SPARSE REPRESENTATION-BASED OPEN SET RECOGNITION

BY HE ZHANG

A thesis submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Prof. Vishal M Patel

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2016

ABSTRACT OF THE THESIS

Sparse Representation-based Open Set Recognition

by He Zhang

Thesis Director: Prof. Vishal M Patel

In this thesis, we study an open set recognition algorithm that is based on the Sparse Representation-based Classification (SRC) method. By modeling the tail distributions of the matched and non-matched reconstruction errors using the statistical Extreme Value Theory (EVT), we simplify the open set recognition problem into a set of hypothesis testing problems. The confidence scores corresponding to the tail distributions of a novel test sample are then fused to determine its identity. The effectiveness of the proposed method is demonstrated using three publicly available image and object classification datasets and it is shown that this method can perform significantly better than many competitive open set recognition algorithms.

Acknowledgements

The whole academic life is somehow like training a model, which adjusts its 'parameter' everyday to fit in the new sample and increase the 'accuracy '. I am lucky that I achieve the 'accuracy' that can make it to the defense with the help of so many talented people. Firstly, I'd like thank my advisor Prof. Vishal M Patel for guiding me and supporting me through my second year of my master degree. He is more like a brother that give me a lot of help both on research and life. I am looking forward to pursuing Phd degree under his advise.

Secondly, I'd like to thank two professors Prof. Anand D. Sarwate and Prof. Predrag Spasojevic who gave me much help on my first year of study. Also, I sincerely thank Shaogang Wang and Shunqiao Sun, who lead me to the gate of research. Many thanks to my lab mates Mahdi Abavisani, Pramuditha Perera and Xing Di. Also, a special thanks to my undergraduate thesis advisor Prof. Le Yang, who greatly influenced my academic career and leads a power-electronic guy into the field of signal processing and computer vision. Meanwhile, great thanks to my best friend Li Liu, who give me a lot of support in life.

Finally, thanks to my parents' love who give me a chance to pursue my study in the US.

Life is still going on. Fighting!

Dedication

Dedicaed to my parents

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

In recent years, sparse representation-based techniques have drawn much interest in computer vision and image processing fields [28], [21]. A number of image classification and restoration algorithms have been proposed based on sparse representations. In particular, sparse representation-based classification (SRC) algorithm [29] has gained a lot of traction. The basic idea of SRC is to identify the correct class by seeking the sparest representation of the test sample in terms of the training. The SRC algorithm was originally proposed for face recognition and later extended for iris recognition and automatic target recognition in [17] and [14], respectively. A simultaneous dimension reduction and classification framework based on SRC was proposed in [30]. Furthermore, non-linear kernel extensions of the SRC method have also been proposed in [31], [13], [6], [24].

The SRC algorithm and its variants are essentially based on the *closed world assumption*. In other words, it is assumed that the testing data pertains to one of K classes that are used during training. But in practice, testing data may come from a class that is not necessarily seen in training. This problem where the testing data corresponds to a class that is not seen during training is known as *open set recognition* [22]. Consider the problem of animal classification. If the training samples correspond to K different animals, then given a test image corresponding to an animal from one of the K classes, the algorithm should be able to determine its identity. However, if the test image corresponds to an animal which does not match one of the K animals seen during training, then the algorithm should have the capability to ignore or reject the test sample [26].

The goal of an open set recognition algorithm is to learn a predictive model that classifies the known data into correct class and rejects the data from open class. As a results, one can view open set recognition as tackling both classification and novelty detection problem at the same time. Novelty detection refers to the problem of finding anomalous behaviors that are inconsistent with the expected pattern. A novelty detection problem can be formulated as a hypothesis testing problem where the null hypothesis, $\mathcal{H}_0$, implies the test sample coming from normal class and the alternative hypothesis, $\mathcal{H}_1$, indicates the presence of anomalies and the objective is to find the best threshold that separates $\mathcal{H}_0$ from $\mathcal{H}_1$.

A number of approaches have been proposed in the literature for open set recognition. For instance, [22] introduced a concept of *open space risk* and developed a 1-vs-Set Machine formulation using linear SVMs for open set recognition. In [23] the concept of Compact Abating Probability (CAP) was introduced for open set recognition. In particular, Weibull-calibrated SVM (W-SVM) algorithm was developed which essentially combines the statistical Extreme Value Theory (EVT) with binary SVMs for open set recognition. Also, the W-SVM framework was recently used in [19] for fingerprint spoof detection. In [3], an open set recognition-based method was developed to identify whether or not an image was captured by a specific digital camera.

In order to reject invalid samples, the notion of Sparsity Concentration Index (SCI) was proposed in [29]. Similarly, a rejection rule based on the ratio of the first two highest projection scores was developed for rejecting non-face images in [15]. The rejection rules defined using sparse representations in [29] and [15] were specifically designed to reject non-face images. As will be shown later, these rules do not work well on general open set recognition problems.

In this thesis, we extend the SRC formulation for open set recognition. Our method relies on the statistical EVT [16] and consists of two main stages. In the first stage, the tail distributions of the matched reconstruction errors and the sum of non-matched reconstruction errors are modeled using the EVT to simplify the open set recognition problem into two hypothesis testing problems. In the second stage, the reconstruction errors corresponding to a test sample from each class are calculated and the confidence

scores based on the two tail distributions are fused to determine the identity of the test sample. Figure 1.1 gives an overview of the proposed Sparse Representation-based Open Set Recognition (SROSR) algorithm.

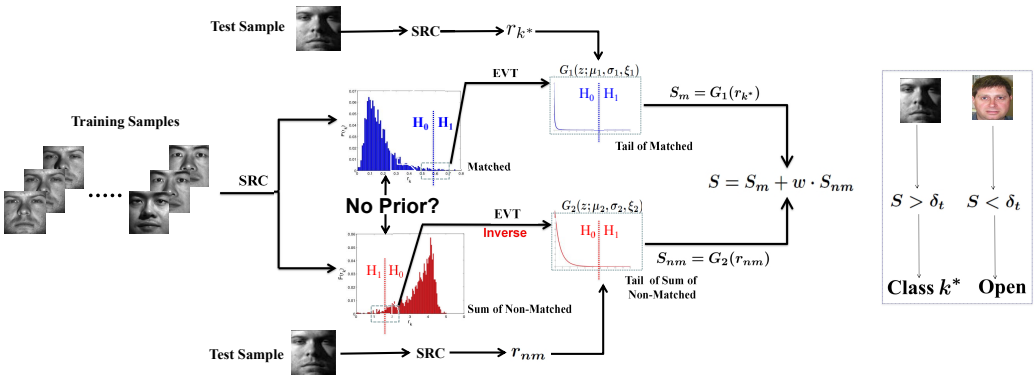


Figure 1.1: Overview of the proposed SROSR algorithm. Given training samples, we model the matched reconstruction error distribution and the sum of non-reconstruction error distribution using the statistical EVT. Given a novel test sample, the modeled distributions and the matched and the sum of non-matched reconstruction errors are used to calculate the confidence scores. Then, these zscores are fused to obtain the final score for recognition.

