AN EXTENDED PROBABILISTIC COLLABORATIVE REPRESENTATION BASED CLASSIFIER FOR IMAGE CLASSIFICATION

Rushi Lan[1,2] *and Yicong Zhou*[2]

[1]Key Laboratory of Intelligent Processing of Computer Image and Graphics,
Guilin University of Electronic Technology, Guilin, China
[2]Department of Computer and Information Science, University of Macau, Macau, China
{rslan2012@gmail.com, yicongzhou@umac.mo}

ABSTRACT

Collaborative representation based classifier (CRC) and its probabilistic improvement ProCRC have achieved satisfactory performance in many image classification applications. They, however, do not comprehensively take account of the structure characteristics of the training samples. In this paper, we present an *extended probabilistic collaborative representation based classifier (EProCRC)* for image classification. Compared with CRC and ProCRC, the proposed EProCRC further considers a prior information that describes the distribution of each class in the training data. This prior information enlarges the margin between different classes to enhance the discriminative capacity of EProCRC. Experiments on two challenging databases, namely CUB200-2011 and Caltech-256, are conducted to evaluate EProCRC, and comparison results demonstrate that it outperforms several state-of-the-art classifiers.

Index Terms— Image classification, collaborative representation based classifier (CRC), prior information

1. INTRODUCTION

Image classification, a fundamental issue in computer vision and pattern recognition, targets at predicting one (or multiple) class label(s) of a query image. Due to the increasing number of digital images generated in our daily life, image classification becomes a quite challenging prediction task and has a large range of practical applications. Generally speaking, a representative image classification system consists of two basic components, namely feature extraction and feature matching.

Feature extraction is to find a discriminative and robust presentation of the image content. So far, numerous features have been developed, which can be roughly classified into two categories, i.e. hand-crafted features and learning based features. The former features are derived from some prior knowledge or observations of the image. Representative features include LBP, SIFT, HOG, and many others. In contrast, learning based features are extracted using some training strategies, exploiting more data-adaptive information. Therefore, the learned features usually outperform the hand-crafted ones in most sitations.

Based on the obtained features, the classifiers are then used to determine the label of a query image. An intuitive idea is that the query image should share the same label with its closest one in the database, yielding the well-known nearest neighbor (NN) classifier [1]. NN is convenient to implement with a fast speed, but it is sensitive to noise. Nearest centroid (NC) and k-NN are two simple yet robust improvements of NN. More complicated extensions also have been developed, such as nearest feature line (NFL) [2] and nearest feature space (NFS) [3]. Promising results have been achieved by these classifiers.

With the success of compressive sensing, a powerful method named sparse representation based classifier (SRC) has been proposed for face recognition [4]. Later, Zhang *et al.* further developed a collaborative representation based classifier (CRC) [5]. These two classifiers are considered as the most typical representation based classifiers. Considerable efforts have been devoted to improving their performance from various perspectives. Recently, Cai *et al.* investigated CRC from a probabilistic view and further designed a probabilistic collaborative representation based classifier (ProCRC) [6]. Though CRC and ProCRC have gained impressive results in many applications, they do not fully apply the structure information of the training samples in their model. The use of this structure information is able to further improve the performance of CRC and ProCRC.

In this paper, we propose a novel method, named *extended probabilistic collaborative representation based classifier (EProCRC)*, for image classification. EProCRC improves CRC and ProCRC by further coupling a prior information when derive the probability that the query image belongs to a specific class. In this way, EProCRC is more flexible and takes

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grants MYRG2014-00003-FST and MYRG2016-00123-FST. The corresponding author is Yicong Zhou.

more characteristics of the training samples into consideration. Hence, it provide a more accurate representation of the query sample for classification. The main contributions of this work are summarized as follows:

- We propose EProCRC as a new classifier for image classification which further considers a prior information derived from the training samples. The closed form of EProCRC is also given.

- We provide a simple way to obtain the prior information from the training samples for EProCRC. Comparison results show that the proposed EProCRC achieves state-of-the-art performance.

The remainder of this paper is organized as follows: Section 2 briefly review some related work. Section 3 presents the proposed EProCRC in detail, and Section 4 provides several experimental results to evaluate EProCRC. Finally, Section 5 concludes the paper.

