# VORONOI CELL SHAPING FOR FEATURE SELECTION WITH DISCRETE HMMS

*Joachim Schenk and Gerhard Rigoll*

Institute for Human-Machine Communication
Technische Universität München
Theresienstraße 90, 80333 München
{schenk,rigoll}@mmk.ei.tum.de

**ABSTRACT**

In this paper, we introduce a novel vector quantization (VQ) scheme for distributing the quantization error equally among the quantized dimensions. Afterwards, the proposed VQ scheme is used to perform feature selection in on-line handwritten whiteboard note recognition based on discrete Hidden-Markov-Models (HMMs).

In an experimental section we show that the novel VQ scheme derives feature sets which contain less than 50 % features, enabling recognition with better performance at less computational costs. Finally, the derived feature set is compared to the quantized features selected within a continuous HMM-based system: the features selected after quantization with the proposed VQ scheme are proved to perform significantly better than those in the continuous system.

***Index Terms*—** handwriting recognition, Hidden-Markov-Models, vector quantization, feature selection

## 1. INTRODUCTION

Automatic Speech Recognition (ASR) and on-line handwriting recognition (HWR) are closely related: using a speech recognizer based on Hidden-Markov-Models (HMMs, [1]) for on-line HWR has been introduced in [2] for the first time — on the ICASSP 1986.

In a common HMM-based HWR system, each symbol (e. g. letter) is represented by a HMM. Words are recognized by combining letter-HMMs using a dictionary [3]. While high recognition rates are reported for *isolated word* recognition systems [4], performance considerably drops when it comes to recognition of whole, unconstrained handwritten *text lines* [5]. An even more demanding task is HWR of whiteboard notes which plays an important role in so-called "smart meeting room" scenarios (see e. g. [6]): when writing on a whiteboard the writer stands rather than sits and the writing arm does not rest introducing additional variation. Furthermore it has been observed that size and width of letters and words vary on a higher degree on whiteboards than on tablets [5].

Literature distinguishes between continuous and discrete HMMs. In case of continuous HMMs, the observation probability is modeled by mixtures of Gaussians [1], whereas for discrete HMMs the probability computation is a simple table look-up. In the latter case vector quantization (VQ) is performed to transform the continuous data to discrete symbols. While in ASR continuous HMMs are increasingly accepted, it remains unclear whether discrete or continuous HMMs should be used in on-line HWR of whiteboard notes. In previous work [7], we studied the use of discrete HMMs for on-line HWR of whiteboard notes and the influence of quantization on the features. It turned out that the pen's pressure information (see Sec. 2) looses significance although this feature has been proved to be important for continuous HMM-based HWR in [8] by applying the "sequential forward selection" (SFS, [9]), a common technique in feature selection [10]. We also showed that due to the design of the quantizer and the distribution of the features, although normalized, the quantization error is not distributed equally among the dimensions, indicating a varying contribution of the features to the quantization.

In this work, we select features for discrete HMM based HWR of whiteboard notes using SFS. Thereby the above mentioned unevenly balanced quantization error is taken into account: when selecting features combined with VQ, the significance of one feature can be influenced either by the quality of its quantization or by its expressiveness for the current recognition task. Hence, we describe a novel VQ scheme based on reshaping the Voronoi regions gained by the $k$-Means algorithm [11]. This is done to achieve an equal contribution of all features to the quantization, i. e. the quantization error of the normalized features is distributed equally among the dimensions.

To that end, the next section gives a brief overview of the HWR system. Section 3 reviews VQ and the SFS for feature selection. The novel VQ scheme for achieving a quantization error which is distributed equally among the dimensions is introduced in Sec. 4. The impact of the novel error shaping on the selected feature sets is shown in the experimental section (Sec. 5). Finally conclusions and discussion are presented in Sec. 6.

## 2. SYSTEM OVERVIEW

This section gives a brief overview of the recognition system used for the experiments in Sec. 5. Further details can be found in [7]. The handwritten whiteboard data is first recorded using the EBEAM-system deriving sample points $\mathbf{s}(t) = (x(t), y(t), p(t))^{\mathrm{T}}$, where $x(t)$ denotes the $x$-coordinate, $y(t)$ the $y$-coordinate, and $p(t)$ the pressure of the pen at time instance $t$. Afterwards, the recorded data is heuristically segmented into lines [5] and resampled in order to achieve a space-equidistant sampling. Then, a histogram-based skew- and slant-correction is performed, and all text lines are normalized to meet a distance of "one" between the corpus and the base line using a histogram-based projection approach.