1.2 Outline

This thesis is organized as follows. In Chapter 2, we give a brief background on the EVT and the SRC algorithm. Details of the proposed SROSR algorithm are given in Chapter 3. Experimental results are presented in Chapter 4 and Section 5 concludes the paper with a brief summary and discussion.

Chapter 2

Background

In this chapter, we review some related work in SRC and EVT.

2.1 Sparse Representation-based Classification

Stack the training samples from the i-th class as columns of a large matrix $\mathbf{Y}_i \in \mathbb{R}^{M \times N_i}$, and write

$$\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_K] \in \mathbb{R}^{M \times N},$$

as the dictionary of training samples from K classes, where $N = \sum_i N_i$ is the total number of training samples and M is the dimension of each training sample. Let $\mathcal{L}^Y$ denote the corresponding label set. If the $\mathbf{Y}_i$ are sufficiently expressive [32], a new input sample from the i-th class, stacked as a vector $\mathbf{y}_t \in \mathbb{R}^M$ will have a sparse representation

$$\mathbf{y}_t = \mathbf{Y}\mathbf{x}$$

in terms of the training data $\mathbf{Y}$: $\mathbf{x}$ will be nonzero only for those samples from class i. The sparse coefficient vector $\mathbf{x} \in \mathbb{R}^N$ can be estimated by solving the following optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{ s.t. } \ \|\mathbf{y}_t - \mathbf{Y}\mathbf{x}\|_2 < \epsilon, \tag{2.1}$$

where we have assumed that the observations are noisy with noise energy ϵ and $\|\mathbf{x}\|_1 = \sum_i |x_i|$. The sparse code $\hat{\mathbf{x}}$ can then be used to determine the class of $\mathbf{y}_t$ based on the class residuals

$$r_k = \|\mathbf{y}_t - \mathbf{Y}_k \hat{\mathbf{x}}_k\|_2, \quad k = 1, \cdots, K, \tag{2.2}$$

where $\hat{\mathbf{x}}_k$ is the part of $\hat{\mathbf{x}}$ that corresponds to class k. Finally, the class k^* that is associated to the test sample $\mathbf{y}_t$, can be declared as the one that produces the smallest

approximation error

$$k^* = \text{class of } \mathbf{y}_t = \arg\min_k r_k.$$

Figure 2.1 gives an overview of the SRC algorithm. This method provides excellent performance on several image classification datasets [29], [17], and is provably robust to errors and occlusion [27]. The basic SRC algorithm is summarized in Algorithm 1.

Algorithm 1 Sparse Representation-based Classification

Input: $\mathbf{Y}$, $\mathcal{L}^Y$, ϵ, $\mathbf{y}_t$

$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \ \text{ s.t. } \ \|\mathbf{y}_t - \mathbf{Y}\mathbf{x}\|_2 < \epsilon$

$r_k = \|\mathbf{y}_t - \mathbf{Y}_k\hat{\mathbf{x}}_k\|_2 \text{ for } k = 1, \cdots, K$

$k^* = \arg\min_k r_k$

Output: $k^*, \mathbf{r} = [r_1, r_2, \cdots, r_K]$

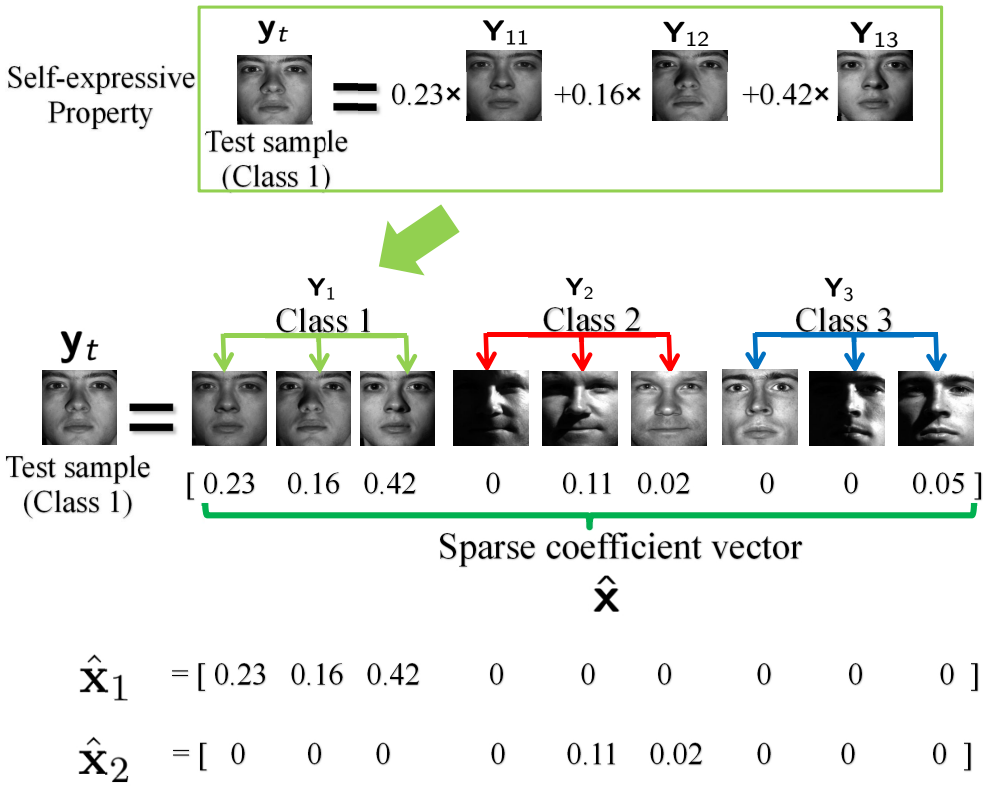


Figure 2.1: An overview of the SRC algorithm.

In order to reject outliers, the following SCI rule was defined in [29]

$$\text{SCI}(\mathbf{x}) = \frac{\frac{K \times \max_k \|\mathbf{x}_k\|_1}{\|\mathbf{x}\|_1} - 1}{K - 1} \in [0, 1]. \tag{2.3}$$

Sparsity coefficient index takes values between 0 and 1. The SCI values close to 1

correspond to the case where the test image can be approximately represented by using only images from a single class. If the SCI value of the recovered coefficient is close to zero, then the coefficients are spread across all classes. Hence, the test vector is not similar to any of the classes and can be rejected. A threshold can be chosen to reject invalid test samples if $\text{SCI}(\hat{\mathbf{x}}) < \alpha$ and otherwise accepted as valid, where α is some chosen threshold between 0 and 1.

2.2 Extreme Value Theory

Extreme value theory is a branch of statistics analyzing the distribution of data of abnormally high or low values. It has been applied in Finance[11], Hydrology[25] and novelty detection problems [20], [2], [8]. In this section, we give a brief overview of the statistical EVT.

