Incoherent Dictionary Pair Learning: Application to a Novel Open-Source Database of Chinese Numbers

Vahid Abolghasemi, *Senior Member, IEEE*, Mingyang Chen, Ali Alameer, *Student Member, IEEE*, Saideh Ferdowsi, *Member, IEEE*, Jonathon Chambers, *Fellow, IEEE*, and Kianoush Nazarpour, *Senior Member, IEEE*

***Abstract*—We enhance the efficacy of an existing dictionary pair learning algorithm by adding a dictionary incoherence penalty term. After presenting an alternating minimization solution, we apply the proposed incoherent dictionary pair learning (InDPL) method in classification of a novel open-source database of Chinese numbers. Benchmarking results confirm that the InDPL algorithm offers enhanced classification accuracy, especially when the number of training samples is limited.**

***Index Terms*—Chinese numbers, classification, incoherent dictionary pair learning.**

I. Introduction

Chinese numbers represent the wealth of China's history and culture. Certain numbers can be considered auspicious. For example, number 6 (八) (Pinyin: *lìu*) is associated with six types of morality and it can be used to express the wish of success. Likewise, number 8 (六) (Pinyin: *bā*) is associated with luck because it sounds similar to the word 发 (Pinyin: *fā*), which means "make a fortune, to be rich." In China, two indigenous number systems, namely simplified and traditional, are used to communicate numeral values. An example of certain simplified Chinese numbers and their Hindu–Arabic counterparts is shown in Fig. 1. Traditional Chinese numerals, also known as *banker's numerals*, are used in commerce because of their robustness against forgery.

Handwritten character recognition is an established pattern recognition problem [1]–[5]. However, despite the wealth of literature in pattern recognition of Chinese characters, e.g., [6]–[8], there is surprisingly little work carried out on the classification of handwritten Chinese numbers [9]. This may be partly due to lack of a user-friendly and compact database of Chinese numbers. Here, we present a new database of handwritten simplified Chinese numbers acquired from 100 Chinese nationals. In addition, to classify these numbers, we introduce a novel concurrent dictionary learning and classification algorithm.

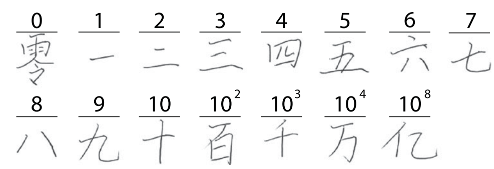

Fig. 1. Example of the simplified Chinese numbers (over the range 0–10^8).

Classic dictionary learning methods do not explicitly embed pattern discrimination within the dictionary construction procedure. Recently, the notion of class-specific dictionary design for classification has been proposed [10]–[15]. For instance, in discriminative K-SVD (D-KSVD) [16], the classification error was incorporated into the objective function. Li *et al.* [17] presented a reference-based objective function that was combined with the K-SVD algorithm for scene image categorization. Similarly, the label consistent K-SVD method attempted to associate label information with columns of the dictionary matrix during learning [18]–[20].

Atoms of a learned dictionary are typically desired to be incoherent [21]–[29]. Several techniques have been proposed to enhance incoherence in dictionary learning. For example, Mailhé *et al.* [21] and Abolghasemi *et al.* [24] added an incoherence penalty to the K-SVD dictionary learning algorithm [30]. In addition, a joint dictionary learning-projection was developed for compressive sensing in [31]. However, existing works that address incoherence in dictionaries for classification tasks are neither sufficient nor application-specific; examples include [32], [34].

We therefore integrated an incoherence penalty term into the dictionary pair learning (DPL) [33] algorithm aiming to minimize similarity (measured by inner product) between the dictionary atoms associated with different classes. Upon deriving a solution by employing an alternating minimization strategy, we verified the efficacy of this approach in classification of our novel dataset of Chinese numbers.

II. Method

A. Data Collection

One hundred Chinese students took part in data collection. Each participant wrote with a standard black ink pen all

Manuscript received October 17, 2017; revised December 22, 2017; accepted January 18, 2018. Date of publication January 25, 2018; date of current version February 15, 2018. The work of A. Alameer was supported by the Higher Committee for Education Development, Iraq (HCED, D1201017). The work of K. Nazarpour was supported by the Engineering and Physical Sciences Research Council, U.K. (EP/M025594/1). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mireille Boutin.
(Corresponding author: Vahid Abolghasemi and Kianoush Nazarpour.)

