Sparse-Representation Classification based Face Recognition

Computer Science Dept.

A. 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.

Because they can be attributed to $y = Ax$ with constraints. In compressive sensing, we called A as measurement matrix but in sparse representation, we call it dictionary.

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

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

The resolved $\hat{x}_2$ 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.

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

We could find the approximate solution of L1 norm instead of L0 norm.

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

L1 norm can be resolved by standard linear programming tool for polynomial. It will be more efficient if the resolution is sparse. There is noise in the real data, $Ax_0 + z = y$, So, x can be resolved by approximate solution of L1 norm.

$$(l^1):\ \hat{x}_1 = \arg\min\|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \varepsilon$$

B. Calculate minimum L1 norm

- Problem: $(l^1):\ \hat{x}_1 = \arg\min\|x\|_1$ subject to $Ax = y$
- Let: $x = u - v$ with $u \geq 0,\ v \geq 0$, then $\|x\|_1 = 1^T u + 1^T v$
- Standard LP: $z = \begin{bmatrix} u \\ v \end{bmatrix}; B = [A \quad -A]$, then $\min 1^T z$ subject to $Bz = y;\ z \geq 0$
- Then we chose Interior-point methods to solve Calculate minimum L1 norm

* 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.

C. 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 if we have many different people's faces and we establish a linear combination of all the faces, then the coefficient matrix of one 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.

The good news is SRC is robust with random noise. Even if 80% face is disturbed by random noise, SRC has an acceptable accuracy. Another pro is when some part is covered, e.g. wearing glasses and scarves, SRC also can work well.

D. Data set

ORL face lib.

There are 40 classes in ORL face lib and each class has ten face images. We chose 7 as training samples in each class and the left 3 are test samples.




E. Methods

- **Dictionary establishing and Test sample processing**

All images are in size of [112, 92], we resized each image to a column vector and combine all column vector to a matrix. Assume k classes, 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.

$$\mathcal{A} = [A_1, A_2, \cdots\cdots, A_k] = \left[v_{1,1}, v_{1,2}, v_{1,3}, \ \cdots\cdots, v_{k,n-1}, v_{k,n}\right]$$

Then we resized one test sample to a column vector too, which means $\mathcal{Y} \in \Re^m$, 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. Then what we want becomes to w_0.

$y = y_0 + e_0 = Ax_0 + e_0$

$y = [A \quad I] \begin{bmatrix} x_0 \\ e_0 \end{bmatrix} = B + w_0$

- **Find solution**

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

$$(l^1)\text{: } \hat{w}_1 = \arg\min \|w\|_1 \quad \text{subject to } Bw = y$$

Using Interior-point methods, we can get solution: $\hat{w}_1 = [\hat{x}_1 \quad \hat{e}_1]$ and revert to original face: $y_r = y + \hat{e}_1$

- **Sparse Representation**

Use $\delta_i(\hat{x}_1)$ 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 $\hat{y}_1$.

$$\arg\min_i r_i(y) = \|y_r - A\delta_i(\hat{x}_1)\|_2 = \|y - \hat{e}_1 - A\delta_i(\hat{x}_1)\|_2$$

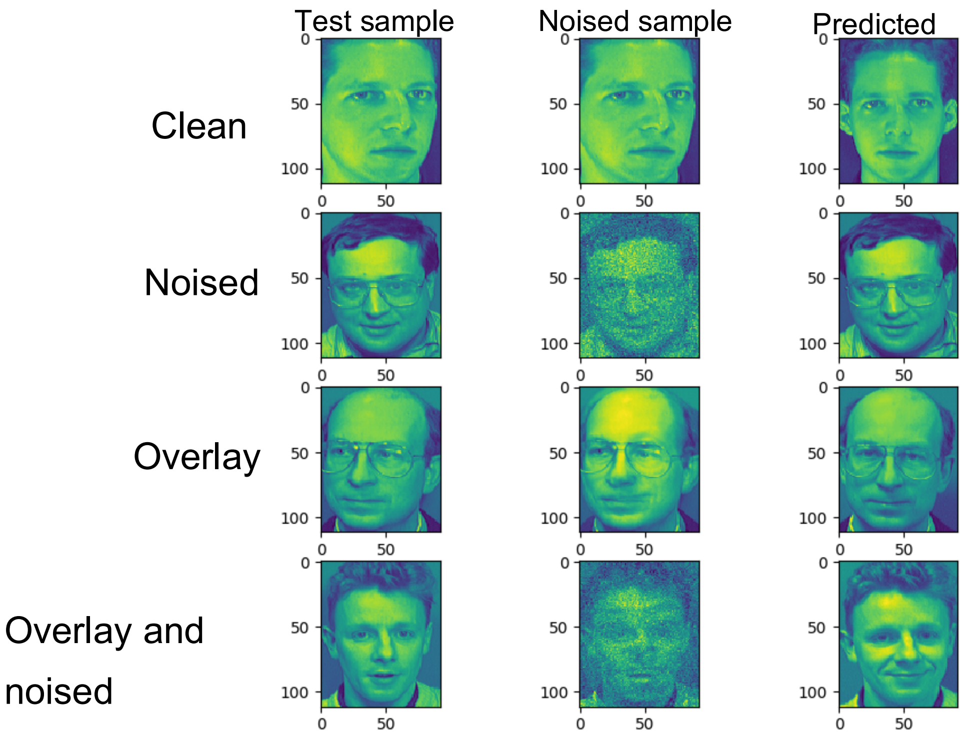
- **Validation**

Definition 1: sparsity concentration index(SCI) of $\hat{x}_1$:

$$\text{SCI}(\hat{x}_1) = \frac{k \cdot \max_i \frac{\|\delta_i(\hat{x}_1)\|_1}{\|\hat{x}_1\|_1} - 1}{k - 1}$$

If SCI($\hat{x}_1$) is larger than τ, then output predicted label, otherwise output "invalid input image"

F. Results



G. Conclusion & Evaluation

- **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
  1. **CVPR2011 LeiZhang Sparse Representation or Collaborative Representation: Which helps Face Recognition**?
  2. **CVPR2011 Meng Yang，Robust Sparse Coding for Face Recognition.**