Image Super Resolution Based on Sparse Representation

Abstract—**With the development of image information technology, people have higher and higher requirements for image quality, so image super-resolution technology was also born of phonology. This paper used a sparse coding method to realize a single-image super-resolution, based on sparse signal to reconstruct a super-resolution image. Some classical super-resolution algorithms were introduced in this paper, and include the feature of these algorithms. The process of sparse reconstruction included dictionary learning, solving sparse coefficient and reconstruction super-resolution image. In this paper, the sparse coefficient was calculated by OMP algorithm and ADMM algorithm, then the algorithm performance was compared by calculated time and reconstruction quality. The reconstruction quality was evaluated by PSNR. The dictionary learning was calculated by K-SVD algorithm. Based these algorithms, the results were reconstructed by changing patch size and dictionary size. It can be found that the reconstruction time increase with dictionary size, and the best quality image can be acquired when the dictionary size was 25*2048. Comparing sparse coding and interpolation it can be significantly found sparse coding quality is better than interpolation, and ADMM reconstruction quality was better than OMP.**

***Index Terms*—image super-resolution (SR), sparse coding, dictionary learning.**

I. INTRODUCTION

With the progress of science and technology, the application of digital image is more and more extensive. Image transmission has gradually become an important role in modern society. The transmission of image information through the Internet and other media has become an important part of modern people's life, and the demand for high-resolution image is growing. The higher the resolution of an image, the richer the information it contains, and the richer the information people can get from it. The resolution of an image is also an important index to judge the quality of an image. At present, image super-resolution technology is widely used in industry, medicine, transportation, remote sensing and other fields. In the industrial field, high-resolution images help to improve production efficiency; in the medical field, high-resolution images can help doctors in clinical diagnosis; in the traffic field, they can help traffic police better find violations or illegal acts; in the remote sensing field, high-resolution images can help to achieve timely update of navigation system information and Location.

However, the resolution of the image obtained in real life is difficult to meet the requirements, which is mainly affected by some hardware sensors and environment, such as the flow of the atmosphere, the change of light conditions, the relative movement of objects. The super resolution problem is called the ill-posed question because we have to increase the image information based on the current limited low resolution image.Therefore, it is of great significance to improve image resolution under the background of information development. High resolution image requires very high hardware, so it greatly increases the cost of current electronic products. In most cases, due to the impact of the environment, the sensor will also be affected. It is difficult to meet the requirements of real life by improving the resolution of image through hardware. Therefore, through the analysis of the collected image information, it is of great significance to process and analyze the image information in the software algorithm level to realize the image super-resolution, which can greatly reduce the hardware cost in the process of image acquisition and meet the requirements of resolution. In additional, the progress in the algorithm will also can have a better behaviour in the application of the super resolution, which encouraged more and more researcher are devoted themself into the development of the field which improve the behaviour of the algorithm.

II. METHOD REVIEW

In recent years, super-resolution has become a more and more hot topic.A large number of scholars have studied image super-resolution, and a number of achievements have been made in this process.According to the information we collected. the mahor image super-resolution technology can be roughly divided into two categories: layer frequency domain and spatial domain according to the processing space. There are wavelet transform, Fourier transform and interpolation in frequency domain. Wavelet transform is usually used to decompose the input image into structure related sub blocks, using the self-similarity of adjacent regions. Due to the lack of wavelet transform in the realization of convolution degradation, it is often combined with Fourier transform to achieve super-resolution reconstruction [5,6].

In the spatial domain, image super-resolution technology can be divided into three categories: spatial interpolation based method, reconstruction based method and learning based method.

In the spatial domain, interpolation is mainly to achieve the super-resolution of the image by interpolating the position of unknown pixels through the relationship between spatial pixels; this method is simple and fast, but the effect of super-resolution is relatively poor[7].

The methods based on reconstruction include convex set projection, iterative back projection and maximum posterior probability estimation[8,9]. On the basis of the image observation model of convex set projection method, it is assumed that every low-resolution image observed is a priori knowledge imposed on the final solution, which is a closed convex set, and the final solution is obtained through continuous iteration, but the final result of this method will be biased when the local image or the motion matrix pre estimation is not accurate. The main idea of the iterative back projection method is: firstly, the image alignment and motion estimation of the observed low-resolution images are carried out and it is worth to get an initial high-resolution image, then the corresponding low-resolution image of the initial high-resolution image is simulated according to the image observation model, and then the error between the simulated low-resolution image and the observed low-resolution image is calculated To reverse adjust the initial high-resolution image, and then iterate until a certain number of iterations or the error is no longer small. The disadvantage of this method is that the result is uncertain. The maximum posterior probability method mainly estimates the reconstructed high-resolution image by Bayesian formula, and obtains the final reconstructed high-resolution result by the maximum posterior probability.

Learning based methods mainly include neural network, sparse expression and so on. Neural network mainly obtains the relationship between high-resolution and corresponding low-resolution images through sample training, and realizes image super-resolution through network learning. With the development of deep learning technology in recent years, this method has made a lot of fruitful achievements in the field of image super-resolution, so it has been widely studied, but the computer hardware requirements are high and large It takes a long time to calculate. Sparse representation is used indirectly in many super-resolution calculations. However, due to the linear relationship between the high-resolution signal and the corresponding low-resolution signal, the high-resolution and low-resolution dictionaries are jointly learned. Compared with other methods, this method requires less training image data for image super-resolution and keeps higher domain characteristics Compared with interpolation, it can get better quality super-resolution image, so this paper uses sparse expression to achieve image super-resolution.

The degeneration image model can be express as:

$$Y=HX+N \tag{1}$$

Where X is original image, N is noise, H is the down sample matrix, Y is the image after degeneration. The observed flow chart can be showed as figure1.

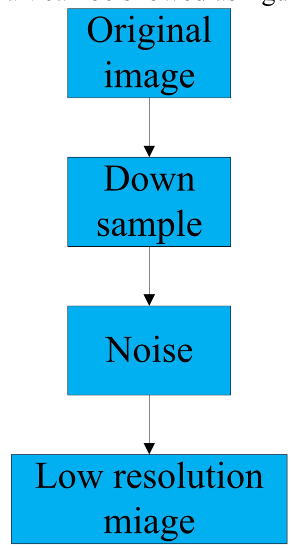

Fig.1. Flow chart of degenerated image

A. Based on interpolation to super-resolution

To realize a single image super resolution can use interpolations, which are nearest neighbor interpolation, bilinear interpolation and bicubic interpolation, respectively. Interpolation is a simple method to realize image super resolution. In this section, three interpolation methods were used to realize image super resolution, and the basic principle was introduced in this section.

