International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 7, No.1 (Jan-2018)
http://dx.doi.org/10.12785/ijcds/070104

Recognition of Arabic Handwritten Words using Gabor-based Bag-of-Features Framework

Mohammed O. Assayony[1] and Sabri A. Mahmoud[1]

[1] *Information and Computer Science Department, KFUPM, Dhahran 31261, Saudi Arabia*

Received 30 Sep. 2017, Revised 9 Nov. 2017, Accepted 11 Dec. 2017, Published 1 Jan. 2018

Abstract: Holistic recognition of isolated words is an essential task in several daily life applications, e.g., bank check processing and postal address reading. In this work we present a system for the automatic recognition of Arabic handwritten words based on statistical features extracted by Bag-of-Features framework that exploits the discriminative power of Gabor features. A handwritten text image is filtered by a set of Gabor filters of different scales and orientations for extracting texture-based local features. The response of the Gabor filters are organized into two layouts, viz. the Statistical Gabor Features and Gabor Descriptors, and fed to the Bag-of-Features in order to produce statistical representations for the handwritten text. The produced features are utilized in a holistic handwritten word recognition system that is applied on handwritten Arabic checks legal amounts public dataset. The effective parameters of the two layouts as well as the Bag-of-Features framework are experimentally evaluated and the optimal values are used in reporting the final recognition accuracies. The best average recognition accuracy achieved by the produced features is 86.44% which is promising in such challenge dataset of large number of classes.

Keywords: Holsitic handwriting recognition, Arabic Handwriting Recognition, Gabor Filters Response features, Bag-of-Features

1. Introduction

Automatic recognition of handwritten words is an essential task encountered in several daily life applications, e.g., bank check processing for validating legal amounts and postal address reading for reading city names. Though the recognition of unconstrained Arabic text is challenging due to the cursive nature of Arabic script [1], the recognition of isolated words from limited vocabulary can be viewed as classification problem where a transcription is assigned to the word image based on the holistic shape of the word [2]. Such systems are known as *holistic recognition systems*, in contrast to *segmented-based systems* that need to segment the text image into parts (words, part of words, characters) and each part is recognized individually. A holistic recognition system involves three main stages, viz. preprocessing, feature extraction and classification. Among these stages, the feature extraction stage is crucial, since it simplifies the classification and improves its performance. Without this stage, the classifier needs to work with raw image representation, i.e., pixel intensities, which is a difficult task and requires huge number of training samples. Despite its importance, deciding appropriate kind of features is hard and requires considerable effort. Vast number of features were handcrafted by experts based on their prior knowledge and experience in the field of handwriting recognition, including structural features e.g. skeleton representation, strokes, loops and statistical features like gradient histograms, projection profiles, energy of Gabor filters and wavelets transforms coefficients [1]. The main shortcoming in most of the *handcrafted features* is that they are script-dependent and so it is hard to adapt them to other scripts. For instance, the features designed for Latin script in [3] might not be suitable for Arabic script while the features proposed for Arabic in [4] [5] [6] might not be suitable for Latin scripts. In the last decade, feature learning and deep learning techniques that automatically produce robust features for handwritten text from raw pixel representation have huge applications in developing handwriting recognition system for different scripts, including Arabic [7] [8] [9] [10] [11].

In this work, we utilize Bag-of-features (BoF) framework in order to produce robust features for Arabic handwritten word images. BoF framework is a feature learning framework that utilizes unsupervised learning algorithms, e.g., clustering, in order to learn robust representations for low-level local features of digital images. The learned representations are used to produce statistical features for the images [12]. Though the framework has been applied in several document image processing applications [8] [13] [14] [15], most of the works relied on gradient histogram features in SIFT format. In this work, however, we consider Gabor features instead of SIFT descriptors as these features achieved superior performance in recognizing Arabic handwritten numerals [16] [17] [18]. Our aim is to

E-mail address: g201102150@kfupm.edu.sa , smasaad@kfupm.edu.sa

investigate the performance Gabor features with the Bag-of-Features framework and compare their performance to the results we achieved by SIFT descriptors in our previous work [19] on the same dataset. In order to formulate the Gabor features in the format of local features appropriate to the BoF framework, the Gabor features are organized into two layouts. The first layout is the traditional format used in handwriting recognition [16] [17] [18] as well as texture analysis [20]. In the second layout, the Gabor filters responses are organized in the format of the visual local descriptors, e.g., SIFT [21] and SURF [22]. The effective parameters of the layouts are experimentally evaluated and the optimal values are used in reporting the final recognition accuracies.

The BoF framework encodes the local features of text images in robust statistical features that are fed to the classifier as a representation for the images. To alleviate the quantization distortion associated with the naïve vector quantization encoding scheme, we apply a simple formula of soft assignment where the local features are assigned to more than one codeword based on the Euclidian distance. The performance of this encoding scheme is compared with the vector quantization encoding and with two other formulas of soft assignment, viz., the Localized Soft-Assignment [23] and the GMM-based encoding scheme that was applied in [8].

