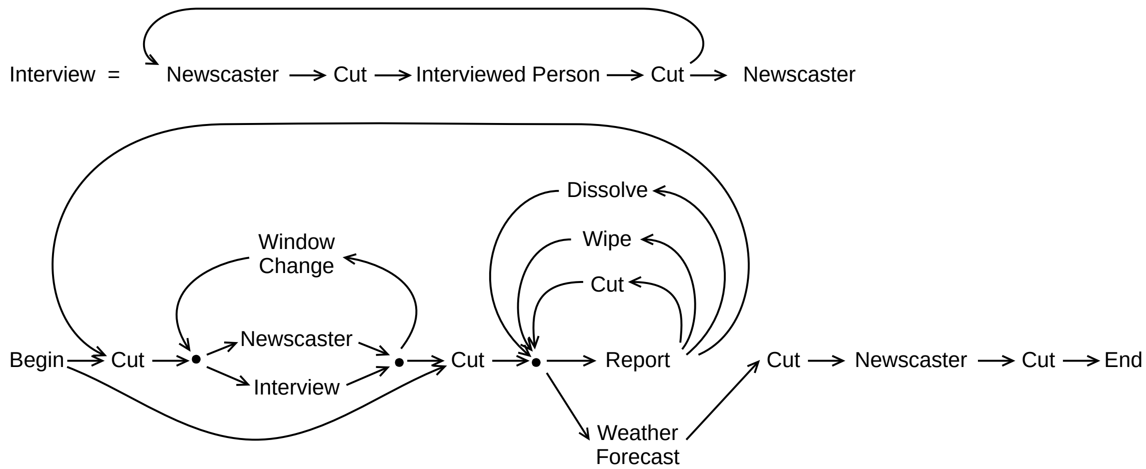

Figure 3: Stochastic video model for content-based indexing of TV news

4. STOCHASTIC IMAGE MODELING USING ONE-DIMENSIONAL HMM'S

As already mentioned, the original most popular use of HMMs has been in the area of time series classification, where mainly one-dimensional signals varying in time can be efficiently modeled using the nonlinear time warping capabilities of HMMs. As will be shown in the following sub-sections, even though computer vision is mainly a problem involving two-dimensional pattern processing techniques, the classical one-dimensional HMM approach can be also applied to patterns varying in space rather than time, and specific computer vision tasks can be solved very elegantly with one-dimensional HMMs. For this purpose, we present a brief introduction into the foundations of 1D-HMMs in the following sub-section.

4.1 Classification of Shapes Using HMM's

One-dimensional Hidden Markov Models can be used in order to classify shapes of objects in natural images or hand-drawn pictograms, where the elastic matching capabilities of the HMMs lead to a deformation tolerant recognition mode. Obviously, when using 1D-HMMs, a feature sequence $(\vec{o}_1, \ldots, \vec{o}_1)$ has to be built from the images or shapes. In order to calculate this sequence, He and Kundu [3] propose the use of radii from the center of gravity (COG) to the shape boundary, (also known as signatures, see e.g. [4]), as feature extraction technique. However, due to the use of these radii, they are limited to classify closed contour shapes, rather than unconstrained shapes such as the one shown in Fig. 4. In this Figure, the shape of a pair of nippers is shown, which is not sufficiently described by its outer geometry only. In order to overcome this limitation, we propose the use of the polar subsampling technique illustrated in Fig. 4, which is also known as a shape matrix [5, 6] and which takes the shapes inner geometry into account as well.

In [3] eight classes of hand-drawn pictograms are classified in a rotation invariant mode. The rotation invariance has been achieved by a number of preprocessing steps which aim to rotate all shapes to the same orientation. These steps are based on geometric properties such as *elongation axis* and *minimum radius point*. Hence, this preprocessing procedure is rather complicated. Later Lee and Lovell [7] carried out

