Journal of Engg. Research **Vol. 4 No. (4) December 2016** pp. 66-80

التعافي من الكلمات البصرية المنفصلة باستخدام علامة جزءا لا يتجزأ من

***مجيد ايرانپور مباركه و**عليرضا احمديفرد**
*قسم الهندسة الحاسب وتقنية المعلومات جامعة شاهرود للتكنولوجيا، شاهرود، ايران.
**قسم الهندسة الكهربائية والروبوتات، جامعة شاهرود للتكنولوجيا، شاهرود، ايران.

الخلاصة

الهدف من كلمة اكتشاف(spotting) ، كحالة خاصة من المحتوى الدلالي على أساس استرجاع الصور (CBIR) هو العثور على حالات استعلام كلمة معينة مثل صورة أو نص (سلسلة) في مستند. في هذه الورقة، فإننا نقترح اتباع نهج شامل لكلمة اكتشاف (spotting) بخط اليد الفارسي. وبالإضافة إلى ذلك، فإن الطريقة المقترحة يمكن أن تستخدم في التعرف على الكلمات المكتوبة بخط اليد. ويتحقق ذلك من خلال مزيج من تضمين التسمية والتصنيف القائم على الانحدار الفضائي الجزئي المشترك. في هذا الفضاء الجزئي، صورة وتمثيل سلسلة من نفس الكلمة هي قريبة من بعضها البعض بحيث أنه يسمح للنظر في التعرف ومهمة الاسترجاع باعتباره أقرب مشكلة الجار. على عكس الطرق الحالية في كلمة اكتشاف والتمثيل المقترح لكلمة لها طول ثابت ومنخفض الأبعاد. من ناحية أخرى، فإن عملية الاستخراج سريعة جدا. كنا استعملنا فارسا وايرانشهر، وهما مجموعتين معروفتين من البيانات للكلمات الفارسية المعزولة المكتوبة بخط اليد، لتقييم الطريقة المقترحة. نتائج التجارب لكلمة اكتشاف واعدة. فمعدلات التعرف على الكلمات المعزولة المكتوبة بخط اليد في قواعد بيانات فارسا وايرانشهر هي 100% و97% على التوالي.

Isolated Persian/Arabic word spotting by label embedding

Majid Iranpour Mobarakeh* and Alireza Ahmadyfard**

**Computer Engineering & Information Technology Department, Shahrood University of Technology, Shahrood, Iran.*

***Electrical Engineering & Robotics Department, Shahrood University of Technology, Shahrood, Iran.*

Corresponding author: Ahmadyfard@shahroodut.ac.ir

ABSTRACT

The aim of word spotting, as a particular case of semantic content based image retrieval (CBIR), is to find instances of a word query given as an image or text (string) in a document. In this paper, we propose a holistic approach for Persian handwritten word spotting. In addition, the proposed method can be used for handwritten word recognition. This is achieved by a combination of label embedding; attribute based classification and common subspace regression. In this subspace, image and string representation of the same word are close together, so it allows for considering recognition and retrieval task as a nearest neighbor problem. Unlike existing methods in word spotting, the suggested representation for a word has fixed length and low dimensionality. On the other hand, the feature extraction process is very fast. We used Farsa and Iranshahr, two common datasets of isolated Persian handwritten words, to evaluate the proposed method. The result of experiments for word spotting is promising. The recognition rates for isolated handwritten words in Farsa and Iranshahr datasets are 100% and 97% respectively.

Keywords: Attribute-based classification; label embedding; word image retrieval; word spotting.

INTRODUCTION

Due to increasing the number of digitized handwritten documents, word indexing has become a crucial task. Handwritten word spotting allows for searching a word image in handwritten document. In primary methods, first optical character recognition (OCR) technique is used to convert images of a handwritten document into the string of text words, and then the text-based search engine is applied to search for a query word within the OCR result. Indeed, the first handwritten word spotting system was designed by Manmatha *et al.* (1996). As reported on this work, searching for a keyword in handwritten document using OCR technique does not provide a good performance.

Although several successful methods for word spotting on handwritten Latin (David *et al.*, 2005; Lu *et al.*, 2008; Srihari *et al.*, 2006) and Chinese scripts (Huang*, et al.*, 2013; Lu *et al.*, 2002; Lu *et al.*, 2004) have been proposed, there is no noticeable work on Persian handwritten. Persian writing is a cursive horizontal script, whose words consist of sub-words or pieces of Persian words (PPWs). Each of these PPWs consists of one or more letters, and each letter includes one or more connected components. In general, the space between two successive words/PPWs may be within a same range, so the boundary of a word is not clearly indicated. This issue alongside with the inherent cursive structure of Persian scripts (which is more unconstrained than other languages) makes word spotting on Persian handwritten documents a challenging task.