2. RELATED WORK

This section briefly introduces some related algorithms as background knowledge. Here we first give some key notations used in this paper. Considering an image classification task with C classes, let $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_C]$ denotes a collection of all training samples. Each column of $\boldsymbol{A}$ represents a sample vector. $\boldsymbol{A}_c$ is the data matrix of the cth class. Suppose $l_{\boldsymbol{A}}$ is the label set of $\boldsymbol{A}$. Spanning all training samples, we can obtain a linear subspace $\mathcal{A}$. Similarly we can achieve the corresponding subspace of each class $\mathcal{A}_c$. Any element $\boldsymbol{a} \in \mathcal{A}$ can be represented by $\boldsymbol{a} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{x}$ is the representation vector. $\boldsymbol{y}$ is the test sample, and our goal is to determine its label $l(\boldsymbol{y})$ from $l_{\boldsymbol{A}}$.

2.1. NN and NC

NN first finds a sample $\boldsymbol{a}'$ from the training set which has the smallest distance with $\boldsymbol{y}$, and considers $l(\boldsymbol{y})$ is the same as $l(\boldsymbol{a}')$. Without any training procedure, NN is sensitive to noise. To improve its performance, NC uses the centroid to represent each class, and then conducts NN to $\boldsymbol{y}$ and the obtained centroids. Compared with NN, NC is more efficient and robust to noise.

2.2. CRC and ProCRC

Zhang *et al.* argued that the collaborative representation of $\boldsymbol{y}$ by $\boldsymbol{A}$ has significant contributions to classification accuracy [5]. Based on this mechanism, they developed a simple yet effective collaborative representation based classifier (CRC). Recently, Cai *et al.* explored the intrinsic mechanism of CRC

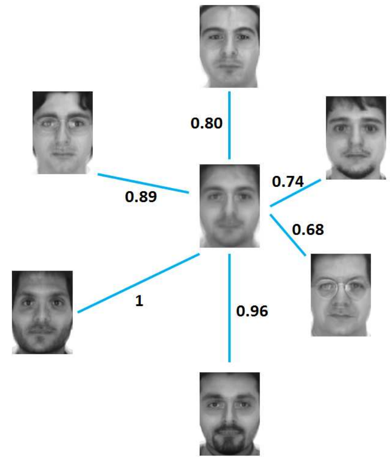

Fig. 1. Illustration of the distribution of some face images. Each face image has a different distance with the centroid of all samples.

from a probabilistic perspective [6]. Considering the probability that $l(\boldsymbol{y})$ belongs to $l_{\boldsymbol{A}}$, they derived the following result:

$$p(l(\boldsymbol{y}) \in l_{\boldsymbol{A}}) \propto \exp(-(\kappa\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \upsilon\|\boldsymbol{x}\|_2^2)), \tag{1}$$

where κ and υ are two constants.

Using the logarithmic operation to maximize $p(l(\boldsymbol{y}) \in l_{\boldsymbol{A}})$, Eq. (1) transforms to:

$$\begin{aligned} \max p(l(\boldsymbol{y}) \in l_{\boldsymbol{A}}) &= \max \ \ln p(l(\boldsymbol{y}) \in l_{\boldsymbol{A}}) \\ &= \min_{\boldsymbol{x}} \kappa\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \upsilon\|\boldsymbol{x}\|_2^2 \\ &= \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2, \end{aligned} \tag{2}$$

where $\lambda = \upsilon/\kappa$. It can be seen that Eq. (2) is the same as the formulation of original CRC. Eqs. (1) and (2) provide a clear probabilistic interpretation of CRC.

Based on Eq. (2), Cai *et al.* further considered $\boldsymbol{y}$'s joint probability $p(l(\boldsymbol{y}) = 1, \cdots, l(\boldsymbol{y}) = C)$ and, after some derivations, finally obtained the following ProCRC model:

$$\begin{aligned} \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \{ &\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2 + \\ &\frac{\gamma}{C}\sum_{c=1}^{C}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}_c\boldsymbol{x}_c\|_2^2 \}, \end{aligned} \tag{3}$$

where γ is a constant. Extensive experimental results have demonstrated that ProCRC achieved satisfactory performance in various pattern classification applications. More information about ProCRC can be found in [6].

