Pattern Recognition Lab
Department Informatik
Universität Erlangen-Nürnberg
Prof. Dr.-Ing. habil. Andreas Maier
Telefon: +49 9131 85 27775
Fax: +49 9131 303811
info@i5.cs.fau.de
www5.cs.fau.de

Writer Identification Using VLAD Encoded Contour-Zernike Moments

Vincent Christlein, David Bernecker, Elli Angelopoulou

To cite this version:

Christlein, Vincent; Bernecker, David; Angelopoulou, Elli, "Writer Identification Using VLAD Encoded Contour-Zernike Moments," in Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, pp. 906-910, 23-26 Aug. 2015

Submitted on February 15, 2015, last revised May 15, 2015

DOI: 10.1109/ICDAR.2015.7333893

Writer Identification Using VLAD Encoded Contour-Zernike Moments

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

***Abstract*—Local feature descriptors in combination with bag of (visual) words have recently become the state of the art in writer identification. In this work we propose the use of Zernike moments evaluated at the contours of the script as local descriptor. We then form a global descriptor by encoding the extracted Zernike moments into Vectors of Locally Aggregated Descriptors (VLAD). This local / global descriptor combination outperforms existing methods: on the ICDAR 2013 benchmark database our Zernike / VLAD method yields 0.880 mAP, a 31% improvement over the 0.671 mAP of the state of the art. We also set a new performance standard on the CVL dataset.**

I. Introduction

Handwritten text can serve as a biometric identifier similar to someone's face, or fingerprints. Typically, experts are consulted to examine the authorship of handwritten text. However, when searching for a specific individual in a large data corpus a manual inspection might not be feasible. This often daunting task of finding an individual writer in a dataset of known authors is formally defined as *writer identification*. Recently, this problem has gained particular interest in the analysis of historical documents [1], [2] especially because of its potential to provide new insights into life in the past.

Two types of writer identification exist, *online* and *offline*, depending on the type of database. Online writer identification is used when the dataset contains temporal information about the text formation. In contrast, *offline* writer identification relies solely on the handwritten text without additional information. Bulacu and Schomaker [3] further categorize offline writer identification into *textural*- and *allograph*-based methods. In textural-based methods the identification is based upon global statistics computed from the handwritten text, e. g., the angle distribution or ink width [1], [4]–[6]. In allograph-based methods, the handwritten text is described by features computed from small letter parts (allographs). In a training step a vocabulary is computed which is further used to collect statistics from local features to form a global descriptor [2], [5], [7]–[9]. Note, the combination of allograph based methods with textural methods have also been proposed [3], [10], [11].

We present an allograph-based method for offline writer identification. It uses Contour-Zernike moments as local descriptors which are encoded using vectors of locally aggregated descriptors (VLAD) and eventually compared using the cosine distance. We evaluate the proposed method on two publicly available datasets, ICDAR13 and CVL, see Figure 1, and show that it improves the retrieval performance of the current state of the art. Finally, we investigate the effect of a dimensionality reduction step and reveal that the reduction of the dimensionality by a factor of up to 100 is accompanied by only a small loss in accuracy.

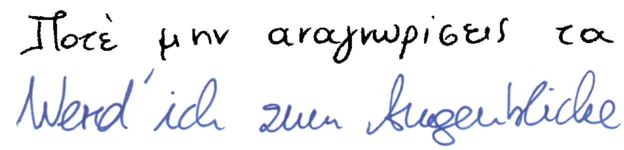


Figure 1. Excerpts of the two datasets: ICDAR13 (top) and CVL (bottom).

The rest of the paper is organized as follows. Section II gives an overview of related work. Contour-Zernike moments and the VLAD technique are presented in Section III. The evaluation of the parameters and the accompanying methods are shown in Section IV. Section V gives a summary and an outlook.

II. Related Work

Textural-based methods do not need to compute a dictionary which makes them typical faster to compute and more interpretable in comparison to allograph based methods. The most recent textural-based method has been proposed by Newell and Griffin [6]. They use oriented Basic Image Features (oBIF) as their descriptors. Hereby, six different Derivative-of-Gaussian filter responses are encoded as histograms. Additionally, they propose the use of *delta encoding*. From a training set the mean oBIF histogram is computed and the difference between this histogram and the query oBIF histogram is used as a discriminating descriptor. The ICFHR 2014 competition on Arabic writer identification [12] reveals that this method is inferior to our previous allograph based method [7].