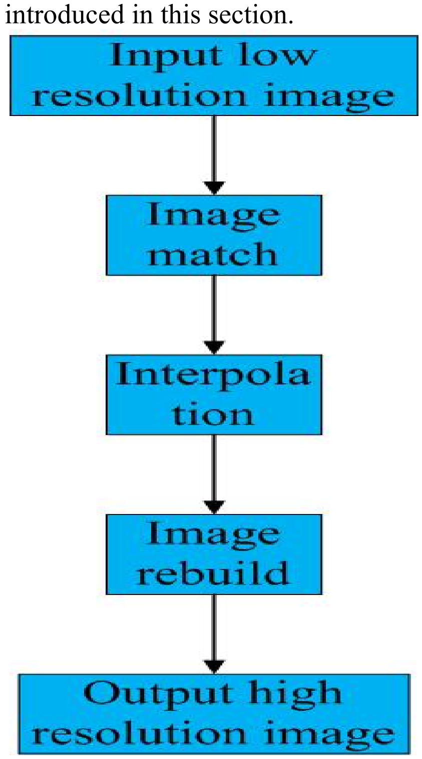

Fig.2. the interpolation process

1.Nearest neighbor interpolation

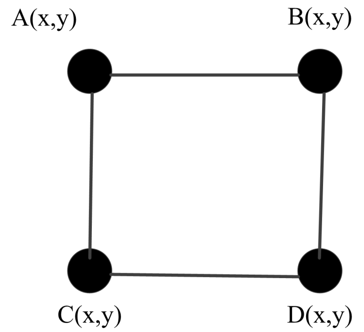

Fig.3. Low resolution image patch

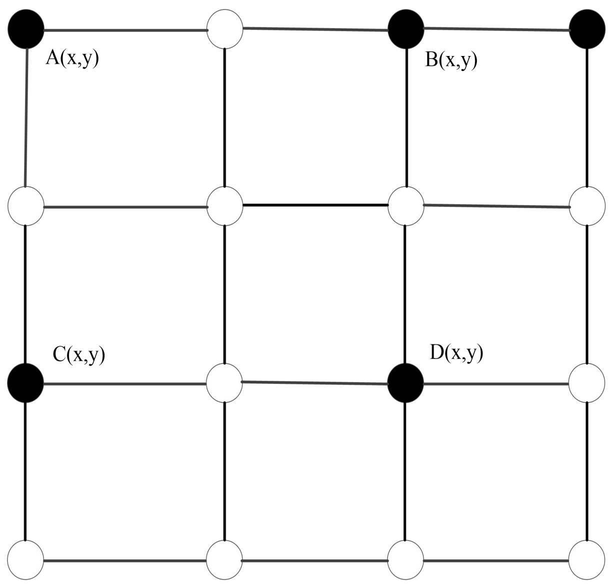

Fig.4. Nearest neighbor interpolation image patch

Nearest neighbor interpolation is used the nearest distance pixel to replace the unknow pixel to realize super resolution. The nearest pixel was chosen by formula(2).

$$d = \min\begin{cases} d[(u,v),d(2i,2j)],d[(u,v),(2i,2j+1)], \\ d[(u,v),d(2i+1,2j)],d[(u,v),(2i+1,2j+1)] \end{cases} \quad (2)$$

2. Bilinear interpolation

Bilinear interpolation is a twice interpolation of combination, figure(7) and figure(8) are the principle of linear interpolation and bilinear interpolation, respectively. The interpolation pixel can be acquire through bilinear interpolation formula(30).

$$f(x,y_1) \approx \frac{x_2-x}{x_2-x_1} f(A) + \frac{x-x_1}{x_2-x_1} f(B) \quad (3)$$

$$f(x,y_1) \approx \frac{x_2-x}{x_2-x_1} f(D) + \frac{x-x_1}{x_2-x_1} f(C) \quad (4)$$

The pixel of bilinear interpolation can be acquire by combining the formula(28) and formula(29), the pixel of bilinear interpolation can be expressed as formula(30).

$$f(x,y) \approx \frac{1}{(x_2-x_1)(y_2-y_1)} (f(A))(x_2-x)(y_2-y) + f(B)(x-x_1)(y_2-y) + f(D)(x_2-x)(y-y_2) + f(C)(x-x_2)(y-y_1) \quad (5)$$

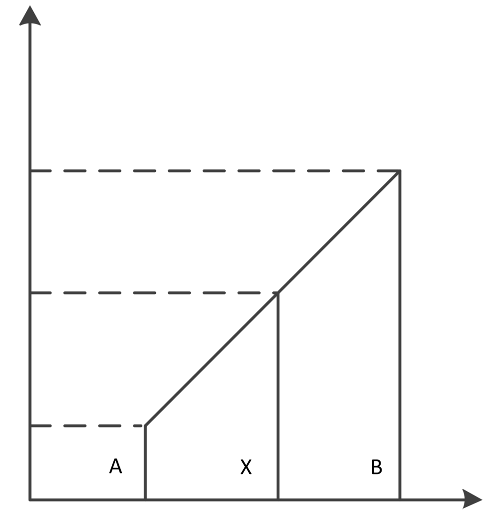

Fig.5. Linear interpolation principle

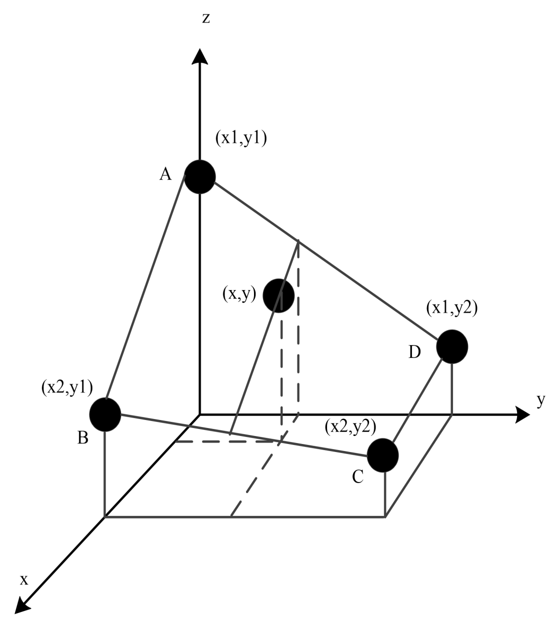

Fig.6. Bilinear interpolation example

3. Bicubic interpolation

The pixel value of bicubic was acquired by 16 pixel surrounding the unknow pixel. The schematic diagram can be expressed as figure6. The unknow pixel can be acquired by formula(31).

$$f(x,y) = ABC^T \quad (6)$$

Where A,B and C can be expressed as formula (6)

$$A = [S(u+1)\ S(u)\ S(u-1)\ S(u-2)]$$

$$B = f(i-1{:}i+2, j-1{:}j+2) \quad (7)$$

$$C = [S(v+1)\ S(v)\ S(v-1)\ S(v-2)]$$

The core function can be expressed as formula(8)

$$s = \begin{cases} 1-2|x|^2+|x|^3 & |x| \le 1 \\ 4-8|x|+5|x|^2-|x|^3 & 1 < |x| \le 1 \\ 0 & \text{ortherwise} \end{cases} \quad (8)$$

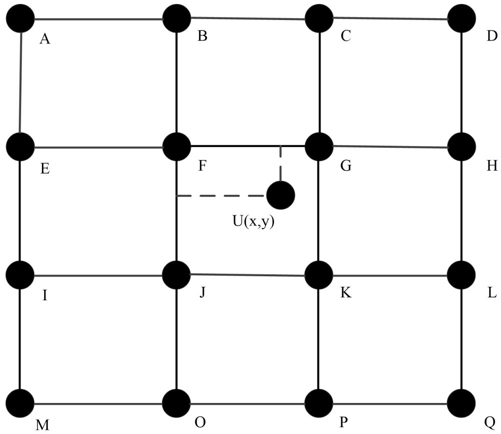

Fig.7. Bicubic interpolation example

Fig.8. Bicubic interpolation example

It can be found that Bicubic interpolation is better than neighbor interpolation and Bilinear interpolation in this photo.

B. Based on reconstruction to super-resolution

There are three kinds of methods based on reconstruction, IBP (iterative Back-projection Method), MAP(Maximum Posterior Probability), and (POCS).

1. IBP algorithm

The basic idea of realizing image SRR is to compare the degraded LR image with the LR image to be reconstructed, correct the HR image reversely according to the error result, and cycle until the convergence condition is met, and finally obtain the HR image.The specific principle is as follows:

The LR image to be reconstructed is Y, the quality reduction model is known to be W, and the initial HR image is X0, which is generally obtained by interpolation of Y, so it can be known that the initial LR image is:

$$\mathrm{Y}^0 = \mathrm{WX}^0 + N \tag{9}$$

Since X0 is not consistent with the original HR image, Y0 and Y obtained must have errors.Error Y0 −Y is used to reverse correct X0, and the formula is as shown in formula :

$$\mathrm{X}^1 = X^0 + W^{BP}(Y - Y^0) \tag{10}$$

WBP is the inverse projection matrix, which is the approximation matrix of W inverse.IBP method is simple in principle and easy to implement, but it has strict requirements on WBP and is difficult to select. In addition, the iterative process of algorithm is not necessarily convergent.