V. Abolghasemi and S. Ferdowsi are with the Faculty of Electrical Engineering and Robotics, Shahrood University of Technology, Shahrood 36155-316, Iran (e-mail: vahidabolghasemi@gmail.com).

M. Chen is with the Department of Electronic Engineering, University of Surrey, Guildford GU2 7XH, U.K.

A. Alameer and K. Nazarpour are with the School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. (e-mail: kianoush.nazarpour@newcastle.ac.uk).

J. Chambers is with the School of Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K., and also with the College of Automation, Harbin Engineering University, Harbin 0086, China.
Digital Object Identifier 10.1109/LSP.2018.2798406

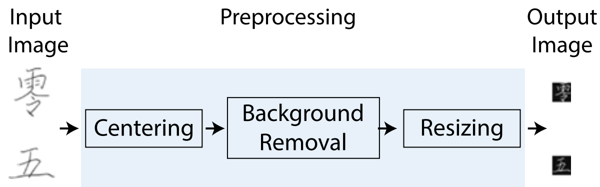

Fig. 2. Preprocessing and resizing (25×25 pixels) input images.

15 numbers in a table with 15 designated regions drawn on a white A4 paper. This process was repeated ten times with each participant. Each sheet was scanned at the resolution of 300×300 pixels.

B. Preprocessing

Subjects were instructed to write the numbers at the center of the designated region. However, deviations were inevitable. To avoid classification error, we adopted a preprocessing procedure comprising: 1) images were scanned vertically and horizontally to determine the center and the bounding box of the number; 2) after centering, background was removed; and 3) images were resized to 25×25 pixels, as depicted in Fig. 2.

C. Discriminative Dictionary Learning

Let matrix $\mathbf{X}_i \in \mathbb{R}^{m\times n}, i = 1,\ldots,K$, comprise all n training samples in class i and $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K]$, where K is the number of classes. If $\mathbf{D}_i \in \mathbb{R}^{m\times p}$ denotes a synthesis dictionary and $\mathbf{S}_i \in \mathbb{R}^{p\times n}$ is a sparse coefficient matrix, discriminative dictionary learning [18]–[20] can be achieved with

$$\left\{\hat{\mathbf{D}}, \hat{\mathbf{S}}\right\} = \arg\min_{\mathbf{D},\mathbf{S}} \left[\|\mathbf{X} - \mathbf{D}\circ\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_1 + \mathcal{R}(\mathbf{X},\mathbf{D},\mathbf{S})\right] \tag{1}$$

where $\lambda > 0$ is a constant scalar, $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_K]$, $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_K]$, and $\mathbf{D}\circ\mathbf{S} = [\mathbf{D}_1\mathbf{S}_1, \mathbf{D}_2\mathbf{S}_2, \ldots, \mathbf{D}_K\mathbf{S}_K]$ denotes the block Hadamard product. In addition, $\|.\|_F^2$ and $\|.\|_1$ are the Frobenius and ℓ_1-norms, respectively. The penalty term $\mathcal{R}(\mathbf{X},\mathbf{D},\mathbf{S})$ is normally defined with the aim of improving classification. Gu *et al.* [34] extended the conventional problem (1) into the DPL model by including a linear decomposition of the sparse matrix as $\mathbf{S} = \mathbf{P}\circ\mathbf{X}$ with $\mathbf{P} \in \mathbb{R}^{m\times nK}$ being an *analysis* dictionary. In this setting, simultaneous learning of $\mathbf{D}$ and $\mathbf{P}$ enabled avoiding direct approximation of the sparse coding coefficients in $\mathbf{S}$. They defined the following:

$$\left\{\hat{\mathbf{D}}, \hat{\mathbf{S}}, \hat{\mathbf{P}}\right\} = \arg\min_{\mathbf{D},\mathbf{S},\mathbf{P}} \sum_{i=1}^{K}\Big[\|\mathbf{X}_i - \mathbf{D}_i\mathbf{S}_i\|_F^2 + \tau\|\mathbf{S}_i - \mathbf{P}_i\mathbf{X}_i\|_F^2 + \lambda\left\|\mathbf{P}_i\overline{\mathbf{X}}_i\right\|_F^2\Big] \tag{2}$$

where $\tau, \lambda > 0$ are constant scalars and $\overline{\mathbf{X}}_i$ denotes a matrix that includes samples from all classes except that of the ith class. Therefore $\mathbf{P}_i$ will best represent samples of the ith class and simultaneously least represent other samples in other classes. The matrices $\mathbf{D}_i$ and $\mathbf{P}_i$ were used for classification.