The rest of this paper is organized as the follows. Section 2 provides an overview of the Gabor filters response features and describes the two layouts utilized in this work. Section 3 presents the Bag-of-Features framework with emphasis on the encoding schemes evaluated on the local features. In Section 4, we present the results of the extensive experimentations we carry out in order to evaluate the effect of the important parameters of the framework. Finally, the conclusions are presented in Section 5.

2. GABOR FILTERS RESPONSE FEATURES

Gabor filter features achieved recognizable performance in handwriting recognition [16] [17] [18] [24] [25], machine-printed text recognition [26], writer identification [27], script identification [28] and handwritten vs. machine-printed text identification [29]. Gabor filters were utilized in implementing early feature learning frameworks inspired by the Hubel and Wiesel model, e.g., the Neocognitron [30], Cresceptron [31] and HMAX [32]. The modern unsupervised feature learning algorithms that were applied directly on the raw pixel intensities (e.g., auto-encoders and RBM) are eventually end by learning local filters similar to Gabor filters [12] [33]. This indicates the power of Gabor filters in providing discriminative representations for vision applications. Recently, Gabor filters' features were utilized in the design of a SIFT-like local descriptor coined "The Biologically Inspired Local Descriptor (BILD)" [34]. Gabor filter response features are obtained by convolving text images with a set of 2-D Gabor filters of different scales and orientations. A 2-D Gabor filter $g(x, y\,;\,\lambda, \theta)$ is a band-pass filter of a bandwidth bounding by a Gaussian envelope [17], and it can be expressed mathematically as:

$$g(x, y\,;\,\lambda, \theta) = \frac{1}{2\pi(k\lambda)^2}\, exp\left(-\frac{x'^2 + y'^2}{2(k\lambda)^2}\right) exp\left(2\pi j\left(\frac{x\cos\theta + y\sin\theta}{\lambda}\right)\right)$$

where λ and θ are respectively the wavelength (the scale) in pixels and the orientation in degrees of the carrier frequency, k is a scalar factor to ensure that the Gaussian envelop is proportional to the filter wavelength, and x' and y' indicate that the Gaussian envelope is rotated towards the carrier orientation θ:

$$x' = \quad x\cos\theta + y\sin\theta$$
$$y' = -x\sin\theta + y\cos\theta$$

In this work, we used a filter bank of three scales (wavelengths of 3, 6 and 12) and six orientations (0^o, 30^o, 60^o, 90^o, 120^o and 150^o) per scale. These orientations were applied earlier for digit recognition [16] [18] as well as in texture image retrieval [20]. In order to produce scale-invariant representations, we applied Gabor filters at different scales. The convolution amplitudes are organized in two layouts that we call Statistical Gabor Features and Gabor Descriptors.

A. Statistical Gabor Features

The statistical Gabor Features (SGF) layout is the format that was applied in the previous works [16] [17] [18]. A text image is sampled horizontally into 8-pixels-width samples and each sample is divided vertically into 8 regions. The mean and variance of the amplitude responses within each region are taken as the features of the region. Therefore, each image sample gives a feature vector of 8×2=16 elements for each scale and orientation in the Gabor filter bank. For 3 scales and 6 orientations per scale, we get 3×6=18 vectors, 16 elements each, for each image sample. The Statistical Gabor Features for an image sample is a 288-d (=16×18).

B. Gabor Descriptors

Gabor Descriptors layout produces features similar to the features produced by common local descriptors like SIFT [21] and SURF [22] as the characteristics of this layout simulates the behavior of the human visual system [21]. A Gabor descriptor is constructed by sampling the image into square patches of P×P pixels. Each patch is divided into 2×2 regions. The amplitude responses of the different Gabor filters within each region are aggregated using a statistical aggregation function. In this work, we applied four such functions, viz., the max, sum, mean and variance. For 3 scales and 6 orientations per scale, the Gabor descriptor for a patch is a 72-d (=2×2×3×6) vector. For a text image sample, the Gabor descriptors are extracted in two configurations: single-scale and multi-scale. In the single-scale configuration, the text image is sampled densely into 32×32 patches with a stride of 8 pixels. Each patch gives a 72-d Gabor descriptor. For multi-scale configuration, the text image is repeatedly sampled into patches of different scales. The 4-scale configuration we evaluated uses four patch sizes (viz. 8, 16, 24 and 32) with a stride of 8 pixels.