2. MAP algorithm

The basic idea of implementing image SRR by MAP method is to treat the transformation from LR image to HR image as a probabilistic statistical problem.The HR and LR images are regarded as two random processes and the reconstructed HR images are determined by calculating the maximum posterior probability.The specific principle is as follows:

The HR image is X, and the LR image sequence is Y1,Y2,...,Yk, We know the prior probability density P(X) and the conditional probability density P(Y1,Y2,..., Yk | X), That is, the solution of X is regarded as the solution of the problem of the maximum posteriori probability P(X |Y1,Y2,...,Yk), so as to determine the HR image:

$$\hat{X}_{map} = \arg\max_{X} [P(X \mid Y_1, Y_2, ..., Y_k)] \tag{11}$$

Use bayes' formula, and then take the natural log,

$$\hat{X}_{map} = \arg\max_{X} [\ln P(Y_1, Y_2, ..., Y_k \mid X) + \ln P(X)] \tag{12}$$

In formula (12), P is the conditional probability, which is determined by the noise in the degradation model, and P(X) is the prior probability, which is generally determined by the gaussian or markov MRF model.The advantage of the MAP method is that it has a unique and definite solution, but its disadvantage is that it is computationally intensive.

3. POCS algorithm

POCS method is based on set theory.The basic idea is to express multiple constrained convex sets by mathematical formulas, such as positive characterization, observational consistency, boundedness of energy and local smoothness.Then the intersection set of many convex sets includes the desired HR image.Starting from a certain point in the HR image, the next point that satisfies all constraints is found through the iterative projection method, and the whole HR image is finally determined.The specific principle is as follows:

Let the initial HR image be x0, and each constrained convex set corresponds to a convex set projection operator kP, then the formula to obtain the HR image is:

$$\mathrm{X}^{\mathrm{n}} = P_N \ldots P_3 P_2 P_1 x^0 \tag{13}$$

The formula (13) is iterated repeatedly until the convergence results in the HR image.POCS method has the advantage of simple principle and easy implementation, but the algorithm is greatly influenced by the initial value of HR image and the reconstruction effect is unstable.

C. Optimization Problem Solving

Since the 0L norm based model problem proposed is an NP problem, and the solution is complex, MP algorithm and OMP algorithm are often used to solve it. The following two algorithms are introduced in detail.

1.MP algorithm

As a greedy algorithm, the basic idea of MP algorithm is that an atom in a dictionary is the best approximation of the signal if its inner product with the signal residual is the largest.Therefore, the core of MP algorithm is to find the atomic vector in the atomic library that satisfies the condition of the maximum inner product of the residual of the current signal.The specific implementation steps of the algorithm are as follows:

Input: an overcomplete dictionary D (Simplified formula, the atoms in D have been standardized, that is, L2 norm value is 1), The image block signal x to be sparse represented, and the sparse representation error e brought by system noise.

Output: Sparse coding coefficient α

1)initialization operation: signal residue $r_0 = x, \alpha = 0$, iteration times k =1;

2)Select the atom with maximum correlation with x:

$$d_1 = \arg\max_{d_n \in D} |< d_n, x >| \tag{14}$$

The signal x is decomposed into:

$$x =< d_1, x > d_1 + r_1 = x_1 + r_1 \tag{15}$$

Sparse representation coefficient:

$$\alpha_1 =< d_1, x > . \tag{16}$$

3) Loop operation: while the stop condition is not met, perform the following loop steps:

a. Select the atom with maximum correlation with the Residual:

$$d_k = \arg\max_{d_n \in D} |< d_n, r_{k-1} >| \tag{17}$$

b. Update the residual and the coefficients:

$$r_{k-1} =< d_k, r_{k-1} > d_k + r_k \tag{18}$$

The signal x is decomposed into:

$$x_k = x_{k+1} + (r_{k-1} - r_k) = x_{k-1} + < d_k, r_{k-1} > d_k \tag{19}$$

After updating, the signal residual is:

$$r_k = r_{k-1} - < d_k, r_{k-1} > d_k \tag{20}$$

Sparse representation coefficient:

$$\alpha_k =< d_k, r_{k-1} > \tag{21}$$

c.Update the iteration times:

$$\text{K=K+1} \tag{22}$$

4)End while

2.OMP algorithm

OMP algorithm is similar to MP algorithm, and its core idea is to find the atom in the atom library that satisfies the condition that the inner product value of the residual with the current signal is the largest.OMP algorithm is improved on this basis. In order to avoid the atoms in D being used repeatedly, and at the same time to improve the convergence speed, orthogonal projection of selected atoms into the span of selected atom vectors is carried out to ensure that the obtained current signal residual remains orthogonal to the atoms previously used.The specific implementation steps of the algorithm are as follows:

Input: the over-complete dictionary D (to simplify the formula, the atoms in D have been standardized, that is, the L2-norm value is 1), the image block signal x to be sparse represented, and the sparse representation error e introduced by system noise.

Output: sparse coding coefficient α

1) Initialization： used to store all used atomic index values, orthogonal projection operators: $P_0 = 0$.

$$\Omega_0 = \Phi,\ r_0 = x, \alpha = 0, k = 1 \tag{23}$$

2) Select the atom with maximum correlation with the x:

$$n_1 = \arg\max_{d_n \in D} |< d_n, x >|, \quad \Omega = \{n_1\} \tag{24}$$

Then the current signal residual obtained after signal x decomposition is:

$$r_1 = r_0 - P_1 r_0 = (I - P_1) r_0 \tag{25}$$

Sparse representation coefficient:

$$\alpha_1 =< D_\Omega, x > \tag{26}$$

3) Loop operation: while the stop condition is not met,perform the following loop steps:

$$P_1 = D_{\Omega_1} (D_{\Omega_1}^T D_{\Omega_1})^{-1} D_{\Omega_1}^T \tag{27}$$

a.Select the atom with maximum correlation with the Residual:

$$n_k = \arg\max_{n} |< d_n, r_{k-1} >| \tag{28}$$

b.Update orthogonal projection operators, the residual and the coefficients:

$$\begin{gathered} \Omega_k = \Omega_{k-1} \cup \{n_k\} \\ P_k = D_{\Omega_k} (D_{\Omega_k}^T D_{\Omega_k})^{-1} D_{\Omega_k}^T \\ r_k = (I - P_k) r_{k-1} \\ \alpha_k =< d_\Omega, r_{k-1} > \end{gathered} \tag{29}$$

c.Update the iteration times:

$$\text{K=K+1} \tag{30}$$

4) End while

D.ADMM algorithm sparse coding

1.Theory of ADMM algorithm

Alternating direction method of multipliers(ADMM) was came out in 70s in 20 century, it is very efficient in solving some large scale sparse coding and can solve a question through dimensionality reduction iteration based on Augmented lagrangian construction. Using ADMM can equivalently decompose original question, and solve the subproblem to solve full question. The ADMM can express as:

$$\min f(x) + g(y) \text{s.t. } Ax + By = b \qquad (31)$$

According to the Augmented Lagrange penalty function can acquire formula as:

$$L(x,y,\lambda) = f(x) + g(x) + \lambda^T(Ax + By - b) + (\rho/2)\|Ax + By - b\|_2^2 \qquad (32)$$

Where ρ is the Penalty parameters, and $\rho > 0$. Classical ADMM algorithm step can express as

$$x^{k+1} = \arg\min_x L_\rho(x,y^k,\lambda^k)$$
$$y^{k+1} = \arg\min_y L_\rho(x^k,y,\lambda^k)$$
$$\lambda^{k+1} = \lambda^k + \rho(Ax^{k+1} + By^{k+1} - c) \qquad (33)$$

It is need set initial variable value then iterate the formula(46) until to the result of iteration can satisfy the condition. It can be significantly see that the ADMM algorithm is consist of x minim, y min and dual problem. The flow chart can show as figure2

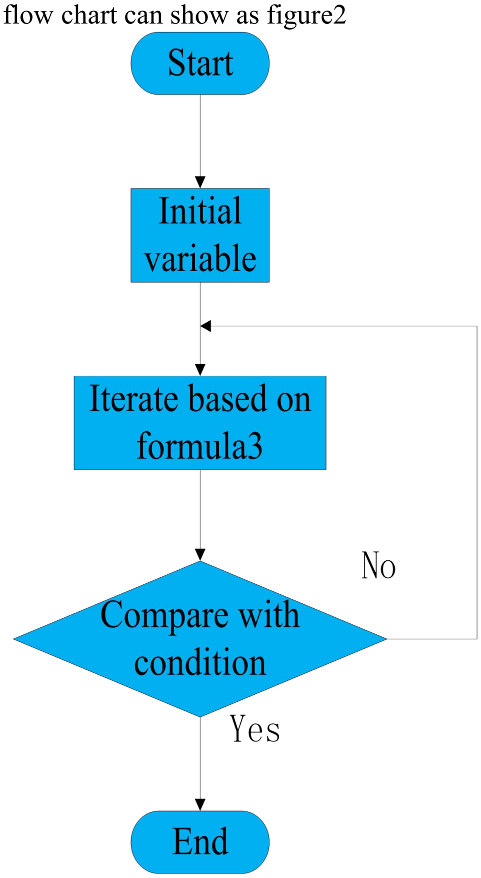

Fig.9. Flow chart of ADMM algorithm.

