Writer Identification and Verification Using GMM Supervectors

Vincent Christlein David Bernecker Florian Hönig Elli Angelopoulou
Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany
{firstname.lastname}@cs.fau.de

Abstract

This paper proposes a new system for offline writer identification and writer verification. The proposed method uses GMM supervectors to encode the feature distribution of individual writers. Each supervector originates from an individual GMM which has been adapted from a background model via a maximum-a-posteriori step followed by mixing the new statistics with the background model.

*We show that this approach improves the TOP-*1 *accuracy of the current best ranked methods evaluated at the ICDAR-2013 competition dataset from* 95.1% *[13] to* 97.1%*, and from* 97.9% *[11] to* 99.2% *at the* CVL *dataset, respectively. Additionally, we compare the GMM supervector encoding with other encoding schemes, namely Fisher vectors and Vectors of Locally Aggregated Descriptors.*

1. Introduction

Similarly to faces [8] or speech [21], handwritten texts can serve as a biometric identifier. Naturally, the question of authenticity of documents plays an important role for law enforcement agencies. However, writer identification and verification have recently also gained attention in the field of historical document analysis [4, 5]. Writer identification attempts to identify the author of a document, given a known set of writers. In contrast, writer verification seeks to answer the question of authenticity, i. e. whether two given documents are written by the same author or not.

Methods for writer identification (and verification) can be roughly categorized into two groups [6]: The first group contains methods which use global statistics to describe handwriting, e. g. the angle of direction and the scribe width [4, 6, 24]. These are also denoted as textural features. The second group consists of methods operating on the level of allographs, i. e. the writer is described by the vocabulary of small parts of letters [6, 11, 13, 23, 24]. The proposed method belongs to this second group.

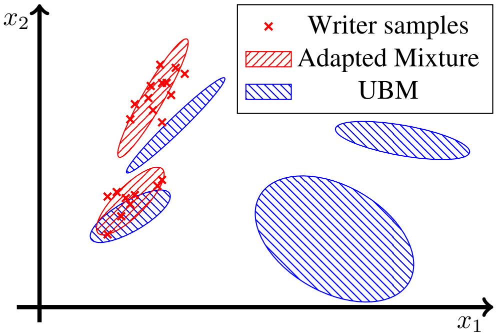


Figure 1: The Universal Background Model (blue) is adapted to samples from one document.

We propose to model the characteristics of each writer by describing the global distribution of feature vectors computed from local image patches by a generative model. We use the so-called GMM-UBM-supervector method, a well-known approach in the field of speaker verification. In speech analysis, the overall distribution of short-time spectral feature vectors of all speakers is modeled by a Gaussian Mixture Model (GMM). The GMM captures the domain's speech style in general, and is therefore termed the Universal Background Model (UBM). In order to describe the speaker of a particular utterance, a maximum-a-posteriori (MAP) adaptation of the UBM to the feature vectors from that utterance is performed [21]. It turns out that stacking the parameters of the adapted GMM (i. e. means, covariances, and weights) in a so-called *supervector* yields an excellent feature vector for characterizing the given speaker, be it for identification [7] or other purposes such as age determination [3]. We adapt this approach by replacing the short time spectral feature vectors by RootSIFT descriptors [1] and using their distribution to characterize the writer of an unknown document.

We show that the combination of RootSIFT descriptors and GMM supvervector encoding outperforms the current state of the art on two publicly available databases, namely the ICDAR13 database [16] and the CVL database [15]. Furthermore, we show that the GMM supervector encoding scheme surpasses other recently proposed encoding schemes

2014 IEEE Winter Conference on Applications of Computer Vision (WACV)

like Fisher vectors [19] or Vectors of Locally Aggregated Descriptors (VLAD) [14, 2] on these databases in most of the evaluated cases.

The paper is organized as follows: Sec. 2 describes related work. The details of our proposed method are presented in Sec. 3. The datasets and evaluation are included in Sec. 4.

2. Related Work

Siddiqi and Vincent [24] have presented a method which uses three different kinds of features: global features, polygon features and codebook features. A universal codebook, computed with the k-means clustering algorithm, and additionally, a local codebook, computed for each document individually by hierarchical clustering, is formed from image patches. The authors achieved higher results by using the universal codebook. In contrast to this work we adapt an UBM to form an individual codebook, which is described by a GMM, for each new document.

The current best performing method proposed by Jain and Doermann [13] also uses document specific codebooks. For each document a vocabulary of contour gradient descriptors is computed using k-means. The gradient descriptors are computed from segmented characters or parts of characters (allograph). The allographs themselves are computed from vertical cuts or seam cuts.