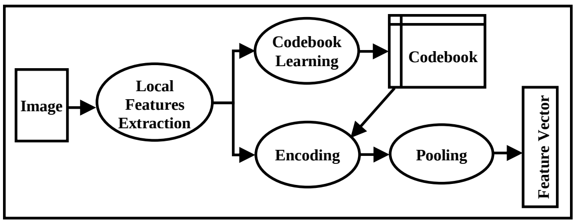


Figure 1. The general structure of the BoF Framework

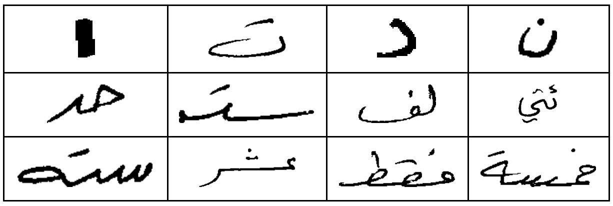


Figure 2. Sample Images for Arabic Legal Amounts Subwords

3. Bag-of-Features Framework

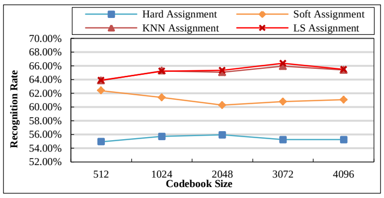
Bag-of-Features (BoF) is a statistical representation in which an image is represented by the occurrences of local features extracted from the image. It is the de-facto standard for statistical features for visual object classification and retrieval applications [35] [36] and is widely applied in text image processing applications such as handwriting recognition [8], word spotting [14], word image query [13] and writer identification and verification [15]. The BoF framework can be viewed as a two-stage framework [12]. The first stage is implemented by local descriptors while the second is an instance of unsupervised feature learning. Figure 1 shows general architecture of the BoF framework. The first stage deals with the image pixel intensity and extracts some sort of local features. Such local features include edges, textures, gradient histograms, and color histograms. Despite the diversity of the local features, most of the text image processing applications relied on gradient histograms in SIFT format. In this work, we utilize Gabor features as they are powerful texture features applied in several applications and they achieved superior performance in Arabic handwritten numerals recognition [16] [17] [18].

The second stage of the framework involves two phases, viz. encoding and pooling. In the encoding phase, the local features are transformed into another domain based on a predefined codebook that is learned in unsupervised manner from the local features of the training samples. The pooling phase aggregates the encoded features into a robust global representation. The encoding of the image's local features is implemented by a non-linear encoding scheme which usually depends on the algorithm used in learning the codebook. In the naïve BoF framework where the codebook is learned by clustering, the vector quantization encoding (Hard Assignment), in which a local feature vector is assigned to the closest codeword in the codebook, is the natural encoding scheme. To attenuate the quantization distortion of the vector quantization scheme, several soft assignment schemes were proposed. Soft assignment schemes assign a local feature vector to more than one codeword [23]. We apply a simple form of soft assignment, coined *K-Nearest-Neighbor Assignment (KNN Assignment)*, in which the local feature is assigned to the nearest K codewords based on the Euclidean distance. The performance of KNN Assignment encoding is compared with two other forms of Soft Assignments, vis. Localized Soft-Assignment (LS Assignment) [23] and GMM-based Soft Assignment [8]. The Localized Soft-Assignment Encoding assigns each local feature to the nearest K codewords based on the code uncertainty (see [23] for the details). In the GMM-based Soft Assignment, each local feature is assigned to every codeword based on the a-posteriori probabilities computed by the estimated Gaussian Mixture Models (GMM).

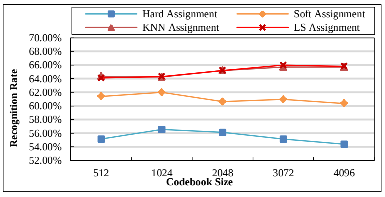
4. Experimental Results

In order to evaluate the quality of the statistical features extracted by the BoF framework, we have implemented a holistic handwriting recognition system and applied it on a dataset of Arabic Handwritten subwords. Our system involves three processing phases, vis., preprocessing, feature extraction and classification. In the preprocessing step, image samples are normalized to 64 pixels height in order to reduce the side effect of the variety in font size. This preprocessing procedure is common and was applied in several earlier works [17] [18] [19]. The height normalization is implemented offline by resizing the images to 64-pixel height while observing the width aspect ratio. For feature extraction, the BoF framework with Gabor filters is applied. The classification step of the recognition system is implemented using Support Vector Machines (SVM) [37]. SVM is a powerful classifier that achieved state-of-the-art performance in different domains, including handwriting recognition [18]. SVM applies the kernel trick to transform the features vectors to a high-dimensional feature space in which they become separable by a high-dimensional hyper-plane. For a two-class classification problem, SVM is trained to find the optimal hyper-plain by solving a convex minimization problem. The main advantage of SVM is that the training always ends up with a global minimum, in contrast to other machine learning techniques, e.g., neural networks, that might stuck in a local minimum value. We use SVM with a linear kernel as implemented in VLFeat library [38]. Though more sophisticated kernels, e.g., Radial Basis Function (RBF), might achieve superior performance, the linear kernel are very efficient for training. We use the default configurations defined for the VLFeat demo of Caltech-101 classification. A 1-vis-all SVM classifier is trained for each class in our dataset.

