In **Proc. Int. Conf. on Document Analysis and Recognition**, Seoul, Korea, 2005.

On Appearance-Based Feature Extraction Methods for Writer-Independent Handwritten Text Recognition

Gernot A. Fink Thomas Plötz
Bielefeld University, Faculty of Technology
33594 Bielefeld, Germany
{gernot,tploetz}@techfak.uni-bielefeld.de

Abstract

Most successful systems for the recognition of unconstrained handwriting currently rely on expert-crafted feature sets that compute local geometric properties from text images. However, by applying appearance based analysis techniques appropriate features could be derived from training data automatically. Therefore, in this paper several different methods for computing appearance-based feature representations are investigated and compared to the performance of a state-of-the-art writer-independent recognition system based on geometric features. In extensive experiments promising results were obtained on a challenging recognition task.

1. Introduction

During the last decades the use of so-called *segmentation-free* methods – based on Hidden-Markov Models (HMMs) – proved to be extremely successful in the recognition of printed and handwritten texts (cf. e.g. [2, 3]). The required time-linear streams of feature vectors are extracted from the word or text line images using a sliding window approach. A narrow analysis window (usually only a few pixels wide and of the height of the text image) is moved along the line to be analyzed creating a sequence of text sub-images – the so-called *frames*. For every frame appropriate features are extracted. The sequence of those feature vectors is then fed into statistical sequence analysis models, as e.g. HMMs, for training the model or for the segmentation of the text image into words or characters.

The features computed for every frame image are frequently some sort of local statistics of the grey-value distribution (cf. [1, 2]) or combinations of expert-crafted geometric properties of the small text-image slices analyzed (cf. [7, 9]). In order to explicitly take into account dynamic aspects of the feature vector sequences sometimes also approximations of discrete time derivatives of the features are computed (cf. [2]).

Interestingly, the feature extraction methods applied to the individual frame images, which are rather crucial for the performance of the subsequent modeling of handwritten text with HMMs, have not been studied systematically in the literature. Especially the use of appearance-based methods has not been investigated for unconstrained handwritten text recognition so far.

In contrast to current feature extraction methods based on the extraction of some local geometric properties appearance-based techniques can be applied in a completely data-driven manner. This is especially favorable for the challenging task of unconstrained handwritten text recognition where models are usually derived from large amounts of training data.

In this paper we, therefore, evaluate different appearance-based methods for feature extraction from frame images within a segmentation-free text recognition framework based on HMMs. We compare the achieved performance to a state-of-the-art handwriting recognition system [12] on a writer-independent recognition task using data from the IAM database [8].

In the following section we first review relevant related work. Before describing the appearance-based feature extraction methods in section 4 we present the recognition system used as a reference in section 3. The results of extensive recognition experiments performed will be presented and discussed in section 5.

2. Related Work

The most well known appearance-based pattern recognition approach is probably the so-called *Eigenface* method [11]. Face images are considered as vectors that span a low-dimensional sub-space of the high-dimensional space of general images. An appropriate representation of this *face*

space via a centroid and a certain number of base vectors can be computed by applying Principle Component Analysis (PCA) to the face images and retaining only a small number of dimensions. The discrimination of individuals – i.e. between different face classes – can be achieved by using the projection vectors as features.

For approaches to appearance-based object recognition besides PCA many different methods for computing abstract representations from example images only have been proposed (cf. e.g. [6]) including LDA, ICA, or Wavelet representations. All approaches rely on a good localization of characteristic object features in the example images. As an example, for face recognition the centers of eyes and mouth are usually aligned prior to the analysis.

The easiest way of using appearance-based techniques for handwriting recognition is to apply them to the classification of isolated characters or numerals (cf. [5]). But also frame images extracted from text lines in sliding window approaches can be subject to an appearance-based analysis, e.g. by computing an Eigen-space representation via PCA. In [4] this approach was applied to the recognition of cursively written isolated words. However, more recent publications on handwriting recognition do not consider purely appearance-based features any more. Especially for the challenging task of unconstrained handwritten text recognition a similar approach has not been investigated, yet.

Due to the much wider variability in shape and style of handwritten texts characteristic features of characters that are usually captured by analyzing geometric primitives are less well localized within the character or frame images. Consequently, appearance-based methods need to take into account a much larger degree of appearance variability which will only to a minor degree be relevant for the discrimination of different characters. It can, therefore, be expected that preprocessing and normalization methods have a much greater impact on the performance of appearance-based feature representation than on those relying on some sort of geometric abstraction.