3. PROPOSED APPROACH

In this section, we will present the proposed EProCRC algorithm in detail.

3.1. Motivation

Observing Eq. (3), it can be seen that the success of ProCRC is attributed to the third term $\sum_{c=1}^{C}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}_c\boldsymbol{x}_c\|_2^2$, which describes the distance between two points, i.e. $\boldsymbol{A}\boldsymbol{x}\in\mathcal{A}$ and $\boldsymbol{A}_c\boldsymbol{x}_c\in\mathcal{A}_c$. ProCRC attempts to find $\boldsymbol{x}$ to minimize the total distances between $\boldsymbol{A}\boldsymbol{x}$ and $\{\boldsymbol{A}_1\boldsymbol{x}_1,\cdots,\boldsymbol{A}_C\boldsymbol{x}_C\}$. In ProCRC model, each $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}_c\boldsymbol{x}_c\|_2^2$ contributes equally to the total distance. This operation, however, ignores the distribution information of the training data. $\mathcal{A}_c$ is located at different position of $\mathcal{A}$, and the distance between $\boldsymbol{A}\boldsymbol{x}$ and different $\boldsymbol{A}_c\boldsymbol{x}_c$ differs too. To verify this observation, we randomly select 6 subjects from the AR database to form a training set, and then calculate the distances between the centroid of all training samples and the centroids of the samples of each class respectively. The centroid here is the mean face of the training samples. The results are depicted in Fig. 1. It can be see that some subjects are close to the centroid of all training samples, vice versa. That is to say, there is an inherent structure in the subspace $\mathcal{A}$. This structure information is not considered in ProCRC. In this work, we will address this limitation to improve the performance of ProCRC by coupling some prior information of $\mathcal{A}$.

3.2. The EProCRC Model

As previously mentioned, the elements in $\mathcal{A}_c$ and $\mathcal{A}$ can be achieved by $\boldsymbol{a}_c=\boldsymbol{A}_c\boldsymbol{x}_c$ and $\boldsymbol{a}=\sum_{c=1}^{C}\boldsymbol{a}_c=\sum_{c=1}^{C}\boldsymbol{A}_c\boldsymbol{x}_c$. Denote $p(l(\boldsymbol{a})=c|l(\boldsymbol{a})\in l_{\boldsymbol{A}})$ the probability that $\boldsymbol{a}$ shares the same class label as $\boldsymbol{a}_c$, which is defined as:

$$p(l(\boldsymbol{a})=c|l(\boldsymbol{a})\in l_{\boldsymbol{A}})\propto\exp(-\delta\beta_c\|\boldsymbol{a}-\boldsymbol{A}_c\boldsymbol{x}_c\|_2^2),\qquad(4)$$

where δ is a constant and β_c is the prior information that contains the structure characteristic between $\boldsymbol{a}_c$ and $\boldsymbol{a}$. Note that if β_c is set to 1, Eq. (4) is equivalent to the probability used in ProCRC [6].

Compared with ProCRC, the proposed EProCRC further considers the prior information β_c, yielding the following significant advantages:

- EProCRC is more flexible than ProCRC. Actually, β_c also can be considered as a weight for the distance between $\boldsymbol{a}$ and $\boldsymbol{a}_c$. In ProCRC, all distances are assigned with an equal weight. In contrast, extending the fixed weight to different values, EProCRC improves the flexibility of the original ProCRC.
- EProCRC takes account of more information of the training samples. As aforementioned, β_c is prior information to explore the structure features of the training data and to reflect the relations between each subspace $\mathcal{A}_c$ and the whole space $\mathcal{A}$. With β_c, we are able to find more accurate coefficients to represent $\boldsymbol{y}$.

Owing to these merits, we further study the collaborative representation of $\boldsymbol{y}$ from the probabilistic view using the improved probability in Eq. (4).

Based on Eq. (4), similarly to ProCRC, we consider the probability that $\boldsymbol{y}$ belongs to the cth class, and obtain the following result:

$$\begin{aligned}&p(l(\boldsymbol{y})=c)=p(l(\boldsymbol{y})\in l_{\boldsymbol{A}})\cdot p(l(\boldsymbol{a})=c|l(\boldsymbol{a})\in l_{\boldsymbol{A}})\\&\propto\exp(-(\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_2^2+\gamma\beta_c\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}_k\boldsymbol{x}_k\|_2^2)),\end{aligned}\qquad(5)$$

where $\gamma=\delta/\kappa$. We can directly maximize $p(l(\boldsymbol{y})=c)$ individually for each class to determine $l(\boldsymbol{y})$. However, this operation will result in the same problem as ProCRC, namely the classification results is unstable and less discriminative.

