Sparse-Representation Classification based Face Recognition

Sang Jae Park
park4s@bu.edu

Mingrui Yang
mryang@bu.edu

Anrui Wang
war0623@bu.edu

1 Compressive Sensing Face Recognition

Traditional signal processing usually follows Shannon and Nyquist's sampling theorem. It states that a signal can be recovered perfectly if the sampling frequency is at least twice the highest frequency present in the signal. However, this theorem assumes that the samples are taken with a uniform time interval, and that the reconstruction happens by interpolating the samples with Sinc functions.

Compressive sensing is a new scheme for sampling and reconstructing which abandons these assumptions. It is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to under determined linear systems. This is based on the principle that, through optimization, the sparsity of a signal can be exploited to recover it from far fewer samples than the traditional theory suggests.

Now, we are implementing compressive sensing into face recognition problem. Face is a complex varied high dimensional pattern. The face recognition problem is a classical one in the area of statistical learning and machine intelligence. This technique is believed having a great deal of potential application in public security, law enforcement, information security, and etc. Walmart already used facial recognition technology to catch shoplifters. The classical way to do face recognition is to first do what is called feature extraction. One can think of feature extraction as a projection to a lower dimensional feature space, such that the requirements for memory and computational power is reduced. There are some regular methods seem to work well when in a controlled environment. However, when parameters such as lighting or noise level changes, or when a subject has occluded parts of his/her face (like the addition of glasses), these methods often begin to struggle. So here we are introducing a new classification scheme called Sparse Representation-based Classification (SRC).

2 What is SR

In compressive sensing, we talked about the recovery problem. Sparse representation and the recovery problem in compressive sensing are the same problem essentially.

We can use a formula $y = Ax$ with constraints to represent the relationship between sparsity signal and truth signal. In compressive sensing, we call A as measurement matrix but in sparse representation problem here, we will call it a dictionary. x is the sparsity signal while y is the truth signal recovered from the sparsity signal x.

- **L2-Norm**

L2 norm is well-known as Euclidean norm. On an n-dimensional Euclidean space Rn, the intuitive notion of length of the vector $x = (x_1, x_2, \ldots, x_n)$ is captured by the formula:

Preprint. Work in progress.

$$\| x \|_2 = \sqrt{x_1^2 + \cdots + x_n^2} \tag{1}$$

We could determine the solution by calculate the minimum L2 norm.

$$(l^2) : \widehat{x_2} = \arg\min \| x \|_2 \text{ subject to } Ax = y \tag{2}$$

The resolved $\widehat{x_2}$ is dense in general. It is hard to be classified. In stead, we will look for the sparsest resolution for the following formula, in which the $Ax = y$. The procedure to solve is a NP problem.

$$(l^0) : \widehat{x_0} = \arg\min \| x \|_0 \text{ subject to } Ax = y \tag{3}$$

Since P0 Problem is an NP hard problem which is really hard to solve, we could find the approximate solution of L1 norm instead of L0 norm.

$$(l^1) : \widehat{x_1} = \arg\min \| x \|_1 \text{ subject to } Ax = y \tag{4}$$

- **L1-Norm**

L1 norm is well-known as Manhattan norm. The name relates to the distance a taxi has to drive in a rectangular street grid to get from the origin to the point x.

$$\| x \|_1 = \sum_{i=1}^{n} | x_i | \tag{5}$$

The set of vectors whose 1-norm is a given constant forms the surface of a cross polytope of dimension equivalent to that of the norm minus 1. The Taxicab norm is also called the l1 norm. The distance derived from this norm is called the Manhattan distance or l1 distance.

L1 norm can be resolved by standard linear programming tool for polynomial. It will be more efficient if the resolution is sparse.

Moreover, there could be noises in the real data, which will turn to the formula of $Ax_0 + z = y$. So, x can be resolved by approximate solution of L1 norm as follows.

$$(l^1) : \widehat{x_1} = \arg\min \| x \|_1 \text{ subject to } \| Ax - y \|_1 \leq \varepsilon \tag{6}$$

3 Calculate minimum L1 norm

- Problem: $(l^1) : \widehat{x_1} = \arg\min \| x \|_1$ subject to $Ax = y$
- Let: $x = u - v$ with $u \leq 0, v \leq 0,$ then $\| x \|_1 = 1^T u + 1^T v$

- Standard LP: $z = \begin{bmatrix} u \\ v \end{bmatrix}; B = [A - A]$, then $\min 1^T z$ subject to $Bz = y; z \geq 0$ [1]

- Then we chose Interior-point methods to solve Calculate minimum L1 norm. [2]

In our code, The module exposes two functions:

```
l1ls(A, y, lmbda, x0=None, At=None, m=None, n=None, tar_gap=1e−3,
quiet=False, eta=1e−3, pcgmaxi=5000)


l1ls_nonneg(A, y, lmbda, x0=None, At=None, m=None, n=None, tar_gap=
1e−3, quiet=False, eta=1e−3, pcgmaxi=5000)
```