Closest to our approach is the work by Schlapbach et al. [22] on online handwriting recognition. At first, they build an UBM by estimating a GMM, and then adapt a GMM for each recorded handwriting. The similarity between two recordings is measured by using the sum of posterior probabilities of each mixture. In contrast to this work we employ RootSIFT descriptors and construct supervectors from the adapted GMMs, c. f. Sec. 3.

Another form of encoding was employed by Fiel and Sablatnig [11]. They first compute slightly modified SIFT features. Then, a GMM is computed from a training set, serving as vocabulary. Using this vocabulary, the data of each document is encoded using improved Fisher vectors [20]. The similarity between handwritten documents is computed using the cosine distance of the corresponding Fisher vectors.

3. Methodology

The general framework of our proposed approach works as follows: first, RootSIFT features for each document are computed. Descriptors from an independent document dataset are used to train a vocabulary, i. e. our UBM, analogous to the typical bag-of-words approaches. Afterwards, the UBM is adapted to each test document individually. The new statistics are stacked into a so-called supervector to form a feature vector for each document. The remainder of this section provides the details of our features, the construction of the UBM, the adaptation process, and the normalization of the supervector.

Features: SIFT descriptors are widely used, e. g. in the related fields of image forensics [9], or object retrieval [1, 2], and have also been used for writer identification [11].

We employ RootSIFT features [1], i. e. a variant of SIFT where the features are additionally normalized using the square root (Hellinger) kernel. Since SIFT vectors are composed of histograms, the Euclidean distance between two vectors can be dominated by the large bin values. This effect is reduced by using the Hellinger distance instead.

In practice this is achieved by applying the L1-norm followed by an element-wise application of the square-root. Note that the descriptors could also be densely sampled [1, 2], however we evaluate the SIFT descriptors at the originally proposed SIFT keypoints [17]. In this way, SIFT features describing the document background are omitted. This can be seen as analogous to the speech activity detector for speaker recognition.

UBM: The UBM is created by estimating a GMM from a set of SIFT descriptors $\mathbf{X} = \{\mathbf{x}_1 \dots \mathbf{x}_T\}$ computed from training documents. Given a feature vector $\mathbf{x}$, its likelihood function is defined as

$$p(\mathbf{x} \,|\, \lambda) = \sum_{i=1}^{N} w_i g_i(\mathbf{x}\,;\, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)\,. \tag{1}$$

It consists of a sum of N mixtures (weighted Gaussians $g_i(\mathbf{x}) := g_i(\mathbf{x}\,;\,\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$), and is described by the set of parameters $\lambda = \{w_i,\, \boldsymbol{\mu}_i,\, \boldsymbol{\Sigma}_i \,|\, i = 1, \dots, N\}$, where $\sum_{i=1}^{N} w_i = 1$.

The GMM parameters are estimated using the expectation-maximization (EM) algorithm [10]. The parameters λ of the UBM are iteratively refined to increase the log likelihood $\log p(\mathbf{X} \,|\, \lambda) = \sum_{t=1}^{T} p(\mathbf{x}_t \,|\, \lambda)$ of the model for the set of training samples $\mathbf{X}$. For computational efficiency, the covariance matrix $\boldsymbol{\Sigma}_i$ is assumed to be diagonal. The vector of diagonal elements will be referred to as $\boldsymbol{\sigma}_i^2$.

GMM adaptation and mixing: The final UBM is adapted to each document individually, using all M SIFT descriptors computed at document W, $\mathbf{X}_W = \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. For the MAP estimation, first the posterior probabilities

$$\pi_j(i) = p(i \,|\, \mathbf{x}_j) = \frac{w_i g_i(\mathbf{x}_j)}{\sum_{k=1}^{N} w_k g_k(\mathbf{x}_j)}\,, \tag{2}$$

are computed for all mixtures i and each feature vector $\mathbf{x}_j$, $j \in \{1 \dots M\}$. Up next, the mixture parameters are adapted. Mixtures with high posteriors are adapted more strongly (c. f.

Fig. 1). This is controlled by a fixed relevance factor r^τ for the adaptation coefficients

$$\alpha_i^\tau = \frac{\sum_{j=1}^{M} \pi_j(i)}{\sum_{j=1}^{M} \pi_j(i) + r^\tau} \tag{3}$$

for each parameter τ $\big(\tau \in \{w,\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}\}\big)$. Note that we set r fixed for each τ as suggested by Reynolds et al. [21] (τ is omitted subsequently). The mixture parameters are adapted according to (for reasons of clarity, let $\mathbf{x}^2 = \mathbf{x} \odot \mathbf{x}$, $\boldsymbol{\mu}^2 = \boldsymbol{\mu} \odot \boldsymbol{\mu}$, and $\hat{\boldsymbol{\mu}}^2 = \hat{\boldsymbol{\mu}} \odot \hat{\boldsymbol{\mu}}$, where $\odot$ denotes the element-wise multiplication):