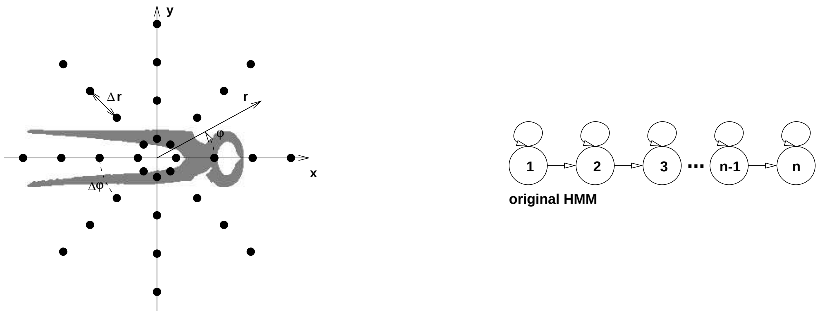

Figure 4: Feature extraction (polar sampling) and corresponding HMM structure for rotation variant recognition of shapes

experiments on a similar eight class closed contour task. Like He and Kundu they use the preprocessing steps to rotate the patterns prior to the feature extraction. The features itself are again the radii from the center of gravity to the shape boundary. The major difference of their approach is that the classification technique is based on a vector quantizer instead of HMMs.

In the approach proposed in the next section, we utilize the polar subsampling technique combined with a HMM classifier. Instead of utilizing a complicated procedure based on rules and geometric properties, we make use of the combined segmentation and classification abilities of the HMMs, which leads to a more elegant solution to the problem of invariant shape recognition.

4.2 Pattern Spotting Using HMMs with Filler States for Rotation Invariant Recognition of Shapes

The polar sampling is performed in an adaptive way as shown in Fig. 4. Samples are taken on polar coordinates (r, φ) where the origin is placed at the COG. The sampling raster in Fig. 4 can be expressed by

$$\begin{aligned} I_s(n \cdot \Delta r, m \cdot \Delta\varphi) &= I\left(\Delta r\left(n - \tfrac{1}{2}\right), m \cdot \Delta\varphi\right) \\ m &= 1, \ldots, T \\ n &= 1, \ldots, k \end{aligned} \tag{1}$$

where I_s denotes the sampled image and Δr and $\Delta\varphi$ the sampling intervals of the radius and the angle, respectively. The interval Δr is determined by dividing the radius R_{max} by a fixed number (the dimension of the feature vector) whereas $\Delta\varphi = \frac{360°}{T}$. Note that the feature extraction itself is rotation variant. Thus, the sampling and modeling scheme in Fig. 4 would lead to a rotation variant recognition mode. In order to achieve rotation invariance, we propose the use of modified HMMs, which have been concatenated with so-called filler models. These filler models are copies of an original or initially trained HMM (as shown in Fig. 4), which are attached in front of and behind the original model, with the first HMM (Filler Model 1) having a modified initial state distribution vector $\vec{\pi}$ and the final one (Filler Model 2) being modified so as to allow the HMM to be in any state, once the end of the feature sequence is reached. The components of the vector $\vec{\pi}$ are uniformly set to $\frac{1}{\mathrm{n}}$, where n denotes the number of states in the original model. This step expresses the assumption, that every (quantized) rotation angle is equally likely. The concatenation of the modified HMMs is illustrated in Fig. 5. If the feature sequence of an object being rotated with respect to the (unrotated) shape used for training the *original* HMM, is presented twice to the concatenated HMMs, the original model will be aligned to the unrotated part of the sequence by the Viterbi algorithm (see also Fig. 5). This leads to a high probability $Pr(\vec{O}|\lambda)$ if the two objects have similar shapes. It can be seen in Fig. 5, that once the alignment of the sequence to the Markov Models has been found, the rotation angle can be estimated by the number of features which have been aligned to the first and second *Filler*-model. These numbers of features are denoted as f_1 and f_2 in the following