D. Incoherent Dictionary Pair Learning (InDPL)

The DPL algorithm has a penalty term corresponding to the analysis subdictionary $\mathbf{P}_i$ for the ith class that projects the samples of all other classes to an approximate-null space. By adding an incoherence penalty to learning of the synthesis subdictionary $\mathbf{D}_i$ we modified the DPL cost function to

$$\left\{\hat{\mathbf{D}}, \hat{\mathbf{S}}, \hat{\mathbf{P}}\right\} = \arg\min_{\mathbf{D},\mathbf{S},\mathbf{P}} \sum_{i=1}^{K}\Big[\|\mathbf{X}_i - \mathbf{D}_i\mathbf{S}_i\|_F^2 + \tau\|\mathbf{S}_i - \mathbf{P}_i\mathbf{X}_i\|_F^2 + \lambda\left\|\mathbf{P}_i\overline{\mathbf{X}}_i\right\|_F^2\Big] + \beta\sum_{j\neq i}\left\|\mathbf{D}_j^T\mathbf{D}_i\right\|_F^2 \tag{3}$$

where $\beta > 0$ is a constant scalar and $(.)^T$ denotes matrix transpose operation. The added penalty attempts to enforce $\mathbf{D}_j^T\mathbf{D}_i \approx \mathbf{0} \quad \forall i \neq j$. To approximate $\mathbf{D}, \mathbf{P}$, and $\mathbf{S}$, we alternately kept two fixed and computed the third. For instance, by fixing $\mathbf{D}$ and $\mathbf{P}$, taking the derivative of (3) with respect to $\mathbf{S}_i$, and equating to zero, $\mathbf{S}_i$ was calculated as

$$\hat{\mathbf{S}}_i = \left[\mathbf{D}_i^T\mathbf{D}_i + \tau\mathbf{I}\right]^{-1}\left[\tau\mathbf{P}_i\mathbf{X}_i + \mathbf{D}_i^T\mathbf{X}_i\right] \tag{4}$$

where $\mathbf{I}$ is the identity matrix. After repeating (4) for all classes, we have $\hat{\mathbf{S}} = [\hat{\mathbf{S}}_1, \hat{\mathbf{S}}_2, \ldots, \hat{\mathbf{S}}_K]$. Similarly, $\mathbf{P}_i$ will be

$$\hat{\mathbf{P}}_i = \tau\mathbf{S}_i\mathbf{X}_i^T\left[\tau\mathbf{X}_i\mathbf{X}_i^T + \lambda\overline{\mathbf{X}}_i\overline{\mathbf{X}}_i^T + \gamma\mathbf{I}\right]^{-1} \tag{5}$$

where γ is typically a small positive parameter to avoid division by zero. We have $\hat{\mathbf{P}} = [\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, \ldots, \hat{\mathbf{P}}_K]$ after repeating this step for all classes. The matrix $\mathbf{D}$ was calculated with the iterative method of alternating direction method of multipliers (ADMM) [35]. An auxiliary matrix $\mathbf{T}$ was introduced into (3)

$$\left\{\hat{\mathbf{D}}, \hat{\mathbf{T}}\right\} = \arg\min_{\mathbf{D},\mathbf{T}} \sum_{i=1}^{K}\|\mathbf{X}_i - \mathbf{D}_i\mathbf{S}_i\|_F^2 + \beta\sum_{j\neq i}\left\|\mathbf{D}_j^T\mathbf{D}_i\right\|_F^2$$
$$\text{s.t.} \quad \mathbf{D} = \mathbf{T}, \ \left\|\mathbf{t}^k\right\|_2^2 = 1 \tag{6}$$

where $k \in \{1, 2, \cdots, p\}$ and $\mathbf{t}^k$ denotes the kth column of $\mathbf{T}$. The columns of $\mathbf{T}$ were normalized to avoid trivial solutions. The solution is then obtained iteratively based on a triple subproblem set

$$\mathbf{D}^{(r+1)} = \arg\min_{\mathbf{D}} \sum_{i=1}^{K}\Big[\|\mathbf{X}_i - \mathbf{D}_i\mathbf{S}_i\|_F^2 + \rho\left\|\mathbf{D}_i - \mathbf{T}_i^{(r)} + \mathbf{U}_i^{(r)}\right\|_F^2\Big] + \beta\sum_{j\neq i}\left\|\mathbf{D}_j^T\mathbf{D}_i\right\|_F^2 \tag{7}$$