To address this problem, we also apply the joint probability $p(l(\boldsymbol{y})=1,\cdots,l(\boldsymbol{y})=C)$ rather than $p(l(\boldsymbol{y})=c)$ individually. Based on this point, the following result can be achieved:

$$\begin{aligned}\max p(l(\boldsymbol{y})=1,\cdots,l(\boldsymbol{y})=C)&=\max\prod_c p(l(\boldsymbol{y})=c)\\&\propto\max\exp(-(\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_2^2+\\&\qquad\frac{\gamma}{C}\sum_{c=1}^{C}\beta_c\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}_c\boldsymbol{x}_c\|_2^2)).\end{aligned}\qquad(6)$$

Applying the logarithmic operation to Eq. (6) and ignoring the constant yield the following result:

$$\begin{aligned}\hat{\boldsymbol{x}}=\arg\min_{\boldsymbol{x}}\ \{&\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_2^2+\\&\frac{\gamma}{C}\sum_{c=1}^{C}\beta_c\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}_c\boldsymbol{x}_c\|_2^2.\}\end{aligned}\qquad(7)$$

The above model is called *extended probabilistic collaborative representation based classifier (EProCRC)* model, which extends the ProCRC by coupling the prior information β_c into the objective function.

For the training data matrix $\boldsymbol{A}$, we set its following elements, $\{\boldsymbol{A}_1,\cdots,\boldsymbol{A}_{c-1},\boldsymbol{A}_{c+1},\cdots,\boldsymbol{A}_C\}$, to $\boldsymbol{0}$ matrix, and keep $\boldsymbol{A}_c$ unchanged. Denote the obtained matrix by $\boldsymbol{A}'_c$ and $\boldsymbol{A}'_c=[\boldsymbol{0},\cdots,\boldsymbol{A}_c,\cdots,\boldsymbol{0}]$. Let $\bar{\boldsymbol{A}}'_c=\boldsymbol{A}-\boldsymbol{A}'_c$. It is convenient to obtain the closed form solution of EProCRC model in Eq. (7). To this end, we first calculate the following matrix:

$$\boldsymbol{M}=(\boldsymbol{A}^T\boldsymbol{A}+\frac{\gamma}{C}\sum_{c=1}^{C}\beta_c(\bar{\boldsymbol{A}}'_c)^T\bar{\boldsymbol{A}}'_c+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T,\qquad(8)$$

where $\boldsymbol{I}$ is the identity matrix. The above projection matrix $\boldsymbol{M}$ can be obtained offline. Given an testing sample $\boldsymbol{y}$, once $\boldsymbol{M}$ is available, the solution of EProCRC in Eq. (7) can be achieved as follows:

$$\hat{\boldsymbol{x}}=\boldsymbol{M}\boldsymbol{y}.\qquad(9)$$

After obtaining $\hat{\boldsymbol{x}}$, we can determine the label of $\boldsymbol{y}$ now. According to Eq. (5), the probability that $\boldsymbol{y}$ is classified to the cth class is:

$$\begin{aligned}p(l(\boldsymbol{y})=c)\propto\exp(-(&\|\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}\|_2^2+\lambda\|\hat{x}\|_2^2+\\&\gamma\beta_c\|\boldsymbol{A}\hat{\boldsymbol{x}}-\boldsymbol{A}_k\hat{\boldsymbol{x}}_k\|_2^2)).\end{aligned}\qquad(10)$$

In this situation, $l(\boldsymbol{y})$ is assigned by:

$$l(\boldsymbol{y}) = \arg\max_c \{p(l(\boldsymbol{y}) = c)\}. \quad (11)$$

However, the experimental results show that the $l(\boldsymbol{y})$, determined by Eqs. (10) and (11), cannot obtain satisfactory performance. The reason lies in that the used of β_c in Eq. (10) will make $\boldsymbol{y}$ overfitting to the training data $\boldsymbol{A}$. As shown in Eq. (8), the class information of the training samples β_c has been considered to obatin the representation coefficients of $\boldsymbol{y}$, and the derived $\hat{\boldsymbol{x}}$ also already includes this prior information. If we further make use of β_c to compute $p(l(\boldsymbol{y}) = c)$ again, the prior information will impose the probability to fit the distribution of the training data, resulting in the the overfitting problem.