Assume that we are given n i.i.d samples $\{Z_1, Z_2, ..., Z_n\}$ drawn from an unknown distribution $F(z)$. Denote

$$Z_m = \max_i Z_i \quad i \in [1, n].$$

The Fisher-Tippett-Gnedenko theorem[5] states that if there exists a pair of parameters (a_n, b_n), subject to the condition $a_n > 0$ and $b_n \in \mathbb{R}$, then

$$\lim_{n\to\infty} P\left(\frac{Z_m - b_n}{a_n}\right) = E(z), \tag{2.4}$$

where $E(z)$ is a non-degenerate distribution that belongs to either Fréchet, Weibull or Gumbel distribution. These distributions can be represented as a Generalized Extreme Value distribution (GEV) as follows

$$E(z; \mu, \sigma, \xi) = \exp^{-p(z)}, \tag{2.5}$$

where

$$p(z) = \left(1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}$$

and μ, σ and ξ are the location, scaling and shape parameters, respectively.

There are two challenges that one has to be overcame before using the GEV distribution to model the tail distribution of data. Firstly, we have to choose which distribution

to use among the three based on prior knowledge. Secondly, we need to segment the data into several parts and model the maximum in each part as a distribution using GEV. However, in practice we have no prior knowledge on what is the best way to separate the data and what is the best distribution to choose. To overcome these challenges, an alternative method based on Generalized Pareto distribution (GPD), denoted as $G(z)$, to estimate the tail distribution of data samples was proposed in [16]. It was shown that given a sufficiently large threshold u, the probability of an observation exceeding u by z conditioned on u can be approximated by

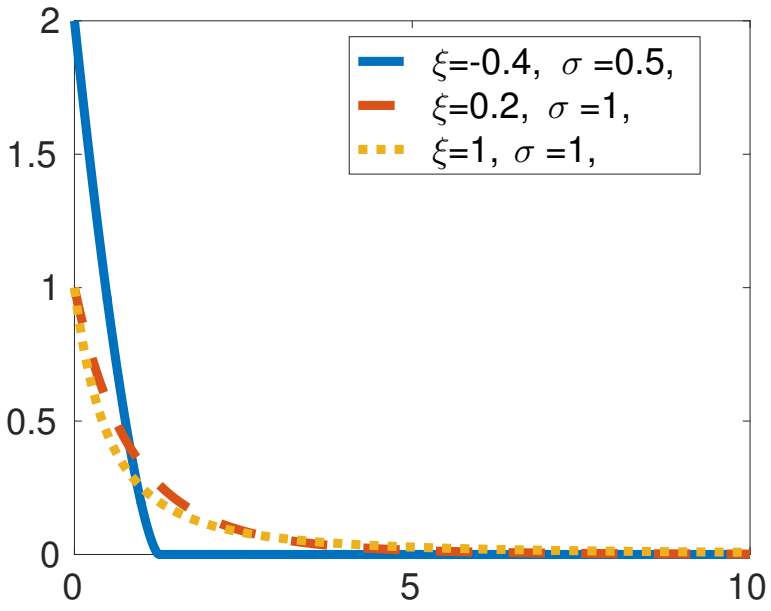
$$\lim_{n\to\infty} P(Z > z + u | Z > u) = 1 - G(z), \tag{2.6}$$

with

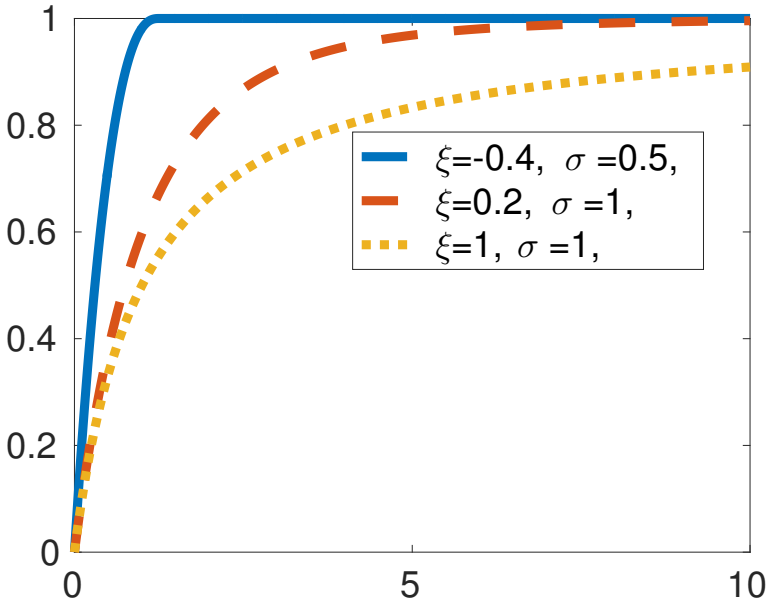
$$G(z) = 1 - \left(1 + \xi\frac{z}{\sigma}\right)_+^{\frac{-1}{\xi}}, \quad z > 0,$$

where $\sigma > 0$, $\xi \in \mathbb{R}$ and $x_+ = \max(x, 0)$. Figure 2.2, shows some GPDs with different parameters.

To estimate the parameters of GPD, one can use the maximum likelihood estimation (MLE) method introduced in [9]. Even though there is the possibility that the parameters of GPD don't exist and that maximum likelihood estimation may not converge when $\xi > 1/2$, it has been shown that these are extremely rare cases in practice [9] [1] .



(a)



(b)

Figure 2.2: Sample PDFs and CDFs of GPD based on different parameters are shown in (a) and (b), respectively.

Chapter 3

SROSR: Sparse Representation-based Open-set Recognition

In [22] the notion of "Open set Risk" was defined as the cost of labeling the open set sample as known sample. Based on this, one can minimize the following cost to develop an open set recognition algorithm

$$\arg\min_{f} C_o(f) + \lambda_r C_\epsilon(f), \tag{3.1}$$

where f is a measurable function, $C_o(f)$ denotes open-set risk, $C_\epsilon(f)$ denotes empirical risk for classification and λ_r is a parameter that balances open-set risk and empirical risk.

The SRC algorithm uses residuals (2.2) for classification which can be used to model f in (3.1) for open set recognition. This is due to the following reason. If the test sample corresponds to class k, then the reconstruction error corresponding to class k should be much lower than that of corresponding to the other classes. As a result, there may be a distinction between matched and non-matched reconstruction errors. To illustrate this, we plot the distributions of matched and non-matched reconstruction errors using the samples from the MNIST handwritten digits dataset [12] in Figure 3.1. Training samples consists of digits 0 to 9 and test samples correspond to digit 9. Matched reconstruction errors here mean that the errors correspond to the sparse coefficients of digit 9 and non-matched reconstruction errors mean that the errors are generated by the sparse coefficients of all other digits. One can see from this figure that matched classes' reconstruction errors follow some underlying distribution. If one can fit a probability model $P(r_k)$ to describe the distribution of the reconstruction errors of the matched class, then one can reformulate the open-set recognition problem as a hypothesis testing

for novelty detection problem as

$$\mathcal{H}_0 : P(r_k) \leq \delta$$
$$\mathcal{H}_1 : P(r_k) > \delta, \tag{3.2}$$

where the null hypothesis $\mathcal{H}_0$ implies that the test data are generated from the distribution $P(r_k)$, and the alternative hypothesis $\mathcal{H}_1$ implies that test data correspond to the classes other than the ones considered in training and $\delta \in [0, 1]$ is the threshold for rejection.

