Single Image Super-Resolution Based on Sparse Representation Via Structurally Directional Dictionaries in Wavelet Domain

Elham Abar

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Electrical and Electronic Engineering

Eastern Mediterranean University
January 2014
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

_____
Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

_____
Prof. Dr. Aykut Hocanın
Chair, Department of Electrical
and Electronic Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Electrical and Electronic Engineering.

_____
Prof. Dr. Hüseyin Özkaramanlı
Supervisor

Examining Committee
_____

1. Prof. Dr. Hüseyin Özkaramanlı   _____

2. Assoc. Prof. Dr. Erhan A. İnce   _____

3. Assoc. Prof. Dr. Hasan Demirel   _____

ABSTRACT

The main aim of super-resolution is reconstructing a higher resolution image by combining a set of lower resolution images (classical approach) or from a single image. In this thesis a framework on Single Image Super-Resolution (SISR) based on sparse coding over structurally directional dictionaries has been presented. The motivation behind structurally directional dictionaries is the fact that images contain directional structures such as edges. A dictionary assigned to a specific direction promises to offer a better representation for directional features. The approach that leads to directional dictionaries is classifying the training data into directional classes.

The design of the directional dictionaries is done in the wavelet domain; by reason of the fundamental theory about the wavelet which it categorizes the data into directional subbands. Here the training set is formed by the patches of the first and second level Discrete Wavelet Transform (DWT) subbands of natural images which are prepared for training process, in two levels of resolution (high and low resolution respectively).

To categorize the training set into directions, several predetermined categories of patches called templates respect to the desired directions as a criterion of comparison are generated. Several classification techniques are studied and the best one which is based on templates matching is chosen. After classifying every high and low resolution training patches into their corresponding categories, K-SVD algorithm is

used to learn several pairs of high and low resolution dictionaries over the categorized data.

On the other hand for reconstructing the high resolution patch given the low resolution one, in order to find sparse coefficients Orthogonal Matching Pursuit (OMP) algorithm is applied to low resolution dictionaries. After choosing the most proper low resolution dictionary among all the presented dictionaries based on the least square error between the main LR patch and reconstructed LR patches, the corresponding high resolution dictionary and same sparse coefficient is used to reconstruct the high resolution patch and finally acquire the super-resolved image.

Quantitative results obtained from simulations has showed that the proposed algorithm indicates an average PSNR raise of 0.2 dB over the Kodak set compared with the images yield by other state of the art methods. Also the qualitative result is shown that the proposed algorithm plays a greater role in reconstructing images with more directional structures, or directional parts of natural images.

Keywords: Single Image Super Resolution; Sparse Representation; dictionary learning; directional dictionaries; classifying patches

ÖZ

Görüntü süper çözünürlüğünün asıl amacı, daha düşük kalitede elde edilmiş görüntüleri klasik yöntem kullanılarak ya da daha yaygın olarak kullanılan tek bir görüntüyü kendisiyle birleştirme yöntemi kullanılarak daha iyi çözünülüğe sahip olan görüntüler oluşturmaktır. Bu tezde, yapısal yönlü sözlüklerin dağınık kodlaması temeline dayanan tek görüntü süper çözünürlük metodu çalışılmıştır. Yapısal yönlü sözlüklerin kullanılmasındaki motivasyon, görüntülerin kenarlarda olduğu gibi yönlü yapılar içermesine dayanmaktadır. Belirli bir yöne atanmış olan bir sözlük, yönlü özellikler için daha iyi bir gösterim sağlamaktadır. Yönlü sözlükler kullanılmasının amacı eğitme verilerinin yönlü sınıflara ayrılmasıdır.

Yönlü sözlüklerin tasarımı, verileri yönlü altbantlara ayıran temel dalga teorisi göz önünde bulundurularak dalga alanında yapılmıştır. Bu çalışmada, eğitme verileri iki çözünürlük seviyesinde (yüksek ve düsük çözünürlükler sırasında) olmak üzere eğitme süreci için hazırlanan doğal görüntülerin birinci ve ikinci seviye DWT altbantları yamalarıyla düzenlenmiştir.

Eğitme verilerini yönlere ayırmak için, istenilen yönlerin bir karşılaştırma kriteri olarak kullanıldığı şablon diye adlandırılan yamaların önceden tanımlanmış birçok kategorisi oluşturulmuştur. Bilim dünyasında çeşitli sınıflandırma teknikleri çalışılmaktadır ancak bu tezde sınıflandırma için en iyi metod olan şablon eşleştirme yöntemi kullanılmıştır. Her bir alçak ve yüksek çözünürlüklü eğitme yamaları ilgili

kategorilere ayrıldıktan sonra, sınıflandırılmış veriler arasından yüksek ve alçak çözünürlüklü sözlük çiftlerini öğrenmek için K-SVD algoritması kullanılmıştır.

Öte yandan, verilen alçak çözünürlüklü yamadan yüksek çözünürlüklü olanı oluşturabilmek ve dağınıklık katsayılarını hesaplayabilmek için, düşük çözünürlüklü sözlüklere OMP algoritması uygulanmıştır. Asıl ve oluşturulmuş LR yamaları arasındaki en küçük kareler hatasına dayanan mevcut sözlüklerin arasından en uygun alçak çözünürlüklü olanlarını seçtikten sonra, yüksek çözünürlüklü yamayı oluşturmak ve nihayetinde süper çözünümlü görüntüyü elde etmek için ilgili yüksek çözünürlüklü sözlük ve aynı dağınıklık katsayısı kullanılmıştır.

Benzetim sonunda elde edilen sayısal sonuçlar, Kodak setiyle diğer bilinen metodların verdiği görüntü sonuçları karşılaştırıldığında, önerilen algoritmanın Kodak setinde ortalama PSNR'de 0.2 dB'lik bir iyilişme yarattığını göstermektedir. Ayrıca, nitel sonuçlar da önerilen algoritmanın cok yönlü yapılarla veya doğal görüntülerin yönlü kısımlarıyla görüntü oluşturmada önemli bir rol oynadığını göstermektedir.

Anahtar Kelimeler: Tek Görüntü Super Çözünürlüğü, Dağınık Gösterim, Sözlük Öğrenimi, Yönlü Sözlükler, Yama Sınıflandırması

Dedicated to

My wonderful parents

For nurturing me with love and affection

And

Supporting me all the way

ACKNOWLEDGMENTS

I wish to thank my supervisor Prof. Dr. Hüseyin Özkaramanlı, for his countless hours of encouraging, reading and most of all his patience through the entire process.