$$\hat{w}_i = \gamma\Big(\frac{\alpha_i}{M} \sum_{j=1}^{M} \pi_j(i) + (1 - \alpha_i) w_i\Big) \tag{4}$$

$$\hat{\boldsymbol{\mu}}_i = \alpha_i \frac{\sum_{j=1}^{M} \pi_j(i) \mathbf{x}_j}{\sum_{j=1}^{M} \pi_j(i)} + (1 - \alpha_i) \boldsymbol{\mu}_i \tag{5}$$

$$\hat{\boldsymbol{\sigma}}_i^2 = \alpha_i \frac{\sum_{j=1}^{M} \pi_j(i) \mathbf{x}_j^2}{\sum_{j=1}^{M} \pi_j(i)} + (1 - \alpha_i)({\boldsymbol{\sigma}^{\mathbf{2}}}_i + \boldsymbol{\mu}_i^2) - \hat{\boldsymbol{\mu}}_i^2 \tag{6}$$

where γ is a scaling factor ensuring that the weights of all mixtures sum up to one.

For the task of writer identification the parameters of the GMM adapted to one document are stacked into the supervector

$$\mathbf{s} = (\hat{w}_1,\, \ldots,\, \hat{w}_N,\, \hat{\boldsymbol{\mu}}_1^T,\, \ldots,\, \hat{\boldsymbol{\mu}}_N^T,\, \hat{\boldsymbol{\sigma}}_1^T,\, \ldots,\, \hat{\boldsymbol{\sigma}}_N^T)^T\,. \tag{7}$$

Normalization: Each GMM supervector is normalized by two normalization steps: first the square root is computed element-wise and then each vector is L2-normalized. Note, this is similar to the normalization scheme proposed by Perronnin et al. [20] for Fisher vectors. For the distance between two encoding vectors the cosine distance is employed.

The GMM supervector of a query document can thus be used for the comparison with other supervectors computed from documents of known authorship. For the identification of the authorship, each distance from the query supervector to all other supervectors of the database is computed. The resulting list of distances is then sorted. Either the list can be further analyzed, e. g. inspecting the first 10 documents, or the author belonging to the smallest distance is assigned to the query document. For writer verification, the distance between the supervector of the query document and a supervector of a document of known authorship is computed and upon a decision threshold it is decided whether the document is written by the same person or not. A suitable decision threshold is typically chosen from the ROC curves, c. f. Sec. 4.2, in a manner so that it fits a certain guideline, e. g. to meet a specific verification rate.

4. Evaluation

We compare our results with two other encoding methods: Fisher vector (FV) encoding [19] and the encoding in Vectors of Locally Aggregated Descriptors (VLAD) [14].

Fisher vectors are derived from Fisher kernels. The idea is to transform samples, whose distribution is described by a generative model, to a high dimensional feature space, namely the gradient space of the model parameters. The Fisher vectors used in this evaluation are in principle the Fisher scores of the samples normalized by the square-root of the Fisher information matrix [19]. These vectors can then be used for classification using a discriminative classifier or can simply be compared using e. g. the cosine distance. As suggested by Perronnin et al. [20] the Fisher vectors are also further L2-normalized to improve their performance.

In a supervised setting, when the generative model contains the class label as a latent variable, Jaakkola and Haussler [12] have shown that a discriminative classifier using Fisher kernels is at least as good as the MAP labeling using the generative model. However, the Fisher vectors we use lose this advantage since the generative model is estimated in an unsupervised manner.

In contrast to Fisher vectors, VLAD is a non probabilistic version of the Fisher kernel encoding only first order statistics [14]. However, Arandjelovi and Zisserman show that additional normalization steps can improve the performance and consequently outperform systems using Fisher vectors [2]. Note that we do not employ cluster center adaptation since we do not have large variety between the training set and the testing set. However we used intra-normalization (component-wise L2-normalization). Also note that the VLAD encodings are computed from the UBM, i. e. the clusters are not quantized in a hard way. Instead, a soft quantization scheme is computed using the posteriors computed from the UBM.

In the following subsections, we first introduce the three evaluation datasets and then describe our evaluation metrics. After the determination of suitable GMM parameters and adaptation parameters, we present our results for writer identification, verification, and compare our methods with the state of the art.

4.1. Benchmark datasets and evaluation procedure

For the evaluation the publicly available CVL, ICDAR13 and IAM datasets were used. Example lines are shown in Fig. 2.