To avoid this problem, we remove the prior information β_c to calculate $p(l(\boldsymbol{y}) = c)$. For simplity, the constant terms in Eq. (10) are also omitted, yielding the following probability that $\boldsymbol{y}$ belongs to the cth class:

$$p_c = \exp(-\|\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{A}_c\hat{\boldsymbol{x}}_c\|_2^2), \quad (12)$$

Then the label of $\boldsymbol{y}$ will be finally determined by:

$$l(\boldsymbol{y}) = \arg\max_c \{p_c\}. \quad (13)$$

The detailed implementation of EProCRC approach is summarized as Algorithm 1.

Algorithm 1 EProCRC

Input: A set of training samples $\boldsymbol{A}$, a test sample $\boldsymbol{y}$, the prior information of each class $\{\beta_1, \cdots, \beta_C\}$, and the parameter λ and γ.
Output: The label of $\boldsymbol{y}$.

(a) Calculate the projection matrix:

$$\boldsymbol{M} = (\boldsymbol{A}^T\boldsymbol{A} + \frac{\gamma}{C}\sum_{c=1}^{C}\beta_c(\bar{\boldsymbol{A}}'_c)^T\bar{\boldsymbol{A}}'_c + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T.$$

(b) Obtain the representation of $\boldsymbol{y}$: $\hat{\boldsymbol{x}} = \boldsymbol{M}\boldsymbol{y}$.

(c) Calculate the probability that $\boldsymbol{y}$ belongs to each class:

$$p_c = \exp(-\|\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{A}_c\hat{\boldsymbol{x}}_c\|_2^2).$$

(d) Predict the label: $l(\boldsymbol{y}) = \arg\max_c\{p_c\}$.

3.3. Strategy to Derive Prior Information

As aforementioned, the prior information β_c plays a key role in EProCRC model. A suitable β_c will benefit the performance of EProCRC. Otherwise, poor β_c may adversely affect the classification accuracy. In this work, we provide the following strategy to obtain the prior information for the proposed EProCRC.

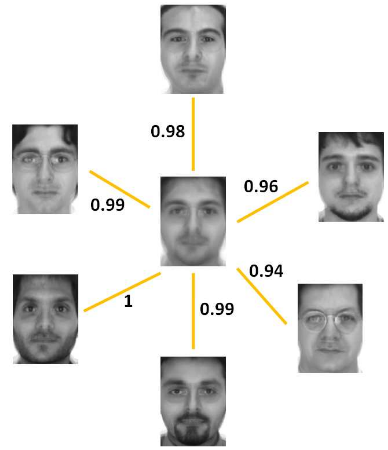

Fig. 2. Illustration of the distribution of the face images with a prior information. All face images now have close distances with the centroid of all samples.

As previously mentioned, the optimal $\hat{\boldsymbol{x}}$ is to minimize the distance between $\boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{A}_c\boldsymbol{x}_c$. As shown in Fig. 1, if one mean face of a subject is too close to the mean face of all samples, it is difficult to find a accurate $\hat{\boldsymbol{x}}$ in this situation. To avoid this problem, we impose the samples, which are close to the centroid of all training samples, to remove away, enlarging their margin. To this end, we introduce the following prior information for the proposed EProCRC model.