In allograph based methods a single global image descriptor is formed to *encode* local descriptors. Usually this involves computing statistics from the local descriptors with respect to a learned vocabulary. This process is also known as Bag of (visual) words (BoW). Early works use zero-order statistics by counting the number of nearest visual words for each cluster center of the vocabulary. This histogram is then used as a general image descriptor. Sanchez et al. propose the use of Fisher Kernels for encoding local descriptors [13]. Hereby, a very high dimensional global descriptor (Fisher Vector) is formed by computing statistics up to the second order from a trained Gaussian Mixture Model (GMM).

In the field of writer identification, Fiel and Sablatnig were the first to use Fisher Vectors as image descriptors [8]. SIFT descriptors are used as local descriptors which are further encoded using Fisher Vectors and compared using the cosine distance. Jain and Doermann also propose to use Fisher Vectors as encoding method [9]. However, they suggest to use different local descriptors which are individually encoded. Eventually, these global descriptors are combined using learned weighting factors. Recently, we proposed using GMM supervectors to encode local SIFT descriptors [7]. They show that this encoding method is superior to other encoding methods like the Fisher Vectors used by Fiel and Sablatnig. However note that since statistics up to the second order are computed, the resulting supervector can be very high dimensional.

In contrast to these approaches we use Zernike moments as sole descriptors and encode those using VLAD. Zernike moments [14] have successfully been used as feature descriptors in different domains. For example they are among the top ranking descriptors in copy-move forgery detection [15]. They have also been used for the classification of music scores [16] and handwritten text [17]. To the best of our knowledge they have not been used for writer identification, yet. However, they have been employed as shape descriptors in the field of signature verification [18]. We choose to use Zernike moments as local feature descriptor because of their great performance and their relative low dimensionality. Note that preliminary tests revealed that computing them at the contours of the script is favorable to other schemes like dense sampling. Furthermore, we choose to use VLAD as our encoding method. In contrast to other encoding methods, VLAD generates feature vectors of significantly lower dimension of the feature vector and can be computed very efficiently [19]. In conjunction with a dimensionality reduction step this makes it a perfect candidate to search through very large datasets. Also in terms of performance, we show that this combination of Zernike moments and VLAD encoding outperforms all of the above methods on all evaluated datasets.

III. Methodology

First, we will present the local descriptors which we use: the Contour-Zernike moments. We will then describe how we encode them using *vectors of locally aggregated descriptors* (VLAD), and how this representation is further improved by means of postprocessing.

A. Contour-Zernike Moments

Zernike moments can be used to extract shape information by mapping an image region onto a sequence of orthogonal polynomials (the Zernike polynomials). A big advantage of the Zernike moments is that they represent image properties with no redundancy or overlap of information between the moments. More specifically, the Zernike moments A_{nm} are defined as [14]:

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq w/2} f(x,y) V^*_{nm}(x,y) \mathrm{d}x \mathrm{d}y \quad (1)$$

for order $n \in \mathbb{Z}^+$ and repetition $m \in \mathbb{Z}$. n and m have to satisfy the constraint that $n - |m|$ is positive and even. The integral in Equation (1) is over a circle with radius $w/2$ around a center point. V_{nm} is a complex valued function and is best defined by using polar coordinates ρ and θ

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{im\theta}. \quad (2)$$

The main part of V_{nm} are the Zernike polynomials

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s [(n-s)!] \rho^{n-2s}}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \quad . \quad (3)$$

To use the Zernike moments as features for describing parts of images, the integrals in Equation (1) are replaced by sums over pixels. We then compose features by concatenating the Zernike moments of different values of n and m. This is done by setting a *degree* d, and for each $n \leq d$ the Zernike moments A_{nm} are calculated for $0 \leq |m| \leq n$ satisfying the condition that $n - |m|$ is positive and even. The feature vector then consists of the concatenation of all possible calculated combinations of n and m.