Besides my supervisor, I would like to thank Assoc. Prof. Dr. Erhan A. İnce and Associate Prof. Dr. Hasan Demirel for agreeing to serve on my committee.

Special thanks goes to all the faculty members of the department of Electrical and Electronic Engineering, and specially the chairman, Prof. Dr. Aykut Hocanın.

Finally I would like to express my appreciation to my parents and my brother who have always given me support and encouragement.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

C_n	Directional categories
D	Dictionary matrix
D_L	Low resolution dictionary matrix
D_H	High resolution dictionary matrix
d_n	Main directions for $n \in [1,8]$
g	Wavelet filter
B	Blurring filter
h	Scaling filter
i_p	Intermediate directions $p \in [1,9]$
L	Number of nonzero entries in sparse coefficients
Q	Scale-up factor
S	Down sampling factor
X_P^H	Categorized high resolution patches
X	High resolution image
x	High resolution patch
$\hat{x}$	Reconstructed high resolution patch
Y_P^H	Categorized low resolution patches
Y	Low resolution image
y	Low resolution patch
$\hat{y}$	Reconstructed low resolution patch
v_i	i^{th} atom of the dictionary matrix

w_h	Horizontal subband
w_v	Vertical subband
w_d	Diagonal subband
α_L	Low resolution sparse coefficient
α_H	High resolution sparse coefficient
α^k	Sparse coefficient of a patch located in k
ϵ	Amount of error
BP	Basis pursuit
CS	Compressive sensing
DWT	Discrete Wavelet Transform
FOCUSS	Focal Underdetermined System Solver
HR	High resolution
K-SVD	An algorithm for learning dictionary
LR	Low resolution
MP	Matching pursuit
OMP	Orthogonal matching pursuit
PSNR	Peak signal-to-noise ratio
MSE	Mean square error
SSIM	Similarity index measurement
SISR	Single image super-resolution

Chapter 1

INTRODUCTION

1.1 Sparse Representation

Sparse representation refers to representing the signal using linear combination of small parts of primary signal named atoms. Atoms are chosen from a matrix which has more columns than rows. Such a representation system is called an over complete dictionary $D \in R^{m \times n}$ $(m < n)$. If X is formed by signal vectors x_i as $X = [x_1, x_2, \ldots, x_j]$ then representation of the signal X is achieved as a linear combination of atoms from the dictionary.

$$X = DA \tag{1.1}$$

Where $A = [\alpha_1, \alpha_2, \ldots, \alpha_j]$ denotes the matrix of representation coefficients vectors of the signal X which can be either accurate or approximate, $X \cong DA$. Under the above condition the number of equations is more than the number of unknowns so it has infinitely many solutions. In other words this system is an *underdetermined* system. The approximation $\|x_i - D\alpha_i\|_p < \epsilon$ should satisfy with typical $\ell^p -$ norms $= 1, 2$ and ∞.

$$\|x_i\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \tag{1.1}$$

$\|.\|_1$, $\|.\|_2$ will be used for l^1 norm and l^2 norm respectively. The l^1 norm refers to sum of the absolute values of the entries in columns and l^2 norm is Euclidean length or energy of the vector.

The assumption of D is a full rank matrix in 1.1 guarantees to have at least one solution, which means columns of D should be linearly independent. If X be in the span of matrix D this equation has many solutions.

Researchers in this field concentrate on two main problems:

1) Dictionary integration methods which refers to designing or learning dictionary atoms. This problem specifies convenient dictionaries.

   Several algorithms have been proposed in this field such as: K-SVD [1], Recursive Least Squares [2] and Online Dictionary learning [3] [4].

2) Performing sparse disintegration and obtaining sparse coefficients (α).

   The most important aim of solving this equation is finding a proper solution among all the possible answers which is the sparsest one, in other words the solution with least number of nonzero entries. The desired sparsest solution has fewest nonzero entries in the vector α which $\| . \|_0$ refers to that.

$$\|\alpha\|_0 \ll L \quad for \quad \alpha \in R^n \tag{1.3}$$

When the equation 1.3 be established α is sparse.

For obtaining the sparsest solution one may require to solve the optimization problem presented in 1.4.

$$\min \|\alpha\|_0 \quad s.t. (subject\ to) \quad \|X - D\alpha\|_2 < \epsilon \tag{1.4}$$

This issue deals with the pursuit problems. Given a signal X and dictionary D, finding the best linear combination of D to reconstruct X is desired. To reach to this aim several algorithms has been proposed such as: FOCUSS [5] [6] [7] [8] [9] [10] (Focal Underdetermined System Solver), the basis pursuit (BP) [11], matching pursuit (MP) [12], orthogonal matching pursuit (OMP) [13] [14].

In this thesis K-SVD algorithm is used for learning dictionaries from part 1 and OMP for obtaining sparse coefficients from the second part and it has discussed more in detail next sections.

1.2 K-SVD: Dictionary Learning Algorithm

The objective of the K-SVD algorithm is to train a dictionary D to sparsely represent the data $\{x_i\}_{i=1}^{P}$ which means solving a given sparsity problem

$$\min_{D,\alpha}\|X - DA\|_F^2 \quad s.t. \quad \|\alpha\|_0 \leq L \tag{1.5}$$

$\|A\|_F$ refers to Frobenius norm (or Euclidean norm) of matrix $A \in R^{m \times n}$, which is defined by

$$\|A\|_F = (\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2)^{1/2} \tag{1.6}$$

K-SVD algorithm minimizes 1.5 iteratively. To start this algorithm, first stage is sparse coding. Dictionary D has to be fixed then initialized coefficients matrices can be found by using any one of the pursuit algorithms mentioned in previous section. The second stage is to search for a better dictionary. In this process all columns are updated one by one in each iteration. To summarize K-SVD algorithm, it fixes every columns of D except one of them, v_i , then finds a new column and the corresponding coefficient which reduces the mean square error.

We shall now perform the following steps for K-SVD:

- Let $\alpha = [\delta_1, ..., \delta_k]^T$ where $\delta_i{}^T$ $(i = 1, ..., k)$ are the rows of α

- We can write (1.7) in the following manner:

$$\min_{D,\alpha}\|E_j - v_j\delta_j{}^T\|_F^2 \quad s.t. \quad \|\alpha\|_0 \leq L \tag{1.7}$$

In this description, $\delta_j{}^T$ stands for the jth row from α. The target is to update v_j and $\delta_j{}^T$ referring to the term

$$E_j = X - \sum_{i \neq j} v_i \delta_i{}^T$$

(1.8)

as a known pre-computed error matrix. After this part every condition are ready to updating the dictionary.