Let $\bar{\boldsymbol{a}}_{tr}$ be the mean vector of all training sample in $\boldsymbol{A}$, while $\bar{\boldsymbol{a}}_{tr}^c$ is the mean vector of the cth training samples $\boldsymbol{A}_c$. The distance between $\bar{\boldsymbol{a}}_{tr}$ and $\bar{\boldsymbol{a}}_{tr}^c$ can be represented as:

$$d_c = \|\bar{\boldsymbol{a}}_{tr} - \bar{\boldsymbol{a}}_{tr}^c\|_2^2. \quad (14)$$

Then the prior information β_c for $\boldsymbol{A}$ and $\boldsymbol{A}_c$ is defined by

$$\beta_c = \exp(-d_c) = \exp(-\|\bar{\boldsymbol{a}}_{tr} - \bar{\boldsymbol{a}}_{tr}^c\|_2^2). \quad (15)$$

That is to say, two close centroids are given a larger weight. For the face images in Fig. 1, we calculate the distance again, multiplying the original distances by the corresponding prior information in Eq. (15). The obtained results are plotted in Fig. 2. It can be seen that these face images now have a more uniform distribution such that they are more separable to give a better representation of $\boldsymbol{y}$.

4. EXPERIMENTS

In this section, we will provide several experimental results to assess the performance of the proposed EProCRC from different perspectives. The databases and used features are first presented. After that, the comparison results on two challenging databases are reported. Finally, a discussion about the running time is given.

Table 1. Classification performance (%) of different classifiers on CUB200-2011 database.

Classifier	CUB200-2011 Database
NN	50.1
NC	60.3
Softmax	72.1
SVM	75.4
Kernel SVM	76.6
NSC	74.5
CRC	76.2
SRC	76.0
CROC	76.2
ProCRC	78.3
PN-CNN	75.7
FV-CNN	66.7
POOF	56.9
EProCRC	**79.3**

Table 2. Classification performance (%) of different classifiers on Caltech-256 database with 30 training samples.

Classifier	Caltech-256 Database
NN	65.2
NC	71.9
Softmax	75.3
SVM	80.1
Kernel SVM	81.3
NSC	80.2
CRC	81.1
SRC	81.3
CROC	81.7
ProCRC	83.3
ZF	70.6
M-HMP	50.7
LLC	41.2
ScSPM	34.0
EProCRC	**84.3**

4.1. Databases and Used Features

To comprehensively evaluate the proposed method, the following two challenging image classification databases are used in this work: *Caltech-UCSD Birds (CUB200-2011)* [7] and *Caltech-256* [8] databases. This first database, consisting of totally 11788 bird images of 200 categories, is popularly used for fine-grained image recognition. It is a challenging database because those images from different categories are quite similar to each other. There is a total number of 30608 images of 256 object categories in the second database, and each category contains at least 80 images. It is commonly applied to evaluate several large-scale image classification algorithms. For these two databases, we also consider the features used in [6] for fair comparison, namely the VGG-verydeep-15 [9] is utilized to extract CNN features. For each image, a feature vector with a size of 4096×1 is obtained for classification.

4.2. Results on the CUB200-2011 Database

We follow the training and testing sets given in the CUB200-2011 database for evaluation. In the training set, each bird category includes about 30 training samples. The following competing algorithms are chosen for comparison: NN, NC, Softmax, SVM, Kernel SVM, NSC [10], CRC, SRC, CROC [11], ProCRC, PN-CNN [12], FV-CNN [13], and POOF [14]. The classification performance of all these classifiers are reported in Table 1.

Observing Table 1, it can be seen that NN achieves the worst performance here because it is a simple and native classifier without a training procedure. Using the centroid to represent each class, NC outperforms NN by about 10 percentages. Some representation-based classifiers, such as CRC, SRC, and CROC, obtain comparable accuracies for this database, and their results are all about 76%. The performance of ProCRC and EProCRC, both derived from a probabilistic perspective, is superior to those of aforementioned representation-based algorithms. The proposed EProCRC, which is coupled with some prior information of the training samples, gains a improvement of 1 percentage compared with ProCRC.

Apart from the representation-based classifiers, EProCRC is also compared with three other state-of-the-art methods, including POOF, FV-CNN, and PN-CNN. These schemes are based on a specially developed CNN architecture for bird specie recognition. The comparable results demonstrate that EProCRC improves the second best method PN-CNN by about 4 percentages.

4.3. Results on the Caltech-256 Database

For this database, we apply the standard experimental setting used in the technical literature, namely N images are randomly chosen from each object to form the training set and the remaining images are considered as testing set. This procedure will be repeated 10 times, obtaining the average classification accuracy. Apart from the representation-based classifiers, the following algorithms are selected here too, including ZF [15], M-HMP [16], LLC [17], and ScSPM [18]. In the following experiments, the number of training images N is set to 15, 30, and 45 respectively.