They can be used as follows:

```
import l1ls as L
import numpy as np
A = np.array([[1, 0, 0, 0.5], [0, 1, 0.2, 0.3], [0, 0.1, 1, 0.2]])
x0 = np.array([1, 0, 1, 0], dtype='f8')  # Original signal
y = A.dot(x0)                            # noise free signal
lmbda = 0.01                             # regularization parameter
rel_tol = 0.01

[x, status, hist] = L.l1ls(A, y, lmbda, tar_gap=rel_tol)
# answer_x = np.array([0.993010, 0.00039478, 0.994096, 0.00403702])
```

4 Why we can use SRC to do face recognition

Sparse representation is based on optical model. In other words, we can represent one person's face in a linear combination of all his faces under different light conditions and at different orientations. The coefficient of other person's face would be zero in theory.

And as the data set (a linear combination of different people's faces) growing larger, the coefficient matrix of any person will be sparse But here is strong assumption: all the faces must be aligned strictly, otherwise sparsity is hard to satisfied. That's means classic SRC cannot deal with significant changes of facial expressions.

However, there's a good news about SRC is that it is robust with random noise. Even if 80% face is disturbed by random noise, SRC has an acceptable accuracy. Another advantage is when some part is overlain, e.g. wearing glasses and scarves, SRC also can work well. In the project we will combine the original image with another random image in some proportion (we used 40% for the random picked image) to simulate the overlay in original image.

One more advantage about SRC is it doesn't need to train a huge data set which will cost several days even months in machine learning methods like neural networks. In some problems SRC need to be modified a little to implement an easy and quick training. The dictionary used by SRC is all training images after aligned, which is a matrix of training data set. Since L0 problem is an NP hard problem, we will use L1-minimization here.

5 Data set

ORL face Lib

We use the data set called ORL face lib. It contains a set of face images taken between

[1] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An Interior-Point Method for Large-Scale l1-Regularized Least Squares, (2007), IEEE Journal on Selected Topics in Signal Processing, 1(4):606-617.

[2] https://github.com/Networks-Learning/l1-ls.py

1992 to 1994 by Cambridge University Computer Lab. There are 10 different images of each person and the whole data set contains images of 40 different people. We choose 7 as training samples in each class and the left 3 are test samples.

Figure 1: ORL Face Images

6 Methods

- **Dictionary establishing and Test sample processing**
All the ORL face lib images are in size of $[112, 92]$. Before implementing those images in our algorithms, we resized each image to a column vector and combine all column vector to a matrix. Assume k classes (here is 40), then A_i is the matrix containing all column vectors of class i and $v_{i,j}$ is column vector of image j in class i.

$$\mathcal{A} = [A_1, A_2, \ldots, A_k] = [v_{1,1}, v_{1,2}, \ldots, v_{k,n-1}, v_{k,n}] \tag{7}$$

Then we resized one test sample to a column vector too, which means $y \in \mathcal{R}^m$, where $m = 112 * 92 = 10304$

- **Normalize dictionary and test sample**
For each column of dictionary and test sample, do normalization and make it to a unit column vector.

- **Robust processing**
In order to increase robustness of overlay and noise, we introduced a noise vector and extended the equation. Here we use randomized value to count for noise. Then what we want becomes to w_0 and we got the dictionary A we need.

$$y = y_0 + e_0 = Ax_0 + e_0 \tag{8}$$

$$y = [A \quad I] \begin{bmatrix} \mathrm{x}_0 \\ \mathrm{e}_0 \end{bmatrix} = B + w_0 \tag{9}$$

- **Find solution**
Now the problem is converted to a L1-Norm minimization problem.

$$(l^1) : \widehat{w_1} = \arg\min \parallel w \parallel_1 \text{ subject to } Bw = y \tag{10}$$

Using Interior-point methods, we can get solution: $\widehat{w_1} = [\widehat{x_1}\ \widehat{e_1}]$ and revert to original face: $y_r = y + \widehat{e_1}$

- **Sparse Representation**

Use $\delta_i(\widehat{x_1})$ as the Eigen function to filter the corresponding coefficient of class i. It is a new vector in which the non-zero values are related to class i. We can classify y by the minimum residual between y and $\widehat{y_1}$.

$$\arg\min_i r_i(y) = \| y_r - A\delta_i(\widehat{x_1}) \|_2 = \| y - \widehat{e_1} - A\delta_i(\widehat{x_1}) \|_2 \tag{11}$$

- **Validation**

Definition 1: sparsity concentration index(SCI) of $\widehat{x_1}$:

$$\mathrm{SCI}(\widehat{x_1}) = \frac{k \cdot \max_i \frac{\| \delta_i(\widehat{x_1} \|_1}{\| \widehat{x_1} \|_1} - 1}{k-1}$$

If SCI $(\widehat{x_1})$ is larger than Υ then output predicted label, otherwise output "invalid input image The recommended value of Υ is 0.85 and we will use this value.