2.*Solve the sparse coding*

The objective function and constraint condition can be clearly description as formula(x) based on OMP algorithm.

$$\hat{\alpha} = \arg\min_\alpha \|\alpha\|_0 \text{ s.t. } y = X\alpha \qquad (34)$$

Thus, based on above the ADMM algorithm theory, the objective function and constraint condition can be expressed as formula(x) by Augmented Lagrange

$$L(y,\alpha,\lambda,v) = \|\alpha\|_0 + < v,y - X\alpha > + (\gamma/2)\|y - X\alpha\|_2^2 \qquad (35)$$

Then, the Augmented Lagrange can be calculate based on formula(x), it can be expressed as formula(x)

$$y_{k+1} = \arg\min_y L(y,\gamma_k,v_k)$$
$$\alpha_{k+1} = \arg\min_\alpha L(y_{k+1},\alpha,\gamma_k,v_k)$$
$$v_{k+1} = v_k + \gamma_k(y_{k+1} - X\alpha_{k+1})$$
$$\gamma_{k+1} = \rho\gamma_k(\rho > 0) \qquad (36)$$

Then, based on the formula(x), we need iterate it until to acquire to a best sparse coding coefficient. The actual calculate steps in the Table1.

1	Set initial the value of y,α,v,γ
2	Repeat
3	$y_{k+1} = \arg\min_y L(y,\gamma_k,v_k)$
4	$\alpha_{k+1} = \arg\min_\alpha L(y_{k+1},\alpha,\gamma_k,v_k)$
5	$v_{k+1} = v_k + \gamma_k(y_{k+1} - X\alpha_{k+1})$
6	$\gamma_{k+1} = \rho\gamma_k(\rho > 0)$
7	Until to can satisfy the stopping condition

Table1. sparse coding algorithm

E.Reconstruction image quality evaluation criteria

Since the development of image SRR method, there is still no authoritative and unified solution to the problem of image quality evaluation after reconstruction.The reasons for this phenomenon can be explained from the image itself, people's subjective consciousness and image application scene. In terms of image, different original images will have different effects after reconstruction due to differences in texture and edge information of the same image SRR algorithm.In terms of people's subjective consciousness, different people have their own subjective evaluation criteria for the HR images obtained after the reconstruction of the same image and the same algorithm. Such factors are uncontrollable and irregular.In terms of the application scenes, the analysis of the application scenes of specific image reconstruction shows that the evaluation results of the same reconstructed image for different needs are different due to the different specific needs during the reconstruction.

The main purpose of image reconstruction is to preserve the effective low-frequency information of the original image and increase the high-frequency information of the image so as to make the texture features of the image more clear at the edge.Image quality includes two meanings: fidelity and intelligibility.In the field of image SRR reconstruction, fidelity refers to the deviation degree of the HR image to be evaluated compared with the reference standard HR image.Intelligibility refers to the ability of the reconstructed HR image to identify people

And the ability of the machine to provide effective information.Fidelity tends to investigate the similarity of high and low resolution of the image, while intelligibility tends to investigate the overall and detailed features of the reconstructed image.At present, the mainstream IQA methods are divided into objective evaluation and subjective evaluation.

Objective quality evaluation is the quantitative calculation of all aspects of image data using mathematical methods and the quantitative evaluation of image pixel and image structure using unified methods.This evaluation method can be used to judge the image from all aspects of information, and the evaluation formula is consistent. The conclusion is more objective, free from subjective and external factors.The main objective quality evaluation indexes include Mean Square Error (MSE), Peak signal-noise Ratio (PSNR) and Structural Similarity Index, (SSIM).

$$\mathrm{MSE}(\hat{x},x)=\frac{1}{N_1N_2}\sum_{k=1}^{N_1}\sum_{j=1}^{N_2}(\hat{x}_{k,j}-x_{k,j})^2 \tag{37}$$

In the field of image, root-mean-square error is used to represent the image quality from the perspective of pixels by calculating the difference between the pixel values corresponding to the reconstructed image and the original image, which represents the deviation degree of the image.The smaller the value of MSE is, the closer the evaluation image is to the reference image, and the better the reconstruction effect of image SRR algorithm is.Among them, the x on behalf of the original image, ˆ x on behalf of the reconstruction image.

$$\mathrm{PSNR}(\hat{x},x)=10*\lg\frac{N^2}{\frac{1}{N_1N_2}\sum_{k=1}^{N_1}\sum_{j=1}^{N_2}(\hat{x}_{k,j}-x_{k,j})^2} \tag{38}$$
$$=10*\lg\frac{255^2}{MSE}$$

The peak signal to noise ratio (PSNR) is from the perspective of pixels, and the difference of corresponding position pixel values between the original image and the reconstructed image is also considered to represent the image quality.It is similar to MSE method, but PSNR method is more complicated.PSNR is by far the most widely used IQA metric.The larger the PSNR value, the image to be evaluated is closer to the reference image, and the SRR algorithm is better.According to experience, the value between 0 and 20dB represents dissatisfaction;20-25db represents acceptable;Between 30 and 40dB, relatively satisfactory;Greater than 40dB is very satisfactory.

III. Sparse Representation

A. *Sparse model*

The sparse model is develoed from the compressed sensing .as for compressed sensing , it's a math method which can generate the high resolution data form the low-resolution data,which is quite useful in the field of reconstruct the old musical recordings , figure out the radio signal in the militarily, get the Magnetic Resonance Imaging(MRI)’s informtaion more efficiently.many famous scientist make a lots of efforts. The compressed sensing theory focues on a highly accurate reconstruction by means of incomplete linear measurements of the height of the signal. The development of the compressed sensing makes the sparse representation applicable and progressed a lot in the image processing application[11].

1. basic structure

2006, Donoho and his team suggests compressed sensing, a well-behavior and advanced signal processing theory, which challenge the traditional Shannon-Nyquist sampling theory. In compressed sensing, the restriction and limit of the Shannon theory has been break, for information can be sensed and stored from the compressed signal[12,13].

1). question describe

in the discrete and real signal space RN, there is a one-dimensional discrete signal X = {x1, x2, …, xN}. N means the dimension, which can regard as a N*1 column vector. Base on the fundamental linear algebra, assuming there are a set of base vector $\{\Psi i\}^{N}_{i=1}$, so any signal in the RN space can be linear represented by the base vector $\{\Psi i\}^{N}_{i=1}$:

$$\mathrm{X}=\sum_{i=1}^{N}s_i\Psi_i=\Psi S \tag{39}$$

In which , S means the x's projection on the Orthogonal basis space, Ψ= [Ψ1,Ψ2,…,ΨN] is the N*N Orthogonal transformation basis matrix. There are only K of the coefficients in S is non zero (K<<N), the left coefficients are 0 or close to 0. So in this situation, X is k-sparse.

Same as the compressed sensing, in sparse representation, signal can be presented as the linear combination of a set of fundamental signal regard as atoms based on an over-complete dictionary. In the following schematic map, Y means the original data, where D is an over-complete dictionary, based on the knowledge of linear algebras, the base vector set have to be bigger than the data, so the dictionary have to be over-complete, and α means the coefficients of the sparse representation.