Regarding word recognition, a lot of work is done in Persian and Arabic scripts (Saeed, 2014; Parvez *et al.*, 2013) but, there are only a few papers about Arabic word spotting (Srihari *et al.*, 2006; Aouadi *et al.*, 2011; Kefali *et al.*, 2011; Khayyat *et al.*, 2014). To the best of our knowledge, there has been no effort to address the problem of word spotting on handwritten Persian documents. Pourasad *et al.* (2012) have introduced a system for word spotting on printed Persian documents (Pourasad *et al.*, 2013; Pourasad *et al.*, 2012). This method is based on font recognition, word shape coding (WSC) representation. Furthermore, sub-letter shape coding schema has shown promising result on word retrieval of Persian printed document (Azmi, 2011). This method was proposed for printed documents with four standard fonts; as reported, it fails for handwritten documents.

Word spotting and word image retrieval aims to make handwritten documents searchable. The main difference between them lies in type of the query word. Query in word image retrieval tasks is a text string which is usually referred to "query by string (QBS)" or "query by text (QBT)" term. On the other hand, in word spotting, a query is an image of the word, known by "query by example (QBE)".

Word spotting is presented as an applicable case of semantic content based image retrieval (CBIR), where the classes are fine-grained. In CBIR, we are interested in exactly one particular word, and a difference of only one character is considered a negative result. The key challenge in word spotting and word image retrieval is the existence of very large intra-class variability, which is caused by writing styles, illumination differences, and typography. This intra-class variability causes samples of same class look very different.

In this paper, we propose a word spotting and retrieval method for Persian handwritten images based on works of Almazán *et al.* (2013). This method uses a joint representation for both QBE and QBS. By employing the mentioned representation, QBE and QBS problems are converted to a simple nearest neighbor problem; furthermore, this structure can be used for word recognition.

First, a label embedding approach is used for text labels, motivated by the bag of characters string kernels (Lodhi *et al.*, 2002; Leslie *et al.*, 2002). This method is common in the machine learning and biocomputing communities. It embeds text strings into a so-called pyramidal histogram of characters or PHOC d−dimensional binary space (Almazán *et al.*, 2013). Persian PHOC (PPHOC), which is a modified version of PHOC, is used. PPHOC encodes appearance of any character in a particular spatial region of the string (Figure 1). Then, this representation is used as the attribute vector. The word images are projected into another more discriminative d−dimensional space. Each of these dimensions encodes how likely a word image contains a specific character in a particular region, which is in evident parallelism with the PPHOC descriptor. By learning attributes independently, training data are used in a better way (since the same training words are used to train several attributes) and out of vocabulary (OOV) spotting and retrieval (i.e., Spotting, retrieval and even recognition of words, which never been seen during training) are straightforward. However, due to some differences (PPHOC is a binary vector, while the attribute scores are not); straight comparison is not ideal and some calibration is needed. We finally propose to learn a low-dimensional common subspace with an associated metric between the PPHOC embedding and the attribute embedding. The advantages of this are twofold (Almazán *et al.*, 2014a). First, it makes direct comparison of word images and text strings meaningful. Second, the attribute score of the same word is brought together since they are guided by their shared PPHOC representation. An overview of the method can be seen in Figure 2 (Almazán *et al.*, 2013).

As mentioned above, word spotting, retrieval and recognition become a simple nearest neighbor problem in a low-dimensional space, because of providing a common subspace for images and text string. We can perform QBE and QBT using exactly the same retrieval framework. The recognition task simply becomes finding the nearest neighbor of the image word in a text dictionary, which is embedded into the PPHOC space and then into the common subspace. Because of employing compact vectors, compression and indexing techniques such as product quantization (Jegou *et al.*, 2011) could now be used to perform spotting in very large datasets. To the best of our knowledge, we are the first ones who suggest a unified framework in Persian, where we can perform out of vocabulary (OOV) QBE and QBS retrieval as well as word recognition using the same compact word representation.

This paper is organized as follows. In Section 2, we review the related work in the literature. The proposed method is explained in Section 3. The experimental validation of our approach on two common Persian handwritten isolated word datasets is reported in section 4 and finally, in Section 5, we conclude the paper.

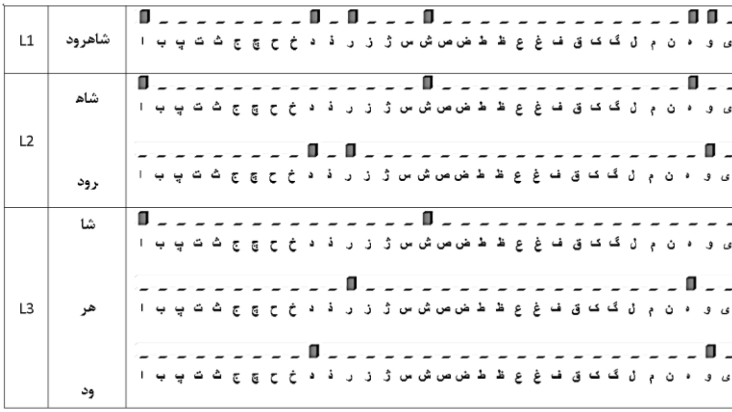


Fig. 1. Persian PHOC (PPHOC). Histogram of word Shahrood “شاهرود” at levels 1, 2 and 3. The final PPHOC histogram is the concatenation of these partial histograms.