7 Results and Conclusion

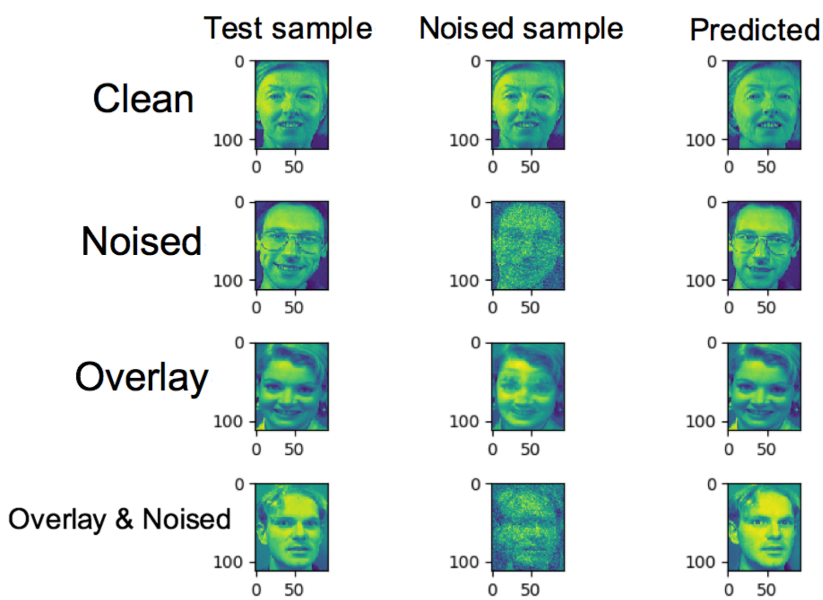

Figure 2: Result

- **We test 120 samples and accuracy is 95%**

- **If we want to apply Sparse representation based face recognition to real-world system. We have to do some improvements in two aspects:** breaking through the limitation of face alignment and figure out a more efficient optimal algorithm.

References

[1]] Vegard Antun.Master thesis code. 2016. URL: https://bitbucket.org/vegarant/code-thesis(visited on 03/23/2017).

[2] Kurt Bryan and Tanya Leise. "Making Do with Less: An Introduction to Com-pressed Sensing." In: Siam Review 55.3 (2013), pp. 547–566.

[3] Farooq Ahmad Bhat and M. Arif Wani. "Performance Comparison of Major Clas- sical Face Recognition Techniques." In: IEEE 13th International Conference on Ma- chine Learning and Applications (ICMLA) (2014), pp. 521–528.

[4] Emmanuel J Candès, Justin Romberg, and Terence Tao. "Robust uncertainty prin- ciples: Exact signal reconstruction from highly incomplete frequency informa- tion." In: IEEE Transactions on informa- tion theory 52.2 (2006), pp. 489–509.

[5] Emmanuel J Candes and Terence Tao. "Near-optimal signal recovery from random projections: Universal encod- ing strategies?" In: IEEE transactions on information theory 52.12 (2006), pp. 5406–5425.

[6] David L Donoho. "Compressed sensing." In: IEEE Transactions on information the- ory 52.4 (2006), pp. 1289–1306.

[7] Yonina C. Eldar and Gitta Kutyniok. Com- pressed Sensing: Theory and Applica- tions. Cambridge university press, 2012. Chap. 12.

[8] Simon Foucart and Holger Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhauser, 2013.

[9] Gareth James et al. An Introduction to Statistical Learning with Applications in R. Springer, 2013.

[10] Kuang-Chih Lee, Jeffrey Ho, and David J. Kriegman. "Acquiring linear subspaces for face recognition under variable light- ing." In: IEEE Transactions on pattern analysis and machine intelligence 27.5 (2005), pp. 684–698.

[11] Øyvind Ryan. Linear algebra, signal pro- cessing and wavelets. A unified approach. 2016. URL: https://github.com/oyvindry/applinalgcode(visited on 04/10/2017).

[12] Robert J. Vanderbei. Linear Programming. 4th ed. Springer, 2014.

[13] John Wright and Yi Ma. "Dense Error Correction Via l1-Minimization." In: IEEE Transactions on Information Theory 56.7 (July 2010), pp. 3540–3560.

[14] John Wright et al. "Robust face recogni- tion via sparse representation." In: IEEE Transactions on Pattern Analysis and Machine Intelligence 31.2 (Feb. 2009), pp. 210–227.

[15] https://blog.csdn.net/xiaoshengforever/article/details/14002843

[16] https://blog.csdn.net/baidu$_3$8060633$/article/details/$70338345

[17] http://www.voidcn.com/article/p-vvolvmgc-ez.html

[18] https://www.cnblogs.com/AndyJee/p/5048235.html