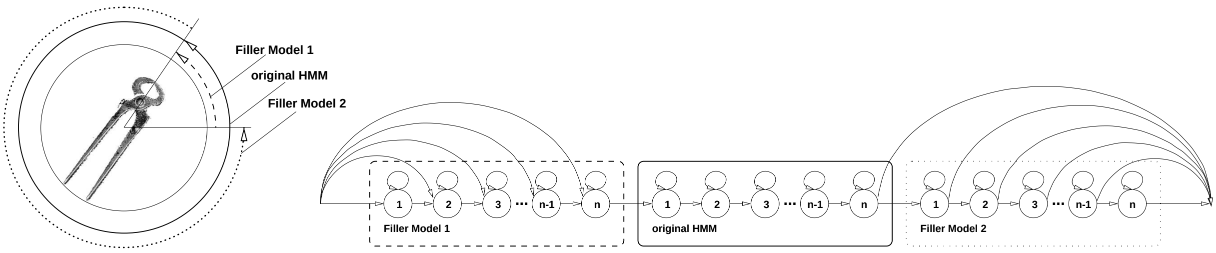

Figure 5: Alignment of the twice presented feature sequence to the original HMM and the Filler models.

and thus the rotation angle $\varphi^\star$ can be estimated as

$$\varphi^\star = \frac{f_1}{f_1 + f_2} \cdot 360^\circ \tag{2}$$

This is a typical pattern spotting problem that can be solved with HMMs more efficiently and elegantly than with most other techniques, due to the fact that an automatic alignment to the filler models is the direct result of the Viterbi decoding procedure. To verify the proposed approach, experiments on two large pictogram databases, both consisting of 20 classes, have been carried out. Each class shows a large intraclass variance and a large amount of contour perturbation. The 20 rotated pictograms per class are split into a test and a training set, consisting of 10 drawings each. The experimental results show a recognition accuracy of up to 99.5%. A detailed description of these experiments can be found in [11].

4.3 Image Retrieval Based on the Use of HMMs

As generally known, the class conditioned probability $Pr(\vec{O}|\lambda)$ can be used to classify unknown patterns. This likelihood can also be used as a score for ordering the elements of an image database according to the similarity with a sketched query image. Thus, the techniques described in the previous section can be used in order to perform content based image database retrieval by user sketch. In this case, every image of the database is represented by a rotation invariant model (see also Fig. 5), whereas the rough sketch of an object is represented by a feature sequence. This sequence can be scored in an efficient way by applying well known pruning techniques within the HMM framework. After the probability score $Pr(\vec{O}|\lambda)$ has been calculated for every database element, those N images with the highest score are retrieved. An experimental system showing the feasibility of the approach is described in [8].

By applying streams, it is possible to integrate inhomogeneous features derived from e.g. color and shape into a single statistical model. Fig. 6 illustrates the generation of features for this case. Stream 1 of the sequence, which is derived from the shape of an object, is calculated by applying the steps depicted in the upper part of Fig. 6. Note that the sampling scheme is identical to the one described in Fig. 4. In order to calculate the color information (represented in an RGB color space), the binarization step is skipped and the remaining steps as illustrated in the lower part of Fig. 6 are carried out. As indicated in the figure, the feature vectors' components are derived from heterogenous information. The influence of the different cues shape and color can be controlled by stream weights, which thus become an integral part of the query. The techniques presented in this section have been evaluated on a color-image database consisting of 120 images of arbitrarily rotated objects. For every image of the database the feature extraction steps presented in Fig. 6 have been applied, followed by the training of an individual model utilizing the Baum-Welch algorithm. As discussed in the previous sub-section, these models are concatenated with the modified filler HMMs and thus represent the image in a rotation invariant mode, thereby integrating shape and color derived features. Fig. 7 presents some results achieved with our system, where in every row the query sketch is shown first (light grey background), followed by those four images (dark grey background) with the highest similarity scores. The numbers of features being aligned to the first and second filler models (f_1 and f_2, respectively) are given below the retrieved images together with an estimated rotation angle calculated from these values according to Eq. 2. A larger number of query sketches and corresponding retrieved images as well as a more detailed description of the experimental retrieval system can be found in [20].