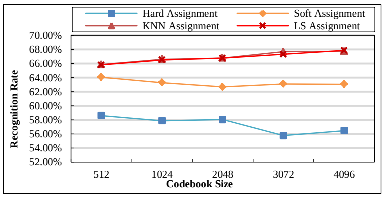
The recognition system is experimentally evaluated on the non-touching Arabic subwords dataset of CENPARMI handwritten Arabic checks database [39]. The dataset contains 27985 samples of Arabic subwords involved in legal amounts of Arabic checks. The dataset contains 96 classes in the training set and 101 in the test set. However, some classes in the training set or in the test set are empty. We consider only the classes that contain at least one sample in the training set and one sample in the test set. There are 84 of such classes. Figure 2 shows sample subwords from the dataset. A complete description of the



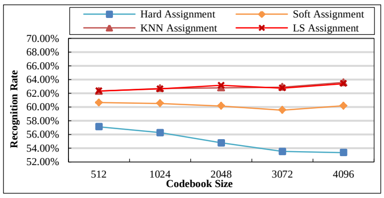
(a) Orientation 1 (0°)



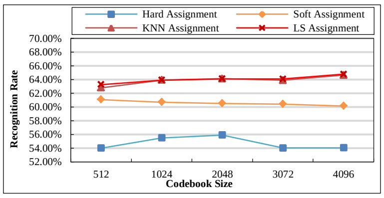
(b) Orientation 2 (30°)



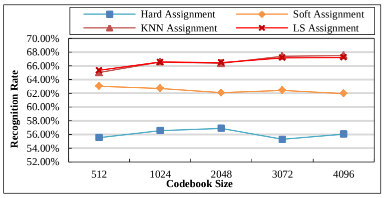
(c) Orientation 3 (60°)



(d) Orientation 4 (90°)

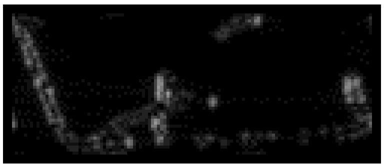


(e) Orientation 5 (120°)



(f) Orientation 6 (150°)

Figure 3. Recognition rates with Statistical Gabor Features using individual orientations of single scale

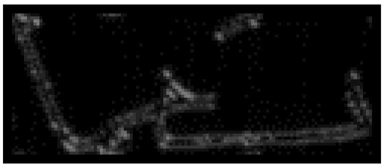


(a) Orientation 1 (0°)
The response on vertical text is strong but it is weak on the horizontal text

(b) Orientation 2 (30°)
The response on the vertical text became low and on the horizontal text became better

(c) Orientation 3 (60°)
The response on both horizontal and vertical text is observed

(d) Orientation 4 (90°)
The response on horizontal text is strong but it is weak on the vertical text



(e) Orientation 5 (120°)
The response on both horizontal and vertical text is observed

(f) Orientation 6 (150°)
The response on both horizontal and vertical text is observed

Figure 4. The response of Gabor filters with different orientations on a sample image

dataset with a comprehensive statistics and a representative sample for each class can be found in our previous work [19]

A. Evaluating Statistical Gabor Features

To evaluate the Statistical Gabor Features (SGF), we use two different Gabor filter banks. The first consists of one scale of wavelength of 3 and six orientations (0°, 30°, 60°, 90°, 120° and 150°) while the second consists of three scales (wavelengths of 3, 6 and 12 pixels) and six orientations (0°, 30°, 60°, 90°, 120° and 150°) per scale. The 6 orientations were applied earlier for digit recognition [16] [18] as well as in texture image retrieval [40]. We use several scales to provide scale-invariant representations.

The first set of experiments is conducted to evaluate the accuracy of the orientations we choose. In these experiments, each image in the training and test sets is filtered by a Gabor filter bank of one scale and six orientations. This gives six Statistical Gabor Feature vectors -each of size 16 elements- for each sample in the image. The 16–D feature vectors corresponding to each orientation are applied individually to the Bag-of-Features learning framework. As the codebook size has crucial impact in the quality of the BoF representation, several codebooks of different sizes are learned by applying k-means clustering on a set of one million local features extracted from the training samples. We generate codebooks of large sizes (up to 4096 codewords) to provide more discrimination, since the non-touching Arabic subwords dataset has large number of classes. Moreover, Gaussian Mixture Models (GMMs) are also estimated for GMM-based soft assignment encoding. The four encoding schemes discussed in Section 3, Hard assignment (the one-to-one encoding), KNN Assignment, LS (both are one-to-K encoding) and GMM-based Soft Assignment (one-to-all encoding) are implemented and evaluated. The final image feature vector is the normalized histogram obtained by averaging the occurrences of the encoded local features.