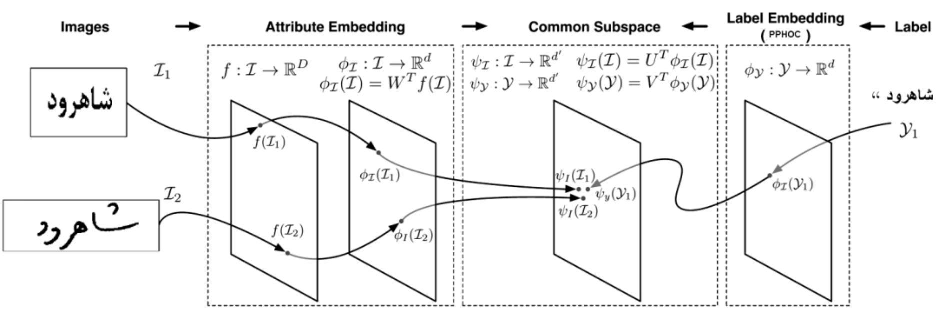


Fig. 2. Overview of method.

RELATED WORKS

In recent years, various methods for word spotting and word image retrieval on document images have been proposed, making it the hallmark of the document image analysis (Fischer *et al.*, 2010; Frinken *et al.*, 2012; Manmatha *et al.*, 1996; Rath *et al.*, 2004; Rath *et al.*, 2007; Rodrıguez *et al.*, 2008; Rodríguez-Serrano *et al.*, 2009; Rodríguez-Serrano *et al.*, 2012). Although word spotting and retrieval on clean, typewritten Latin, Arabic and Persian documents have been almost solved (Pourasad *et al.*, 2013), it is still a challenging problem on handwritten documents, especially for Persian and with Arabic script. These challenges arise from different writing styles of the writers in handwritten documents and imperfections in historical documents. Noise is another issue in this type of documents. Due to these challenges, most of the popular methods are based on describing word images as a sequence of variable length features, which are used by techniques like dynamic time warping (DTW) or hidden Markov models (HMM) for classification. Variable-length features are more flexible than fixed-length feature vectors and have been shown promising results in difficult word spotting tasks. The good performance is rooted in the fact that they can adapt to the different styles and word length (Fischer *et al.*, 2010; Frinken *et al.*, 2012; Manmatha *et al.*, 1996; Rodrıguez *et al.*, 2008; Rodríguez-Serrano *et al.*, 2012; Yalniz *et al.*, 2011).

Uncomfortably, this causes two unsatisfying results. First, due to the difficulty of learning with sequences, many supervised methods cannot perform out of vocabulary (OOV) spotting, i.e., only a limited number of keywords, which are modeled at the training time, can be searched as queries. Second, because the methods deal with sequence of features, the computation of distance between words is typically time consuming at the test time, usually quadratic with respect to the number of features. Therefore, dealing with large volume of data at the test time is not feasible.

No doubt, with the solid increase of dataset size, there has been a renewed interest in compact, fast to-compare word representations. In Manmatha *et al.* (1996), a distance between binary word images based on the XOR result of the images, is defined. In Keaton *et al.*(1997), a set of projections and profile features is extracted and used to compare the images. Both methods are limited to small datasets and a few query keywords. A recent work (Perronnin *et al.*, 2009) exploits the Fisher kernel to construct the Fisher vector of a HMM, which has a fixed length and can be used for inefficient retrieval tasks, but the authors focused only on10 different keywords. Recently, the methods, which are not restricted to a limited number of keywords have been proposed (Gatos *et al.*, 2009; Rusinol *et al.*, 2011; Almazán *et al.*, 2012; Almazán *et al.*, 2014b). Gatos *et al.* (2009) used template matching and block-based image descriptors, Rusinol *et al.* (2011) combined SIFT descriptors into a bag of visual words (BOVW) to describe images, while Almazan´ *et al.* (Almazán *et al.*, 2012; Almazán *et al.*, 2014b) used HOG descriptors (Dalal *et al.*, 2005) combined with an exemplar-SVM framework (Malisiewicz *et al.*, 2011). Although the results on simple datasets are encouraging, these methods do not have enough flexibility to perform well on more complex datasets, especially in a multi-writer scenario.