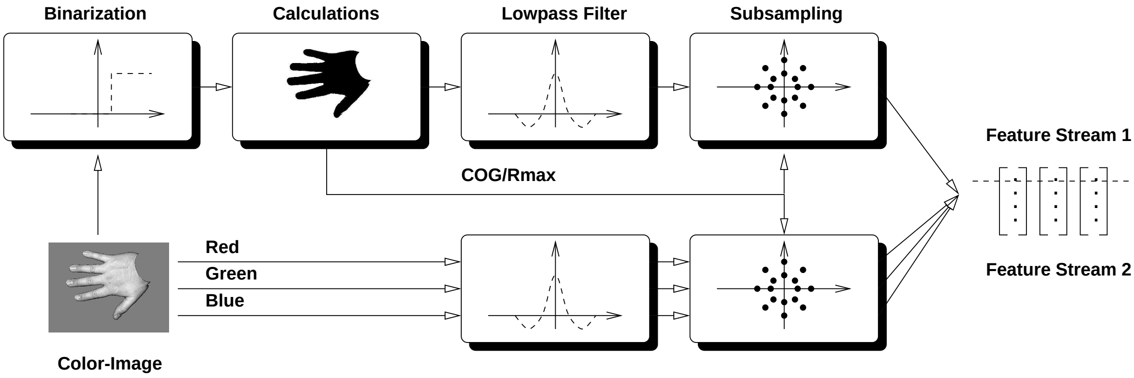

Figure 6: Blockdiagram of the feature extraction steps

Figure 7: Query sketches and retrieved images. Note that the colors match very well, the grips of the tong (sketch 2) are red and so are the grips of the first three retrieved tongs.

5. STOCHASTIC MODELING OF IMAGES USING P2DHMMs

Pseudo 2-D HMMs (P2DHMMs) are an extension of the one-dimensional HMM paradigm, which have been developed in order to model two-dimensional data. They are called *pseudo* due to the fact that the state alignment of consecutive columns is calculated independently of each other.

5.1 A Brief Introduction to Pseudo 2-D HMMs

P2DHMMs, which are also known as planar HMMs, are stochastic automata with a two-dimensional arrangement of the states, as outlined in Fig. 8. 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 [9] and face recognition [10, 12]. If one considers e.g. the image of a face in Fig. 8 subdivided into vertical stripes, it is possible to use P2DHMMs for modeling two-dimensional data 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 possible to model data, which is considered as consisting of horizontal stripes.

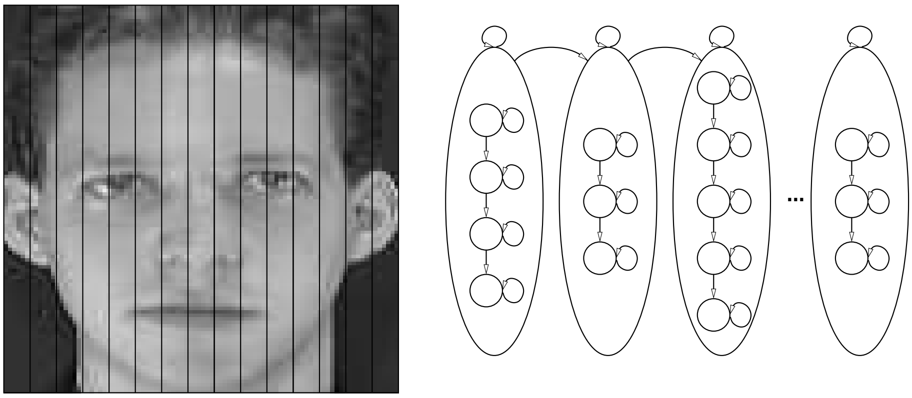

Figure 8: Basic structure of a Pseudo 2-D HMM for image classification

The P2DHMM shown in Fig. 8 can be trained from data, after features have been extracted, using the segmental k-means algorithm. 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 [9]. Alternatively, Samaria shows in [10], that a P2DHMM can be transformed into an equivalent one-dimensional HMM by the insertion of special *start-of-line* states and features. Fig. 9 shows an augmented 6×6 P2DHMM with start-of-line states

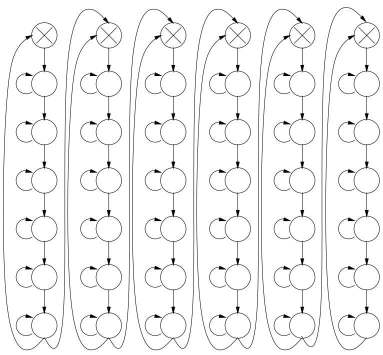

Figure 9: Augmented 6×6 P2DHMM with start-of-line marker states

(indicated by a cross). These states generate a high probability for the emission of start-of-line features. When using the structure in Fig. 9 one has to take care of the fact that the value for the start-of-line feature is different from all possible ordinary features. These equivalent HMMs can be trained by the standard Baum-Welch algorithm and the recognition step can be carried out using the standard Viterbi algorithm. P2DHMMs can be utilized for the classification of images, which is demonstrated in the following section on a face recognition task.