- Define the group of examples that use the atom v_j (those where $\delta_j(i)$ is nonzero) :

$$w_j = \{p | 1 \leq p \leq P, \delta_j{}^T(p) \neq 0\}, 1 \leq j \leq k$$

- Let Ω_j be a $P \times |w_j|$ matrix with ones on the $(w_j(i), i)$ entries and zero elsewhere then (1.9) can be written as: $\min_{D,\alpha} \left\| E_j \Omega_j - v_j \delta_j{}^T \Omega_j \right\|_F^2$.

- Let $U\Lambda V^T$ be the SVD (singular value decomposition) of $E_j{}^R = E_j \Omega_j$ and define $\delta_{j,R}{}^T = \delta_j{}^T \Omega_j$.

  To obtain the largest singular value σ_1 of $E_j{}^R$ and the corresponding left and right singular vectors u_1, v_1

$$\min_{D,\alpha} \left\| U\Lambda V^T - D_j \delta_{j,R}{}^T \right\|_F^2$$

(1.9)

- Solutions are: $v_j = u_1$ and $\delta_{j,R} = \sigma_1 v_1$.

1.3 Orthogonal Matching Pursuit (OMP)

OMP is a repetitious greedy algorithm [14] [15] [16] [17] [18] like MP algorithm. This algorithm selects an atom of the known dictionary which has the best correlation with the residual signal in each stage. For the following problem :

$$\min_{\alpha \in R^n} \|X - D\alpha\|_2^2 \quad s.t. \quad \|\alpha\|_0 \leq L$$

(1.10)

where dictionary $D \in R^{m \times n}$ have normalized atoms $\{v_i\}_{i=1}^m$. Initialized parameters are $\alpha = 0$, $r = X$, then we perform the following steps:

- Select the dictionary atom with the maximum correlation with the residual.

$$l = arg \max_{i=1,..,P} |v_i{}^T r|$$

- Update active set $\Omega = \Omega \cup l$

- MP $\quad \alpha[\hat{l}] = \alpha[\hat{l}] + v_{\hat{l}}{}^T r$

  OMP $\quad \alpha[\hat{l}] = (v_{\hat{l}}{}^T v_{\hat{l}})^{-1} v_{\hat{l}}{}^T r$

- $r = r - (v_{\hat{l}}{}^T r)v_{\hat{l}} = r - <v_{\hat{l}}, r> v_{\hat{l}}$

The above routine will continue for L iterations where L is a predetermined value so the process will stop after selecting fixed number of atoms.

To understand the content better, it has provided a figure about updating the residual in MP [19].
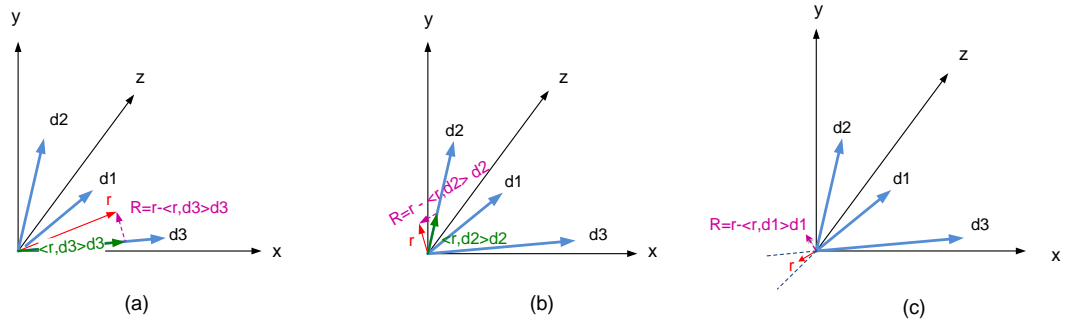
Figure 1.1: How to update residual in MP algorithm

In Figure 1.1 (a) vector r and $d3$ have the maximum correlation among other vectors with each other. So $r - <r, d_3> d_3 \rightarrow r$ and this routine is continued in Figure 1.1 (b) and (c) for updating r. But unlike MP, an atom can be selected only once with OMP.

1.4 Sparsity Based Applications

Sparsity-based algorithm, recently have been proven to be a successful method in many states of the art signal processing, machine learning as well as computer vision tasks. In image processing we can refer to super resolution [20] [21] [22], inpainting

[23] [24] and denoising [25] [26] [27] issues. Sparse representation also plays a main role in object recognition such as face recognition [28] and text recognition [29].

In this thesis we will consider super resolution problem and the assumption through this job is natural image patches can represented well using a sparse linear combination of a proper dictionary atoms.

1.5 Super Resolution

Enhancing the resolution of images is an active area in recent years. Different kind of methods such as frequency domain [30] [31] [32], Bayesian [33], example-based [34] [35] [36], set theoretic [37] [38], and interpolation, have been applied to super resolution techniques. Also in this field there has been a growing interest in the study of image resolution enhancement in the wavelet domain and many new algorithms have been proposed in this area [39]. The most considerable aspects of wavelet based super resolving is the capability in modeling the regularity of natural images [40].

Another category of super resolution is machine learning based, which intent to learn the occurrence of LR and HR image patches simultaneously.

A significant achievement in this area is the patch based algorithm presented to reduce the computational cost [20] [21]. According to equation 1.1, patches of images can be represented by sparse coding over a dictionary. In [20] and [21] this method is applied into HR and LR image patches, and LR sparse coefficients are used to generate the HR patches with respect to their own dictionary. R. Zeyde.et.al. [22] modified the mentioned method by changing some parts such as using a different training procedure for learning pair of dictionaries and changing the scale-up phase.

In this thesis a strategy similar to R. Zeyde.et.al [22] is followed. However instead of learning one single condensed and big dictionary for the whole training set, we propose to learn structurally directional dictionaries over clustered training datasets in the wavelet domain. The motivation behind learning several structurally directional dictionaries is based on two ideas: 1) A plurality of sparse representation is better than the sparsest one alone [41] and 2) The signal can be represented better after clustering. So by learning several dictionaries over clustered data, it will have the potential for better representation of structurally directional features.