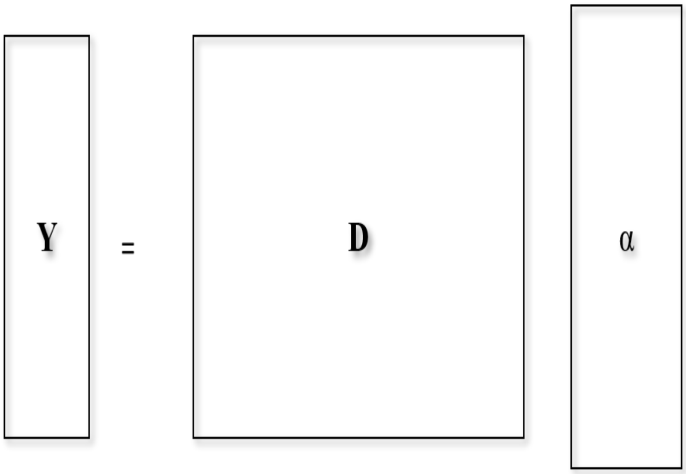

Fig.10. The schematic map of the Sparse model

So called sparse representation means the following equation

Find α s.t. Dα = Y.

But because the dictionary is uncertain, the liear system will lead to many possible solves. We need to find out the most simple one, which represent the one with fewest non-zeros and the simplest way to search the dictionary. In the convex optimization field, the L0-norm will return the zero numbers of the vector. So the question can be transformed to :

$$\min||\alpha||_0 \ s.t. \ Y = D\alpha \tag{40}$$

By allowing some error tolerance to the last equation, the equation can be :

$$\min||\alpha||_0 \ s.t. \ ||Y - D\alpha||_2^2 \leqslant \sigma \tag{41}$$

And there are some challenge in computing the L0 norm, and the L1,L0' Optimal convex approximation, is easier to calculate. Under this situation, The L0 norm is replace by the L1 norm and the objective function is:

$$\min||\alpha||_1 \ s.t. \ ||Y - D\alpha||_2^2 \leqslant \sigma \tag{42}$$

Last the Lagrange multipliers can used to get the final objective function:

$$\min_{\alpha} \ ||Y - D\alpha||_2^2 + \lambda||\alpha||_1 \tag{43}$$

It is clear to get the sparse representation is a optimization problem, which can be inferred step by one step . so it's quite useful to help us understand the whole procedure. Even though, nowdays the deep learning has become a hot topic , but there are still many mysterious in the whole deep learning system. That's quite of the reason why we choose topic.

B. How to use the sparse representation

In the former sections, we have already talked about the super resolution problem and the sparse representation. So how to take advantage of the sparse presentation model to deal with the super resolution problems. According the sparse representation and compressed sensing , there are an Sparse Signal Recovery Theories:Under Mild Conditions, A High-resolution Signal Can Be Recovered From Its Low-resolution Version If The Signal Has A Sparse Representation In Terms Of Some Dictionary. In addition,We have already mentioned that the sparse representation is a convex optimization problem. For the low-resolution image, we can use the low-resolution dictionary and the sparse coefficients to represent the image. For the high-resolution image, we can use the high-resolution dictionary and the sparse coefficients to represent the image. The key point of applied the sparse representation into the super-resolution problem is normalization and joint dictionary learning, which can help us force the high resolution and low resolution get the same sparse coefficients. Joint dictionary learning will be introduced in the next section. Consequently, the whole technique process can be summarized as that we use the low resolution image and low resolution dictionary to get the sparse coefficients, and take the same sparse coefficient and the high resolution dictionary to get the high resolution image.So the ill-posed sper-resolution problem is now become a mathmetic problem, a convex optimization problem which can be calculated by the greedy algorithm.

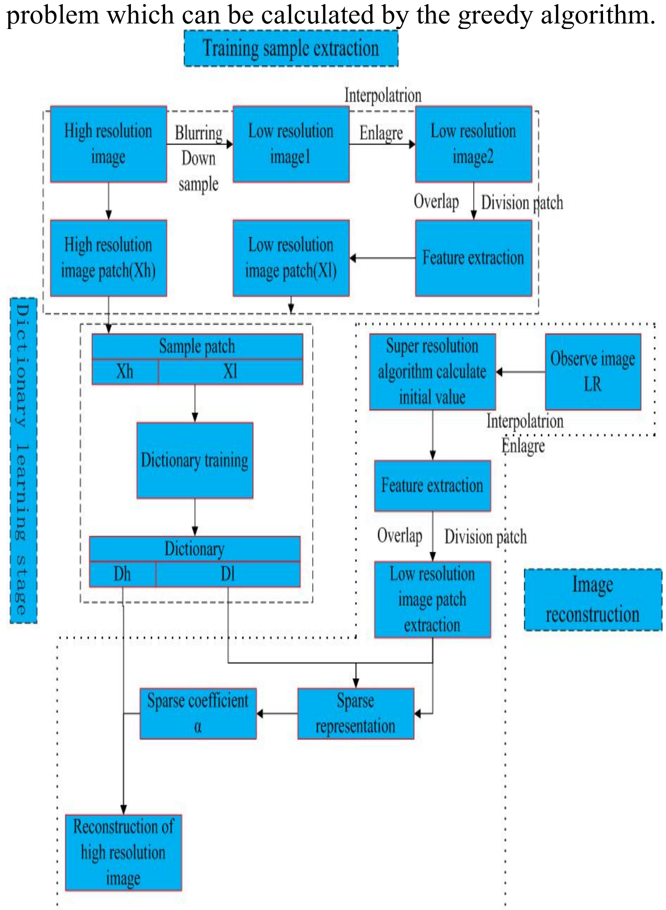

Fig.11. Training sample and Dictionary learning.

1. Feature extraction method

The image feature extraction algorithm can help us to get image high frequency information, which can affect the output of the dictionary we get the result of sparse representation. Base on the literature review, there are many people tried different feature extraction algorithm, such as Sun [1] take advantage of the Gaussian derivative filters to get the contour information of the image.whereas in 2004,Chang [2] used the first and second order gradients operators to extract the feature because they are quite simple and effect. For further feature extraction, Li tried the wavelet decomposition algorithm to get more detail feature.

C. feature extraction based on order gradient

The first order, second order gradient feature can represent the basic structure feature information of the image, and it's quite simple to achieve based on knowledge we learned from class, the following is the first and second order gradient operator.The first order gradient operator:

$$f_1 = [-1,0,1],\ f_2 = f_1^T$$

The second order gradient operator: (44)

$$f_3 = [1,0,-2,0,1],\ f_4 = f_3^T$$

We applied the first and second order gradient to lena image, so the output is :

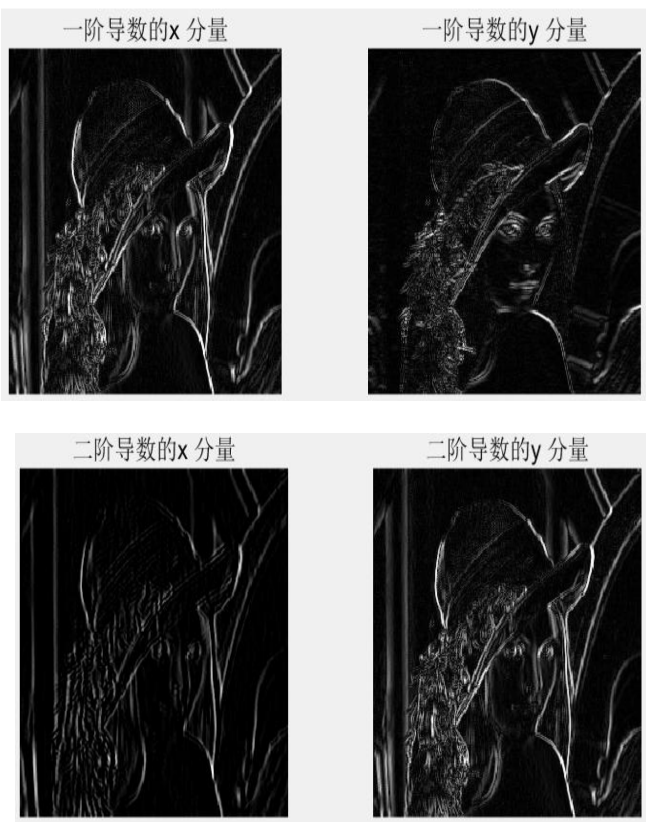

Fig.12. The first and second order gradient operator.