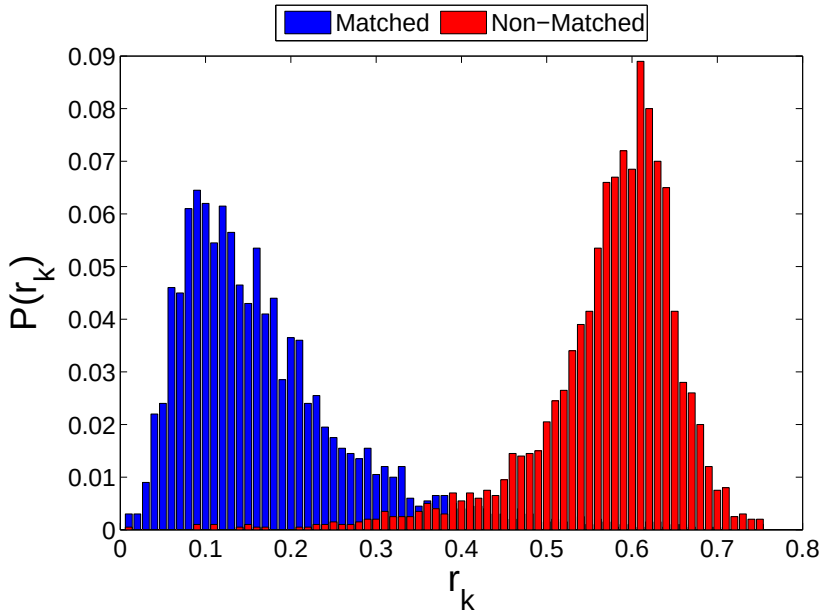


Figure 3.1: Histogram of the matched and non-matched reconstruction errors. Matched reconstruction errors are the errors corresponding to the sparse coefficients of digit 9 and non-matched reconstruction errors are the errors that are generated by the sparse coefficients of all other digits when training samples consists of digits 0 to 9 and the test samples correspond to digit 9. All samples are from the MNIST dataset.

However, as we have no prior knowledge on the underlying distribution of the matched reconstruction errors, we cannot fit a proper distribution on them. Instead, we can apply the EVT on the tail of the matched distribution as we are only concerned about the right tail of this distribution for hypothesis testing. As the implementation of GEV on real data is difficult, we instead use the GPD to model the tail of the matched distribution. Once we learn the distribution of the tail, we can modify the hypothesis

testing problem (3.2) to the following

$$\mathcal{H}_0 : G(r_k) \leq \delta_g$$
$$\mathcal{H}_1 : G(r_k) > \delta_g, \tag{3.3}$$

where $G(r_k)$ is the learned GPD distribution for fitting the right tail of r_k and δ_g is the rejection threshold.

When SRC is used for classification, we cannot only get the information of the matched reconstruction errors but we can also have access to the non-matched reconstruction errors. Suppose that the training data only contains digits 0 to 5 and the test samples consist of closed set digits 0 to 5 and open set digits 6 to 9. As one can see from Figure 3.2, the sum of the non-matched reconstruction errors from the closed set digits 0 to 5 also follow a certain distribution that is very different from the distribution that one obtains from the errors corresponding to the open set digits.

As a result, we can formulate another hypothesis testing problem similar to (3.2) for the sum of non-matched reconstruction errors. We can combine the two hypothesis testing problems together to make the open set recognition algorithm more accurate. As we are only interested in the right tail of the matched distribution and the left tail of the sum of non-matched distribution, we apply an inverse procedure on the random variable Z as

$$Z_I = -Z.$$

So the right tail of Z_I is the left tail of Z.

3.0.1 Training

In the training phase, we have to estimate the parameters for fitting the tail distribution based on the GPD. Estimating the parameters based on MLE requires the availability of multiple reconstruction errors. To deal with this issue, we propose the following iterative procedure. For each iteration, we first randomly order the training samples from each class $\mathbf{Y}_i$ and then partition them into cross-test $\mathbf{Y}_i^{te}$ and cross-train $\mathbf{Y}_i^{tr}$ sets. The cross-test and cross-train sets contain 20 and 80 percent of the training samples in $\mathbf{Y}_i$, respectively. Let $\mathcal{L}_i^{tr}$ and $\mathcal{L}_i^{te}$ denote the associated label sets corresponding to $\mathbf{Y}_i^{tr}$

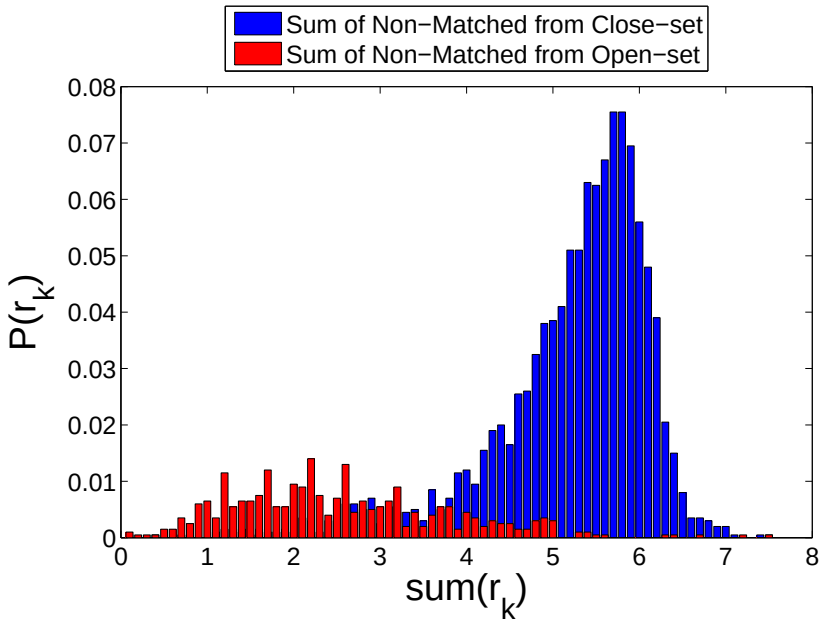


Figure 3.2: Histogram of the sum non-matched reconstruction errors corresponding to the closed set classes 0 to 5 and the sum of non-matched reconstruction errors corresponding to the open set digits 6 to 9 . All samples are from MNIST dataset.

and $\mathbf{Y}_i^{te}$, respectively. Once the training samples from all classes are partitioned into cross-train and cross-test sets, combine the cross-train samples from all K classes into a cross-train matrix $\mathbf{Y}^{tr} = [\mathbf{Y}_1^{tr}, \mathbf{Y}_2^{tr}, \cdots, \mathbf{Y}_K^{tr}]$ and their associated labels into a label set $\mathcal{L}^{tr} = \{\mathcal{L}_1^{tr}, \mathcal{L}_2^{tr}, \cdots, \mathcal{L}_K^{tr}\}$. Similarly, combine the cross-test sets into a cross-test matrix $\mathbf{Y}^{te} = [\mathbf{Y}_1^{te}, \mathbf{Y}_2^{te}, \cdots, \mathbf{Y}_K^{te}]$ and their labels into a label set $\mathcal{L}^{te} = \{\mathcal{L}_1^{te}, \mathcal{L}_2^{te}, \cdots, \mathcal{L}_K^{te}\}$. Use $(\mathbf{Y}^{tr}, \mathbf{Y}^{te}, \mathcal{L}^{tr}, \mathcal{L}^{te}, \epsilon)$ as the inputs to the SRC algorithm and obtain the reconstruction error vector $\mathbf{r}_i$. We repeat this process for L times and gather the matched $\mathbf{R}_i^m$ and the sum of non-matched reconstruction errors $\mathbf{R}_i^{nm}$, respectively for $i = 1, \cdots, K$, for fitting the tail distribution based on the GDP. The entire training phase of our method is summarized in Algorithm 2, where ρ indicates the tail size. The choice of ρ depends on the overlap between matched and non-matched reconstruction errors.