Following the preprocessing and normalization, 24 state-of-the-art *on-line* and *off-line* features are extracted. The extracted on-line features are: the pen's "pressure", indicating whether the pen touches the whiteboard surface ($f_1$); a velocity equivalent, which is computed *before* resampling ($f_2$); the $x$- and $y$-coordinate after resampling ($f_{3,4}$); the "writing direction", i. e. the angle $\alpha$ of the strokes, coded as $\sin\alpha$ and $\cos\alpha$ ($f_{5,6}$); and the "curvature", i. e. the difference of consecutive angles $\Delta\alpha = \alpha(t) - \alpha(t-1)$, coded as $\sin\Delta\alpha$ and $\cos\Delta\alpha$ ($f_{7,8}$). Besides, on-line features describing the relation between the sample point $\mathbf{s}(t)$ to its neighbors are used:

a logarithmic transformation of the "vicinity aspect" $v$, $\text{sign}(v) \cdot \lg(1+|v|)$ ($f_9$); the "vicinity slope", i. e. the angle $\varphi$ between the line $[\mathbf{s}(t-\tau), \mathbf{s}(t)]$, whereby $\tau < t$ denotes the $\tau^{\text{th}}$ sample point before $\mathbf{s}(t)$, and the bottom line, coded as $\sin\varphi$ and $\cos\varphi$ ($f_{10,11}$); and the "vicinity curliness", the length of the trajectory normalized by $\max(|\Delta x|; |\Delta y|)$ ($f_{12}$). Finally the average square distance to each point in the trajectory and the line $[\mathbf{s}_(t-\tau), \mathbf{s}_(t)]$ is given ($f_{13}$). The off-line features are: a $3 \times 3$ "context map" to incorporate a $30 \times 30$ partition of the currently written letter's image ($f_{14-22}$); and "ascenders" and "descenders" (the number of pixels above respectively beneath the current sample point) ($f_{23,24}$). As the values of the features vary in different ranges, each dimension $d$ of the feature vector is normalized to a mean of $\mu_d = 0$ and variance of $\text{var}_d = 1$.

After feature extraction, the handwritten data is recognized by a discrete HMM-based recognizer: each symbol (each letter in this paper) is modeled by one HMM. For comparability, the HMM topology is mainly adopted from [5]. Training of the HMMs is performed by the EM algorithm. Using the Viterbi algorithm, the handwritten data is recognized and segmented [1].

## 3. VECTOR QUANTIZATION AND FEATURE SELECTION

**Vector Quantization** For using discrete HMMs, all continuous observations $\mathbf{F}$ are assigned to a stream of discrete observations $\hat{\mathbf{f}}$ via quantization, whereby the continuous, $D$-dimensional sequence $\mathbf{F} = (\mathbf{f}(1), \ldots, \mathbf{f}(T))$, $\mathbf{f}(t) \in \mathbb{R}^D$ of length $T$ is mapped to a discrete, one dimensional sequence of codebook indices $\hat{\mathbf{f}} = (\hat{f}(1), \ldots, \hat{f}(T))$, $\hat{f}(t) \in \mathbb{N}$ provided by a codebook $\mathbf{C} = (\mathbf{c}(1), \ldots, \mathbf{c}(N_{\text{cdb}}))$, $\mathbf{c}(k) \in \mathbb{R}^D$ containing $|\mathbf{C}| = N_{\text{cdb}}$ centroids $\mathbf{c}(i) \in \mathbb{R}^D$ [11]. For $D = 1$ this mapping is called *scalar*, and in all other cases ($D \geq 2$) *vector* quantization (VQ). The codebook $\mathbf{C}$ and its entries $\mathbf{c}(i)$ are derived from a training set containing $N_{\text{train}}$ training sequences $\mathbf{F}_j$, by partitioning the $D$-dimensional feature space defined by $\mathcal{S}_{\text{train}}$ into $N_{\text{cdb}}$ so-called Voronoi cells $V_i$ represented by the centroids $\mathbf{c}(i)$ [11]. In this paper, this is performed by the well known $k$-Means algorithm as described e. g. in [11]. The $N_{\text{cdb}} = 16$ Voronoi cells partitioning the space spanned by the two features $f_6$ and $f_9$ (see Sec. 2) are shown in Fig. 1 (left).