TABLE 1 THE BEST RECOGNITION RATES ACHIEVED BY EACH ORIENTATION

Filter Orientation	Recognition Rate	Codebook Size	Encoding
0°	65.98%	3072	KNN Assignment
30°	65.98%	3072	KNN Assignment
60°	67.87%	4096	KNN Assignment
90°	63.55%	4096	LS Assignment
120°	64.78%	4096	KNN Assignment
150°	67.50%	4096	LS Assignment

Figure 3 shows the recognition rates achieved by the Statistical Gabor Features of individual orientation. The results show that for all the orientations, the KNN Assignment and LS Assignment encodings produce comparable results which outperform the two other encoding schemes (i.e., the Hard Assignments and GMM-based Soft Assignments). The KNN Assignment and LS Assignment encoding schemes get the benefits of larger codebooks while the performance of the Hard Assignment and GMM-based Soft Assignment dropped down when the codebook size went beyond 2048. This indicates that to get the benefit of sophisticated soft assignments it is important to generate codebooks of large sizes.

The best results achieved by each orientation are shown in Table 1. The table shows that the Statistical Gabor Features of the third orientation (60°) achieved the best recognition rate (67.87%) followed by the sixth orientation (150°) that achieved 67.50%. This is attributed to the property of the Gabor filter that the filter responds to the variation towards its orientation. For handwritten text, the variation in the horizontal and vertical directions is high due to the fact that most of the text components are horizontal or vertical. While the filters with horizontal and vertical orientations strongly respond to the variations in one orientation, the filters with diagonal orientations like 60° and 150° can observe discriminative variations in both directions. This explains the higher recognition accuracies achieved by the Gabor filters with the diagonal orientations 60° and 150°. Figure 4 visualizes the responses of the six filters on a sample sub-word image.

In the next set of experiments, Gabor filter banks of three scales (wavelengths of 3, 6 and 12 pixels) and six orientations (0°, 30°, 60°, 90°, 120° and 150°) per scale are used to evaluate the effectiveness of increasing the wavelength of the Gabor filters. Therefore, we get 3×6=18 Statistical Gabor Features vectors -each of size 16 elements- for each image sample.

The 16–D Statistical Gabor Features vectors corresponding to the individual orientation of each scale are applied to the BoF framework using codebook of size 4096 and the KNN Assignment encoding. Figure 5 shows the recognition accuracy achieved by the orientations of the different scales. The results show that increasing the filters' scale negatively affected the recognition accuracies. This is due to the fact that large filters cover large text portions and hence the response became indiscriminative. Figure 6 visualizes the response of Gabor filters of the three scales on a sample image. The three filters have 60° orientation. Concatenating the Statistical Gabor Feature vectors of the three scales per orientation significantly improved the recognition accuracies as the statistics of the multi scale filters produce scale-invariant representation. In all the cases, the best accuracies were achieved using the third orientation (60°), followed by the sixth (150°), first (0°), second (30°), fourth (90°) and fifth (120°). Highest recognition rate of 74.06% was obtained by concatenating the scales of the third orientation (60°).

Finally, we concatenated the Statistical Gabor Feature vectors of the three scales and the six orientations into a single vector of 288 (16×3×6) elements. The resulting features achieve recognition rate of 85.08% using codebook size of 4096 and the KNN Assignment encoding. This indicates that using the information of the whole scales and orientations significantly improved the discrimination of the BoF representation. This is inline with the results of the previous studies that applied the traditional statistical Gabor features in handwriting recognition [16] [18].

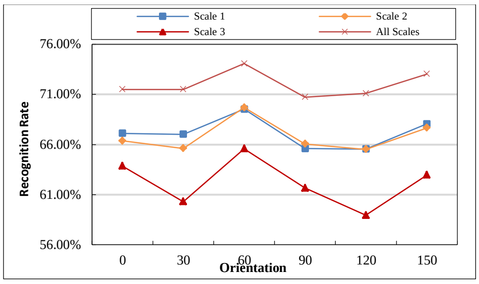
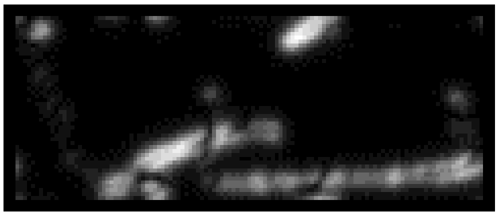
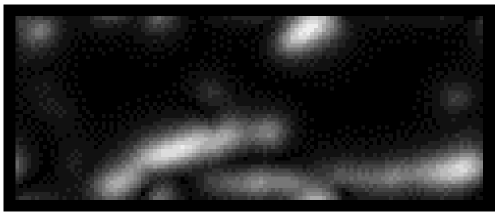


Figure 5. Recognition Rates with Statistical Gabor Features of individual orientations of three scales

(a) Scale 1 (3 pixels)



(b) Scale 2 (6 pixels)



(c) Scale 3 (12 pixels)

Figure 6. The response of Gabor filters of the three scales on a sample image. The three filters have 60° orientation

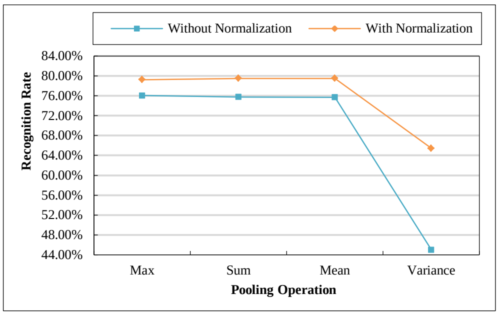


Figure 7. Recognition Rates with different Aggregation Operations applied to Gabor Descriptors