1. feature extraction based on wavelet decomposition

Wavelet Analysis is originated from Fourier Analysis. Unlike the Fourier transform, which can only observe the signal in the time domain. In the frequency domain, we can only analyze the whole signal Statistical characteristics. However, wavelet transform can be used to analyze the time and frequency characteristics of local signals. An adaptive time-frequency window is introduced, through the expansion, translation and other transformations of the signal for multi-scale processing, which known as multi-resolution analysis. Based on those advantages, wavelet analysis is used to process image signals, so the two-dimensional discrete wavelet transform is mainly introduced. For the discrete two-dimensional signal f (x, y), perform the wavelet transform. For certain one dimensional scaling function $\Phi(x)$ and wavelet function $\Psi(x)$, and we combine them to get different two dimensional scaling function $\Phi(x,y)$ and wavelet function $\Psi^H(x,y)$, $\Psi^V(x,y)$, $\Psi^D(x,y)$, the formulasis following:

$$\begin{aligned} \Phi(x,y) &= \Phi(x)\,\Phi(y) \\ \Psi^H(x,y) &= \Psi(x)\,\Phi(y) \\ \Psi^V(x,y) &= \Phi(x)\,\Psi(y) \\ \Psi^D(x,y) &= \Psi(x)\,\Psi(y) \end{aligned} \tag{45}$$

So the signal f(x,y) can be decomposed into the linear combination of the former formulas:

$$f(x,y) = \frac{1}{\sqrt{MN}} \sum_m \sum_n W_\varphi(0,m,n)\varphi_{0,m,n}(x,y) +$$

$$\frac{1}{\sqrt{MN}} \sum_{j=0}^{\infty} \left\{ \begin{array}{l} \sum_m \sum_n W_\psi^H(j,m,n)\psi_{j,m,n}^H(x,y) + \\ \sum_m \sum_n W_\psi^V(j,m,n)\psi_{j,m,n}^V(x,y) + \\ \sum_m \sum_n W_\psi^D(j,m,n)\psi_{j,m,n}^D(x,y) \end{array} \right\} \tag{46}$$

In the former formula, $W_\Phi(0,m,n)$ is the approximation coefficient, and $W_\Psi^H(j,m,n)$, $W_\Psi^V(j,m,n)$, $W_\Psi^D(j,m,n)$ is the detail coefficients:

$$W_\varphi(0,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\varphi_{0,m,n}(x,y)$$

$$W_\psi^i(j,m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\psi_{j,m,n}^i(x,y) \quad i=\{H,V,D\} \tag{47}$$

And the orthogonal basis $\Phi_{0,m,n}(x,y)$, $\Psi^i_{0,m,n}(x,y)$ is like the following equation:

$$\varphi_{0,m,n}(x,y) = 2^{j/2}\varphi(2^j x - m, 2^j - n)$$

$$\psi_{j,m,n}^i(x,y) = 2^{j/2}\psi^i(2^j x - m, 2^j - n) \quad i=\{H,V,D\} \tag{48}$$

It is clear to get that the 2-dimensional signal can be decomposed to the smooth part and the high-frequency detail part. The detail part can be divided to horizontal, vertical and diagonal three parts. As for the smooth part ,which can be further analysis by the advanced wavelet decomposition to get the next smooth part and detail part , finally get the muliti-scale decomposition of the signal.

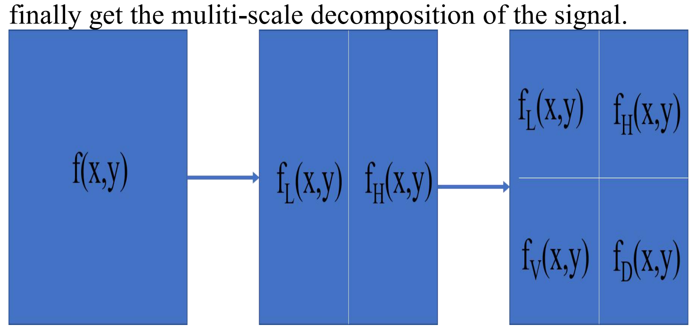

Fig.13. The 1-dimensional wavelet decomposition

According to the multi-scale analysis process of two-dimensional discrete wavelet, the original image decomposition process to be processed is summarized as follows: get the f_l(x y) and f_h(x y) are obtained by horizontal sampling and filtering of LR images f(x, y). We do the same thing vertically, and finally we get the f_L(x y), f_H(x y), f_V(x y), f_D(x y). f_L(x y) is the smooth part of the original image, f_H(x y), f_V(x y), f_D(x y) present the vertical , horizontal and diagonal part of the detail image. The process is following.

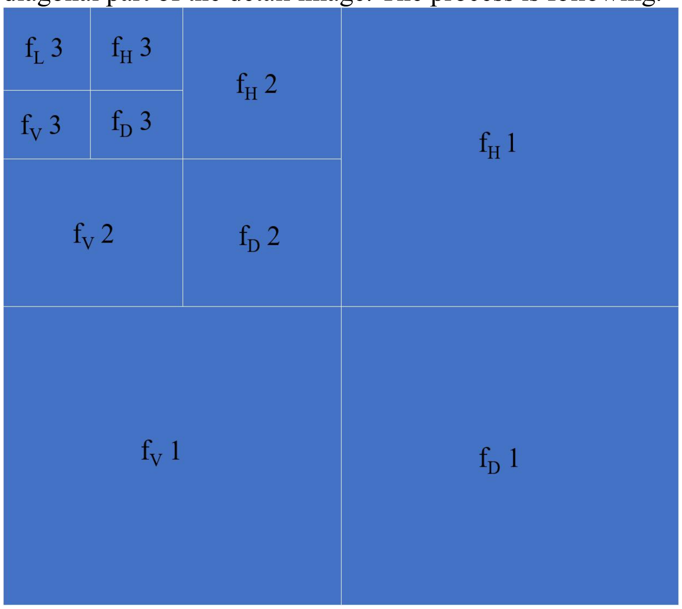

Fig.14. The 3-dimensional wavelet decomposition

D. the dictionary learning algorithm

In the former section, we have already discussed the dictionary learning and the feature extraction applied in the image reconstruct process, after we get the all the features of the low-resolution and high-resolution image. We compact them and train the dictionary together to get the low-resolution dictionary D_H and D_L. before we trained the dictionary, we tried take a random dictionary which is not based on the certain date base. The reconstruct image is quite low-behavior and full of noise.so it is quite necessary for us to take the dictionary to improve the output the sparse representation[4]. The result based on random dictionary is following:

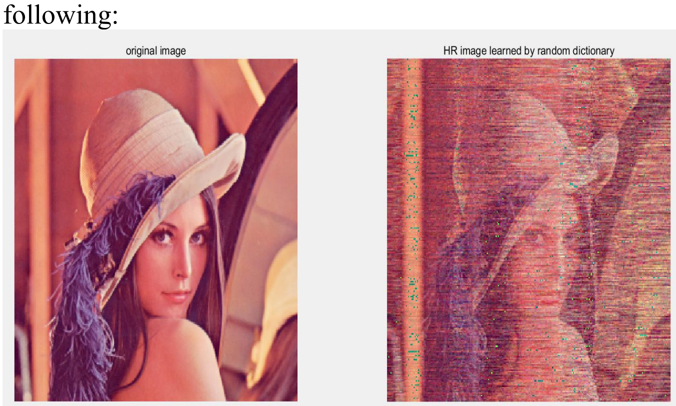

Fig.15. the reconstruct image based on an random dictionary

1. the dictionary learning model

For the training set X = {x_i},base on the image's sparse representatio=n model . the dictionary generated question can be described as:

$$D=\underset{D,A}{\operatorname{argmin}}|X-DA|_2^2+\lambda|A|_1 \quad \text{s.t. } \|d_j\|^2_2\leq 1 \;\; j=1,2,\ldots,k \quad (49)$$

In the formula, A = {α_i} is the sparse representation coefficients matrix, α_i is column vector. L1 norm can increase the sparse ability. d_j is the atoms of dictionary D. it is clear to get the dictionary generating problem is not a convex optimization problem. So, we can assume one of the variable fixed and the objective function can be regarded as a convex optimization problem. The specific process is following:

Input: the training data set

Output: learned dictionary D

1) Initialize the dictionary: D = D0, we generate the dictionary based on an online dataset, which contains 69 high resolution pictures, and by bicubic interpolation, we get the high resolution and low-resolution pairs. generating the feature map the former feature extraction methods. And we divide the images into 5*5 image patches and compact the high-resolution and low-resolution image pairs by normalization.
2) Fix the dictionary D, to get the sparse representation coefficients α_i of x_i :

$$\alpha_i=\underset{a_i}{\operatorname{argmin}}||x_i-D\alpha_i||_2^2 \;\; s.t.||a_i||_0\leqslant s \quad (50)$$

The sparse representation coefficients can be solved by the MP and OMP algorithm.