Once a codebook $\mathbf{C}$ is generated, the assignment of the continuous sequence to the codebook entries is a minimum distance search

$$\hat{f}(t) = \underset{1 \leq k \leq N_{\text{cdb}}}{\text{argmin}} \; d(\mathbf{f}(t), \mathbf{c}(k)), \tag{1}$$

where $d(\mathbf{f}(t), \mathbf{c}(k))$ is commonly the square Euclidean distance. The quality of the VQ is measured by its distortion. In this paper, the signal-to-noise ratio (SNR) is used:

$$\text{SNR} = 10 \lg \frac{\bar{S}}{\bar{E}} = 10 \lg \frac{\sum_{j=1}^{N_{\text{train}}} \sum_{t=1}^{T_j} ||\mathbf{f}_j(t)||^2}{\sum_{j=1}^{N_{\text{train}}} \sum_{t=1}^{T_j} ||\mathbf{f}_j(t) - \mathbf{c}(\hat{f}_j(t))||^2}, \tag{2}$$

with $\bar{S}$ the average, square signal amplitude, $\bar{E}$ the average, square quantization error of all observations $\mathbf{F}_j$ and $\mathbf{f}_j(t)$ the $t^{\text{th}}$ of $T_j$ feature vectors in the $j^{\text{th}}$ of $N_{\text{train}}$ sentences in the training set. Hence, the SNR is the average signal energy normalized by the distortion on a logarithmic scale [12]. As mentioned in Sec. 2, each dimension of the continuous feature vector, and therefore each feature, is normalized by its mean and variance value, yielding an average square signal amplitude of $\bar{s}_d = 1$ in each dimension. The SNR can then be expressed by the average square quantization error $\bar{e}_d$ of each feature:

$$\text{SNR} = 10 \lg \frac{\bar{S}}{\bar{E}} = 10 \cdot \left[ \lg D - \lg \left( \sum_{d=1}^{D} \bar{e}_d \right) \right], \tag{3}$$

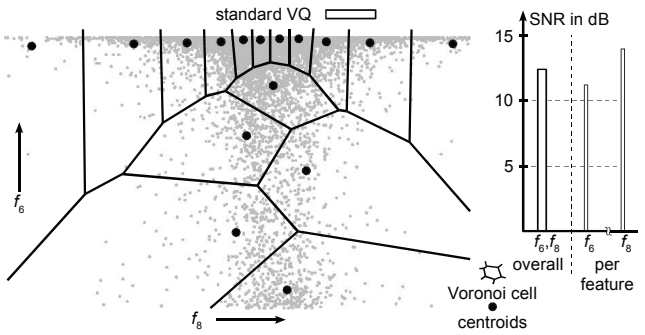


**Fig. 1**. Voronoi cells and centroids for joint VQ of the features $f_6$ and $f_9$ (left), overall and (unevenly distributed) per feature SNR (right).

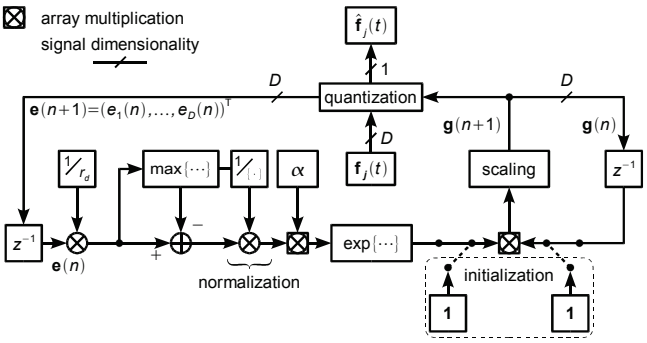


**Fig. 2**. VQ post processing depicted as control loop for achieving an equally distributed quantization error.

with $\bar{e}_d = \sum_{j=1}^{N_{\text{train}}} \sum_{t=1}^{T_j} (f_{j,d}(t) - c_{\hat{f}_j,d}(t))^2$. The overall as well as the per-dimension SNR when quantizing the features $f_6$ and $f_9$ with the centroids $\mathbf{c}(i)$ in Fig. 1 (left) is shown in Fig. 1 (right). As can be seen, although normalized, the per-feature SNR and hence, the quantization error are not equal. This has also been shown for different vector quantizers and a higher number of features in [7].