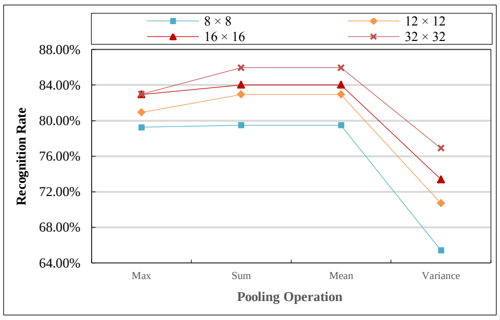


Figure 8. Recognition Rates with different Sampling Scales

B. Evaluating Gabor Descriptors

In this work, we evaluate four pooling operations: Max, Sum, Mean and Variance. We also evaluate the normalization of the final vector to unit length, the approach applied by SIFT and SURF. In all experiments, we use Gabor filter bank of three scales (wavelengths of 3, 6 and 12) and six orientations (0^o, 30^o, 60^o, 90^o, 120^o and 150^o) per scale. Therefore, we obtain a vector of 72 (2×2×3×6) elements for each sample. These features are applied to the BoF framework using codebook of size 4096 and the KNN Assignment encoding.

In the first set of experiments, the performance of the four aggregation operations and the normalization is evaluated on samples of size 8×8 pixels. Figure 7 shows the Recognition accuracies. The recognition accuracies of Max, Sum and Mean operations were comparable while the Variance was worse. The normalization significantly improve the performance of all aggregation operations. The impact of the normalization on the Variance is recognizable. Best recognition rate of 79.49 % is achieved by the Sum and Mean with normalization.

In the next set of experiments, we evaluate the effect of increasing the sample size. We draw samples of four scales: 8×8 , 12×12 , 16×16 and 32×32. Figure 8 shows the recognition rates of the four aggregation operations on the normalized Gabor Descriptors of individual scale. The results show that increasing the sample size improves the results of all aggregation operations. Highest recognition rate of 85.95% was achieved by the Sum and Mean operations.

In the last experiment, we use the normalized Gabor Descriptors of the three scales together. This combination achieves recognition rate of 86.44% using the Sum and Mean operations and the rate of 85.21% using the Max operation. Comparing these results with results of the Statistical Gabor Features we note that Gabor descriptors with the Mean aggregation function achieves better recognition accuracy than the Statistical Gabor Features, yet the dimensionality of the Gabor descriptors are much less than that of the Statistical Gabor Features (72-D viz. 288-D). The lower dimensionality of the Gabor descriptors has noticeable impact on the clustering and encoding computation performance.

C. Discussion

The best recognition accuracies achieved by the Statistical Gabor Features and Gabor descriptors are 85.08% and 86.44%, respectively. Both results are lower than the accuracy of SIFT descriptors we achieved in our previous work [19] where a recognition accuracy of 89.93% was reported. Despite that, the Gabor descriptor format enabled us augmenting the performance of Gabor features by applying techniques proposed to enhance local descriptors, e.g., patch sub-regions and normalization. Applying sophisticated normalizations such as those proposed in [34] would improve the performance of the Gabor descriptors format.

This work is motivates us to look for different type of features that might bring better performance to the BoF framework. There are many low-level features proposed in the format of local descriptors that claimed superior performance than SIFT in visual object recognition and matching [41] [22] [42]. Utilizing algorithms that show prominent performance in BoF framework would enhance the quality of the produced features. Further, the characteristics of the text images and handwritten text could be also utilized in improving these algorithms as we did with SIFT in our previous work [19].

5. Conclusions

In this paper, we utilize Bag-of-Features (BoF) framework and Gabor features for producing robust statistical features for holistic handwriting recognition systems. The Gabor filters responses are organized in two layouts coined *Statistical Gabor Features* and *Gabor Descriptors* that are fed to BoF framework as low-level local features. The framework exploits the low-level features in order to produce mid-level global features that represent the word image to the classifier. Several critical parameters of the Gabor features including the number of spatial scales, number of orientations per scale in the *Statistical Gabor Features* as well as the pooling operations, final normalization and sample size of the *Gabor Descriptors* are experimentally evaluated. The codebook size of the BoF framework is adjusted and several encoding schemes are assessed in order to come up with the optimal configurations for the framework. The produced features are utilized in a holistic handwriting recognition systems which is evaluated on the non-touching Arabic subwords dataset of CENPARMI handwritten Arabic checks database. The Statistical Gabor Features of 3 scales and 6 orientations per scale achieves average recognition accuracy of 85.08% using a codebook size of 4098 and the *KNN Assignment* encoding. Gabor descriptors achieve average recognition accuracy of 86.44% using Mean pooling function and multi-scale configuration. In addition to their superior accuracy, Gabor Descriptor layout produce local features with lower dimensionality which in turn reduce the computational time of clustering and encoding steps of the BoF framework. The analysis of the recognition accuracies of individual classes shows that the main source of misclassification is the lack of enough training samples and the challenging writing styles.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their invaluable comments. We would also acknowledge the support provided by King Fahd University of Petroleum & Minerals (KFUPM) through project number RG 1313-1/2. The first author is supported by Hadhramout Establishment for Human Development, Yemen Graduate Scholarship.