$$\mathbf{T}^{(r+1)} = \arg\min_{\mathbf{T}} \sum_{i=1}^{K}\rho\left\|\mathbf{D}_i^{(r+1)} - \mathbf{T}_i + \mathbf{U}_i^{(r)}\right\|_F^2$$
$$\text{s.t.} \quad \left\|\mathbf{t}^k\right\|_2^2 = 1 \tag{8}$$

$$\mathbf{U}^{(r+1)} = \mathbf{U}^{(r)} + \mathbf{D}^{(r+1)} - \mathbf{T}^{(r+1)} \tag{9}$$

where r is the iteration index and $0 < \rho < 1$ is a scalar that gradually increases at rate $\rho_{\text{rate}} \geq 1$.

Closed-form solutions for (7) and (8) can be obtained by taking the derivatives of every subdictionary and equating

Algorithm 1: Incoherent dictionary pair learning (InDPL).

Input: $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K$ and parameters
$\lambda = 0.005, \tau = 1, \beta = 0.08, \gamma = 10^{-4}, \rho = 1, \rho_{rate} = 1.2$,
and synthesis sub-dictionary size $p = 30$.
Initialize $\mathbf{D}^{(0)}$, $r = 0$.
Output: $\mathbf{D}$, $\mathbf{P}$, and $\mathbf{S}$
for $l \leftarrow 1, Iter$ **do**
  for $i \leftarrow 1, K$ **do**
    perform (4) for $\mathbf{S}_i$
  end for
  $\mathbf{S} \leftarrow [\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_K]$
  for $i \leftarrow 1, K$ **do**
    perform (5) for $\mathbf{P}_i$
  end for
  $\mathbf{P} \leftarrow [\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_K]$
  for $i \leftarrow 1, K$ **do**
    repeat
      perform (10) to solve for $\mathbf{D}_i$
      $\mathbf{T}_i \leftarrow \mathbf{D}_i + \mathbf{U}_i$
      Normalize columns of $\mathbf{T}_i$
      $\mathbf{U}_i \leftarrow \mathbf{U}_i + \mathbf{D}_i - \mathbf{T}_i$
      $\rho \leftarrow \rho * \rho_{rate}$
      $r \leftarrow r + 1$
    until convergence
    end for
  $\mathbf{D} \leftarrow [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_K]$
  end for
end for

to zero

$$\mathbf{D}_i^{(r+1)} = \left[\rho(\mathbf{T}_i^{(r)} - \mathbf{U}_i^{(r)}) + \mathbf{X}_i \mathbf{S}_i^T\right] \left[\rho \mathbf{I} + \mathbf{S}_i \mathbf{S}_i^T + \beta \sum_{j \neq i} \mathbf{D}_j^T \mathbf{D}_j\right]^{-1} \quad (10)$$

$$\mathbf{T}_i^{(r+1)} = \mathbf{D}_i^{(r+1)} + \mathbf{U}_i^{(r)}. \quad (11)$$

The pseudocode of the proposed approach is given in Algorithm 1.

For any two distinct synthesis dictionaries $\mathbf{D}_i$ and $\mathbf{D}_j$, three incoherence measures, namely, $\mu_{\min}, \mu_{\max}$, and μ_{average} can be calculated as follows:

$$\mu_{\text{average}} = \frac{1}{K} \sum_{i \neq j} \left\| \mathbf{D}_j^T \mathbf{D}_i \right\|_F^2$$
$$\mu_{\max} = \max_{i \neq j} \left\| \mathbf{D}_j^T \mathbf{D}_i \right\|_F^2$$
$$\mu_{\min} = \min_{i \neq j} \left\| \mathbf{D}_j^T \mathbf{D}_i \right\|_F^2 . \quad (12)$$

Smaller values of $\mu_{\min}, \mu_{\max}$, and μ_{average} indicate that higher incoherence between dictionaries is achieved.

In both DPL and InDPL algorithms, if a test sample $\mathbf{y}$ belongs to class i, then the error $\|\mathbf{y} - \mathbf{D}_i \mathbf{P}_i \mathbf{y}\|_2^2$ would be the smallest.