**Feature Selection** Given a set $\mathcal{F} = \{f_1, \ldots, f_D\}$ of $D$ features $f_d$, feature selection aims at deriving a new set $\mathcal{X}_k = \{x_1, \ldots, x_k\}$ containing $k \leq D$ features out of $\mathcal{F}$, in a way such that the performance of the underlying recognition system stays the same or even rises while $k$ declines [10]. To avoid the computationally infeasible number of combinations when forming all possible sets of features deducible from $\mathcal{F}$, in this paper the sequential forward selection (SFS, [9]) is used: starting with a feature set $\mathcal{X}_1 = f_d$, $1 \leq d \leq D$, which contains one single feature, the set is iteratively augmented and evaluated until all features are added, i. e. $\mathcal{X}_D = \mathcal{F}$. The evaluation objective is the character-accuracy of the system, measured on the validation set (see Sec. 5).

## 4. VORONOI CELL SHAPING

As pointed out in Sec. 3, the quantization error introduced by the quantization is not distributed equally among the dimensions. The significance of the features during feature selection is therefore either influenced by (im)proper quantization or by the feature's expressiveness: a well-suited feature quantized with high quantization error may appear less significant than a poorer feature. Hence, this section describes our approach for distributing the quantization error equally

among the dimensions, consisting of two stages: First, the centroids estimated as described in Sec. 3. Once the centroids are found, the Voronoi cells are shaped in order to achieve a distinct distribution of the quantization error.

### 4.1. Preliminaries

The vector $\mathbf{r} = (r_1, \ldots, r_D)^{\mathrm{T}}$ is introduced, which contains coefficients $r_d$ corresponding to the features $f_d$. The goal of our VQ is to provide average per dimension quantization errors $\bar{e}_d$ with

$$\bar{e}_1/r_1 = \bar{e}_2/r_2 = \ldots = \bar{e}_D/r_D. \tag{4}$$

Throughout this paper $r_d = 1$ is chosen for $1 \leq d \leq D$. The shaping of the Voronoi cells is achieved by choosing the distance measure

$$d(\mathbf{f}_j(t), \mathbf{c}(k)) = (\mathbf{f}_j(t) - \mathbf{c}(k))^{\mathrm{T}} \cdot \mathbf{G} \cdot (\mathbf{f}_j(t) - \mathbf{c}(k)), \tag{5}$$

with $\mathbf{G}$ a diagonal weight-matrix containing the weights $g_d$ of the features $f_d$ in Eq. 1. By selecting the weights $g_d$, the average quantization error $\bar{e}_d$ of the feature $f_d$ can be influenced. To show this property, the following relations between the weights $g_d$ are exemplary assumed:

$$g_1 = x \cdot g_2 = x^2 \cdot g_3 = \ldots = x^{D-1} \cdot g, \quad x, g > 1. \tag{6}$$

The feature $\mathbf{f}_j(t)$ is then assigned to $\mathbf{c}(k)$ instead of $\mathbf{c}(k')$ if

$$d(\mathbf{f}_j(t), \mathbf{c}(k)) < d(\mathbf{f}_j(t), \mathbf{c}(k')) \Rightarrow \tag{7}$$
$$\sum_{d=1}^{D} (f_{j,d}(t) - c_d(k))^2 x^{D-d} g < \sum_{d=1}^{D} (f_{j,d}(t) - c_d(k'))^2 x^{D-d} g.$$

Dividing Eq. 7 by $x^{D-1} \cdot g > 1$ yields

$$\begin{aligned} (f_{j,1}(t) - c_1(k))^2 + \ldots + \frac{(f_{j,D}(t) - c_D(k))^2}{x^{D-1}} < \\ (f_{j,1}(t) - c_1(k'))^2 + \ldots + \frac{(f_{j,D}(t) - c_D(k'))^2}{x^{D-1}}. \end{aligned} \tag{8}$$

As can be seen from the example in Eq. 8, choosing $x = 1$ lets each feature contribute to the overall distortion $d(\mathbf{f}_j(t), \mathbf{c}(k))$ by its actual distance to the corresponding centroid dimension $c_d(k)$. However, when rising the value of $x$ the contribution of higher numbered features decays. Finally, when choosing $x \to \infty$ only the distance of the first feature contributes to the distortion and is therefore minimized.