ICDAR13 [16] is part of the ICDAR 2013 Writer Identification Competition. It consists of an experimental dataset and a benchmark dataset. The experimental dataset consists of 100 and the benchmark set of 250 writers with four documents per writer. Two documents were written in Greek, the

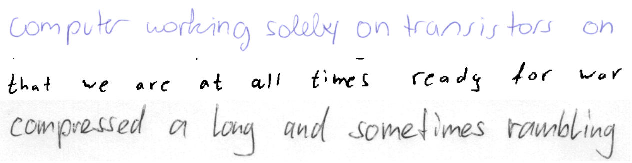

computer working solely on transistors on
that we are at all times ready for war
compressed a long and sometimes rambling

Figure 2: Example lines of the three datasets, from top to bottom: CVL, ICDAR13 and IAM.

other two in English. The documents of the dataset are in binary image format.

CVL [15] consists of handwritten English texts by 309 writers, with five documents per writer. One document is written in German and the remaining four are written in English. In contrast to the ICDAR13 dataset the documents are captured in color. However, in the evaluation the color of the handwriting is ignored and grayscale versions of the images are used.

CVL +ICDAR13: from the ICDAR13 benchmark set and the CVL dataset we created a combined dataset consisting of 559 scribes with four documents per writer resulting in 2236 documents, i. e. we omitted one document for each writer of the CVL database. Note, in the remainder of the text we refer to this dataset as MERGED.

Vocabularies: For the generation of the UBM for the CVL and the ICDAR13 evaluation, we used both times the ICDAR13 experimental dataset. To see the influence of the background model, we also trained a second UBM using samples from the IAM dataset [18]. There, a subset of 356 writers, which contributed one page each, was used. Thus, for the experiments conducted on the MERGED database we show the results using a) the UBM estimated from the ICDAR13 experimental and b) estimated from the IAM database.

4.2. Evaluation Metrics

We will express the results of our evaluation in mean average precision (mAP), receiver operating characteristic (ROC) and cumulative match characteristic (CMC) curves.

Mean average precision is a measure used in the context of information retrieval. First, the average precision (aP) for each query of size Q, of which R documents are relevant, is calculated by

$$\mathrm{aP} = \frac{1}{R} \sum_{r=1}^{Q} \Pr(k) \cdot \mathrm{rel}(k) \,, \qquad (8)$$

where $\Pr(k)$ is the precision at rank k of the query (i. e. number of relevant documents in the first k query items divided by k), and $\mathrm{rel}(k)$ is a binary function that is 1 when the document at rank k is relevant, and 0 otherwise. Consequently, the soft TOP-k accuracy refers to the aP at a specific rank k.

[TOP-1]	SV	FV	VLAD
SIFT	0.943	0.828	0.861
RootSIFT	0.972	0.945	0.931

Table 1: Comparison of SIFT and RootSIFT for the three proposed encoding schemes.

The mAP is calculated by averaging over the aP values of all queries. mAP assigns higher values to methods that return relevant documents at low ranks in a query. Note, this is related to the writer retrieval criterion which has recently been used [11, 15].

The ROC curve describes the verification rate, i. e. it compares two adapted GMMs in a two class classification problem, where the question is whether the GMMs describe the same writer or not. Hereby, the verification rate (or true positive rate) is plotted as a function of the false positive rate for different cut-off points. Each point on the ROC curve represents a pair corresponding to a particular decision threshold, i. e. a cut-off threshold on the distances between the GMMs. The closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test.

The CMC curve depicts the $1 : k$ identification rate. It plots the identification rate as a function of the rank k, i. e. it shows the probability that a subject is under the first k scribes with the corresponding k highest scores. Note, this is related to to the soft Top-k criterion of the ICDAR 2013 competition [16].

4.3. SIFT vs. RootSIFT

This and the following experiment have been conducted on the ICDAR13 dataset. Unless being evaluated, the parameters are set to: $r = 16$, $N = 64$ and a full supervector, i. e. weights, means, and variances are taken.

Arandjelovic and Zisserman [1] showed that the SIFT descriptor profits from an additional normalization scheme using the square root (Hellinger) kernel. Table 1 shows the benefit of using this Hellinger normalization over the normal SIFT descriptor. Especially the Fisher vectors and VLAD encodings improve by a large margin.