Table 2 reports the performance of all competing schemes when $N = 30$. It can be seen that with deep learning based features, NN and NC obtain much better performance than some traditional algorithms, namely M-HMP, LLC, and ScSPM. SVM and NSC get quite close classification results, outperforming NC about 5 percentages. The performance of

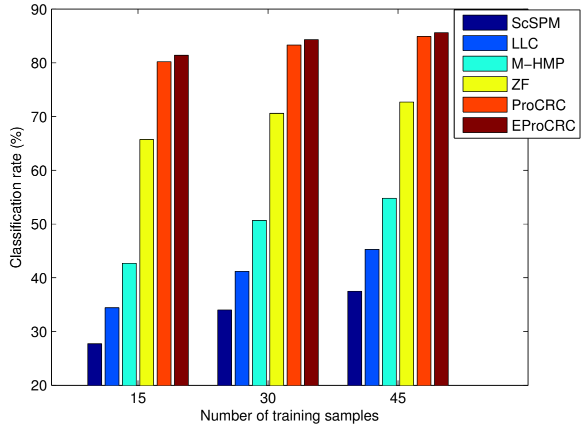

Fig. 3. Classification rates of different classifiers on Caltech-256 database using 15, 30, and 45 training samples.

Kernel SVM, CRC, SRC, and CROC is comparable, which is about 2 percentages lower than that of ProCRC. The proposed EProCRC achieves the highest classification accuracy in this situation.

Fig. 3 further depicts the results of some selected methods when N is set to 15, 30, and 45 respectively. We can find that EProCRC and PRoCRC both significantly surpass those competing schemes by different degrees. Compared with ProCRC, the proposed EProCRC leads an improvement of $\{1.2\%, 1\%, 0.7\%\}$ in corresponding N, which demonstrates the effectiveness of the used prior information.

4.4. Discussion

In this part, we discuss the time complexity of the proposed method. In [6], the authors pointed out that ProCRC and CRC take the same runing time, and their speeds are faster than those of SRC and CROC. For EProCRC, it also has an analytical solution and further considers the prior information in contrast with ProCRC. In this work, the prior information is derived from the distances between the centroids of each training class and that of all training samples, which can be efficiently obtained. Therefore, we can conclude that the proposed methods is of a slightly larger time complexity than that of ProCRC and CRC, but also is faster that SRC and CROC.

5. CONCLUSION

In this paper, we developed the EProCRC as a novel classifier for image classification. As an variant of ProCRC, the proposed classifier further takes account of a prior information to obtain a more accurate representation of the query sample for classification. We also provided a simple way to derive a reasonable prior information for EProCRC. Experimental results on two benchmark databases, namely CUB200-2011 and Caltech-256 databases, demonstrated the advantages of the proposed EProCRC over other existing classifiers.

6. REFERENCES

[1] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. on IT*, vol. 13, no. 1, pp. 21–27, 1967.

[2] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. on NN*, vol. 10, no. 2, pp. 439–443, 1999.

[3] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. on PAMI*, vol. 32, no. 11, pp. 2106–2112, 2010.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on PAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[5] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *ICCV*, 2011.

[6] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *CVPR*, 2016.

[7] C. Wah, S. Branson, P. Perona, and S. Belongie, "The caltech-ucsd birds-200-2011 dataset," 2011.

[8] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2014.

[10] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Trans. on PAMI*, vol. 24, no. 12, pp. 1644–1649, 2002.

[11] Y. Chi and F. Porikli, "Classification and boosting with multiple collaborative representations," *IEEE Trans. on PAMI*, vol. 36, no. 8, pp. 1519–1531, 2014.

[12] S. Branson, G. Van Horn, S. Belongie, and P. Perona, "Bird species categorization using pose normalized deep convolutional nets," *BMVC*, 2014.

[13] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *CVPR*, 2015.

[14] T. Berg and P. Belhumeur, "Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation," in *CVPR*, 2013.

[15] M. D Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *ECCV*, 2014.

[16] L. Bo, X. Ren, and D. Fox, "Multipath sparse coding using hierarchical matching pursuit," in *CVPR*, 2013.

[17] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *CVPR*, 2010.

[18] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *CVPR*, 2009.