In our application, the feature vector consists of both the real and imaginary parts of the Zernike moments. Furthermore we restrict m to be positive which leads to a feature vector of dimension

$$N = 2 \sum_{i=0}^{d} \left(\left\lfloor \frac{i}{2} \right\rfloor + 1 \right). \quad (4)$$

An alternative is to only use the absolute value of the Zernike moments. In that case, the feature vector is invariant to rotations of the image. In the way we are using the Zernike moments, they are susceptible to rotations and scalings. We compute the Zernike-based feature vectors at pixels centered on the contour of the handwritten text. The contour of the text is determined by finding the contour of the connected components in a binarized version of the image.

B. VLAD Encoding

We will use *vectors of locally aggregated descriptors* (VLAD) to form a global image descriptor from the Contour-Zernike moments. VLAD aggregates the residuals of each local descriptor and its nearest cluster center. Thus, VLAD can be seen as a non-probabilistic version of the Fisher Kernel [20]. In conjunction with improvements like intra-normalization [21] or whitening [22] it achieves state of the art performance on several benchmark datasets. VLAD encoding yields a more compact image representation than Fisher Vectors, but showed a performance similar to Fisher Vectors or Gaussian supervectors in preliminary tests. Additionally, we evaluate different normalization strategies like power normalization and intra-normalization, see Section III-C for more details.

Formally a VLAD is constructed as follows. Let $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ denote T local image descriptors $\mathbf{x} \in \mathbb{R}^N$. First, a codebook $\mathbf{D} = \{\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K\}$, consisting of K clusters $\boldsymbol{\mu} \in \mathbb{R}^N$, is computed using k-means. Each local descriptor is then assigned to its nearest cluster center. For each cluster all differences between the cluster center and the assigned local descriptors are accumulated [20]:

$$\mathbf{v}_k = \sum_{\mathbf{x}_t :\, \mathrm{NN}(\mathbf{x}_t) = \boldsymbol{\mu}_k} (\mathbf{x}_t - \boldsymbol{\mu}_k) \,, \quad (5)$$

where $\mathrm{NN}(x_t)$ refers to the nearest neighbor of $\mathbf{x}_t$ in the dictionary $\mathbf{D}$. The concatenation of all $\mathbf{v}_k$ forms the VLAD encoding:

$$\mathbf{v} := \left(\mathbf{v}_1^\top, \dots, \mathbf{v}_K^\top\right)^\top . \quad (6)$$

After a normalization step (see Section III-C), the final VLAD representations of each document are compared with each other using the cosine distance [20].

C. Postprocessing

We propose the employment of two postprocessing steps: a) normalization to counter burstiness (see below) and b) whitening to reduce the effect of co-occurences.

Normalization can help to reduce the effect of visual bursts. Burstiness occurs when a few large components of the VLAD representation dominate the similarity computation between two vectors [21]. We can address this problem on two different levels: i) at each element of the vector, or ii) at each component $\mathbf{v}_k$ of the representation.

The most widely used normalization is the *power* normalization. Each element of the VLAD representation $\mathbf{v}$ is normalized by applying the signed square root [20] resulting in the normalized vector $\hat{\mathbf{v}}$:

$$\hat{\mathbf{v}}_i := \operatorname{sign}(\mathbf{v}_i)|\mathbf{v}_i|^\rho \qquad \forall i = \{1,\dots,|\mathbf{v}|\}, 0 < \rho \leq 1\,, \tag{7}$$

where ρ is typically set to 0.5. This method is also known as *signed square root* (ssr) normalization.

Arandjelovic and Zisserman proposed the use of *intra-normalization* [21]. They show that this dampens the influence of dominant components. Hereby, each $\mathbf{v}_k$ is l_2 normalized individually:

$$\hat{\mathbf{v}} := \left(\frac{\mathbf{v}_1^\top}{\|\mathbf{v}_1\|_2}, \dots, \frac{\mathbf{v}_K^\top}{\|\mathbf{v}_K\|_2} \right)^\top . \tag{8}$$

After one of these normalization steps the full VLAD representation $\hat{\mathbf{v}}$ is l_2 normalized. In the following section we evaluate the impact of the different normalization steps.