### 4.2. Weight Estimation

The analytic relation between the actual quantization error $\bar{e}_d$ of each feature $f_d$ and the corresponding weight $g_d$ is unknown. Hence, a control loop is used for recursively fitting the weights $g_d$, $1 \leq d \leq D$ to achieve an error distribution as defined by Eq. 4. The weight $g_{d_{\max}}$ of the feature $f_{\max}$ which differs most from the distribution defined by Eq. 4, $e_{d_{\max}}$ is not changed, all other weights are lowered by a factor depending on $\bar{e}_{d_{\max}} - \bar{e}_d > 0$. After an initialization, $g_d(0) = 1/D$, the weights $g_d(n)$ are recursively updated:

$$\tilde{g}_d(n+1) = g_d(n) \cdot \exp\left[\alpha \cdot \frac{\bar{e}_d(n)/r_d - \max_{1 \leq \delta \leq D} \bar{e}_\delta(n)/r_\delta}{\max_{1 \leq \delta \leq D} \bar{e}_\delta(n)/r_\delta}\right], \tag{9}$$

with $1 \leq d \leq D$ and $\alpha$ an experimentally chosen step size. The normalization by $\max(\cdot)$ in Eq. 9 is necessary, due to the variation in the absolute value of the quantization errors. In order to prevent an infinite growing of the updated weights $\tilde{g}_d(n+1)$ their values are normalized so that they meet $\sum_{d=1}^{D} g_d(n+1) = 1$.

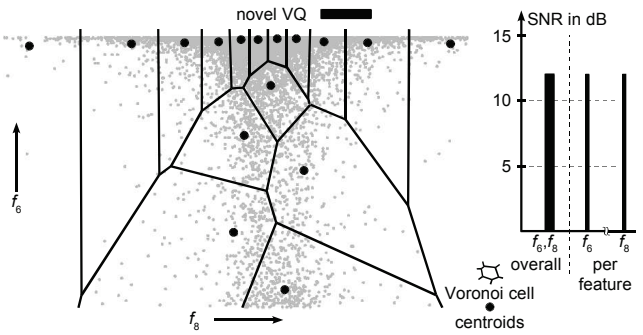


**Fig. 3**. Shaped Voronoi cells around the same centroids as in Fig. 1 (left), overall and (evenly distributed) per feature SNR (right).

The weight adaptation is continued until the change in the individual quantization errors $\bar{e}_d$ falls below a threshold. Figure 2 shows the control loop in order to find the desired weighting values $g_d$. The result of applying the proposed VQ scheme on the centroid and feature distribution as depicted in Fig. 1 is shown in Figure 3: after shaping the Voronoi cells, the SNR of both features is the same. However, this is accompanied by a slight drop in the overall SNR.

## 5. EXPERIMENTS

The experiments presented in this section are conducted on the IAM-OnDB database, containing handwritten, heuristically line-segmented whiteboard notes [13]. Comparability of the results is provided by using the settings of the writer-independent IAM-onDB-t1 benchmark, which consists of 56 different letters and provides well-defined writer-disjunct sets (one for training, two for validation, and one for testing). Statistical significance of the results is proved by the one-sided $t$-test, giving the probability $p_N$ of rejecting the hypothesis "both approaches perform equally." Two experiments are conducted.