Instead of the spatial domain the SISR problem is formulated in wavelet domain. By incorporating wavelet domain benefits such as compactness and directionality with advantages of sparse coding, a progress in super resolution enhancement is predicted. All patches of 2 level wavelet subbands are classified into several categories according to their directions. K-SVD algorithm is applied to low resolution categorized patches to obtain low resolution dictionaries and sparse coefficients, then with regard to the assumption that the LR and HR sparse coefficients are approximately equal [21], obtained sparse coefficients are used for learning the corresponding high resolution dictionaries. Low resolution dictionaries are used to reconstruct LR patches. By observing least square error between reconstructed LR patches and the main LR patch, the most proper LR dictionary is chosen then the corresponding HR dictionary and sparse coefficient are used for reconstructing the super resolved patch.

By the increase of 1.36 dB and 0.28 dB PSNR than bicubic interpolation and the state of the art algorithms respectively over benchmark images, the effectiveness of

the proposed algorithm which is mainly based on sparsely representing patches by structurally directional dictionaries is proved.

1.6 Thesis Description

In this thesis we mainly focus on structurally directional dictionaries. Due to this, it is necessary to classify the training dataset into several directional categories. To reach to this aim, several pre-determined directional templates are generated to recognize the direction of the patches according to the similarity of patches and templates.

The contents of this thesis have been organized as follows. We will talk about image super resolution from sparsity and two fundamental approaches in this area, Yima.Et.al [21] and R.Zeyde.Et.al [22] more in details in chapter 2. Chapter 3 is included the proposed method mainly about how to classify training set into directions and learning directional dictionaries. Chapter 4 provides details about simulation and comparing results with state of the art algorithms and chapter 5 appertains to conclusion of the thesis and future work.

Chapter 2

SUPER RESOLUTION FROM SPARSITY

We mentioned about various super resolution algorithms from the literature in part 1.5. We shall talk about one of the modified patch based algorithms suggested by R. Zeyde.et al [22], which the main idea had been proposed by J.Yang.et.al [21] .

2.1 Relationship Between High and Low Resolution Images

The patch based single image super resolution inspired from new approaches in compressive sensing (CS), which suggests that HR signals can be reconstructed from their low-dimensional projections. Although the super-resolution problem is an ill-posed problem, and for a HR image so many LR images can be formed, sparse coding is   effective in adjusting the problem.

We start by explaining about the problem and relation between high and low resolution images. The given low resolution image Y, can be established from high resolution image X. to avoid complexity of mismatching sizes of LR and HR images and boundary issues the low resolution image is scaled –up by factor Q which is an interpolation factor (bicubic interpolation is chosen in [22]) .

$$Y \approx Q.S.B.X \approx \psi.X \qquad (2.1)$$

where $S \in R^{n \times m}$ is the down sampling factor as a projection matrix, *B* represents a known blurring filter and to simplify, Ψ characterizing the multiplication of *Q, S* and *H*. The aim of this algorithm is estimating high resolution patches of *X*, and to

reach to this aim operating on the corresponding low resolution image patches which they are extracted from low resolution image is needed. Let $x_k = R_k X$, k refers to the location of central pixel of high resolution image patch, x, in high resolution image X and R_k is an operator for extracting x_k.

2.2 Equality of Sparse Coefficients in HR and LR Patches

As we explained in chapter 1 each patch in an image can be represented sparsely by a proper dictionary:

$$x_k \approx D^H \alpha_H^k \tag{2.2}$$

$$y_k \approx D^L \alpha_L^k \tag{2.3}$$

x_k and y_k are the corresponding HR and LR patches which we call them patches and features briefly in this thesis and their central pixels are located in k. equation 2.1 can be written for HR and LR patches of the image:

$$y \approx \psi x \approx \psi D^H \propto_H \tag{2.4}$$

It has assumed in [42], operator ψ not only relates patches and features but also can relate HR and LR dictionaries. So the equation 2.4 can be written as:

$$y \approx D^L \propto_H \tag{2.5}$$

From 2.3 and 2.5 the assumption of sparse coefficients equality of features and patches is obvious.

$$\propto_L \approx \propto_H \tag{2.6}$$

It means each super resolved image patch is reconstructed by multiplying the corresponding LR sparse coefficient and HR dictionary. This is the main idea behind the single image super resolution algorithm proposed by J. Yang.et.al in [20] [21].

2.3 State of The Art Sparse Representation

Here in this part the proposed algorithm by R. Zeyde.et.al [22] is studied. For

representing image textures rather than absolute intensity, for each patch the mean value of pixels is subtracted. The mean value of each Patch is predicted from its LR version which is the feature. This algorithm consists of two main phases: training phase and reconstruction phase.

2.3.1 Training Phase

Before extracting patches and features there need to apply the high-pass filter to low and high resolution images in order to remove low frequencies and extract local features respectively. After the steps mentioned above patches and features are extracted according to the corresponding location in HR and LR images to form high and low resolution dataset $\{x_k, y_k\}$. However features fit in the data set after interpolating by factor Q. Consider that the location of an intended Patch in a high resolution image is the same as the location of its feature in the low resolution image.

Next step is learning dictionaries. The starting point is features. In order to learn dictionaries K-SVD algorithm is used. A dictionary D_L is learned for features according to solve this optimization problem:

$$D^L, \{\alpha^k\} = \arg \min_{D^L, \{\alpha^k\}} \sum_k \|y_k - D^L \alpha^k\|^2 \quad s.t. \|\alpha^k\|_0 \leq L \qquad (2.7)$$

where α^k refers to vectors of sparse representation coefficients.

As we mentioned before, sparse representation coefficients can be equal for HR and LR patches so to learn high resolution dictionaries the corresponding low resolution sparse coefficient is used in each patch, with respect to the optimization problem below:

$$D^H = \arg \min_{D^H} \sum_k \|x_k - D^H \alpha^k\|^2 \qquad (2.8)$$

The training phase is finished by learning high and low resolution dictionaries.