5.2 Classification of Images Using Pseudo 2-D HMMs

The statistical modeling capabilities of the P2DHMMs have been demonstrated in [12], where a recognition accuracy of up to 99.5% could be achieved on the Olivetti Research Laboratory (ORL) face database. This result has been achieved by applying the following steps: The first step, namely the feature extraction, is based on the discrete cosine transform (DCT). The image is scanned with a sampling window (block) top to bottom and left to right. The pixels in this sampling window of the size 8×8 are transformed using the DCT according to the equation:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{7}\sum_{y=0}^{7} f(x,y)\cos\left(\frac{(2x+1)u\pi}{16}\right)\cos\left(\frac{(2y+1)v\pi}{16}\right) \tag{3}$$

A triangle shaped mask extracts the first 15 coefficients ($u + v \leq 4$), which are arranged in a vector. The result of the feature extraction is a two-dimensional array of vectors.

An overlap between adjacent sampling windows improves the ability of the HMM to model the neighborhood relations between the sampling blocks. The effect of this overlap is somehow comparable to the use of delta-features in speech recognition, and includes redundant information into the features. Experiments showed that an overlap of 75% (6 pixel) in each direction gives the best recognition accuracy.

The next step is the statistical classification based on P2DHMMs. A single P2HMM, with a model structure as illustrated in Fig. 9, is trained for every person in the database using the Baum-Welch algorithm. For the recognition the Viterbi algorithm or the forward-backward algorithm is used in order to determine the probability of each face model for the test image. On the ORL database, both algorithms showed similar recognition accuracies, however the Viterbi algorithm is faster than the forward-backward algorithm and allows an automatic segmentation of the face. The image to be recognized is assigned to the person, whose model has the highest production probability on the test image.

The Baum-Welch algorithm, which is used for training the P2DHMM for each person, provides the HMM parameters corresponding to a local maximum of the likelihood function depending on the initial model parameters [1]. Therefore, it is very important to provide a good initial model for the training. We exploit the similarity of all faces compared to other objects and train a common initial model on all 200 faces of the training set. This common model is refined on the five faces of one person to obtain the face model for that person.

The main improvements of our approach, compared to e.g. [10], are the use of DCT features instead of grey values, and the use of a common initial model. Additionally, the columns of the images are modeled by the superstates, while in [10] the rows are modeled by the superstates. We think that this has advantages for the recognition of a person tilting the head. The eyes, which are very important points in the human face, are not on the same level in this case. This effect can be compensated by modeling the columns by the superstates. Fig. 10 shows the segmentation (Viterbi alignment) of a tilted face. Those regions of the face, that have been aligned to the same state, are framed by white lines.

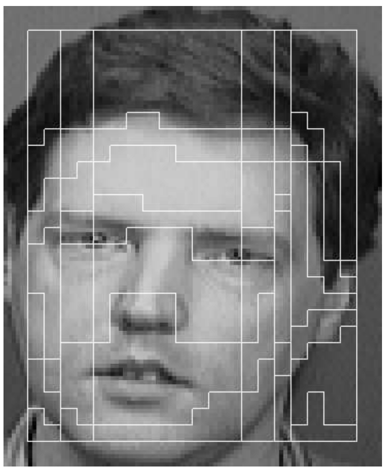

Figure 10: Segmentation (Viterbi alignment) of the image of a human face

6. SUMMARY AND CONCLUSION

The aim of this paper has been the demonstration of the use of statistical pattern recognition techniques - mainly based on Hidden Markov Models - for other purposes than speech recognition. From the outlines in this paper, it may be obvious that this approach has a tremendous potential in many application areas of human-computer interaction and multimedia information processing. We therefore believe that the research results presented in this paper may be helpful in stimulating other researchers to contribute to the further exploration of this powerful paradigm for future demanding pattern recognition tasks.