3. Reference Recognition System

The system for unconstrained handwritten text recognition that we use as a reference for our experiments is a state-of-the-art segmentation-free recognition system based on HMMs which was successfully applied to challenging writer-independent recognition tasks [12, 13, 14].

After text line extraction the handwriting is normalized with respect to skew, baseline orientation, and slant. Additionally, a re-sizing of the line images is performed that tries to normalize the character width by scaling the image such that the average distance between local minima of the text contour equals a certain parameter (25 pixels). After binarization of the normalized text lines frames of constant width (4 pixels) and of the (varying) height of the text line are extracted with some overlap (2 pixels). On each of these frames 9 geometric features (see [12] for details) together with a discrete approximation of their first order derivatives are computed. The handwriting model consists of semi-continuous HMMs with Bakis-topology and a varying number of states for context independent characters (both upper and lower case), numerals, punctuation symbols, and white space (75 models in total). The emissions of these models in the 18-dimensional feature space are described by state-specific continuous mixture densities based on a shared set of component densities (Gaussians with diagonal covariance matrices).

4. Appearance-Based Features

In a segmentation-free text recognition framework appearance-based analysis methods can be applied for extracting features from the individual frame images that result from sliding-window processing of the text-lines. In order for an analytic transformation, as e.g. PCA, to produce useful results on such data it needs to be assured that related elements of the writing appear at roughly the same position in the frame images. Therefore, when extracting frames from normalized text lines the position of the estimated baseline is mapped to a specific position in the frame image.

Due to variation in writing style size normalization based on estimated parameters of the writing, as e.g. average character width or core height, will still produce normalized text-line images with a large variation in overall height. As appearance-based analysis techniques require input images of constant size the height variations have to be coped with during frame extraction. We investigated two possibilities: In the first configuration, for which the majority of experiments were performed, a scaling factor for the mapping of normalized text images to frames was determined such that all image content above the baseline was mapped exactly to the upper portion of the extracted frame image. The same scaling factor was used for mapping the descenders accordingly. In the case that the size of the descender area was not big enough for filling the corresponding area in the extracted frame completely the remaining pixels were assigned the maximum grey value in the source image, i.e. the background intensity. In the second configuration the frame image were not re-scaled but merely cropped from the normalized text lines. The vertical position of the cropped image region was determined by the baseline estimate. Pixels not defined via the source image were again mapped to the background intensity.

Both frame extraction procedures described above generate a sequence of frame images of constant size from the normalized text lines. These can then be directly sub-

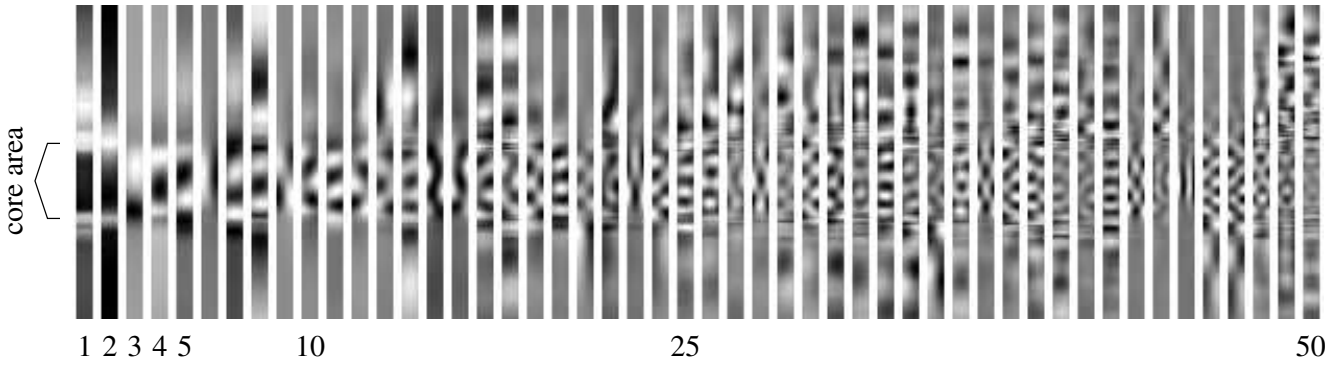

Figure 1. The first 50 Eigenvectors of the frame images

ject to an analytic image transform in order to compute appearance-based feature sets. In the work reported here we considered Principle Component Analysis (PCA) and the Discrete Wavelet Transform (DWT).