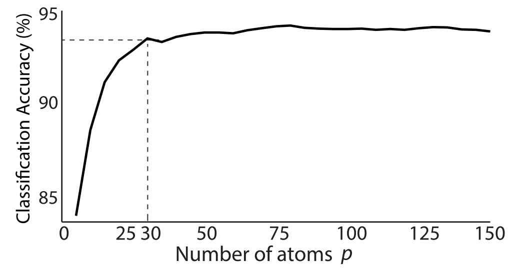

Fig. 3. Classification accuracy of the proposed method versus the number of atoms p within the synthesis dictionary.

E. Benchmarking

We compared the performance of the proposed InDPL algorithm with the performance of the DPL algorithm [34]. In addition, as a benchmark, we compared the results to the case when only a *k*-nearest neighbor (*k*NN) classifier was applied to the resized data without any dimensionality reduction or dictionary learning. Inputs to the *k*NN classifier were vectorized images and their class labels.

F. Cross Validation

Three different cross-validation techniques were implemented. In the interest of clarity, we use the following notation: number of subjects $n_s = 100$, number of repetitions $n_r = 10$, and number of Chinese numbers $n_c = 15$.

1) *Conventional:* All n_r repetitions from all n_s subjects were pooled. Therefore, the data set has $(n_s \times n_r)$ images for each of the n_c classes; total: $n_s \times n_r \times n_c = 15,000$. We then performed a conventional 10-fold cross-validation.
2) *Between-Subjects:* In each fold of this cross-validation method, the training set comprised data from $n_s - 1$ subjects, n_r repetitions, and n_c classes. The testing set included the remaining $n_s = 1$ subject, n_r repetitions, and n_c classes. We repeated this process 100 times. Each time all 10×15 images from a distinct subject were left out for testing.
3) *Within-Subject:* In each fold of this cross-validation method, the training set comprised data from n_s subjects, $n_r - 1$ repetitions, and n_c classes. The testing set included the remaining n_s subjects, $n_r = 1$ repetition, and n_c classes. We repeated this process 10 times, each time all 100×15 images from a distinct repetition were left out for testing.

III. Results

To choose an optimum number of dictionary atoms, we plotted the overall classification accuracy, achieved with the InDPL method, versus the number of atoms in each dictionary p. Fig. 3 shows that as p reaches 30, essentially peak accuracy is achieved. In order to avoid unnecessary combinational complexity, we therefore chose $p = 30$ beyond which the improvement in classification accuracy was negligible.

Fig. 4 illustrates an example of learned dictionary atoms for all 15 classes. For the purpose of clarity in visualization we have included only 25 of the 30 atoms in this figure. The dictionary atoms of each class distinctly represent one specific number. This observation suggested that the added penalty for

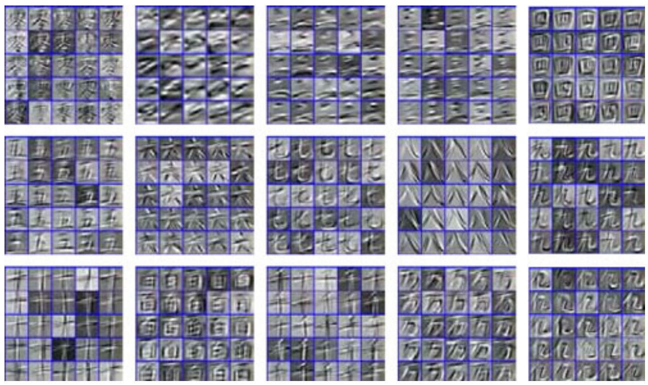

Fig. 4. Illustration of 25 learned atoms of all 15 synthesis dictionaries. Atoms in each class clearly reflect the class label. In analysis, 30 atoms were used.

TABLE I
COHERENCE VALUES AMONG PAIRS OF SYNTHESIS DICTIONARIES

	$\mu_{\min}$	$\mu_{\max}$	μ_{average}
InDPL	**3.02**	**6.72**	**4.51**
DPL	3.34	7.19	5.30

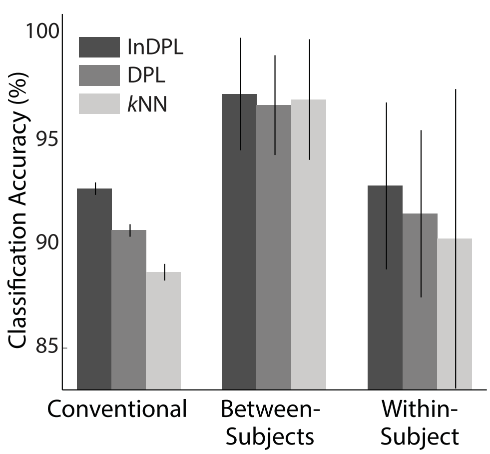

Fig. 5. Average classification accuracy with standard deviation for the three cross-validation cases.