4.4. Method Parameters

In this experiment the optimal values for the GMM parameters are evaluated. The experiment is divided into evaluating the number of mixtures N, the relevance factor r and the feature combinations for the GMM supervector (SV) construction, i. e. a supervector solely formed by the covariances (c), or means (m), or by the combination of both (mc), or by adding the weights (w). The optimal parameters in terms of

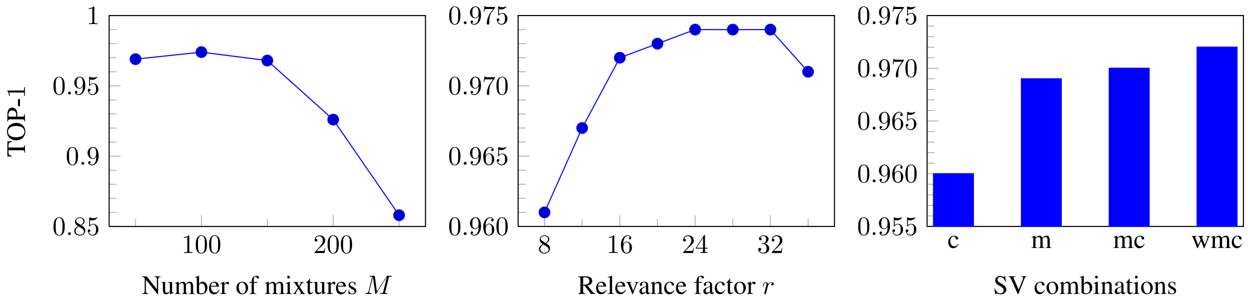


Figure 3: Evaluation of the GMM and mixing parameters on ICDAR13

	TOP-1	mAP
SV	0.97	0.666
SV + SR normalization	0.971	0.671

Table 2: Influence of GMM supervector (SV) normalization. Top: plain GMM supervector + L2 normalization, bottom: GMM supervector + (element wise) square root normalization + L2 normalization.

TOP-1 accuracy on the ICDAR13 dataset are found to be a number of mixtures $N = 100$ with a relevance factor $r = 28$ and a supervector consisting of all the weights, means and covariances (c. f. Fig. 3).

4.5. GMM Supervector Normalization

We conducted a small experiment to show the benefit of a appropriate normalization. The GMM supervectors are first normalized by taking the square root element-wise (SR) followed by the L2 normalization of the whole vector, we compare this with the results when omitting the SR normalization, i. e. the supervectors are only normalized by their L2 norm. Table 2 shows that in terms of mAP the method greatly improves using this normalization scheme.

4.6. Results

We now present the results, first in terms of writer identification and then in terms of writer verification. Where possible, we compare our approach with the current state of the art. Furthermore, we compare to the two other encoding schemes: Fisher vectors and VLAD.

4.6.1 Writer identification

Writer identification attempts to find the author of a document from a set of known authors. The identification rate can be described in terms of a *soft* evaluation, i. e. under the first k documents at least one document of the author in question must occur. This is depicted in the CMC curves of Fig. 4. Comparing Fig. 4a with Fig. 4b it can be noticed that in general the ICDAR13 dataset is more challenging than the CVL database. In all three experiments the proposed GMM supervectors achieve the highest identification rates. At the CVL database the other two encoding schemes (Fisher vectors and VLAD) outperform the current best performing method by Fiel and Sablatnig [11] (CV) as well. In contrast, when evaluating the methods on ICDAR13, the current best performing method, denoted as CS, by Jain et al. [13] is better than using Fisher vectors and VLAD. Still, our proposed method performs overall the best on ICDAR13.

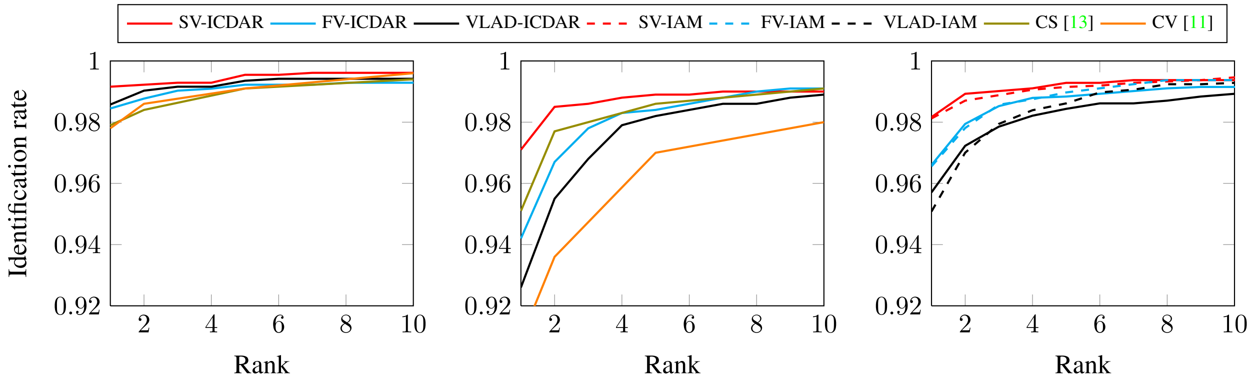
	TOP-1	TOP-2	TOP-3	TOP-4	mAP
CV [11]	0.978	0.956	0.894	0.758	–
CS [13]	0.979	0.90	0.71	0.483	–
VLAD	0.986	0.954	0.871	0.720	0.936
FV	0.984	0.952	0.880	0.756	0.940
SV	**0.992**	**0.981**	**0.958**	**0.887**	**0.971**

Table 3: Hard criterion evaluated on CVL. The proposed method (SV) is compared with the state of the art CV, CS and with two other encoding methods FV, VLAD.