References

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. IEEE*, Vol. 77, No. 2, Feb. 1989, pp. 257–286.

[2] V. N. Gupta, M. Lenning, and P. Mermelstein, Integration of Acoustic Information in a Large Vocabulary Word Recogni zer, In *Proceedings, IEEE Intern. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, 1997, pp. 697–700.

[3] Y. He and A. Kundu, "2-D Shape Classification Using Hidden Markov Model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 11, 1991, pp. 1172–1184.

[4] R. F. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley, 1992.

[5] A. Goshtasby, "Description and Discrimination of Planar Shapes Using Shape Matrices", *IEEE Transactions on PAMI*, Vol. 7, No. 6, 1985, pp. 738–743.

[6] R. Sabourin, J.-P. Drouhard, E.S. Wah, "Shape Matrices as a Mixed Shape Factor for Off-line Signature Verification", In *Proc. Intern. Conference on Document Analysis and Recognition (ICDAR)*, 1997, Ulm (Germany), pp. 776–779.

[7] S. Lee and B. Lovell, "Modelling and Classification of Shapes in Two-Dimensions Using Vector Quantisa tion",*Proc. IEEE-ICASSP*, Vol. 5, 1994, pp. 141–144.

[8] S. Müller, S. Eickeler, G. Rigoll, "Image Database Retrieval of Rotated Objects by User Sketch," In *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL)* in conjunction with CVPR '98, Santa Barbara, 1998, pp. 40–44.

[9] S. Kuo and O. Agazzi, "Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 11, 1991, pp. 1172–1184.

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

[11] S. Müller, G. Rigoll, A. Kosmala, D. Mazurenok, "Invariant Recognition of Hand-Drawn Pictograms Using HMMs with a Rotating Feature Extraction", in 6th Intern. Workshop on Frontiers in Handwriting Recognition (IWFHR), Taejon, Korea, 1998, pp. 25-34.

[12] S. Eickeler, S. Müller, G. Rigoll, "Improved Face Recognition Using Pseudo-2D Hidden Markov Models," in Workshop on Advances in Facial Image Analysis and Recognition Technology in conjunction with 5th European Conference on Computer Vision (ECCV'98), Freiburg, Germany.

[13] S.-J. Young, "The General Use of Tying in Phoneme-Based HMM Speech Recognisers", In *Proc. IEEE Intern. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1992, pp. 569–572.

[14] S. Müller, G. Rigoll, "Searching an Engineering Drawing Database for User-specified Shapes" Submitted to *Intern. Conference on Document Analysis and Recognition (ICDAR)*, 1999.

[15] S. Eickeler, A. Kosmala, G. Rigoll, "Hidden Markov Model Based Continuous Online Gesture Recognition", In Proc. Int. Conference on Pattern Recognition (ICPR), Brisbane, 1998, pp.1755–1757.

[16] G. Rigoll, A. Kosmala, D. Willett, "An Investigation of Context-Dependent and Hybrid Modeling Techniques for Very Large Vocabulary On-Line Cursive Handwriting Recognition", 6th International Workshop on Frontiers in Handwriting Recognition (IWFHR), Taejon, Korea, August 1998.

[17] G. Rigoll, "Hybrid Speech Recognition Systems: A Real Alternative to Traditional Approaches?", Survey Lecture, Int. Workshop "Speech and Computer", pages 33-42, St. Petersburg, Russia, October 1998.

[18] S. Eickeler, S. Müller, "Content-Based Video Indexing of TV Broadcast News Using Hidden Markov Models", in Proc. IEEE Intern. Conference on Acoustics, Speech, and Signal Processing (ICASSP), Phoenix, 1999.

[19] S. Müller, S. Eickeler, G. Rigoll, "Pseudo 3-D HMMs for Image Sequence Recognition", IEEE Intern. Conference on Image Processing (ICIP), Kobe, Japan, 1999.

[20] S. Müller, G. Rigoll, "Improved Stochastic Modeling of Shapes for Content-Based Image Retrieval", IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL-99), in conjunction with CVPR-99, pp. 23-27, Fort Collins, June 1999.