incoherence effectively enforced the dictionaries to be as discriminative as possible. The coherence between the learned dictionaries is reported in Table I - smaller numbers mean higher incoherence. The values associated with the proposed InDPL algorithm were consistently smaller than those achieved with the DPL algorithm.

Fig. 5 shows the average classification accuracy that was achieved with each of the three InDPL, DPL, and *k*NN algorithms. These scores are plotted with respect to the cross-validation method. The InDPL algorithm outperformed the DPL and the *k*NN algorithms in both conventional and within-subject cross validations. The three algorithms exhibited comparable performance in the between-subjects cross validation, suggesting that when a large training dataset is available the choice of algorithm is less important from the accuracy point of view.

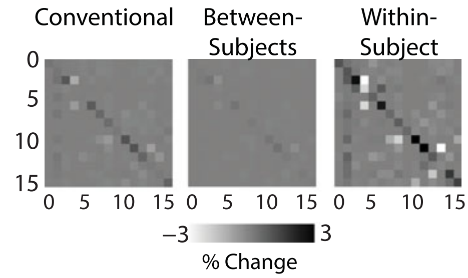

Fig. 6. Confusion matrices; reflecting improvements in accuracy by using InDPL versus DPL.

In addition, we observed a significantly larger standard deviation in the case of classification with *k*NN in the within-subject cross-validation scheme.

Fig. 6 shows the improvement in classification accuracy achieved by using the InDPL algorithm instead of the DPL method. Darker colors on the diagonal reflect clearly the better performance of the InDPL in the conventional and within-subject cross-validation conditions. Visual inspection of the digits in Fig. 1 led to the prediction that distinguishing between numbers 10 and 10^3 would be most challenging. Class-specific accuracy showed that the most difficult pair to decode were 2 and 3, reflecting the sparsity of image data before encoding.

IV. CONCLUSION

We augmented the DPL algorithm by adding an incoherence penalty term. With the resulting InDPL algorithm class-specific dictionaries were achieved. The InDPL cost function was broken to three sub-problems, of which two were solved using closed-form solutions. The third was solved by utilizing the ADMM method. It was applied to a novel database of Chinese handwritten numbers. We developed three cross-validation techniques to verify the efficacy of the proposed incoherent dictionary pair learning methodology. Results showed that in the conventional and within-subject cross-validation conditions, the classification accuracy achieved with the InDPL algorithm exceeds that obtained with the DPL and the kNN methods. When the number of training samples increases, the three classification methods result in comparable scores, reaffirming the hypothesis that dictionary learning-based techniques are suited better to cases in which the number of training data is limited. The availability of this novel dataset allows machine learning and signal processing researchers to develop further pattern recognition algorithms and utilize the proposed algorithm and cross-validation methods for benchmarking.

ACKNOWLEDGMENT

Data supporting this publication is openly available under an "Open Data Commons Open Database License." Additional metadata are available at: http://dx.doi.org/10.17634/137930-3. Please contact Newcastle Research Data Service at rdm@ncl.ac.uk for access instructions.

References

[1] H. Khosravi and E. Kabir, "Introducing a very large dataset of handwritten Farsi digits and a study on their varieties," *Pattern Recog. Lett.*, vol. 28, no. 10, pp. 1133–1141, 2007.

[2] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.

[3] Q. F. Wang, F. Yin, and C. L. Liu, "Handwritten Chinese text recognition by integrating multiple contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 8, pp. 1469–1481, Aug. 2012.

[4] Y. Wang, X. Ding, and C. Liu, "Topic language model adaption for recognition of homologous offline handwritten Chinese text image," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 550–553, May 2014.

[5] D. Keysers, T. Deselaers, H. A. Rowley, L. L. Wang, and V. Carbune, "Multi-language online handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1180–1194, Jun. 2017.

[6] R. Dai, C. Liu, and B. Xiao, "Chinese character recognition: history, status and prospects," *Frontiers Comput. Sci. China*, vol. 1, no. 2, pp. 126–136, 2007.