3.0.2 Testing

Given a novel test sample $\mathbf{y}_t$, we compute its sparse coefficient $\hat{\mathbf{x}}$ by solving the ℓ_1-minimization problem (2.1). We then obtain K reconstruction errors as required by

Algorithm 2 Pseudocode for SROSR Training

Input: $\mathbf{Y}, \rho, \epsilon, L, \mathcal{L}^Y$
  Initialization
  for $i = 1 : K$ **do**
    for $j = 1 : L$ **do**
      $\tilde{\mathbf{Y}}_i$ = randomly ordered $\mathbf{Y}_i \in \mathbb{R}^{M \times N_i}$
      $N_{tr} = N_i \times 0.8$
      $\mathbf{Y}_i^{tr} = \tilde{\mathbf{Y}}_i(:, 1 : N_{tr})$
      $\mathcal{L}_i^{tr}$ = Labels of $\mathbf{Y}_i^{tr}$
      $\mathbf{Y}_i^{te} = \tilde{\mathbf{Y}}_i(:, N_{tr} + 1 : \text{end})$
      $\mathcal{L}_i^{te}$ = Labels of $\mathbf{Y}_i^{te}$
      $\mathbf{r}_i(j, :) \leftarrow \text{SRC}\ (\mathbf{Y}^{tr}, \mathbf{Y}^{te}, \mathcal{L}^{tr}, \mathcal{L}^{te}, \epsilon)$
    end for
    $\mathbf{R}_i^m = [\mathbf{r}_i(1, i), \cdots, \mathbf{r}_i(L, i)]$
    $\mathbf{R}_i^{nm} = [\sum_{p:p \neq i} \mathbf{r}_i(1, p), \cdots, \sum_{p:p \neq i} \mathbf{r}_i(L, p)]$
    $\boldsymbol{\sigma}_m(i), \boldsymbol{\xi}_m(i) \leftarrow \text{GPDfit}(\mathbf{R}^m, \rho)$
    $\boldsymbol{\sigma}_{nm}(i), \boldsymbol{\xi}_{nm}(i) \leftarrow \text{GPDfit}(-\mathbf{R}^{nm}, \rho)$
  end for
Output: $\boldsymbol{\sigma}_m, \boldsymbol{\xi}_m, \boldsymbol{\sigma}_{nm}, \boldsymbol{\xi}_{nm}$

the SRC algorithm. We choose the class with the minimum reconstruction error as the candidate class. We then obtain two probability scores by fitting matched and sum of non-matched reconstruction errors to their corresponding GPDs. As the two raw reconstruction errors are all normalized into probabilities by their corresponding GPDs, we can add the two probability scores together with appropriate normalization to obtain the final score. Since the non-matched classes can be regarded as a very small part in the open set world for the matched classes, we can specify a coefficient w as the proportion that the number of non-matched classes occupy in the total number of open set classes. In [22], the notion of openness was defined as

$$\text{Openness} = 1 - \sqrt{\frac{2 \times N_{TA}}{N_{TG} + N_{TE}}}, \tag{3.4}$$

where N_{TA} is the number of training classes, N_{TG} is the number of target classes to be identified and N_{TE} is the number of testing classes. We can set the weight as

$$w = 1 - \text{Openness}.$$

If 'Openness = 0', then our setting reduce to the traditional classification problem (i. e., a completely closed problem). With the growth of 'Openness', more and more

unknown classes will appear during testing. As a result the weight on the non-matched probability scores will decrease.

Our testing algorithm is summarized in Algorithm 3. The inputs required during testing are the test sample $\mathbf{y}_t$, training samples $\mathbf{Y}$, the estimated parameters for the matched $(\boldsymbol{\sigma}_m, \boldsymbol{\xi}_m)$ and the sum of non-matched distributions $(\boldsymbol{\sigma}_{nm}, \boldsymbol{\xi}_{nm})$, rejection threshold δ_t and the weight w. The output of the testing phase is one of the following classes $\{1, 2, \cdots, K, \mathcal{O}\}$, where $\mathcal{O}$ represents the open class.

Algorithm 3 Pseudocode for SROSR Testing

Input: $\mathbf{y}_t, \mathbf{Y}, \boldsymbol{\sigma}_m, \boldsymbol{\xi}_m, \boldsymbol{\sigma}_{nm}, \boldsymbol{\xi}_{nm}, \delta_t, w, \epsilon$

1: $\mathbf{r} \leftarrow$ SRC $(\mathbf{Y}, \mathbf{y}_t, \mathcal{L}^Y, \epsilon)$

3: $k^* = \arg\min_i r_i$

4: $r_m = r_{k^*}$, $r_{nm} = \sum_{i=1, i\neq k^*}^{K} r_i$

5: $S_m = G(r_m; \boldsymbol{\sigma}_m(k^*), \boldsymbol{\xi}_m(k^*))$,
$S_{nm} = G(r_{nm}; \boldsymbol{\sigma}_{nm}(k^*), \boldsymbol{\xi}_{nm}(k^*))$

6: $S = S_m + w \cdot S_{nm}$

if $S > \delta_t$ **then**
  Class of $\mathbf{y}_t = \mathcal{O}$
else
  Class of $\mathbf{y}_t = k^*$
end if

Output: k^* or $\mathcal{O}$

Chapter 4

Experimental Results

In this chapter, we present several experimental results demonstrating the effectiveness of the proposed SROSR method on open set recognition. In particular, we present the open set recognition results on the Caltech101 object dataset [4], the MNIST handwritten digits dataset [12] and the Extended Yale B face dataset [7]. The comparison with other existing open set recognition methods such as 1-vs-All Multi-class RBF SVM with Platt Probability Estimation [18] and Pairwise Multi-class RBF SVM [10] in [23] suggests that the W-SVM algorithm is among the best. Hence, we treat it as state-of-the-art and use it as a benchmark for comparisons in this paper. Furthermore, we compare the performance of our method with two other methods based on sparse representation for rejecting invalid samples - SCI[29] and Ratio method [15].