Whitening and dimensionality reduction can be applied to the resulting vectors in order to decorrelate the representation and to find a more compact descriptor, respectively [22]. This can be computed jointly be means of PCA whitening. Note that we add a small regularization factor (0.001) on the eigenvalues to counter numerical instability.

Following the approach of Jégou and Ondřej [22] we use multiple vocabularies which are jointly decorrelated. This has been shown to be very beneficial for image retrieval [22], [23]. In practice this means that we compute several dictionaries by using k-means. Note that we use the mini-batch version of k-means [24] for a faster computation. Consequently, we compute multiple VLAD representations using these dictionaries, which are then concatenated and jointly decorrelated and dimensionality reduced. We will show that the dimensionality reduction reduces the accuracy only marginally.

IV. Evaluation

We first introduce the two datasets and the metrics we use for the evaluation of our proposed approach. Next, we evaluate the effect of the different parameters of our methodology. The behaviour of the Zernike moments is influenced by two parameters, the Zernike degree d and the window size w. Furthermore, we evaluate the effect of the different normalization schemes presented in the previous section. In the last part of the evaluation we present the results for writer identification on both datasets.

A. Datasets

Two publicly available datasets have been used in the evaluation, namely CVL, and ICDAR13. Example lines from the two datasets can be seen in Figure 1.

1) ICDAR13 [25]: This dataset was part of the ICDAR 2013 Writer Identification Competition and contains 350 scribes. Each scribe contributed four documents, two written in Greek, and two written in English. 100 of the writers are part of the training set, while the others make up the benchmark dataset.

2) CVL [26]: The CVL dataset contains 310 writers. The training set consists of 27 writers, who contributed seven documents each. The independent test set contains 283 writers who contributed 5 documents each, resulting in a total of 1415 documents for testing. The documents in the test set contain different texts, one written in German, the others in English. Note that we converted the documents to grayscale.

B. Evaluation metrics

As metrics we use the *mean average precision* (mAP) and the hard TOP-k rates. Both metrics are commonly used in information retrieval. Given a reference document, a query is made and the documents in the database are returned in an order, where the first returned document is the one that best matches the query document. A returned document is called *relevant*, when it is written by the same author as the query document.

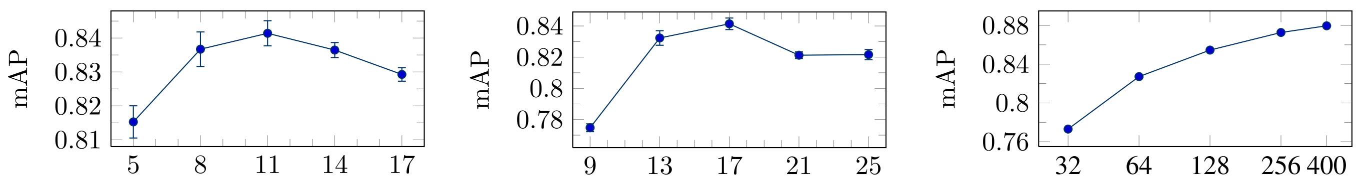
Mean average precision is the mean over all queries of the average precision. The latter is calculated by averaging over the precision values at different ranks of a query. When n documents are retrieved, the average precision aP is calculated by

$$\mathrm{aP} = \frac{\sum_{k=1}^{n} P(k) \cdot \mathrm{rel}(k)}{\text{number of relevant documents}} \quad , \tag{9}$$

where $P(k)$ is the precision at rank k of the retrieved documents (i. e., the number of relevant documents in the query up to rank k divided by k), and $\mathrm{rel}(k)$ is a relevance function that is 1 when the returned document at rank k is relevant and zero otherwise. The hard TOP-k rate is calculated by determining the percentage of queries, where the k highest ranked retrieved documents were relevant ones.