Persian is a right to left horizontal script, which is inherently cursive. One of the important characteristics of Persian writing style is that the shape of a Persian letter changes depending on its position within the word. A Persian word consists of a number of sub-words, which we call pieces of Persian words (PPWs). Figure 3 shows a sample of Persian word and its PPWs. The Persian writing has a naturally cursive structure and is more unconstrained than other languages. These different structures make Persian word spotting a challenging problem. To the best of our knowledge, the Persian word spotting has been done only on printed documents (Pourasad *et al.*, 2013; Pourasad *et al.*, 2012) by word shape coding (WSC) with structural features such as ascenders, descenders, holes and dots. As mentioned earlier in section 2, some researches have been done in other languages, such as Arabic and Urdu, whose structures are very similar to Persian.

Fig. 3. A sample of Persian handwritten word. Different PPWs are illustrated with different colors. Green and red PPW are contained two characters and others have one character each.

In Fischer *et al.* (2012) a script independent segmentation free method for word spotting has been proposed. This method is based on HMM and the experimental result shows that a learning-based method outperforms the standard template matching. As reported in this work, the performance for the Arabic language is the poorest in comparison with other languages (Khayyat *et al.*, 2014). DTW is a matching technique, which has been extensively used for word spotting. In order to measure the similarity between two connected components the Euclidean distance has been enhanced by rotation along with DTW (Moghaddam *et al.*, 2009). Furthermore, sub-words are clustered by self-organizing map (SOM) based on the shape complexity. Recently, a model-based similarity measure between vector sequences is proposed (Rodríguez-Serrano *et al.*, 2012). In this method, semi-continuous hidden Markov model (SC-HMM) is used for every sequence, then HMMs similarity is computed. The result of applying this measure to three different Arabic datasets shows that the proposed similarity outperforms DTW and ordinary continuous HMMs. Saabni *et al.* (2011) suggested a word matching approach for Arabic script, which embeds the extracted contour features into Euclidean space; finally an active-DTW (Sridha *et al.*, 2006) is applied to match two patterns. A word image retrieval approach based on codebook is proposed in Sari *et al.* (2008); Saykol *et al.* (2004) and Shahab *et al.* (2006). In this approach, codes of symbols, characters or sub-words are represented by meaningful features. Then a final match is performed by computing distance between the codes and the codebook.

In this paper, we follow these recent works (Almazán *et al.*, 2013; Almazán *et al.*, 2014a; Rusinol *et al.*, 2011; Almazán *et al.*, 2012; Almazán *et al.*, 2014b) and focus on fixed-length representation, which are faster to compare and store. Furthermore, they can be used in large-scale scenarios. This approach is based on attributes and directly addresses the aforementioned problems. Opposed to HMM-based methods, this attribute based framework deals with OOV query words very naturally and produces discriminative and compact signatures that are fast to compute, compare, and store. In the next section, we discuss details of our proposed method.

CURRENT WORK

In this paper, we introduce an attribute-based label embedding approach for word spotting and word image retrieval on isolated Persian words. The details of our method are explained below.

Supervised word representations by attributes

In this subsection, the attribute based-representation of a word image is described. We start by introducing the representation using pyramidal histogram of characters (PHOC) (Almazán *et al.*, 2013; Almazán *et al.*, 2014a). This representation embeds label strings into a d−dimensional space. Based on PHOC, we create a Persian pyramidal histogram of characters (PPHOC). In a word spotting system at the learning stage, a set of particular keywords is used to learn word models (usually an HMM). At the test stage, the probability that a given word matches each of keyword models is computed and used as a score. This approach can work when keywords are learned offline. Some researchers attempt to solve the mentioned problem by using semi-continuous HMM (SC-HMM) (Rodríguez-Serrano *et al.*, 2012). After the SC-HMM's parameters are learned, the model can represent the query. These approaches, which learn at the word level, cannot share information between similar words. For example, to learn an HMM model for a word, all words which have only one letter difference are considered as negative samples, such as "كتاب" and "كباب". Hence, such approach cannot be used to share information between words. While, sharing information between words is extremely important to learn good discriminative representations, and that the use of attributes is the best solution. Attributes are semantic properties, which describe images (Farhadi *et al.*, 2009). Recently, a lot of image retrieval and classification approaches have been proposed based on this idea (Farhadi *et al.*, 2009; Akata *et al.*, 2013a; Akata *et al.*, 2013b; Ferrari *et al.*, 2008; Lampert *et al.*, 2013). Since the attributes must be selected based on the task in hand, for word spotting we should define them appropriately.

PPHOC Attribute

This approach embeds text strings, which construct a binary histogram of characters (Almazán *et al.*, 2013; Almazán *et al.*, 2014a). Each dimension shows the existence of a particular character. However, this embedding is not discriminative because it does not inform the location of letters. Therefore, we propose to use a Persian pyramid version of this histogram of characters, which was inspired by PHOC (Almazán *et al.*, 2013; Almazán *et al.*, 2014a) (see Figure 2).