For PCA the frame images are considered as vectors in high-dimensional space. From the training data their mean and covariance matrix are computed. The Eigenvectors for the covariance matrix belonging to the largest Eigenvalues represent those directions in frame-image space that represent the largest variations in the data. Those variations are also considered to be the most characteristic aspects of the frame images with respect to the recognition of handwriting. Therefore, the projection of the frame images on those first few Eigenvectors can be used as features.

An interesting aspect of PCA is that the Eigenvectors used for the analysis can be visualized easily as if they were elements of the source data, i.e. as frame images. Such a visualization of the first 50 Eigenvectors where the vector components were re-scaled to the range of 256 grey-values is shown in Fig. 1. Especially in the first few of these "Eigenframes" one can easily see the structures corresponding to the core area of the writing analyzed.

Discrete Wavelet analysis of 2-dimensional data is based on a certain type of mother Wavelet – we use *Daubechies* of 2nd-order – and produces a representation split up into approximation and detail coefficients for the vertical and horizontal direction (cf. e.g. [10]). For a source image four blocks of coefficients – each one fourth the image size – are obtained. On the approximation coefficients obtained the Wavelet analysis can be applied recursively. As the frame images considered here are only a few pixels wide (8x128 pixel frames) only two steps of this multi-resolution analysis could reasonably be performed. Usually, when applying Wavelet transforms for feature extraction the approximation coefficients together with some of the detail coefficients are used as features. In order to obtain a certain target feature vector dimension in a more flexible way we performed an PCA on the Wavelet coefficients themselves. The projection vectors obtained from this final transform were used as features.

5. Results

In order to evaluate the performance of different appearance-based feature extraction methods we conducted a series of writer-independent recognition experiments on the IAM database of handwritten texts [8]. The database consists of several hundred documents scanned at 300 dpi which were generated by having subjects write short paragraphs of text from several different text categories. The documents collected represent truly unconstrained handwriting as no instructions concerning the writing style were given.

As in our previous experiments (cf. [12, 13, 14]) we used all documents from text categories A to D (485 documents, 4222 extracted text lines) for training and the documents from categories E and F (129 documents, 1076 extracted text lines) for testing.

After feature extraction semi-continuous HMMs with a codebook of approximately 2k densities were trained for the 75 symbol models used. During recognition the use of a lexicon or a statistical language model was deliberately avoided in order to be able to observe the effect of different feature representations without a possible bias resulting from higher order models. Consequently, no restrictions were imposed on the hypothesized character sequences. As performance measure we computed the Character Error Rate (CER) of the recognition results with respect to the reference transcription of the data.

The results of the extensive experiments are summarized in Table 5. In the upper section various configurations of appearance-based feature extraction methods are listed. The results of the reference system using geometric features are shown in the lower section of the table. Compared to the best configuration of the reference system with a CER of only 26% all appearance-based feature extraction methods

No.	Feature Type	Size Normalization[1]	Frame Extraction	Frame Size	Feature Dimension	CER
1.	PCA	avg. char. width	baseline	1x128	25	41.5%
2.		(25 pixels)	at 75%,		25+25Δ	37.5%
3.			re-scaling	4x128	25	38.2%
4.					50	38.7%
5.				8x128	25	33.8%
6.					25+25Δ	33.7%
7.					50	34.5%
8.					128	35.1%
9.	DWT + PCA			8x128	25	34.0%
10.					25+25Δ	32.8%
11.					50	33.6%
12.	PCA	avg. core height	baseline	8x128	25	40.8%
13.		(30 pixels)	at 67%,		25+25Δ	37.7%
14.			cropping		50	40.6%
15.					50+50Δ	38.0%
16.	geometric (30 pixels)	avg. core height	–	4x?	9+9Δ	38.3%
17.		avg. char. width	–	4x?	9	29.2%
18.		(25 pixels)			9+9Δ	26.0%
19.				8x?	9+9Δ	27.0%

Table 1. Comparison of different appearance-based feature sets

perform significantly worse.[2] However, the results are quite promising as for the design of these feature sets no expert knowledge was required. Both PCA-based features and those derived by Wavelet analysis can be obtained in a completely data-driven manner.

From a closer investigation of the figures the following conclusions can be drawn.

The size of the analysis window, i.e. the width of the frame, is an important parameter and affects appearance-based and geometric features differently. From frames of only a single pixel width and PCA features (experiment 1 in Table 5) to the same configuration with 8-pixel wide frames (experiment 5) the error rate is reduced by almost 20% relative (As in experiments 1 to 8 the average character width is normalized to 25 pixels we did not consider frames wider than 8 pixels). This reduction is much smaller when additional context is considered via dynamic features (10% relative reduction from experiment 2 to 6). However, for the geometric features an increase of the frame size to 8 pixels width decreases the performance. Due to the nature of this feature set frames of a single pixel width can not be used at all.