C. Parameter evaluation

We first evaluate the impact of the parameters of the Zernike moments and of the different normalization schemes presented in Section III-C. All of the parameters are evaluated using leave-one-out cross validation on the ICDAR13 training dataset using 100 clusters for k-means and power normalization. The experiments were performed five times and Figure 2 shows the mean values and the standard deviation of the mAP for the different runs. This is necessary due to the random initialization of the k-means and the random selection of the mini-batches for k-means clustering used in training the dictionaries. The power normalization has been applied as normalization technique. The two parameters in the calculation of the Zernike moments, the degree d and the window size w, both show distinct peaks at $d = 11$ and $w = 17$ (cf. Figure 2a and Figure 2b, respectively). Thus, these values were chosen for the remaining experiments.



(a) Zernike degree (b) Window size (c) Number of components

Figure 2. Evaluation of different parameters: (a) different values of ther Zernike degree d, with a fixed window size of 17, (b) different window sizes, with a fixed Zernike degree of 11, and (c) different retained number of components after the dimensionality reduction.

Table I. Influence of the normalization of the VLAD representation evaluated on Icdar13 (training set).

	TOP-1	mAP
global l_2	0.988	0.815
ssr + global l_2	0.991	0.841
intra. + global l_2	0.992	0.844

Table I shows the evaluation of the normalization schemes averaged over five runs. The baseline method consists of only applying a l_2 normalization to the encoded feature vector. For the other two methods, first the respective normalization method presented in Section III-C was applied, and then additionally the baseline normalization was applied. The intra-normalization (*intra*) and power normalization (*ssr*) both show improved results compared to the baseline method. In the remainder of the experiments we choose to use intra-normalization. Note that we also examined residual normalization [27]. However, it does not show a performance increase in contrast to the other normalization techniques. This could be explained by the fact that the Zernike moments themselves are not normalized.

D. Results

For the final evaluation of the datasets we used Zernike moments up to the degree of 11, a window size of 17, 100 clusters estimated per k-means and intra-normalization. The results are presented in three different ways: *Proposed* refers to our proposed method excluding whitening averaged over five runs, while for *Proposed + W.-256* the vectors are decorrelated using PCA whitening and dimensionality reduced to 256 components. *Proposed + W.-full* refers to the decorrelated method without any dimensionality reduction. Additionally, we give the current best results for each dataset to the best of our knowledge. The decorrelation matrices are computed on the training sets of Icdar13 and Cvl, respectively. For the dictionary training we used the Icdar13 training set for both evaluations of the Icdar13 and the Cvl test set, since the Cvl training set is rather small (similar to our previous work [7]).

Contour-Zernike vs. other Local Descriptors: Table II shows the hard criterion and mAP evaluated on Icdar13. To compare Contour-Zernike moments with other feature descriptors we ran the same pipeline (normalized VLAD encoding plus joint decorrelation and whitening) with two other local descriptors: RootSIFT and SURF descriptors. Both have been used successfully for writer identification by Christlein et al. [7] and Jain and Doermann [9], respectively. Interestingly, both descriptors perform equally well and achieve about 71% mAP. Thus, being slightly better than the GMM supervector approach [7]. In contrast, our proposed Contour-Zernike moments give a significantly higher mAP.

Table II. Hard criterion and mAP evaluated on Icdar13 (test set).

	TOP-1	TOP-2	TOP-3	mAP
SV [7]	0.971	0.428	0.238	0.671
RootSIFT + VLAD + W.-full	0.961	0.517	0.291	0.707
SURF + VLAD + W.-full	0.956	0.506	0.282	0.705
Proposed	0.975	0.707	0.481	0.808
Proposed + W.-256	0.993	0.798	0.596	0.873
Proposed + W.-full	**0.994**	**0.810**	**0.618**	**0.880**

Table III. Hard criterion and mAP evaluated on Cvl (test set).