11

2.3.2 Reconstruction Phase

First step of reconstruction part is almost the same as training part. This phase has tested for the images which they have down sampled and blurred by the same factors used in training phase. Then before extracting patches, the same high pass filter is applied to extract features. According to the location of k, patches are extracted. Using OMP algorithm sparse coefficient for each low resolution patch is found and then the corresponding high resolution patch is reconstructed by multiplying HR dictionary and the corresponding sparse coefficient which obtained in low resolution part by OMP. The super resolved patches $\{\hat{x}_k\}_{k=1}^{T}$ (T shows the number of patches) concatenated to form the high resolution image $\hat{X}$ by solving an optimization problem:

$$\hat{X} = \arg\min \sum_k \left\| R_k(\hat{X} - Y) - \hat{x}_k \right\|_2^2 \tag{2.9}$$

This optimization problem means the patch extracted by R_k from the difference image $\hat{X} - Y$ should be close enough to approximated patches. This problem has a closed-form Least-Squares solution, given by

$$\hat{X} = Y + \left[\sum_k R_k^{\ T} R_k \right]^{-1} \sum_k R_k^{\ T} \hat{x}_k \tag{2.10}$$

Solving this equation is equivalent to putting $\hat{x}_k$ in their proper locations, averaging in overlap regions and adding Y to get the final super resolved image $\hat{X}^0$. In this thesis the approach of R. Zeyde.et.al [22] is adopted, however there are fundamental differences. First difference is the fact that we work in the wavelet domain instead of the spatial domain. The most important motivation for this is the fact that wavelet transform itself sparsifies the signal of interest, furthermore it limits directionality.

Due to the fact that wavelet subbands are high-passed filtered version of input signal, the need for feature extraction is eliminated.

Instead of learning single big dictionary, several directional dictionaries all learned by classifying data in wavelet subbands into directional clusters and employing K-SVD algorithm for learning.

In the testing side for each feature the most proper dictionary is found and the sparse coefficient is obtained by OMP algorithm. The representation coefficient is multiplied by the corresponding high resolution dictionary. We will talk about it precisely in next chapter.

Chapter 3

PROPOSED SINGLE IMAGE SUPER-RESOLUTION ALGORITHM VIA STRACTURALLY DIRECTIONAL DICTIONARIES

The proposed algorithm has two main parts: training part and testing part. Both training and testing parts are done in the wavelet domain. Every operation in training phase leads to learn several HR and LR pairs of structurally directional dictionaries. Patches in training set are classified into several categories according to their directions to form a training set for learning the dictionaries. Then these dictionaries are used to reconstruct HR patches from LR patches of a test image to form the super resolved image at the end. This chapter starts with a short preface about wavelet decompositions of an image.

3.1 Wavelet Decomposition

In this thesis two level wavelet decomposition stages constitute HR and LR sets. Figure 3.1 shows two level wavelet decomposition for a 2D signal. g and h denote wavelet and scaling filters. $\downarrow 2$ shows the down sampling factor by 2. The outcomes of first level wavelet decomposition are $LH1, HL1$ and $HH1$ namely Low-High, High-Low and High-High which refer to horizontal, vertical and diagonal details of the input image respectively. Sizes of the subbands after one level wavelet decomposition are half of the size of the input image while Sizes of $LH2, L2$ and $HH2$ after two level wavelet decompositions are one-fourth of the input image.
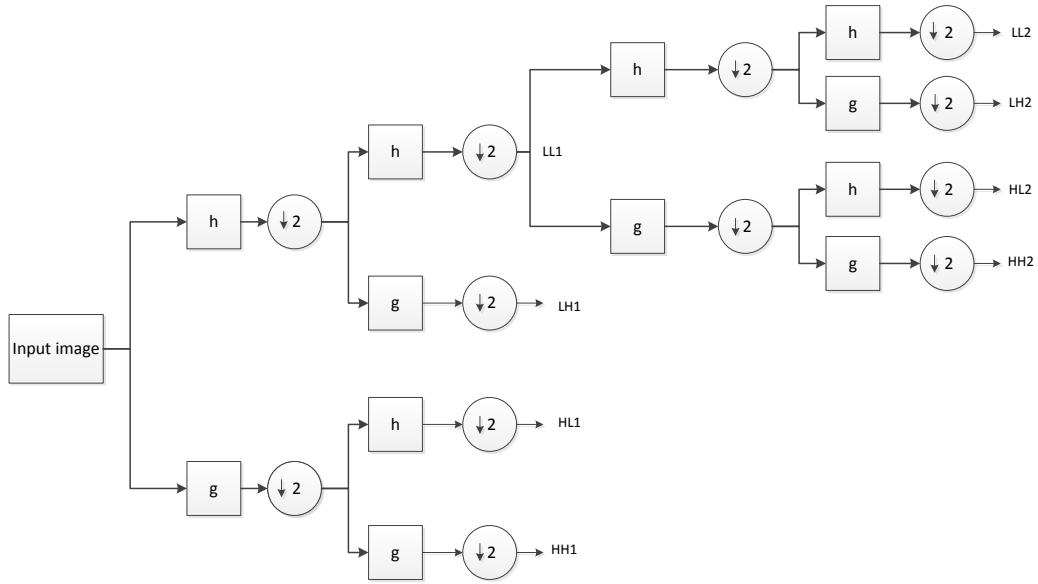
Figure 3.1: Block diagram of two-level forward wavelet transform

The subbands yielded from first level wavelet decomposition is formed the HR dataset. To avoid complexity created by mismatching the sizes of first and second level subbands, LH2, HL2 and HH2 organize the required set of low resolution patches after one level wavelet interpolation which we will talk about it in the following sections. Also Figure 3.2 shows a block diagram of one-level inverse wavelet transform.
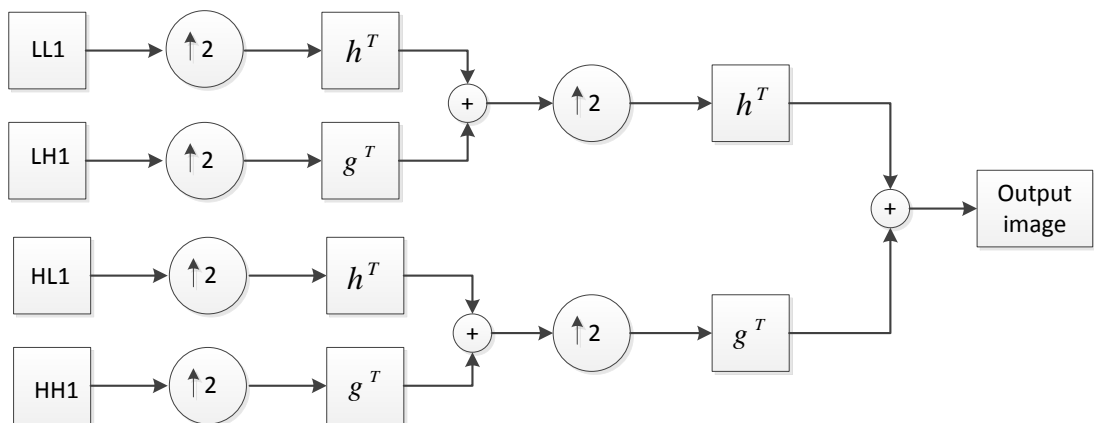

Figure 3.2: Block diagrams of one-level inverse wavelet transform

g^T and h^T denote the transpose of wavelet and scaling filters which had been used in forward wavelet transform in Figure 3.1.