Calibration

In the previous section, we proposed a writer-independent attribute based representation of the word images. Since the scores of one attribute may dominate over the scores of other attributes, special care has to be put for comparing different words. Therefore, some calibration on attribute scores is necessary. This is particularly crucial in QBS, whose attribute scores are not comparable to the binary PPHOC representations. Platts scaling is a popular approach. It consists of fitting a sigmoid over the output scores to obtain calibrated probabilities, P(y = 1| s) = (1 + exp (αs + β))− 1, where α and β can be estimated using MLE. In the more recent paper (Scheirer *et al.*, 2012), extreme value theory is used to fit better probabilities to the scores and to find a multi-attribute space similarity. Although the similarity measure involves all the attributes, the calibration of each attribute is done individually.

Because of its better correlation exploitation capability between different attributes, we propose to perform the calibration of the scores jointly. To achieve this goal, we make use of canonical correlation analysis (CCA) to embed the attribute scores and the binary attributes in a common subspace, where they are maximally correlated. CCA is a tool to exploit available information from different data sources, used for example in retrieval (Hardoon *et al.*, 2004) and clustering (Blaschko *et al.*, 2008). In Gong *et al.* (2013), CCA was used to correlate image descriptors and their labels, which brought significant benefits for retrieval tasks. We believe this is the most similar use of CCA to our approach. However, while Gong *et al.* (2013) combined images and labels with the hope of bringing some semantic consistency to the image representations, our goal here is to bring the imperfect predicted scores closer to their perfect value.

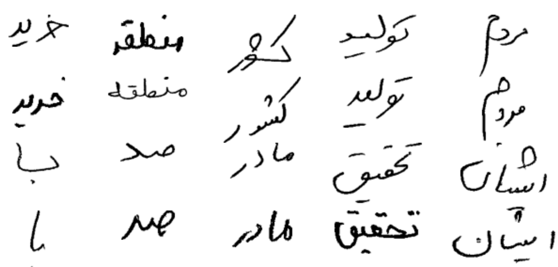
Let us assume that we have N labeled training samples, A ∈ RD× N is the D-dimensional attribute score representation of those samples, and B $\in \{0,1\}^{D\times N}$ is their binary attribute representation. Let μ_a and μ_b denote sample means of A and B. Let $C_{aa} = (1/N)\,(A-\mu_a)(A-\mu_a)^T + \rho I$, $C_{bb} = (1/N)\,(B-\mu_b)(B-\mu_b)^T + \rho I$, $C_{ab} = (1/N)\,(A-\mu_a)(B-\mu_b)^T$, and $C_{ba} = C_{ab}^T$, where ρ is a regularization factor used to avoid numerically ill-conditioned situations and I is the identity matrix. The goal of CCA is to find a projection, which maximizes projected representations. This can be expressed as:

$$\operatorname{argmax}_{w_a, w_b} \frac{w'_a C_{ab} w_b}{\sqrt{w'_a C_{aa} w_a}\sqrt{w'_b C_{bb} w_b}} \tag{1}$$

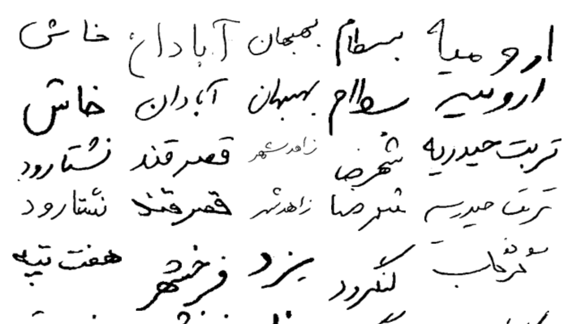
This CCA embedding can be seen as a way to exploit the correlation between different attributes to correct the scores predicted by the model. Furthermore, after performing CCA, the attribute scores and binary attributes lie in a more correlated space, which makes the comparison between the scores and the PPHOCs for our QBS problem more principled. CCA can also be seen as a label embedding method, inherently similar to the approach proposed in Rodriguez-Serrano *et al.*(2013). CCA is also used as a dimensionality reduction tool.

EXPREMENTAL RESULT

We evaluate our method on two isolated Persian word datasets: Farsa and Iranshahr. Farsa dataset (Imani *et al.*, 2013), contains 30000 images of 300 common words of Farsi formal language that scanned. Some sample images of Farsa dataset are shown in Figure 3(a) Iranshahr dataset consists of 17000 images from 503 Iranian city names. For each word, at least 20 samples were provided which we use 20 samples per class. Some sample images of Iranshahr dataset are shown in Figure 3(b). The images of both datasets were divided into training (60%), and testing (40%) sets.



(a) Farsa



(b) Iranshahr

Fig. 4. Some samples of (a) Farsa and (b) Iranshahr Datasets.