REFERENCES

[1] M. T. Parvez and S. A. Mahmoud, "Offline Arabic Handwritten Text Recognition: A Survey," *ACM Comput. Surv.*, vol. 45, no. 2, pp. 1–35, Feb. 2013.

[2] S. Madhvanath and V. Govindaraju, "The role of holistic paradigms in handwritten word recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 149–164, 2001.

[3] A. J. Elms and J. Illingworth, "Modelling Polyfont Printed Characters with HMMs and a Shift Invariant Hamming Distance," in *Proceedings of 3rd International Conference on Document Analysis and Recognition (ICDA'95)*, 1995, vol. 1, pp. 504–507.

[4] I. S. I. Abuhaiba, S. A. Mahmoud, and R. J. Green, "Recognition of handwritten cursive Arabic characters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 664–672, Jun. 1994.

[5] S. Alma'adeed, C. Higgins, and D. Elliman, "Off-line Recognition of Handwritten Arabic Words using Multiple Hidden Markov Models," *Knowledge-Based Syst.*, vol. 17, no. 2–4, pp. 75–79, May 2004.

[6] S. Haboubi, S. Maddouri, N. Ellouze, and H. El-Abed, "Invariant Primitives for Handwritten Arabic Script: A Contrastive Study of Four Feature Sets," in *10th International Conference on Document Analysis and Recognition (ICDAr 22009)*, 2009, pp. 691–697.

[7] M. Hamdani, A. E.-D. Mousa, and H. Ney, "Open Vocabulary Arabic Handwriting Recognition Using Morphological Decomposition," in *12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, 2013, pp. 280–284.

[8] L. Rothacker, S. Vajda, and G. a. Fink, "Bag-of-Features Representations for Offline Handwriting Recognition Applied to Arabic Script," in *International Conference on Frontiers in Handwriting Recognition*, 2012, pp. 149–154.

[9] A. AbdulKader, "A Two-Tier Arabic Offline Handwriting Recognition Based on Conditional Joining Rules," in *Arabic and Chinese Handwriting Recognition*, D. S. Doermann and S. Jaeger, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 70–81.

[10] A. Graves and J. Schmidhuber, "Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks," in *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Vancouver, B.C., Canada, 2009, pp. 545–552.

[11] R. Maalej and M. Kherallah, "Improving MDLSTM for Offline Arabic Handwriting Recognition Using Dropout at Different Positions," in *Artificial Neural Networks and Machine Learning – ICANN 2016*, A. E. P. Villa, P. Masulli, and A. J. P. Rivero, Eds. Springer International Publishing, 2016, pp. 431–438.

[12] Y. LeCun, "Learning Invariant Feature Hierarchies," in *Computer Vision – ECCV 2012*, A. Fusiello, V. Murino, and R. Cucchiara, Eds. Florence, Italy: Springer Berlin Heidelberg, 2012, pp. 496–505.

[13] R. Shekhar and C. V. Jawahar, "Word Image Retrieval Using Bag of Visual Words," in *2012 10th IAPR International Workshop on Document Analysis Systems*, 2012, pp. 297–301.

[14] L. Rothacker, M. Rusiñol, and G. A. Fink, "Bag-of-Features HMMs for Segmentation-Free Word Spotting in Handwritten Documents," in *2013 12th International Conference on Document Analysis and Recognition (ICDAR)*, 2013, pp. 1305–1309.

[15] S. Fiel and R. Sablatnig, "Writer Identification and Writer Retrieval Using the Fisher Vector on Visual Vocabularies," in *12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, 2013, pp. 545–549.

[16] S. A. Mahmoud, "Arabic (Indian) Handwritten Digits Recognition using Gabor-based Features," in *2008 International Conference on Innovations in Information Technology*, 2008, pp. 683–687.

[17] S. A. Mahmoud, "Recognition of Arabic (Indian) Check Digits using Spatial Gabor Filters," in *5th IEEE-GCC Conference & Exhibition*, 2009, pp. 1–5.

[18] S. A. Mahmoud and W. G. Al-Khatib, "Recognition of Arabic (Indian) Bank Check Digits using Log-Gabor Filters," *Appl. Intell.*, vol. 35, no. 3, pp. 445–456, May 2010.

[19] M. O. Assayony and S. A. Mahmoud, "An Enhanced Bag-of-Features Framework for Arabic Handwritten Sub-words and Digits Recognition," *J. Pattern Recognit. Intell. Syst.*, vol. 4, no. 1, pp. 27–38, 2016.