Recognition accuracy and F-measure are used to measure the performance of different algorithms on open set recognition. The F-meaure is defined as a harmonic mean of Precision and Recall

$$\text{F-measure} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{4.1}$$

where Recall is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$$

and Precision defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}.$$

Here TP, FN, and FP denote true positive, false negative and false positive, respectively. F-measure is always between 0 and 1. The higher the F-measure the better the performance of an object recognition system. Accuracy is defined as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TN} + \text{TP} + \text{FP} + \text{FN}},$$

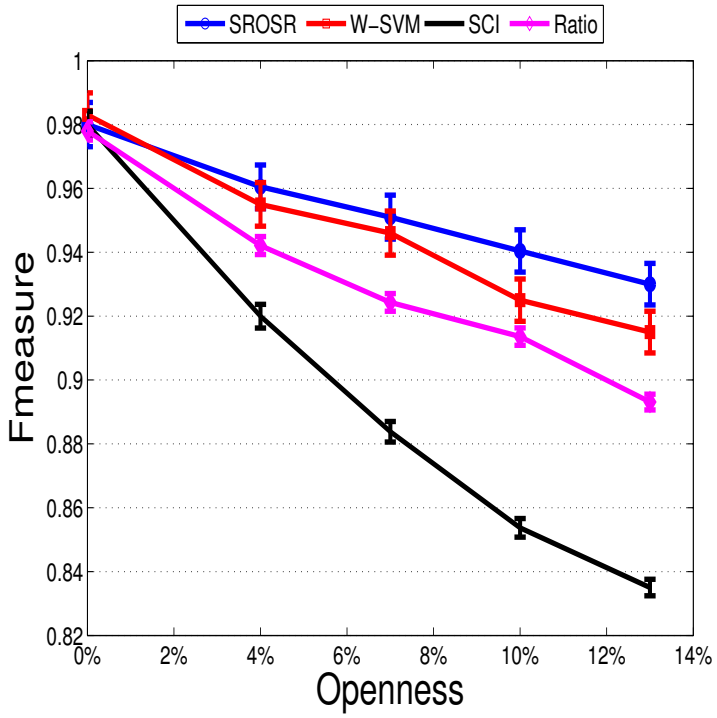
where TN denotes true negative. The rejection threshold, δ_t was empirically determined. In our experiments, we have used $\delta_t = 0.5 \cdot (1 + w), 0.15 \cdot (1 + w), 0.05 \cdot (1 + w)$ for the simulations with the MNIST dataset, Extended YaleB dataset and Caltech101 dataset, respectively.

4.1 Results on the MNIST Dataset

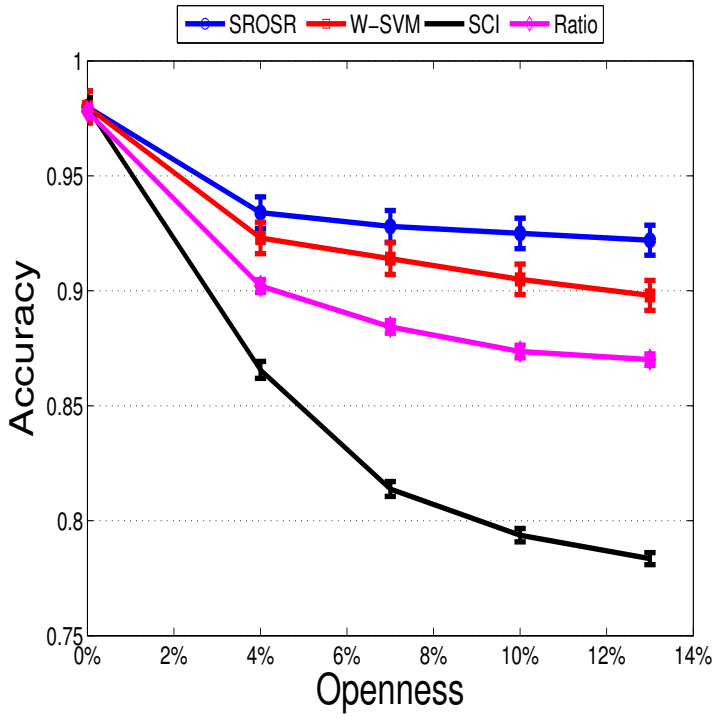
The MNIST dataset contains gray scale images of handwritten digits of size 28×28. There are about 60,000 training images and 10,000 testing images corresponding to 10 classes in this dataset. Sample images from the MNIST dataset are shown in Figure 4.1 Following the experimental setting described in [23], we randomly choose 6 classes for training and alter the openness by the remaining 4 classes. We repeat this experiment 50 times and record the average F-measure and Accuracy. Finally, we plot the Openness vs F-measure and Openness vs Accuracy curves to validate our approach.

Figure 4.1: Sample images from the MNIST handwritten digits dataset.

The Openness vs F-measure and Openness vs Accuracy curves corresponding to this experiment are shown in Figure 4.2(a) and Figure 4.2(b), respectively. Furthermore, the maximum and minimum F-measure values of different methods as we vary openness are summarized in Table 4.1. It can be seen from these results that the proposed SROSR method performs better than W-SVM and sparsity-based rejection methods.



(a)



(b)

Figure 4.2: Results on the MNIST dataset: (a) Openness vs F-Measure results. (b) Openness vs Accuracy results. The proposed SROSR method outperforms the other three methods.

Our method achieves the highest F-measure and accuracy among all the four methods as we vary openness. The rejection methods such as SCI and Ratio are based on the sparsity of the test vector with respect to the training samples. If an open set sample has a sparsity pattern similar to that corresponding to one of the training samples, then the SRC method based on SCI will not reject that sample. This demonstrates that incorporating matched as well as non-matched reconstruction errors can significantly enhance the performance of a sparsity-based classification method on open set recognition.

Table 4.1: F-measure Results on the MNIST dataset.

Openness		SROSR	W-SVM [23]	SCI [29]	Ratio [15]
13%	Max	95.35	95.32	85.69	93.44
	Min	89.33	89.13	80.21	81.14
10%	Max	95.63	95.35	87.52	93.85
	Min	90.79	90.11	81.32	85.14
7%	Max	95.92	96.12	89.40	94.47
	Min	92.01	89.11	88.20	88.14
4%	Max	96.35	96.32	92.84	95.14
	Min	92.32	89.11	89.59	92.5

4.2 Results on the Extended YaleB Dataset

The Extended Yale B Dataset consists of 2,414 frontal images of 38 individuals. These images were captured under various controlled indoor lighting conditions. Each class contains about 64 images. They were cropped and normalized to the size of 32×32 pixels. Samples images from the Extended YaleB dataset are shown in Figure 4.3. We randomly choose 10 classes for training and vary the openness by the remaining 28 classes. The following steps summarize our data partition procedure on the Extended Yale B dataset

1. Randomly select 10 classes among the 38 classes.

2. Randomly choose 80% of the samples in each of the 10 selected classes as training samples.

3. Select the remaining 20% of the samples from step 2 and all the samples from the

other 28 classes as testing samples.

We repeat the above procedure 50 times and report the average F-measure and accuracy of different methods.