The results of experiments using the proposed approach on word retrial (QBS) and word spotting (QBE) are shown in Table 1. For Farsa and Iranshahr datasets, we compare our method with Fisher vector (FV), which can only be used in QBE; attribute based classification (Attribute) without calibration, calibrated with Platt Scaling (Attribute + Platt Scaling), one way regression (Attribute + Regression), Common Subspace Regression (Attribute + CSR) and kernelized common subspace regression (Attribute + KCSR).

From Table 1 it can be seen that the use of attributes effectively increases the performance, even when no calibration is conducted. This is not surprising, since the attribute space has been learned using significant amounts of labeled training data. It emphasizes the idea that exploiting labeled data during training is vital to obtain competitive results. Conversely, the QBS results are not particularly good, since the direct comparison between attribute scores and PPHOCs is not well principled.

Table 1. Word spotting and retrieval results on Farsa and Iranshahr datasets.

Method	**Farsa**		**Iranshahr**	
	QBE	***QBS***	***QBE***	***QBS***
Fisher Vector	96.15	-	62.40	-
Attribute	98.67	98.33	93.81	91.04
Attribute + Platt Scaling	94.85	100.00	93.91	99.60
Attribute + Regression	98.51	100.00	93.56	97.61
Attribute + CSR	97.71	100.00	94.68	99.20
Attribute + KCSR	98.88	100.00	96.00	99.20

Word recognition result on Farsa and Iranshahr datasets are shown in Table 2 and 3 respectively. In Farsa dataset, we compared our proposed method with three other HMM-based word image modeling approaches. We also compared the proposed method with two recent works (Ebrahimpour et al., 2011a; Ebrahimpour *et al.*, 2011b) which are based on ensemble classifiers. This experiment has been performed on Iranshahr dataset. The result shows that attribute based label embedding by making common subspace between class's sample and class's label along with attributes can change the viewpoint of classification in this problem and improves the recognition results.

Table 2. Word recognition results on Farsa dataset.

Method	**Lexicon Size**	**Top-1**	**Top-5**	**Top-10**	**Top-20**
DHMM+smoothing (Dehghan, *et al.*, 2001a)	198	65.05	76.09	90.83	95.00
FVQ/HMM (Dehghan, *et al.*, 2001b)	198	67.18	87.55	92.06	96.5
SOM+DHMM+smoothing (Imani, *et al.*, 2013)	198	68.88	87.54	91.75	95.63
Proposed Method	**300**	**99.03**	-	-	-

Table 3. Word recognition results on Iranshahr dataset.

Method	Lexicon size	Recognition rate
Divide and conquer (Ebrahimpour Reza, 2011)	503	90.50
Mixture of Expert (Ebrahimpour*, et al.*, 2011)	503	91.11
Proposed Method	**503**	**96.94**

CONCLUSION

In this paper, we proposed an approach to represent and compare word images. Our approach is applicable in word spotting, word image retrieval and word recognition. We showed how an attributes-based approach based on a PPHOC can be used to embed the word images and their textual class labels into a shared, more discriminative space. In this discriminative space, the similarities between words are independent of the writing and font style. This attribute representation leads to a unified representation of word images and strings. As a result, the proposed method allows one to perform query-by-example or query-by-string searches, as well as word recognition, in a unified framework. We evaluated our proposed method on two public Persian datasets. Experimental results show that our proposed method outperforms other methods in both QBS and GBE. For future work, we have observed empirically that the quality of the attribute models is quite dependent on the available number of training samples, and our model does not perform well for less used characters in unlikely positions. We believe that larger training sets could improve the quality of those models and lead to better overall results. Furthermore, our learning approach is currently based on whole word images and does not require segmenting the individual characters of the words during training or test, which is an advantage of our model. However, it has been shown that learning on characters can lead to large improvements in accuracy. We aim to study how we could learn on individual segmented characters and transfer that information into our system without manually segmenting characters of words in the training set.

REFERENCES

Akata, Z. & Perronnin, F. & Harchaoui, Z. & Schmid, C. 2013a. Label-embedding for attribute-based classification. Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on.

Akata, Z. & Perronnin, F. & Harchaoui, Z. & Schmid, C. 2013b. Attribute-Based Classification with Label-Embedding. NIPS 2013 Workshop on Output Representation Learning.

Almazán, J. & Gordo, A. & Fornés, A. & Valveny, E. 2014a. Word Spotting and Recognition with Embedded Attributes. 12: 2552-2566.

Almazán, J. & Gordo, A. & Fornés, A. & Valveny, E. 2013. Handwritten Word Spotting with Corrected

Attributes. ICCV 2013-IEEE International Conference on Computer Vision (2013).

Almazán, J. & Gordo, A. & Fornés, A. & Valveny, E. 2012. Efficient exemplar word spotting. Proceedings of the British Machine Vision Conference.