	TOP-1	TOP-2	TOP-3	TOP-4	mAP
Comb. [9]	**0.994**	0.983	0.948	0.829	0.969
SV [7]	0.992	0.981	0.958	0.887	0.971
Proposed	0.988	0.976	0.953	0.862	0.960
Proposed + W.-256	0.992	0.987	0.975	0.925	0.978
Proposed + W.-full	**0.994**	**0.989**	**0.974**	**0.927**	**0.979**

Influence of Post Processing: Table III reveals that the decorrelation step is critical for an improved accuracy. Without decorrelation our baseline (*Proposed*) is inferior to the GMM supervector approach [7]. In contrast, if we decorrelate the VLAD representation, the accuracy in terms of mAP improves drastically. Notably, a dimensionality reduction to 256 components does not reduce the accuracy by much (0.004 mAP in average). Consequently, the resulting global representations can be compared much more efficiently which is beneficial for very large datasets. Figure 2c shows different numbers of retained components using the Icdar13 training set to compute the decorrelation matrix[1].

Proposed vs. State of the Art: Table II shows that the proposed method improves by more than 30% in terms of mAP in contrast to the GMM supervector method [7]. Interestingly, the TOP-2 and TOP-3 rate improved significantly. This means that the chance to recognize documents written by the same author but in a different script style (English / Greek) is much higher. For the Cvl dataset, the improvement is not that large but still noticeable, cf. the TOP-3 rate of Table III. Consequently, we can conclude that especially in non homogeneous datasets, i. e., datasets containing more than one script style, our proposed combination of Zernike moments and VLAD with an additional decorrelation is superior to other methods.

[1]The maximum number is limited due to the size of the Icdar13 training set.

Table IV. SOFT CRITERION OF THE PROPOSED METHOD (W.-FULL).

	TOP-2	TOP-3	TOP-5	TOP-10
ICDAR13	0.995	0.995	0.996	0.997
CVL	0.994	0.994	0.994	0.995

For reference, the soft TOP-k criterion, i. e., the average precision at rank k, are given in Table IV.

V. CONCLUSION

We have presented a new method for offline writer identification which uses local Contour-Zernike moments and multiple VLAD representations which are subsequently decorrelated using PCA whitening. We show that this greatly improves the retrieval rate. Furthermore, a joint dimensionality reduction may give a very compact image representation with only a slight loss in accuracy. Since VLAD can be computed very fast [19], it enables an efficient large scale writer identification.

As part of future work, we would like to evaluate the impact of feature combinations in conjunction with Contour-Zernike moments. Furthermore, the VLAD encoding could be improved, e. g., by augmenting the representation with higher order statistics [28] or by computing local coordinate systems [27].

ACKNOWLEDGMENT

This work has been supported by the German Federal Ministry of Education and Research (BMBF), grant-no. 01UG1236a. The contents of this publication are the sole responsibility of the authors.

REFERENCES

[1] A. Brink, J. Smit, M. Bulacu, and L. Schomaker, "Writer Identification using Directional Ink-Trace Width Measurements," *Pattern Recognition*, vol. 45, no. 1, pp. 162–171, 2012. 1

[2] T. Gilliam, R. Wilson, and J. Clark, "Scribe Identification in Medieval English Manuscripts," " in *Pattern Recognition, 20th International Conference on*, 2010, pp. 1880–1883. 1

[3] M. Bulacu and L. Schomaker, "Text-Independent Writer Identification and Verification using Textural and Allographic Features," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 701–17, 2007. 1

[4] C. Djeddi, L.-S. Meslati, I. Siddiqi, A. Ennaji, H. E. Abed, and A. Gattal, "Evaluation of Texture Features for Offline Arabic Writer Identification," " in *Document Analysis Systems, 2014 11th IAPR International Workshop on*, 2014, pp. 8–12. 1

[5] I. Siddiqi and N. Vincent, "Text Independent Writer Recognition using Redundant Writing Patterns with Contour-Based Orientation and Curvature Features," *Pattern Recognition*, vol. 43, no. 11, pp. 3853–3865, 2010. 1

[6] A. J. A. Newell and L. D. L. Griffin, "Writer Identification using Oriented Basic Image Features and the Delta Encoding," *Pattern Recognition*, vol. 47, no. 6, pp. 2255–2265, 2014. 1

[7] V. Christlein, D. Bernecker, F. Honig, and E. Angelopoulou, "Writer Identification and Verification using GMM Supervectors," " in *Applications of Computer Vision, IEEE Winter Conference on*, 2014, pp. 998–1005. 1, 2, 4