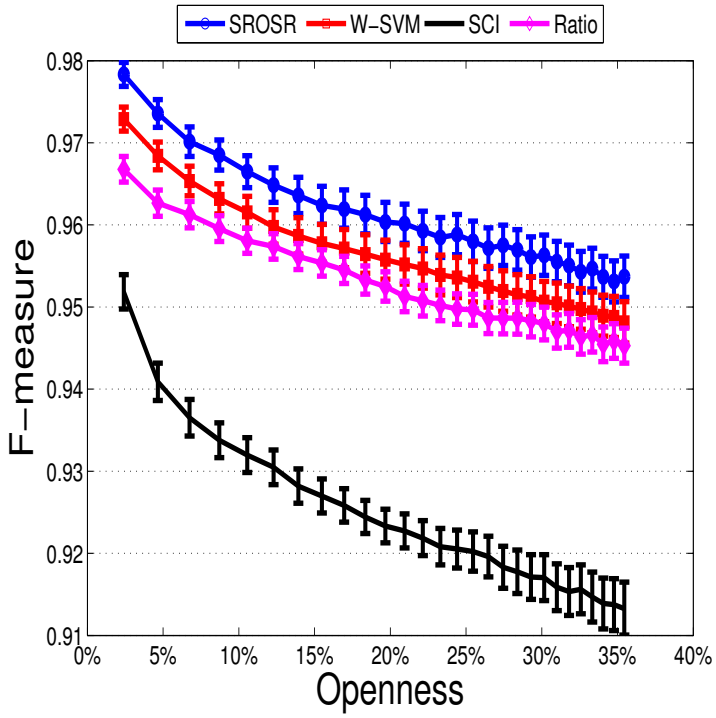
Figure 4.3: Sample images from the Extended YaleB face dataset.

Figure 4.4(a) shows the average F-measure results on this dataset. The face images in this dataset are cropped and well-aligned. Furthermore, the images contain almost the same background. As a result, all compared methods achieve very high F-measures on this dataset. Figure 4.4(b) shows the average accuracy of different methods as we vary openness. As can be seen from both of these plots, the proposed SROSR method outperforms the other compared methods.

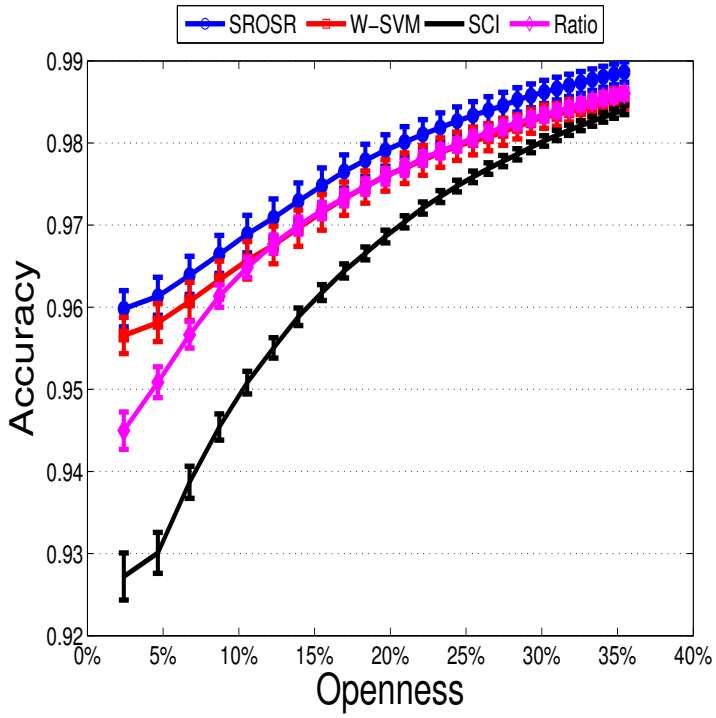
By comparing Figure 4.2(b) with Figure 4.4(b), we see that the accuracy in Figure 4.4(b) increases while the accuracy in Figure 4.2(a) decreases. This is mainly due to the fact that open class samples dominate the accuracy on the Extended YaleB dataset while the closed class samples dominate the accuracy on the MNIST dataset.

4.3 Results on the Caltech101 Dataset

The Caltech101 dataset contains 102 categories including one background class. Each category has about 40 to 80 images and most of the categories have about 50 images. Sample images from the Caltech101 dataset are shown in Figure 4.5. We extracted the spatial pyramid features from these images. We follow an evaluation protocol that is very similar to the previous two experiments. We randomly select 10 categories as training classes and vary the openness by randomly selecting 21 to 40 classes out of the other 92 classes. For all the selected classes either from closed set or from open set, we randomly choose 31 samples for evaluation. So the openness of our experiments on



(a)



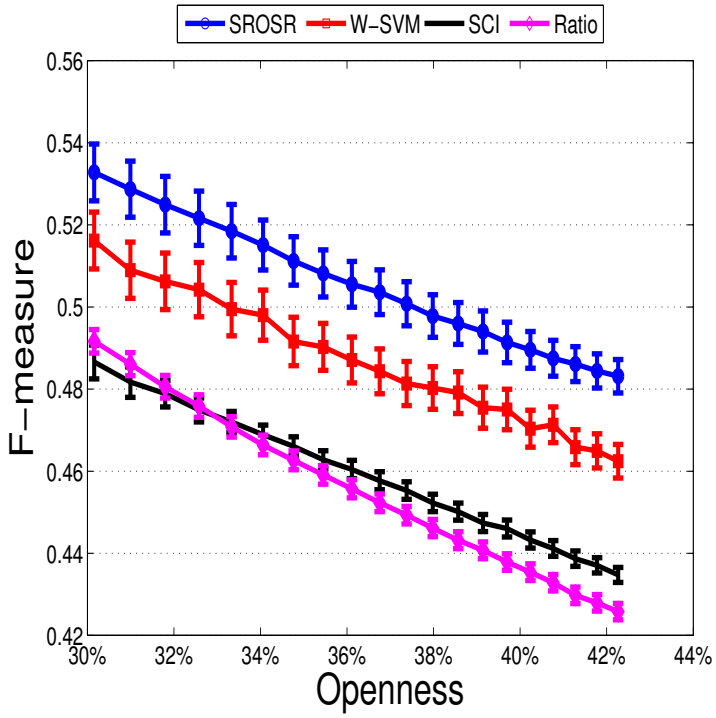
(b)

Figure 4.4: Results on the Extended YaleB dataset: (a) Openness vs F-Measure results. (b) Openness vs Accuracy results. The proposed SROSR method outperforms the other three methods.