The use of dynamic features always improves performance, though this effect is much more pronounced for geometric features (10% relative reduction of CER from experiment 17 to 18) than for the best performing appearance-based configuration using PCA-transformed Wavelet coefficients (only 4% improvement from experiment 9 to 10).

Compared to the rather simple PCA-based feature sets the improvement achieved by applying a Wavelet transform is rather small (only approximately 3% improvement from experiment 6 to 10). The reason for this is most likely that the multi-resolution analysis can not be exploited fully due to the extremely small width of the frame images analyzed.

The most important observation is, however, that the combination of appropriate size normalization and frame extraction methods is crucial for the performance of both appearance-based and geometric feature sets. When nor-

1 Both normalization methods produce approximately the same total number of frames for the training set.

2 The rather high character error rates reported in the experiments are due to the extremely challenging task of unconstrained handwritten text recognition considered and the fact that *no* restrictions whatsoever were imposed on the hypothesized character sequences.

malizing the average core height instead of the average character width the performance of geometric features degrades by more than 45% relative (experiments 18 to 16). With this normalization and cropped frame images PCA-based features even outperform the geometric feature set (experiment 13). This observation suggests that more research is required with respect to a robust method for size normalization and frame extraction that optimally complements the appearance-based features computed from the frame images by PCA or DWT on highly varying writer independent handwriting data.

6. Conclusion

In this paper we presented an experimental analysis of different methods for computing appearance-based feature sets – namely using PCA or discrete Wavelet transforms – for the writer independent recognition of handwritten texts in a segmentation-free framework based on HMMs. The extensive experiments performed show that promising results with respect to the state-of-the-art reference system using geometric features could be achieved. The still existing performance gap and the observed strong impact of normalization steps indicate that these aspects need to be optimized together with the appearance-based methods applied in order to reach the performance of the geometric feature set. This would allow to design powerful feature sets for handwriting recognition in a completely data-driven manner.

7. Acknowledgment

We would like to thank the Institute of Informatics and Applied Mathematics, University of Bern, namely Horst Bunke and Urs-Viktor Marti, who allowed us to use the IAM database of handwritten forms [8] for our recognition experiments.

References

[1] K. Aas and L. Eikvil. Text page recognition using grey-level features and hidden Markov models. *Pattern Recognition*, 29(6):977–985, 1996.

[2] I. Bazzi, R. Schwartz, and J. Makhoul. An omnifont open-vocabulary OCR system for English and Arabic. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(6):495–504, 1999.

[3] H. Bunke, M. Roth, and E. G. Schukat-Talamazzini. Off-line cursive handwriting recognition using Hidden Markov Models. *Pattern Recognition*, 28(9):1399–1413, 1995.

[4] W. Cho, S.-W. Lee, and J. H. Kim. Modeling and recognition of cursive words with hidden Markov models. *Pattern Recognition*, 28(12):1941–1953, 1995.

[5] S. E. N. Correia, J. M. de Carvalho, and R. Sabourin. On the performance of wavelets for handwritten numerals recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 127–130, Québec, 2002.

[6] R. B. Fisher. CVonline: The evolving, distributed, non-proprietary, on-line compendium of computer vision. http://homepages.inf.ed.ac.uk/rbf/CVonline/.

[7] U.-V. Marti and H. Bunke. Handwritten sentence recognition. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 467–470, Barcelona, 2000.

[8] U.-V. Marti and H. Bunke. The IAM-database: An english sentence database for offline handwriting recognition. *Int. Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.

[9] A. W. Senior and A. J. Robinson. An off-line cursive handwriting recognition system. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(3):309–321, 1998.

[10] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. Wavelets for computer graphics: A primer, part 1. *IEEE Computer Graphics and Applications*, 15(3):76–84, 1995.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuro Science*, 3(1):71–86, 1991.

[12] M. Wienecke, G. A. Fink, and G. Sagerer. Experiments in unconstrained offline handwritten text recognition. In *Proc. 8th Int. Workshop on Frontiers in Handwriting Recognition*, Niagara on the Lake, Canada, August 2002.

[13] M. Wienecke, G. A. Fink, and G. Sagerer. Towards automatic video-based whiteboard reading. In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 87–91, Edinburgh, 2003.

[14] M. Wienecke, G. A. Fink, and G. Sagerer. Video-based whiteboard reading. *Int. Journal on Document Analysis and Recognition*, 2005. to appear.