In the first experiment (*Exp. 1*), feature selection is performed on quantized data applying the novel VQ scheme as proposed in Sec. 4. Thereby, five different codebook sizes $N_{\text{cdb}}$, with $N_{\text{cdb}} \in \{10, 100, 500, 1000, 2000\}$ are used, yielding an even distribution of the quantization error, i. e. all features contribute equally. The character-accuracy (ACC) derived from the validation set with a discrete HMM-based classifier which parameters are trained on the validation set, is shown in Fig. 4. As can be seen, for each codebook size peak performance is reached using less than the maximum of $D = 24$ features. The best performing feature set is shown as feature map: the features are placed in a $4 \times 6$ "matrix," where the feature number rises from left to right and top to bottom beginning with the feature $f_1$ in the upper left corner. A solid square (■) indicates that the current feature is part of the feature set (see Fig. 4). The best performing parameters and feature sets are used to conduct a final test on the test set of the IAM-onDB-t1 database; hence, an implicit adaptation to the test set is avoided. Table 1 summarizes the results and gives the relative improvement $\Delta r$ and the significance $p_N$ compared the results presented in [7]. Please note that the results presented here are given as *character*-ACC. In all cases, the system proposed in [7], which uses the same number of codebook entries and the standard VQ on the whole feature set *without* the novel Voronoi cell shaping, is outperformed. All improvements are highly

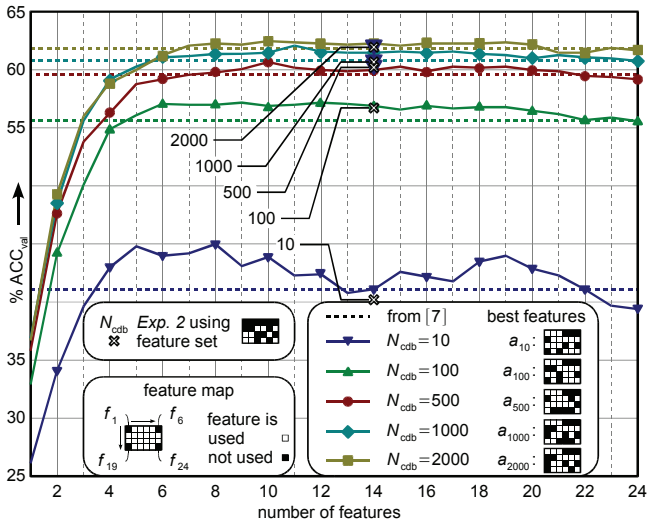


**Fig. 4**. Character-ACC for varying codebook sizes and number of selected features as well as the feature maps when feature selection is performed with and without the novel VQ. An explanation of the feature map is given and the results as presented in [7] are shown.

statistically significant. The peak performance of $a_{\mathrm{t},2000} = 68.4\,\%$, a relative improvement of $\Delta r = 1.5\,\%$ is achieved with only $k = 10$ features, i. e. less than half of the $D$ features used in [7].

Features are selected using the SFS as explained in [8] in the second experiment (*Exp. 2*) using a continuous HMM-based HWR system on the continuous features, hence the features are not vector quantized. On the derived feature set, VQ using a conventional quantizer without the novel shaping of the Voronoi Cells is applied. The peak character-ACC measured on the validation set for each codebook size as well as the feature map are shown in Fig. 4. Table 1 summarizes the results as character-ACC achieved on the test set, the relative improvement $\Delta r$, and the statistical significance of the improvement. Again, to avoid an implicit adaptation to the test set, the parameters yielding the best performance on the validation set are used for the final test. As can be seen, a different, optimal feature set is derived, achieving higher word-accuracies for all codebook sizes. The improvement is significant for $N_{\mathrm{cdb}} \in \{100, 1000, 2000\}$. As pointed out in [7], the pressure information loses significance during VQ. However, Fig. 4 shows that this feature ($f_1$) is part of the feature set used in *Exp. 2*, due to its importance in continuous systems [8]. This is the main reason for the drop in performance.

## 6. OUTLOOK

In this paper we first introduced a novel VQ scheme which is capable of deriving an even distribution of quantization errors among the quantized dimensions, i. e. each dimension contributes to the quantization process equally. We then used this VQ scheme for feature selection using the SFS. In an experimental section it has been shown that with the novel VQ scheme feature sets can be derived which perform better and utilize less features than a recently published system [7]. It turned out that systems which utilize the features of the found feature sets outperform systems which use the quantized features of a feature set which has been found by a continuous HMM system.