[20] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, 1996.

[21] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[22] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[23] L. Liu, L. Wang, and X. Liu, "In defense of soft-assignment coding," in *2011 International Conference on Computer Vision (ICCV)*, 2011, pp. 2486–2493.

[24] U. Porwal, Y. Zhou, and V. Govindaraju, "Handwritten Arabic Text Recognition using Deep Belief Networks," in *21st International Conference on Pattern Recognition (ICPR)*, 2012, pp. 302–305.

[25] M. Elzobi, A. Al-Hamadi, Z. Al Aghbari, L. Dings, and A. Saeed, "Gabor Wavelet Recognition Approach for Off-Line Handwritten Arabic Using Explicit Segmentation," in *Image Processing and Communications Challenges 5*, R. S. Choras, Ed. Springer International Publishing, 2014, pp. 245–254.

[26] A. Zaafouri, M. Sayadi, and F. Fnaiech, "A Vision Approach for Expiry Date Recognition using Stretched Gabor Features," *Int. Arab J. Inf. Technol.*, vol. 12, no. 5, pp. 448–455, 2015.

[27] B. Helli and M. E. Moghaddam, "A Text-Independent Persian Writer Identification Based on Feature Relation Graph (FRG)," *Pattern Recognit.*, vol. 43, no. 6, pp. 2199–2209, Jun. 2010.

[28] G. G. Rajput and H. B. Anita, "Handwritten Script Identification from a Bi-Script Document at Line Level using Gabor Filters," in *International Workshop on Soft Computing Applications and Knowledge Discovery*, 2011, pp. 94–101.

[29] A. K. Echi and A. Saidani, "How to Separate Between Machine-Printed / Handwritten and Arabic / Latin Words ?," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 13, no. 1, pp. 1–16, 2014.

[30] K. Fukushima, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[31] J. Weng, N. Ahuja, and T. S. Huang, "Cresceptron: A Self-Organizing Neural Network which Grows Pdaptively," in *IJCNN International Joint Conference on Neural Networks*, 1992, vol. 1, pp. 576–581.

[32] T. Poggio and M. Riesenhuber, "Hierarchical Models of Object Recognition in Cortex," *Nat. Neurosci.*, vol. 2, no. 11, pp. 1019–1025, Nov. 1999.

[33] A. Coates, H. Lee, and A. Y. Ng, "An Analysis of Single-Layer Networks in Unsupervised Feature Learning," in *International Conference on AI and Statistics*, 2011, pp. 215–223.

[34] Y. Zhang, T. Tian, J. Tian, J. Gong, and D. Ming, "A Novel Biologically Inspired Local Feature Descriptor," *Biol. Cybern.*, vol. 108, no. 3, pp. 275–290, Jun. 2014.

[35] S. O'Hara and B. A. Draper, "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval," *arXiv Prepr. arXiv1101.3354*, pp. 1–25, Jan. 2011.

[36] M. T. Law, N. Thome, and M. Cord, "Bag-of-Words Image Representation: Key Ideas and Further Insight," in *Fusion in Computer Vision, Advances in Computer Vision and Pattern Recognition*, B. Ionescu, J. Benois-Pineau, T. Piatrik, and G. Quénot, Eds. Switzerland: Springer International Publishing, 2014, pp. 29–52.

[37] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[38] A. Vedaldi and B. Fulkerson, "VLFeat: An Open and Portable Library of Computer Vision Algorithms," in *18th ACM international conference on Multimedia*, 2010, pp. 1469–1472.

[39] Y. Al-Ohali, M. Cheriet, and C. Y. Suen, "Databases for Recognition of Handwritten Arabic Cheques," *Pattern Recognit.*, vol. 36, pp. 111–121, 2004.

[40] B. S. Manjunath and W. Y. Ma, "Texture Features for Browsing and Retrieval of Large Image Data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, 1996.

[41] K. Mikolajczyk and C. Schmid, "Performance Evaluation of Local Descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–30, Oct. 2005.

[42] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of Interest Regions with Local Binary Patterns," *Pattern Recognit.*, vol. 42, no. 3, pp. 425–436, Mar. 2009.

Mohammed O. Assayony received Ph. D. degree in computer science and engineering from King Fahd University of Petroleum & Minerals, Saudi Arabia in 2017 and MS degree in computer science from the Universiti Sains Malaysia in 2009. His research interests include Arabic Document Analysis and Recognition, Automatic Feature Learning for Arabic handwriting as well as parallel algorithms for scientific applications

Sabri A. Mahmoud is a Professor of computer science in Information & Computer Science Department, King Fahd University of Petroleum & Minerals, Saudi Arabia. His research interests include Arabic Document Analysis and Recognition, Arabic NLP, Image Analysis and applications of Pattern Recognition. Dr. Mahmoud is a life senior member of IEEE. He published over 90 papers in refereed journals and conference proceedings.