Fig. 4c depicts that GMM supervectors are better than Fisher vectors followed by VLAD. It also shows that using the UBM trained on ICDAR13 is slightly better than the UBM computed from the IAM.

For a better comparison with the state of the art we show the measurements in terms of hard TOP-1–3 and hard TOP-1–4 for the ICDAR13 and CVL databases, respectively. Given a query document, the hard TOP-k rates denote that under the first k nearest retrieved documents from the database exactly k documents are written by the same author. Furthermore, we present the mAP value for all three datasets to express the overall retrieval performance of each method.

Table 3 shows the results of the hard criterion evaluated on the CVL database. We compare the three encoding schemes: GMM supervectors (SV), Fisher vectors (FV), and VLAD. Additionally, we list the best ranked method of the ICDAR13 competition, denoted as "CS" [13], as well as the best ranked



(a) CVL (b) ICDAR13 (c) Merged

Figure 4: CMC Curves showing the writer identification performance of the proposed method (SV) evaluated at three different datasets in comparison to two other encoding schemes FV and VLAD. Additionally, the identification rater of the current best method of the ICDAR13 competition by Jain et al. [13] (CS), taken from the ICDAR13 competition evaluation [16], and the current best method of the CVL database by Fiel and Sablatnig [11] (CV) are included. Note that only the values for TOP-1, TOP-2, TOP-5, and TOP-10 were given for these two methods. For the MERGED database (c) we compare the encoding schemes using two different vocabularies, computed from i) ICDAR13 and ii) IAM.

	TOP-1	TOP-2	TOP-3	mAP
CV [11]	0.909	0.448	0.245	–
CS [13]	0.951	0.196	0.071	–
VLAD	0.926	0.429	0.248	0.651
FV	0.942	**0.475**	**0.25**	**0.677**
SV	**0.971**	0.428	0.238	0.671

Table 4: Hard criterion evaluated on ICDAR13. The proposed method using GMM supervectors (SV) is compared with the methods by Fiel and Sablatnig [11] (CV) and Jain et al. [13] (CS) as well as with Fisher vectors (FV) and VLAD.

	TOP-1	TOP-2	TOP-3	mAP
VLAD-ICDAR	0.957	0.713	0.571	0.814
FV-ICDAR	0.966	0.738	0.588	0.833
SV-ICDAR	**0.982**	0.733	0.624	0.841
VLAD-IAM	0.951	0.719	0.593	0.818
FV-IAM	0.966	**0.751**	0.608	0.841
SV-IAM	0.981	0.745	**0.638**	**0.849**

Table 5: Hard criterion evaluated on the MERGED DB. In the top three experiments, the UBM has been created using the ICDAR13 experimental, in the bottom three, the UBM has been created using the IAM dataset. The proposed method using GMM supervectors (SV) is compared with Fisher vectors (FV) and VLAD.

method of the CVL database, denoted as “CV” [11]. As Table 3 shows, the proposed method clearly surpasses the current best ranked methods. It is followed by the other two encoding schemes Fisher vectors and VLAD which do not differ much in their performance.