3) Fix the sparse representation coefficients α_i and update the dictionary D

$$D=\underset{D}{\operatorname{argmin}}||X-DA||_2^2 \;\; s.t. \;\; ||d_j||_2^2 \leqslant 1, \;\; j=1,2,...,K \quad (51)$$

As for dictionary solved algorithm, we take the KSVD introduced in the next section, we can update the dictionary by one column each time [3] .

The above dictionary learning method is single, to get the high-resolution dictionary D_H and low-resolution dictionary D_L, in the former section, we have already setup the whole technique key point is to make the high resolution image and low resolution image get the same sparse representation coefficients, beside the normalization to compact the high and low resolution image, we also used the joint dictionary method. The method's description is following:

As the high resolution image, low resolution image set as X={x_i},Y={y_i}. we can get those two dictionaries individually by the following equations:

$$D_H=\underset{D_H,\;A}{\operatorname{argmin}}||X-D_HA||_2^2 \;+\; \lambda||A||_1$$
$$D_L=\underset{D_L,\;A}{\operatorname{argmin}}||X-D_LA||_2^2 \;+\; \lambda||A||_1 \quad (52)$$

We can combine these two objective functions and get a objective function:

$$\{D_H,D_L\}=\underset{D_H,\;D_L,A}{\operatorname{argmin}}\frac{1}{N}||X-D_HA||_2^2 \;+\; \frac{1}{M}||Y-D_LA||_2^2+\lambda\left(\frac{1}{N}+\frac{1}{M}\right)||A||_1 \quad (53)$$

In this section, N and M represent the HR and LR image patch length. We take the 1/N and 1/M to normalize these two functions, we can get the simpler version:

$$D_C = \underset{D_C,\ A}{\operatorname{argmin}} ||X_C - D_C A||_2^2 \ + \ \lambda ||A||_1 \tag{54}$$

Where

$$X_c = \left[\frac{1}{\sqrt{N}}X,\ \ \frac{1}{\sqrt{M}}Y\right]^T, D_c = \left[\frac{1}{\sqrt{N}}D_H,\ \ \frac{1}{\sqrt{M}}D_L\right]^T . \tag{55}$$

The process of train the joint dictionary

2. KSVD dictionary learning algorithm

KSVD dictionary learning algorithm was first introduced by Aharon and Elad, which is a training algorithm in the original dictionary. On the basis of this, the dictionary updating method is improved, and the dictionary updating method is realized by K times of SVD decomposition [4]. This method can reduce the number of atoms in the dictionary and have the complete information of the dictionary. The algorithm is processed as following:

The training image data set X = {x_i}, i= 1,2,..,P, xi is the n-dimensional vector transformed from image patch (assuming the upscale factor is m, if the low resolution image patch is p*p ,so n = p*p. and high resolution image patch is mp*mp , so n = mp*mp). For random x_i, given the initial dictionary D = D{dj},j= 1,2,…,k, atom dj is the n-dimensional vector , k is the atom number. So the sparse coefficients can be solved the former OMP or ADMM algorithm:

$$\min ||X - DA||_2^2 = \sum_{i=1}^{P} ||x_i - D\alpha_i||_2^2 \ s.t. \ ||\alpha_i||_0 \leqslant s, \ i \ = \ 1,2,...,P \tag{56}$$

The solution algorithm have already mentioned in the former section and we can take the KSVD algorithm to update the dictionary. The algorithm is analyzed as following. Assuming α_T^i is the i column of the A, so the objective function can be transform into :

$$\begin{aligned} ||X - DA||_2^2 &= \ ||X - \sum_{j=1}^{K} d_j \alpha_T^j||_2^2 \\ &= ||\left(X - \sum_{j=1, j\neq k}^{K} d_j \alpha_T^j\right) - d_k \alpha_T^k||_2^2 \\ &= ||E_k - d_k \alpha_T^k||_2^2 \end{aligned} \tag{57}$$

In this formula, E_k is the error without atom dk. Define ω_k is the atom index set of dk used by {x_i} .

$$\omega_k \ = \ \{i | 1 \leqslant i \leqslant P, \alpha_T^k(i) \neq 0\} \tag{58}$$

Choosing every row in the Ek which can fit the ω_k ‘s condition and make them E_K^R , so we can use the Singular value decomposition (SVD) on the E_K^R , and $E_K^R = U\Delta V^T$, in the singular value decomposition U and V are the unitary matrix, Δ is the rectangular matrix with non-negative real numbers on the diagonal.

Task: Find the best dictionary to represent the data samples y as sparse compositions, by solving

$$\min_{D,X} \{||Y - DX||_F^2\} \ s.t. \ \forall i \ ||x_i||_0 \leqslant T_0 \tag{59}$$

Initialization: Set the dictionary matrix $D^{(0)} \in R^{n*k}$ with l^2 normalized columns;

Set J=1.

Repeat until convergence (stopping rule):

Sparse Coding Stage: Use any pursuit algorithm to compute the representation vectors xi for each example yi, by approximating the solution of:

$$i \ = \ 1,2,...,N, \ \min\{|y_i - Dx_i|_2^2\} \ s.t. \ ||x_i||_0 \ \leqslant \ T_0 \tag{60}$$

Codebook Update Stage: For each column k=1,2,...,K in $D^{(J-1)}$,

update it by

- Define the group of examples that use this atom

$$\omega_k \ = \ \{i | 1 \leqslant i \leqslant N, x_T^k(i) \neq 0\} \tag{61}$$

- Compute the overall representation error matrix, Ek, by

$$E_k \ = \ Y \ - \sum_{j \neq k} d_j X_T^j \tag{62}$$

-Restrict Ek by choosing only the columns corresponding to ω_k and obtain E_k^R

-Apply SVD decomposition $E_k^R \ = U\Delta V^T$. Choose the updated dictionary column dk to be the first column of U. Update the coefficient vector X_k^R to be the first column of V multiplied by Δ（1，1）

set J=J+1

IV. RESULTS

This chapter we realize above algorithm based on MATLAB and verify the performance of the algorithm in many aspects. The reconstruction effect of the proposed algorithm is compared with other image SR algorithm,including Bicubic interpolation algorithm and OMP algorithm. SR experiments were carried out with dictionaries with different atomic numbers and the differences were analyzed and summarized.

We find some images to build our Image Library, it include 69 high quality faces photos, and then, we use these photos to train our dictionary.We spend amount of time in finding image library and training our library.

Fig.16. Image Library.

In the following figure two methods are used to realize image super-resolution reconstruction.

PSNR	child	girl	old man	lena
Sparse representation based on OMP	33.5071	33.2244	32.9313	31.6493
Sparse representation based on ADMM	34.6031	34.0236	34.0476	33.3277

Table.2. The PSNR of different algorithm

Fig.17. The results of different algorithms

By comparing the table2, it can be found that ADMM algorithm is better than OMP algorithm in PSNR.

Fig.18. The results of different patch

Patch Size	t(s)	PSNR(dB)
3*3	4.036	33.91986
4*4	0.594	33.91955
5*5	82.454	34.66837
9*9	9.322	33.93614

Table.3. The PSNR of different patch

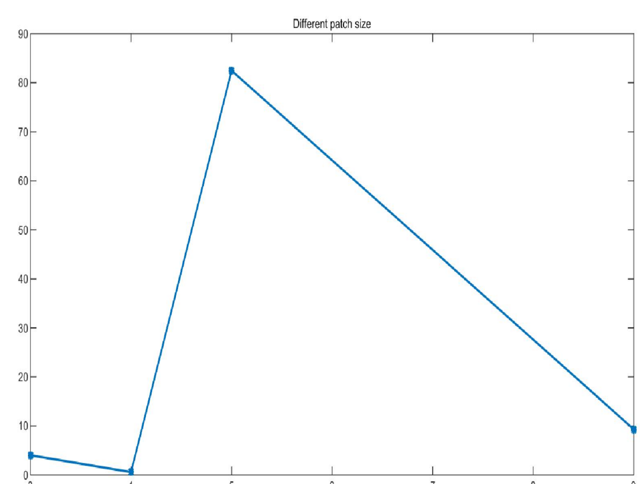

Fig.19. The computation time of different patch

By comparing the table3, it can be found that the maximum value of PSNR occurs at 5*5

Fig.20. The results of different dictionary size

Dictionary size	t(s)	PSNR(dB)
25*256	45.642	34.47038
25*512	58.397	34.46680
25*1024	82.457	34.66837
25*2048	263.562	34.68539

Table.4. The PSNR of different dictionary size.