Almazán, J. & Gordo, A. & Fornés, A. & Valveny, E. 2014b. Segmentation-free word spotting with exemplar SVMs. Pattern Recognition. **47**:3967-3978..

Aouadi, N. & Kacem, A. 2011. Word spotting for Arabic handwritten historical document retrieval using generalized hough transform. PATTERNS 2011, The Third International Conferences on Pervasive Patterns and Applications.

Azmi, R. 2011. Farsi/Arabic Document Image Retrieval through Sub-Letter Shape Coding for mixed Farsi/Arabic and English text. International Journal of Computer Science Issues (IJCSI). **8**:166-172.

Blaschko, M.B. & Lampert, C.H. 2008. Correlational spectral clustering. Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.

Dalal, N. & Triggs, B. 2005. Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.

David, J.E.Y.W.T. & Maire, F.R.B.M. & Vesom, G. 2005. Making latin manuscripts searchable using gHMM's. Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference.

Dehghan, M. & Faez, K. & Ahmadi, M. & Shridhar, M. 2001a. Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM. Pattern Recognition. **34**:1057-1065.

Dehghan, M. & Faez, K. & Ahmadi, M. & Shridhar, M. 2001b. Unconstrained Farsi handwritten word recognition using fuzzy vector quantization and hidden Markov models. Pattern Recognition Letters. **22**:209-214.

Ebrahimpour Reza & Sarhangi, S. & Javadi M. & Sharifizadeh F. 2011a. Handwritten Word Recognition by Multiple Classifiers: A Divide-and-Conquer Approach. International Journal of Applied Research on Information Technology and Computing. **2**:9-21.

Ebrahimpour, R. & Sarhangi, S. & Sharifizadeh, F. 2011b. Mixture of Experts for Persian handwritten word recognition. Iranian Journal of Electrical & Electronic Engineering. **7**:217-224.

Farhadi, A. & Endres, I. & Hoiem, D. & Forsyth, D. 2009. Describing objects by their attributes. Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, Pp 1778-1785.

Ferrari, V. & Zisserman, A. 2008. Learning visual attributes. Advances in Neural Information Processing Systems. Pp

Fischer, A. & Keller, A. & Frinken, V. & Bunke, H. 2010. HMM-based word spotting in handwritten documents using subword models. Pattern recognition (icpr), 2010 20th international conference on, Pp 3416-3419.

Frinken, V. & Fischer, A. & Manmatha, R. & Bunke, H. 2012. A novel word spotting method based on recurrent neural networks. Pattern Analysis and Machine Intelligence, IEEE Transactions on. Pp34: 211-224,34.

Gatos, B. & Pratikakis, I. 2009. Segmentation-free word spotting in historical printed documents. Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on.

Gong, Y. & Lazebnik, S. & Gordo, A. & Perronnin, F. 2013. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. Pattern Analysis and Machine Intelligence, IEEE Transactions on. **35**:2916-2929.

Hardoon, D.R. & Szedmak, S. & Shawe-Taylor, J. 2004. Canonical correlation analysis: An overview with application to learning methods. Neural computation. **16**:2639-2664.

Huang, L. & Yin, F. & Chen, Q.H. & Liu, C.L. 2013. Keyword spotting in unconstrained handwritten Chinese documents using contextual word model. Image and Vision Computing. **31**:958-968.

Imani, Z. & Ahmadifard, A.R. & Zohrevand, A. 2013. Introduction to Database FARSA: digital images of handwritten Farsi words. Eleventh Iranian Conference on Intelligent Systems (ICIS2013).

Imani, Z. & Ahmadyfard, A.R. & Zohrevand, A. & Alipour, M. 2013. Offline handwritten Farsi cursive text recognition using hidden Markov models. Machine Vision and Image Processing (MVIP), 8th Iranian Conference on.

Jegou, H. & Douze, M. & Schmid, C. 2011. Product quantization for nearest neighbor search. Pattern Analysis and Machine Intelligence, IEEE Transactions on. **33**:117-128.

Keaton, P. & Greenspan, H. & Goodman, R. 1997. Keyword spotting for cursive document retrieval. Document Image Analysis, 1997.(DIA'97) Proceedings., Workshop on.

Kefali, A. & Chemmam, C. 2011. A Semi-Automatic Approach of old Arabic Documents Indexing. CIIA.

Khayyat, M. & Lam, L. & Suen, C.Y. 2014. Learning-based word spotting system for Arabic handwritten documents. Pattern Recognition. **47**:1021-1030.

Lampert, C. & Nickisch, H. & Harmeling, S. 2013. Attribute-based classification for zero-shot learning of object categories. **36**:453:465.

Leslie, C.S. & Eskin, E. & Noble, W.S. 2002. The spectrum kernel: A string kernel for SVM protein classification. Pacific symposium on biocomputing.

Lodhi, H. & Saunders, C. & Shawe-Taylor, J. & Cristianini, N. & Watkins, C. 2002. Text classification using string kernels. The Journal of Machine Learning Research. **2**:419-444.