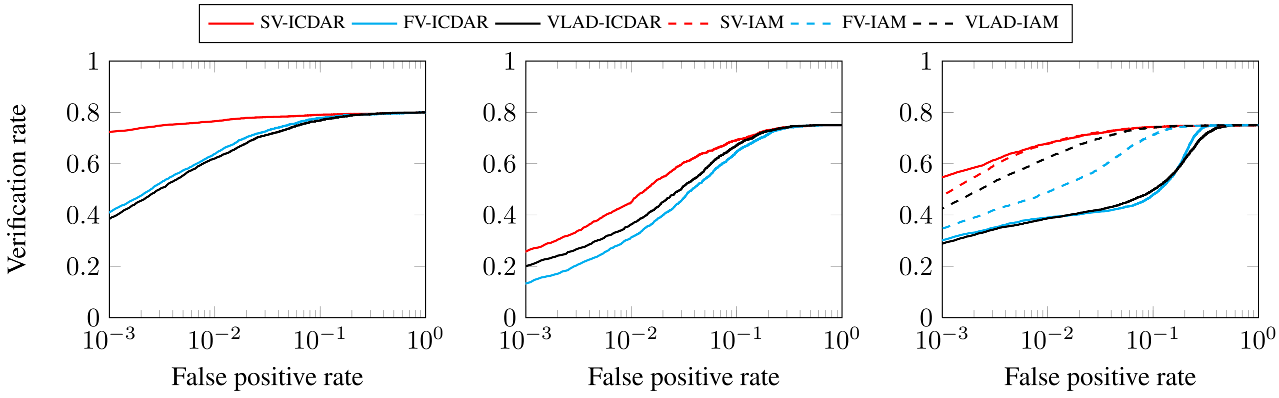
The results from the evaluations using the ICDAR13 benchmarking set are shown in Table 4. Here, a slightly different outcome can be noted. The proposed method using GMM supervectors is still the best in terms of TOP-1 accuracy. However Fisher vector encoding is slightly better in terms of mAP, TOP-2 and TOP-3.

Naturally, the results on the MERGED dataset slightly differ depending on whether the vocabulary was computed from the ICDAR13 experimental set or the IAM dataset, c. f. Table 5. In terms of mAP the vocabulary computed from the completely independent dataset IAM slightly improved the results of all three methods. However, the TOP-1 rates of VLAD and the GMM supervectors deteriorate slightly.

4.6.2 Writer verification

Writer verification decides for each pair of documents whether they were written by the same author or not. In contrast to the identification problem, this is a binary classification and thus can be evaluated with a ROC curve.

The ROC curve of Fig. 5a depicts the verification rates when evaluating the GMM supervectors (SV), Fisher vectors (FV) and VLAD on the CVL database. While the pro-



(a) CVL

(b) ICDAR13

(c) MERGED

Figure 5: ROC curves showing the writer verification performance of the proposed encoding scheme (SV) evaluated at three datasets. They are compared with Fisher vectors (FV) and VLAD. For the MERGED database we compare the encoding schemes using two different vocabularies, computed from i) ICDAR13 and ii) IAM. Please note the logarithmic scaling of the x-axis.

posed method is by a large margin better than Fisher Vectors and VLAD at lowe false positive rates, these two encoding schemes perform very similar.

The results for the evaluation of the ICDAR13 dataset are depicted in Fig. 5b. As earlier in Fig. 5a, the proposed method exceeds the other two encoding schemes.

Comparing the two databases ICDAR13 and CVL with each other, the ICDAR13 database seems to be much more competitive than the CVL database. Two explanations may be given for this: on one hand, the ICDAR13 dataset contains only binary data, while the CVL database is composed of colored images. Although the color information is lost due to the grayscale conversion, the gradient magnitudes are still more variable among the dataset, and thus, the SIFT descriptors might be more discriminative for a single author. On the other hand, the ICDAR13 images contain less text than the CVL database. Consequently, fewer feature vectors per images are computed for the ICDAR13 database.

As Fig. 5c shows, the performance of the methods depends on the dataset used for training the UBM when evaluating on the MERGED dataset. While the IAM dataset seems to be favorable in terms of mAP, c. f. Table 5, the GMM supervectors slightly lose the accuracy at low false positive rates. In contrast, Fisher vector encoding, and especially VLAD, improve using this database as vocabulary. VLAD even outperforms the Fisher vector encoding, which corroborates the results of Arandjelovi et al. [2], i. e. with an additional center normalization the results of VLAD could be improved if the vocabulary has been trained on a dataset with different characteristics from the test set.

If the ICDAR13 experimental set is used as UBM, the results for the supervector encoding slightly improve at very low false positive rates, while the other two databases lose performance. Please also note that in terms of equal error rate (EER), i. e. where the false positive rate is equal to the verification rate, no big difference can be determined, all three methods have an EER of a) CVL 0.20 b) ICDAR13 0.25 and c) MERGED 0.25 (note the x-axis is denoted with logarithmic scale, thus the EER does not lie on the diagonal of the plot).

5. Conclusion

We have presented a system for the problem of writer identification and writer verification that is based on building a generative model from similar data. The model is used to extract features for individual documents following the supervector idea, which is adapted from the field of speaker recognition. As an universal background model a GMM is used that is later adapted to each query document individually. Adaptation is achieved by first performing one MAP step followed by mixing the new computed statistics with the ones of the UBM. The parameters of the adapted GMM are stacked together to form the GMM supervector. After suitable normalization the supervector is used as a discriminating feature vector for one document.