3.2 Training Phase

We shall start this phase by preparing HR and LR patches in the wavelet domain as we mentioned in Figure 3.1.

3.2.1 Preparing LR and HR Patches for Training

To form HR and LR data sets we used two level wavelet decomposition. First level subband images form HR dataset and the second level is used to form LR data set. In order to avoid mismatch in dimensions between LR and HR datasets, due to difference between sizes in first and second wavelet subbands images, an inverse wavelet transform is taken from each of the second level subbands by appropriately padding corresponding subband with zeroes as shown in Figure 3.3. After one level wavelet decomposition, the wavelet subbands w_h^H, w_v^H, w_d^H are used to extract the high resolution patches. After the 2nd level wavelet transform wavelet subbands w_h^L, w_v^L, w_d^L are interpolated via zero padded inverse wavelet transform to obtain w_h^M, w_v^M, w_d^M from which LR patches (features) are extracted. The superscript M in w_h^M, w_v^M, w_d^M refers to the medium level of resolution, which we call them mid-resolution.

After extracting HR and LR patches from the mentioned subbands, each of the patches and features should place into a category according to their direction. Eight directions are intended and several templates are generated for all the eight directions. Each patch will place into a category according to the best match of correlation with the corresponding template.
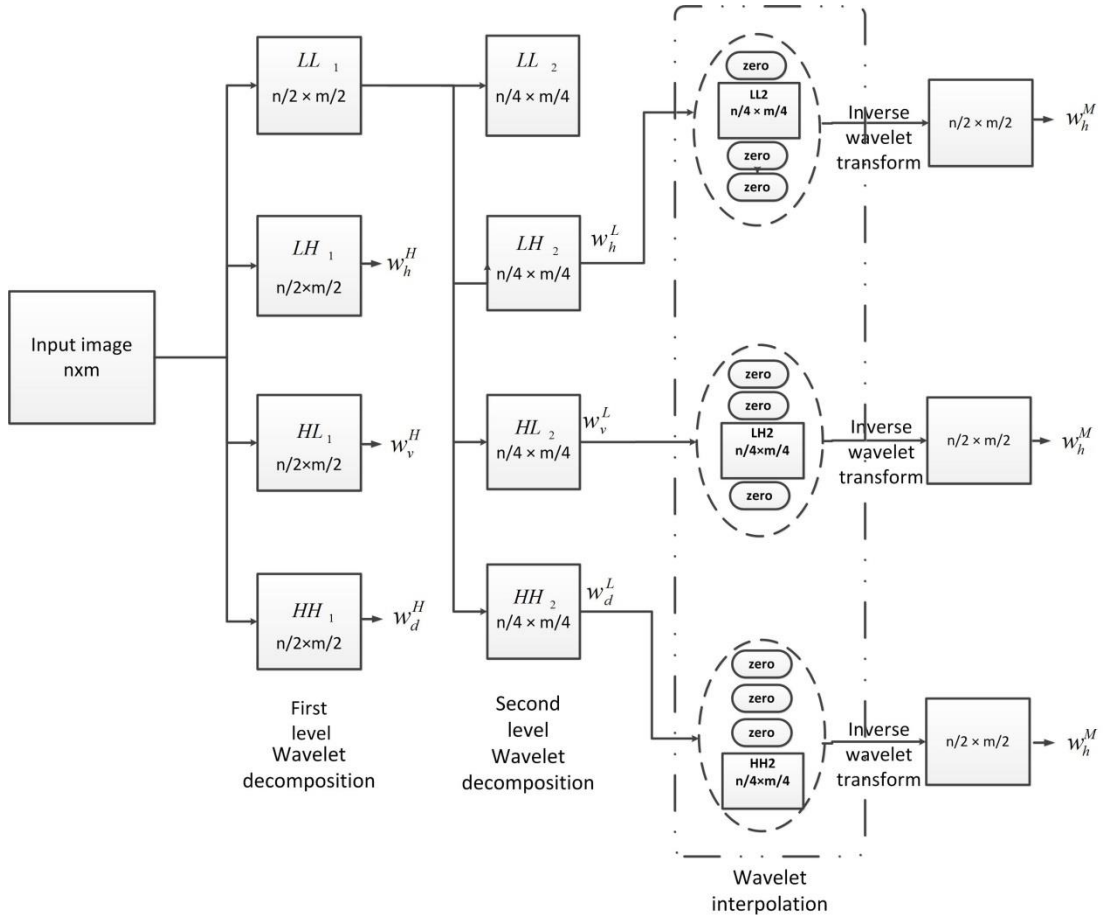
Figure 3.3: Preparing LR and HR training images using two-level wavelet decomposition.

3.2.2   Pre-Determined Templates

Eight categories of patches which are compatible with the corresponding template for each direction have intended and for better comparing, Size of each template is chosen equal to the selected patch size.

Figure 3.4 is showing directions of templates $\{d_1, d_2, \dots, d_8\}$. Directions $\{i_1, i_2, \dots, i_9\}$ are assumed as intermediate directions. The area between each two intermediate directions belongs to the corresponding main direction. For example if a patch is directional and its direction is in the area between i1 and i2 it will place in the category of d1. Every directional patches of HR and LR dataset is categorized

into these eight directions and will put into ninth category which is non-directional if there is no recognizable direction in it.
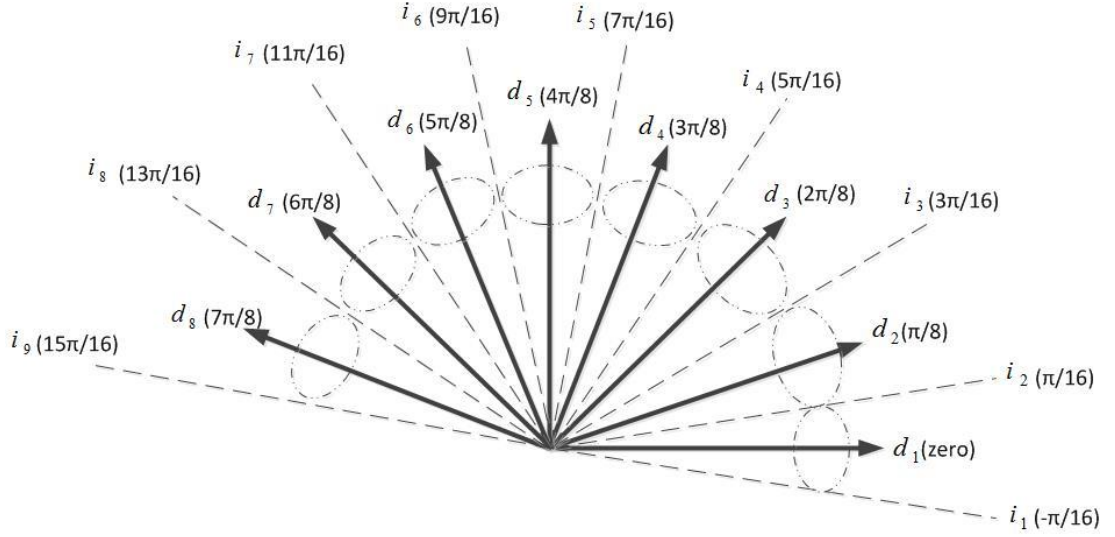