As pointed out in Sec. 4, our novel VQ scheme consists of two consecutive steps. First the centroids are computed, then the Voronoi cells are shaped in order to meet a certain distribution. In future work, we aim at combining the centroid computation and Voronoi cell shaping in a one stage approach in order to raise the quantization performance. The ACC-curves in Fig. 4 (especially for $N_{\mathrm{cdb}} = 10$) do not run smooth indicting the "nesting" effect of the SFS: once a feature is selected it cannot be discarded from the set $\mathcal{X}_k$. In future work, we plan to combine more sophisticated feature selection approaches such as the sequential forward floating selection (SFFS, see [10]) in conjunction with our novel VQ scheme.

| | codebook size $N_{\mathrm{cdb}}$, $a_{\mathrm{t},N_{\mathrm{cdb}}}$ | | | | |
|---|---|---|---|---|---|
| | 10 | 100 | 500 | 1000 | 2000 |
| *Exp. 1* | 50.0 % | 62.9 % | 66.4 % | 67.2 % | 68.4 % |
| [7] | 47.0 % | 61.3 % | 65.1 % | 66.4 % | 67.4 % |
| $\Delta r$ | 6.0 % | 2.5 % | 2.0 % | 1.2 % | 1.5 % |
| $p_N$ | $> 0.99$ | $> 0.99$ | $> 0.99$ | 0.99 | $> 0.99$ |
| *Exp. 2* | 45.5 % | 62.8 % | 66.1 % | 66.5 % | 67.8 % |
| $\Delta r$ | -3.3 % | 2.4 % | 1.5 % | 0.2 % | 0.6 % |
| $p_N$ | $> 0.99$ | $> 0.99$ | $> 0.99$ | 0.62 | 0.89 |

**Table 1**. Results of the experiments *Exp. 1*, *Exp. 2*, and those as presented in [7] measured in *character*-ACC on the *test* set as well as the relative improvement $\Delta r$ and the significance $p_N$.


### ACKNOWLEDGMENTS

The authors sincerely thank H. Schenk for his vital comments.


# References


[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257 – 285, 1989.

[2] R. Nag, K. Wong, and F. Fallside, "Script Recognition Using Hidden Markov Models," *ICASSP*, vol. 11, pp. 2071 – 2074, 1986.

[3] R. Plamondon and S.N. Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey," *TPAMI*, vol. 22, no. 1, pp. 63 – 84, 2000.

[4] J. Schenk and G. Rigoll, "Novel Hybrid NN/HMM Modelling Techniques for On-Line Handwriting Recognition," *IWFHR*, pp. 619 – 623, 2006.

[5] M. Liwicki and H. Bunke, "HMM-Based On-Line Recognition of Handwritten Whiteboard Notes," *IWFHR*, pp. 595 – 599, 2006.

[6] A. Waibel, T. Schultz, M. Bett, I. Rogina, R. Stiefelhagen, and J. Yang, "SMaRT: The Smart Meeting Room Task at ISL," *ICASSP*, vol. 4, pp. 752 – 755, 2003.

[7] J. Schenk and G. Rigoll, "Neural Net Vector Quantizers for discrete HMM-Based On-Line Handwritten Whiteboard-Note Recognition," *ICPR*, p. in press, 2008.

[8] M. Liwicki and H. Bunke, "Feature Selection for On-Line Handwriting Recognition of Whiteboard Notes," *IGS*, pp. 101 – 105, 2007.

[9] A. W. Whitney, "A Direct Method of Nonparametric Measurement Selection," *IEEE TC*, vol. 20, no. 9, pp. 1100 – 1103, 1971.

[10] M. Kudo and J. Sklansky, "Comparison of Algorithms that Select Features for Pattern Classifiers," *PR*, vol. 33, no. 1, pp. 25 – 41, 2000.

[11] J. Makhoul, S. Roucos, and H. Gish, "Vector Quantization in Speech Coding," *Proc. IEEE*, vol. 73, no. 11, pp. 1551 – 1588, 1985.

[12] R.M. Gray, "Vector Quantization," *IEEE ASSP*, pp. 4 – 29, 1984.

[13] M. Liwicki and H. Bunke, "IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard," *ICDAR*, vol. 2, pp. 1159 – 1162, 2005.