We could show that this approach achieves the best scores in terms of writer identification and verification, outperforming the current best methods. In particular, we could show that thanks to the adaptation and mixing step the proposed supervector encoding outperforms the two other tested encoding schemes: Fisher vectors and VLAD. On the ICDAR13 and CVL datasets our method improves the TOP-1 accuracy from the best previously published results from 95.1% to 97.1%, and from 97.9% to 99.2%, respectively.

As part of our future work, we want to evaluate the effect of replacing SIFT with other feature descriptors. Furthermore, the GMM supervector encoding could be further improved by combining multiple GMM supervectors, adapted from different vocabularies.

Acknowledgments

This work has been supported by the German Federal Ministry of Education and Research (BMBF), grant-nr. 01UG1236a. The contents of this publication are the sole responsibility of the authors.

References

[1] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2911–2918, Providence, RI, June 2012. 1, 2, 4

[2] R. Arandjelovic and A. Zisserman. All about VLAD. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1578 – 1585, Portland, OR, June 2013. 2, 3, 7

[3] T. Bocklet, A. Maier, and E. Nöth. Age and Gender Recognition for Telephone Applications Based on GMM Supervectors and Support Vector Machines. In *Acoustics, Speech and Signal Processing. ICASSP 2008. IEEE International Conference on*, pages 1605 – 1608, Las Vegas, NV, Mar. 2008. 1

[4] A. Brink, J. Smit, M. Bulacu, and L. Schomaker. Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 45(1):162–171, Jan. 2012. 1

[5] M. Bulacu and L. Schomaker. Automatic Handwriting Identification on Medieval Documents. In *14th International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 279–284, Modena, Sept. 2007. 1

[6] M. Bulacu and L. Schomaker. Text-independent writer identification and verification using textural and allographic features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):701–17, Apr. 2007. 1

[7] W. M. Campbell, D. E. Sturim, and D. A. Reynolds. Support Vector Machines Using GMM Supervectors for Speaker Verification. *Signal Processing Letters, IEEE*, 13(5):308–311, May 2006. 1

[8] V. Christlein, C. Riess, E. Angelopoulou, G. Evangelopoulos, and I. A. Kakadiaris. The Impact of Specular Highlights on 3D-2D Face Recognition. In *Defense, Security + Sensing (Biometric and Surveillance Technology for Human and Activity Identification)*, volume 1, pages 8712–8719, Baltimore, MD, May 2013. 1

[9] V. Christlein, C. Riess, J. Jordan, and E. Angelopoulou. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on Information Forensics and Security*, 7(6):1841–1854, 2012. 2

[10] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. 2

[11] S. Fiel and R. Sablatnig. Writer Identification and Writer Retrieval using the Fisher Vector on Visual Vocabularies. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 545–549, Washington DC, NY, Aug. 2013. 1, 2, 4, 8, 5, 6

[12] T. S. Jaaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Advances in Neural Information Processing Systems II*, pages 487–493, Denver, CO, 1999. 3

[13] R. Jain and D. Doermann. Writer Identification Using an Alphabet of Contour Gradient Descriptors. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 550–554, Washington, NY, Aug. 2013. 1, 2, 8, 5, 6

[14] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid. Aggregating local image descriptors into compact codes. *IEEE transactions on pattern analysis and machine intelligence*, 34(9):1704–16, Sept. 2012. 2, 3

[15] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig. CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 560 – 564, Washington DC, NY, Aug. 2013. 1, 4

[16] G. Louloudis, B. Gatos, A. Stamatopoulous, and A. Papandreou. ICDDAR2013 Competition on Writer Identification. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 1397 – 1401, Washington DC, NY, 2013. 1, 3, 4, 6

[17] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004. 2

[18] U.-V. Marti and H. Bunke. The IAM-Database: An English Sentence Database for Offline Handwriting Recognition. *International Journal on Document Analysis and Recognition*, 5:39–46, 2002. 4

[19] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. IEEE Conference on*, pages 1–8, Minneapolis, MN, June 2007. 2, 3

[20] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *European Conference on Computer Vision (ECCV) 2010*, pages 143–156, Hersonissos, Heraklion, Crete, Sept. 2010. 2, 3

[21] D. a. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker Verification Using Adapted Gaussian Mixture Models. *Digital Signal Processing*, 10(1-3):19–41, Jan. 2000. 1, 3

[22] A. Schlapbach, M. Liwicki, and H. Bunke. A writer identification system for on-line whiteboard data. *Pattern Recognition*, 41(7):2381–2397, 2008. 2

[23] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6):787–798, 2004. 1

[24] I. Siddiqi and N. Vincent. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognition*, 43(11):3853–3865, Nov. 2010. 1, 2