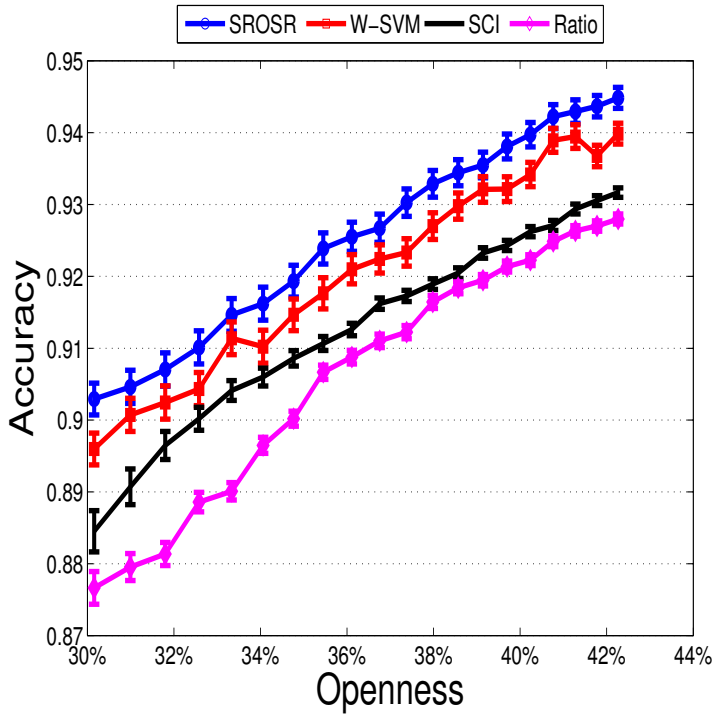
the Caltect101 dataset varies from 30.12% to 42.46%. We average the results over 50 random trails.

Figure 4.5: Sample images from the Caltech101 object dataset.

Figures 4.6(a) and 4.6(b) show the average F-measure and accuracy curves of different methods as we vary the openness, respectively. It can be seen that the proposed SROSR method performs better than the other compared methods. This dataset is more challenging and diverse as it contains objects like animals, natural scenes, buildings, and musical instruments. As a result, the average F-measures and accuracies on this dataset are not as high as compared to the other two datasets for all methods.



(a)



(b)

Figure 4.6: Results on the Caltech101 dataset: (a) Openness vs F-Measure results. (b) Openness vs Accuracy results. The proposed SROSR method outperforms the other three methods.

Chapter 5

Conclusion and Future Work

The SRC algorithm classifies a test sample by seeking the sparsest representation in terms of the training data and does not work well under the open world assumption. In this thesis, we have introduced a training stage to the SRC algorithm so that it can be extended to the open set recognition problems. The resulting algorithm makes use of the reconstruction error distributions modeled by the EVT. Various experiments on popular image and object classification datasets have shown that our method can perform significantly better than many competitive open set recognition algorithms.

Even though, in this thesis, we extended the SRC algorithm for open set recognition, it is also possible to extend many other dictionary learning-based classification algorithms that use reconstruction errors for classification to open set recognition. Furthermore, it remains an interesting topic for future work to develop a sparse representation, dictionary learning-based and deep learning based open set recognition algorithm by directly minimizing the open risk criteria. Also, it will be interesting to explore how could we automatically decide the tail size.

Vita

He ZHang

2010-2014 B. Sc. in Electrical Engineering and its Automation from Jiangnan University

2016.1-Now Graduate assistant, Department of Electrical and Computer Engineering, Rutgers University

References

[1] V Choulakian and MA Stephens. Goodness-of-fit tests for the generalized pareto distribution. *Technometrics*, 43(4):478–484, 2001.

[2] David Andrew Clifton, Samuel Hugueny, and Lionel Tarassenko. Novelty detection with multivariate extreme value statistics. *Journal of signal processing systems*, 65(3):371–389, 2011.

[3] Filipe de O. Costa, Ewerton Silva, Michael Eckmann, Walter J. Scheirer, and Anderson Rocha. Open set source camera attribution and device linking. *Pattern Recognition Letters*, 39:92 – 101, 2014.

[4] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

[5] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.

[6] Shenghua Gao, I.W. Tsang, and Liang-Tien Chia. Sparse representation with kernels. *IEEE Transactions on Image Processing*, 22(2):423–434, Feb 2013.

[7] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.

[8] X. Gibert-Serra, V. M. Patel, and R. Chellappa. Sequential score adaptation with extreme value theory for robust railway track inspection. In *IEEE International Conference on Computer Vision (ICCV) workshop on Computer Vision for Road Scene Understanding and Autonomous Driving (CVRSUAD)*, 2015.

[9] Scott D Grimshaw. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics*, 35(2):185–191, 1993.

[10] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415–425, 2002.

[11] Samuel Kotz and Saralees Nadarajah. *Extreme value distributions*, volume 31. World Scientific, 2000.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[13] H. Van Nguyen, V. M. Patel, N. M. Nasrabadi, and R. Chellappa. Design of non-linear kernel dictionaries for object recognition. *IEEE Transactions on Image Processing*, 22(12):5123–5135, 2013.

[14] V. M. Patel, N. M. Nasrabadi, and R. Chellappa. Sparsity-motivated automatic target recognition. *Applied Optics*, 50(10):1425 –1433, Apr 2011.

[15] Vishal M Patel, Tao Wu, Soma Biswas, P Jonathon Phillips, and Rama Chellappa. Dictionary-based face recognition under variable lighting and pose. *IEEE Transactions on Information Forensics and Security*, 7(3):954–965, 2012.

[16] James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

[17] J. K. Pillai, V. M. Patel, R. Chellappa, and N. K. Ratha. Secure and robust iris recognition using random projections and sparse representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(9):1877–1893, Sept 2011.

[18] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. 1999.

[19] A. Rattani, W.J. Scheirer, and A. Ross. Open set fingerprint spoof detection across novel fabrication materials. *IEEE Transactions on Information Forensics and Security*, 10(11):2447–2460, Nov 2015.

[20] Stephen J Roberts. Novelty detection using extreme value statistics. *IEE Proceedings-Vision, Image and Signal Processing*, 146(3):124–129, 1999.

[21] R. Rubinstein, A. M. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, June 2010.

[22] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(7):1757–1772, 2013.

[23] Walter J. Scheirer, Lalit P. Jain, and Terrance E. Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, November 2014.

[24] A. Shrivastava, V. M. Patel, and R. Chellappa. Multiple kernel learning for sparse representation-based classification. *IEEE Transactions on Image Processing*, 23(7):3013–3024, July 2014.

[25] Richard L Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377, 1989.

[26] M.J. Wilber, W.J. Scheirer, P. Leitner, B. Heflin, J. Zott, D. Reinke, D.K. Delaney, and T.E. Boult. Animal recognition in the mojave desert: Vision tools for field biologists. In *IEEE Workshop on Applications of Computer Vision*, pages 206–213, Jan 2013.

[27] J. Wright and Yi Ma. Dense error correction via ℓ^1 -minimization. *IEEE Transactions on Information Theory*, 56(7):3540–3560, July 2010.

[28] J. Wright, Yi Ma, J. Mairal, G. Sapiro, T. S. Huang, and Shuicheng Yan. Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044, June 2010.

[29] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[30] D. Zhang, Meng Yang, Zhizhao Feng, and D. Zhang. On the dimensionality reduction for sparse representation based face recognition. In *International Conference on Pattern Recognition*, pages 1237–1240, Aug 2010.

[31] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Jing Liu, Zhe Yan, Ting Wang, and Fan-Zhang Li. Kernel sparse representation-based classifier. *IEEE Transactions on Signal Processing*, 60(4):1684–1695, April 2012.

[32] Yuqian Zhang, Cun Mu, Han-Wen Kuo, and John Wright. Toward guaranteed illumination models for non-convex objects. In *IEEE International Conference on Computer Vision*, pages 937–944, 2013.