Figure 3.4: Directions of pre-determined templates

But how the templates are defined to cover every positions of a directional line in a patch? It has used Gabor filter to create Gabor patches.

3.2.2.1 Gabor Filter

Gabor filter has been used in many applications such as line and edge detection [43], enhancement [44], segmentation [45] and object detection [46]. In order to define directional templates we propose to use the Gabor filter as well. For 2D signal, it named complex Gabor function and it is calculated by multiplying sinusoidal wave and Gaussian kernel which is the case in our method. The real part of a Gabor filter is defined as a multiplication of complex sinusoidal plane and Gaussian kernel and can be written as [47] :

$$g(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \omega\right) \tag{3.1}$$

where

$$x' = (x - x_0) \cos \alpha + (y - y_0) \sin \alpha \qquad (3.2)$$

$$y' = -(x - x_0) \sin \alpha + (y - y_0) \cos \alpha \qquad (3.3)$$

(x_0, y_0) shows the center of the filter, γ, σ, λ refer to spatial aspect ratio, sigma of the Gaussian envelope and wavelength of sinusoidal factor respectively and ω is the phase offset [48].

By 2D Gabor function Gabor patches are generated as the pre-determined templates for each desired directions as shown in Figure 3.5.
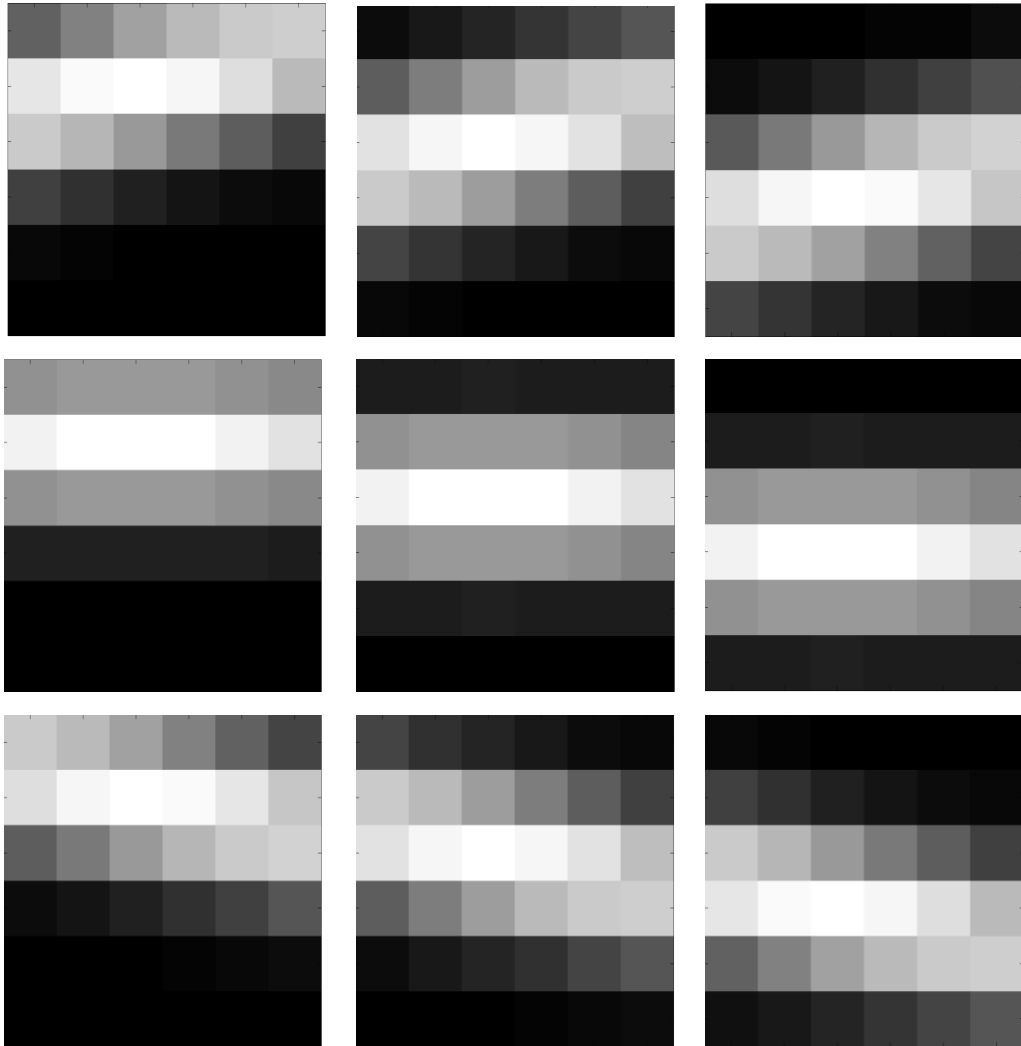
Figure 3.5: Pre-determined Gabor templates of size 6×6 for zero degree direction