[7] S. Wang, L. Chen, L. Xu, W. Fan, J. Sun, and S. Naoi, "Deep knowledge training and heterogeneous CNN for handwritten Chinese text recognition," in *Proc. 15th Int. Conf. Frontiers Handwriting Recog.*, 2016, pp. 84–89.

[8] Y. Shao, G. Gao, and C. Wang, "A connection reduced network for similar handwritten Chinese character discrimination," in *Proc. 15th Int. Conf. Frontiers Handwriting Recog.*, 2016, pp. 54–59.

[9] C. Qiang, S. Yu-jun, and X. De-shen, "A novel segmentation method of handwritten chinese number character strings," in *Proc. 8th Control, Autom., Robot. Vis. Conf.*, 2004, vol. 2, pp. 1123–1128.

[10] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[11] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2010, pp. 3501–3508.

[12] W. Li, Y. Zhou, N. Poh, F. Zhou, and Q. Liao, "Feature denoising using joint sparse representation for in-car speech recognition," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 681–684, Jul. 2013.

[13] M. Mao, Z. Zheng, Z. Chen, H. Liu, X. He, and R. Ye, "Group and collaborative dictionary pair learning for face recognition," in *Proc. 23rd Int. Conf. Pattern Recog.*, 2016, pp. 4107–4111.

[14] K. Wang, L. Lin, W. Zuo, S. Gu, and L. Zhang, "Dictionary pair classifier driven convolutional neural networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 2138–2146.

[15] Z. Zhang *et al.*, "Jointly learning structured analysis discriminative dictionary and analysis multiclass classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, doi: 10.1109/TNNLS.2017.2740224. to be published.

[16] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2691–2698.

[17] Q. Li, H. Zhang, J. Guo, B. Bhanu, and L. An, "Reference-based scheme combined with K-SVD for scene image sategorization," *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 67–70, Jan. 2013.

[18] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Red Hook, NY, USA: Curran Associates, Inc., 2009, pp. 1033–1040.

[19] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 543–550.

[20] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[21] B. Mailhé, D. Barchiesi, and M. D. Plumbley, "INK-SVD: Learning incoherent dictionaries for sparse representations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 3573–3576.

[22] H. Tang, X. Zhang, H. Chen, L. Zhu, X. Wang, and X. Li, "Incoherent dictionary learning method based on unit norm tight frame and manifold optimization for sparse representation," *Math. Problems Eng.*, vol. 2016, 2016, Art. no. 9407503.

[23] C. Rusu, "Design of incoherent frames via convex optimization," *IEEE Sig. Process. Lett.*, vol. 20, no. 7, pp. 673–676, Jul. 2013.

[24] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Fast and incoherent dictionary learning algorithms with application to fMRI," *Signal Image Video Process.*, vol. 9, no. 1, pp. 147–158, 2015.

[25] Z. Li, S. Ding, T. Hayashi, and Y. Li, "Incoherent dictionary learning with log-regularizer based on proximal operators," *Dig. Sig. Process.*, vol. 63, pp. 86–99, 2017.

[26] B. Dumitrescu and P. Irofti, "Regularized K-SVD," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 309–313, Mar. 2017.

[27] S. Wilson and C. K. Mohan, "Coherent and noncoherent dictionaries for action recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 698–702, May 2017.

[28] M. Sadeghi and M. Babaie-Zadeh, "Incoherent unit-norm frame design via an alternating minimization penalty method," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 32–36, Jan. 2017.

[29] A. Alameer, G. Ghazaei, P. Degenaar, J. A. Chambers, and K. Nazarpour, "Object recognition with an elastic net-regularized hierarchical MAX model of the visual cortex," *IEEE Signal Proc. Letters*, vol. 23, no. 8, pp. 1062–1066, 2016.

[30] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[31] W. Chen and M. R. D. Rodrigues, "Dictionary learning with optimized projection design for compressive sensing applications," *IEEE Signal Process. Lett.*, vol. 20, no. 10, pp. 992–995, Oct. 2013.

[32] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 186–199.

[33] Y. Suo, M. Dao, U. Srinivas, V. Monga, and T. D. Tran, "Structured dictionary learning for classification," arXiv:1406.1943, 2014. [Online]. Available: http://arxiv.org/abs/1406.1943

[34] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Adv. Neural Inform. Process. Syst.*, 2014, pp. 793–801.

[35] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.