Lu, Y. & Tan, C.L. 2002. Word spotting in Chinese document images without layout analysis. Pattern Recognition, 2002. Proceedings. 16th International Conference on.

Lu, Y. &Tan, C.L. 2004. Chinese word searching in imaged documents. International Journal of Pattern Recognition and Artificial Intelligence. **18**:229-246.

Lu, S. & Tan, C.L. 2008. Retrieval of machine-printed Latin documents through Word Shape Coding. Pattern Recognition. **41**:1799-1809.

Malisiewicz, T & Gupta, A. & Efros, A.A. 2011. Ensemble of exemplar-svms for object detection and beyond. Computer Vision (ICCV), 2011 IEEE International Conference on.

Manmatha, R. & Han, C. & Riseman, E.M. 1996. Word spotting: A new approach to indexing handwriting. Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on.

Moghaddam, R.F. & Cheriet, M. 2009. Application of multi-level classifiers and clustering for automatic word spotting in historical document images. Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on.

Parvez, M.T. & Mahmoud, S.A. 2013. Offline Arabic handwritten text recognition: a survey. ACM Computing Surveys (CSUR). **45**:23-45.

Perronnin, F. & Rodriguez-Serrano, J.A. 2009. Fisher kernels for handwritten word-spotting. Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on.

Pourasad, Y. & Hassibi, H. & Ghorbani, A. 2013. A word spotting method for Farsi machine-printed document images. Turkish Journal of Electrical Engineering & Computer Sciences. **21**:734-746.

Pourasad, Y. & Hassibi, H. & Ghorbani, A. 2012. A Farsi/Arabic Word Spotting Approach for Printed Document Images. International Journal of Natural & Engineering Sciences, **6**:15-18.

Rath, T.M. & Manmatha, R. & Lavrenko, V. 2004. A search engine for historical manuscript images. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval.

Rath, T.M. & Manmatha, R. 2007. Word spotting for historical documents. International Journal of Document Analysis and Recognition (IJDAR).

Rodrıguez, J.A. & Perronnin, F. 2008. Local gradient histogram features for word spotting in unconstrained handwritten documents. Int. Conf. on Frontiers in Handwriting Recognition.

Rodriguez-Serrano, J.A. & Perronnin, F. & Meylan, F. 2013. Label embedding for text recognition. Proceedings of the British Machine Vision Conference.

Rodríguez-Serrano, J.A. & Perronnin, F. 2009. Handwritten word-spotting using hidden Markov models and universal vocabularies. Pattern Recognition. **42**:2106-2116.

Rodríguez-Serrano, J.A. & Perronnin, F. 2012. A model-based sequence similarity with application to handwritten word spotting. Pattern Analysis and Machine Intelligence, IEEE Transactions on. **34**: 2108-2120.

Rusinol, M. & Aldavert, D. & Toledo, R. & Lladós, J. 2011. Browsing heterogeneous document collections by a segmentation-free word spotting method. Document Analysis and Recognition (ICDAR), 2011 International Conference on.

Saabni, R. & Bronstein, A. 2011. Fast key-word searching via embedding and Active-DTW. Document Analysis and Recognition (ICDAR), 2011 International Conference on.

Saeed, U. 2014. Automatic Recognition of Handwritten Arabic Text: A Survey. Life Science Journal. 11:.

Sari, T. & Kefali, A. 2008. A search engine for Arabic documents. Dixième Colloque International Francophone sur l'Ecrit et le Document.

Saykol, E. & Sinop, A.K. & Gudukbay, U. & Ulusoy, O. & Çetin, A.E. 2004. Content-based retrieval of historical Ottoman documents stored as textual images. Image Processing, IEEE Transactions on. **13**:314-325.

Scheirer, W.J. & Kumar, N. & Belhumeur, P.N. & Boult, T.E. 2012. Multi-attribute spaces: Calibration for attribute fusion and similarity search. Computer Vision and Pattern Recognition (CVPR), IEEE Conference on.

Shahab, S. & Al-Khatib, W.G. & Mahmoud, S.A. 2006. Computer aided indexing of historical manuscripts. Computer Graphics, Imaging and Visualisation, 2006 International Conference on.

Sridha, M. & Mandalapu, D. & Patel, M. 2006. Active-dtw: a generative classifier that combines elastic matching with active shape modeling for online handwritten character recognition. Tenth International Workshop on Frontiers in Handwriting Recognition.

Srihari, S.N. & Srinivasan, H. & Huang, C. & Shetty, S. 2006. Spotting words in Latin, Devanagari and Arabic scripts. VIVEK-BOMBAY.

Yalniz, I.Z. & Manmatha, R. 2011. An efficient framework for searching text in noisy document images. Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on.

Submitted: 05/07/2015
Revised: 10/07/2016
Accepted: 26/09/2016