[8] S. Fiel and R. Sablatnig, "Writer Identification and Writer Retrieval using the Fisher Vector on Visual Vocabularies," " in *Document Analysis and Recognition, 12th International Conference on*, 2013, pp. 545—-549. 1

[9] R. Jain and D. Doermann, "Combining Local Features for Offline Writer Identification," " in *Frontiers in Handwriting Recognition, 14th International Conference on*, 2014, pp. 583–588. 1, 2, 4

[10] 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*, vol. 26, no. 6, pp. 787–798, 2004. 1

[11] X. Wu, Y. Tang, and W. Bu, "Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 3, pp. 526–536, 2014. 1

[12] F. Slimane and S. Awaida, "ICFHR2014 Competition on Arabic Writer Identification Using AHTID/MW and KHATT Databases," " in *Frontiers in Handwriting Recognition, 14th International Conference on*, 2014, pp. 797 – 802. 1

[13] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image Classification with the Fisher Vector: Theory and Practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013. 1

[14] F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode," *Physica*, vol. 1, no. 7, pp. 689–704, 1934. 2

[15] V. Christlein, C. Riess, J. Jordan, C. Riess, and E. Angelopoulou, "An Evaluation of Popular Copy-Move Forgery Detection Approaches," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 6, pp. 1841–1854, 2012. 2

[16] A. Fornés, J. Lladós, and G. Sánchez, "Primitive Segmentation in Old Handwritten Music Scores," in *Graphics Recognition. Ten Years Review and Future Perspectives*, ser. Lecture Notes in Computer Science, 2006, vol. 3926, pp. 279–290. 2

[17] A. Wahi, S. Sundaramurthy, and P. Poovizhi, "Handwritten Tamil Character Recognition Using Zernike Moments and Legendre Polynomial," in *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, ser. Advances in Intelligent Systems and Computing, 2015, vol. 325, pp. 595–603. 2

[18] S. Chen and S. Srihari, "Use of Exterior Contours and Shape Features in Off-Line Signature Verification," " in *Document Analysis and Recognition, Eighth International Conference on*, 2005, pp. 1280–1284. 2

[19] K. van de Sande, C. Snoek, and A. Smeulders, "Fisher and VLAD with FLAIR," " in *Computer Vision and Pattern Recognition, 2014 IEEE Conference on*, 2014, pp. 2377–2384. 2, 5

[20] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1704–16, 2012. 2, 3

[21] R. Arandjelovic and A. Zisserman, "All about VLAD," " in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2013, pp. 1578 – 1585. 2, 3

[22] H. Jégou and C. Ondřej, "Negative Evidences and Co-occurences in Image Retrieval: The Benefit of PCA and Whitening," " in *Computer Vision – ECCV 2012*, no. 1, 2012, pp. 774–787. 2, 3

[23] E. Spyromitros-Xioufis, S. Papadopoulos, I. Kompatsiaris, G. Tsoumakas, and I. Vlahavas, "A Comprehensive Study Over VLAD and Product Quantizationin Large-Scale Image Retrieval," *Multimedia, IEEE Transactions on*, vol. 16, no. 6, pp. 1713–1728, 2014. 3

[24] D. Sculley, "Web-scale K-means Clustering," in *World Wide Web, 19th International Conference on*, ser. WWW '10, 2010, pp. 1177–1178. 3

[25] G. Louloudis, B. Gatos, N. Stamatopoulos, and A. Papandreou, "ICDAR 2013 Competition on Writer Identification," " in *Document Analysis and Recognition, 12th International Conference on*, 2013, pp. 1397–1401. 3

[26] 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, 12th International Conference on*, 2013, pp. 560 – 564. 3

[27] J. Delhumeau, P.-H. Gosselin, H. Jégou, and P. Pérez, "Revisiting the VLAD Image Representation," " in *21st ACM international conference on Multimedia - MM '13*, 2013, pp. 653–656. 4, 5

[28] X. Peng, L. Wang, Y. Qiao, and Q. Peng, "Boosting VLAD with Supervised Dictionary Learning and High-Order Statistics," " in *Computer Vision–ECCV 2014*, 2014, pp. 660–674. 5