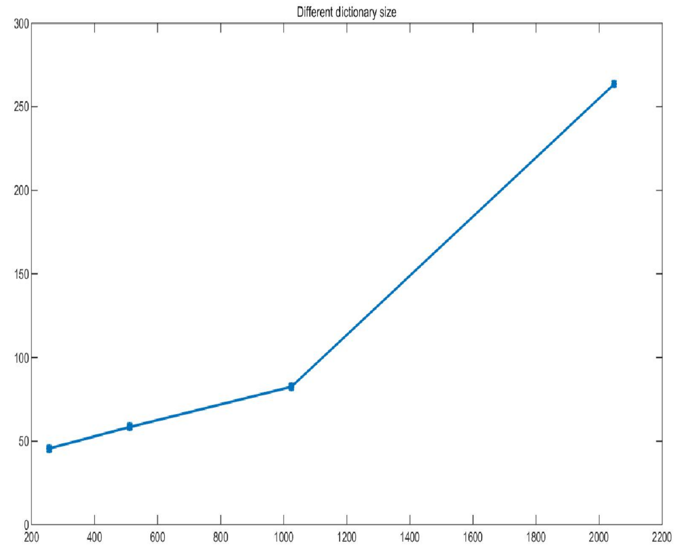

Fig.21. The computation time of different dictionary size.

Fig.22. The results of different overlap size

Overlap size	t(s)	PSNR(dB)
2	8.818	34.399903
3	19.029	34.635321
4	0.594	33.91955
5	0.514	33.919904

Table.5. The PSNR and computation time of different overlap size

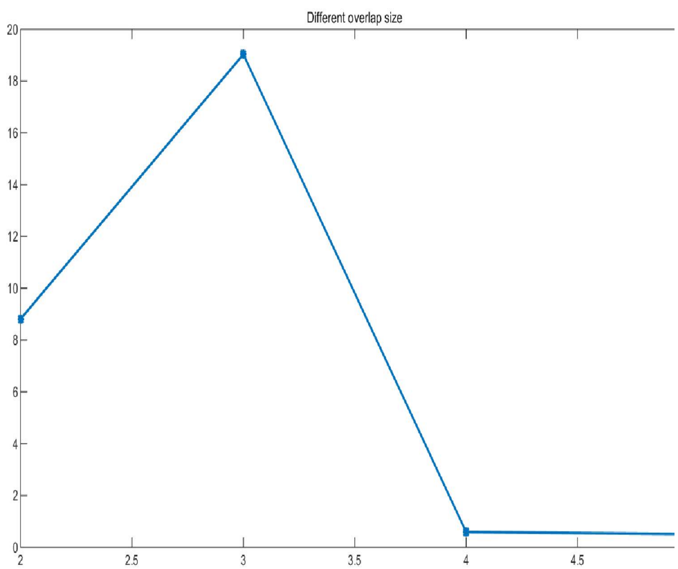

Fig.23. The computation time of different overlap size.

As we can see from the results,The size of the block is not linear with the computation time, but the computation time increases with the increase of the dictionary, PSNR increase as the dictionary increase.

The main content of this chapter is the experiment and verify the performance of the algorithm.The verification mainly includes the performance analysis of the algorithm in this paper compared with other algorithms, the influence of the algorithm in this paper on the number of dictionary atoms, and the anti-noise performance of the algorithm in this paper. The analysis of the experimental results shows that the algorithm proposed in this paper improves the performance of the algorithm on the basis of the traditional algorithm based on sparse representation.In contrast, the reconstruction effect of the algorithm is less affected by the dictionary, indicating that it has certain stability.And through the experiment of LR image contrast with noise, the proposed algorithm also has better anti-noise performance.

As for the Patch size's result, we can figure out that the best patch size is the 5*5, because the feature extraction step will extraction the useful information of the image. but after the patch step, there are many patch is all-zero , and will increase the complex of the algorithm. So those all-zeros part have been disgard by us. And by different patch-size trail we figure out that 5*5 maybe the best suitable patch-szie under this condition.

V. Conclusion

As we mentioned before, sparse representation is one of the most popular optimization applications before the deep learning. The intrinsic procedure of how it works is quite clear and many scientists devoted themselves into its development. Whereas there still many unknown mysterious in the neuron network. So, it' s quite important for us to take this into application. We choose the super – resolution as our topic. Based on the major technique idea: to make the high-resolution and low-resolution image have the same sparse representation coefficients. So, we get the sparse representation coefficients form the low-resolution image and dictionary and use the sparse representation coefficients with high-resolution dictionary to get the high-resolution image. In this paper, we compare two sparse coefficient solving algorithm (OMP and ADMM algorithm) based on their process efficiency and the reconstructed results. The results turn out to ADMM can have a better optimization result. As for the dictionary learning is based on the KSVD algorithm. We also set the dictionary size as the variable to check how it affect our optimization results. We also compare our super-resolution algorithm with the traditional interpolation method. Consequently, our super-resolution can have better behavior.

VI. Aknowledgement

According to this semester's learning, we do expand our knowldge about the image process.we learned a lot thing about the digital image process .what most impressive in this semester is the final project. We three do tons of the research, collect the inforamation , program the code, debug the algorithm. It's not only one simple project, it can be one of the begin for our postgraduate life. we have the whole research procedure for this project . we meet many quetsion time by time, but we tried and discussed until the better results is out. Even though For a one tiny small normalization operation, we can have a all-night discussion. It's a wonderful experience in the begin of our postgraduate life. And for about half semester's hard working, we have a basic understanding about the sparse representation adn super-resolution,which in our opinion will be useful in our future research life.what's more,Thank you professor Wang and TA Wang Sai's sincere help.

REFERENCES

[1] [J. Sun, N. N. Zheng, H. Tao, and H. Shum, “Image hallucination with primal sketch priors,” in Proc. IEEE Conf. Comput. Vis. Patter]

[2] [H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in Proc. IEEE Conf. Comput. Vis. Pattern Class., 2004, vol. 1, pp. 275–282.]

[3] Yang, J. , Wright, J. , Huang, T. S. , & Ma, Y. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 19(11), 2861-2873.. (2010).

[4] Aharon, M. , Elad, M. , & Bruckstein, A. . k -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Press.(2006).

[5] Yang, J. .. Sparse modeling of high-dimensional data for learning and vision. (2011)

[6]S. Baker, T. Kanade, "Limits on super-resolution and how to break them", IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 9, pp. 1167-1183, Sep. 2002.

[7] 肖杰, 卢涛, 万永静, et al. 基于稀疏表达的图像超分辨率算法实现[J]. 计算机与数字工程, 2014.

[8] 马强. 基于稀疏表示的单幅彩色图像超分辨率重建研究[D].兰州理工大学,2019.

[9] Irani M,Peleg S,Motion Analysis for Image Enhancement: Resolution, Occlusion, andTransparency[J]. Journal of Visual Communication & Image Representation, 24(4):324-335,2019

[10]Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, usingconvex projections.[J]. Journal of the Optical Society of America A Optics & ImageScience,6(11):1715-1726 ,1989

[11] E. J. Candes, "Compressive Sampling", Proc. Int. Congr. Mathetmaticians, vol. 3, pp. 1433-1452, 2006.

[12] D. L. Donoho, "Compressed sensing", IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[13] H. Rauhut, K. Schnass, P. Vandergheynst, "Compressed sensing and redundant dictionaries", IEEE Trans. Inf. Theory, vol. 54, no. 5, May 2008.