It has prepared samples of templates for zero degree, 45 degree and 135 degree in Figure 3.5, Figure 3.6 and Figure 3.7 respectively. First and third row is covering intermediate directions with two level shifts from set $\{i_p\}$, where p refers to the number of direction, and it is obvious second row is shown exact zero degree from the set of main directions $\{d_n\}$ with two levels of shifts. For each category of directions nine sample patches are formed. It is clear that by increasing the selective size of high resolution patch the number of samples for each category will increase, because it needs more number of samples to cover all the shifts for each direction in a patch with bigger size.
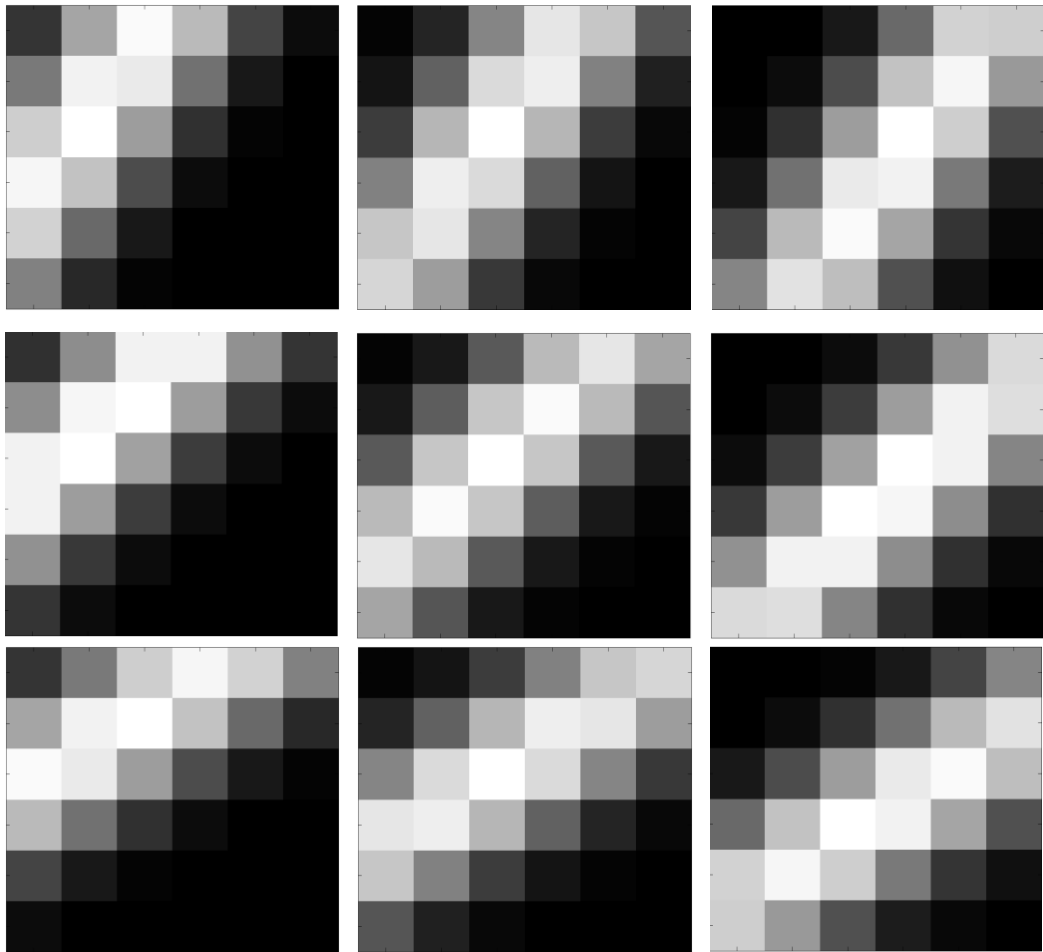
Figure 3.6: Pre-determined Gabor templates of size 6×6 for 45 degree direction
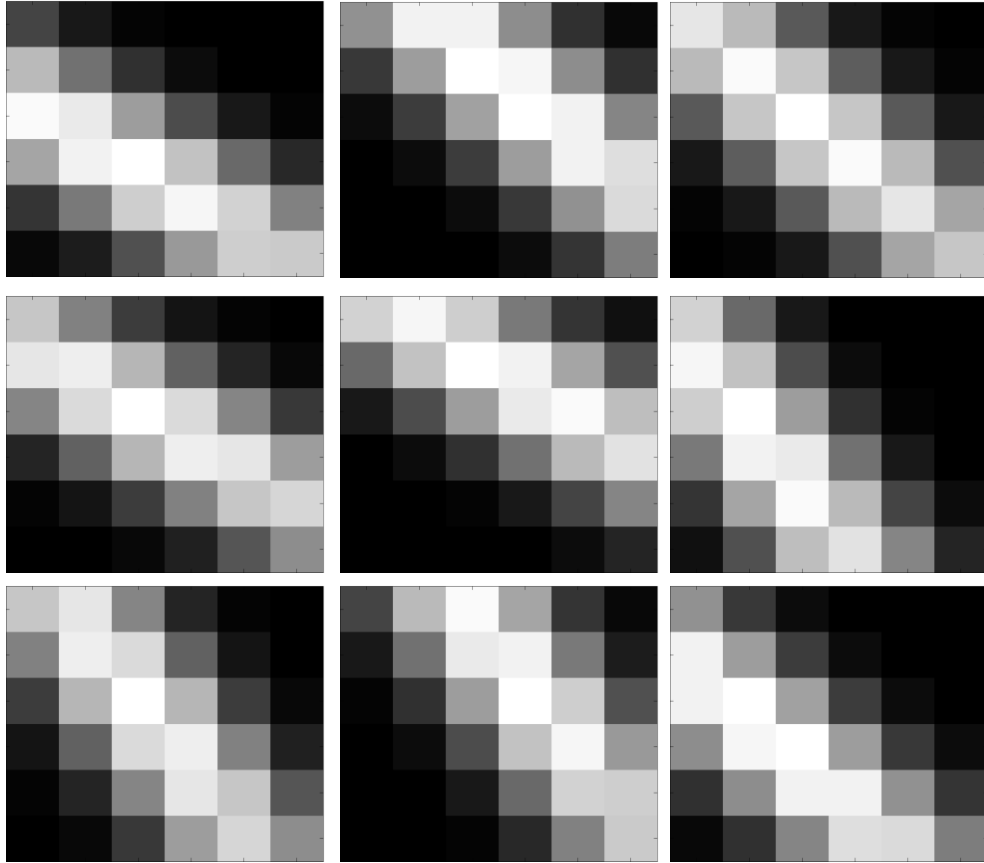
Figure 3.7: Pre-determined Gabor templates of size 6×6 for 135 degree direction

3.2.3 Classification

Next step is classifying HR and LR patches extracted from w_h^H, w_v^H, w_d^H and w_h^M, w_v^M, w_d^M respectively into several directional categories. If the direction of the patch is compatible with one of the pre-determined templates, the patch places into that category, but if the patch has no specific direction, it places into non-directional category.

The criterion of compatibility between the patch and templates is correlation match. To find the best match of correlation, the shape of patches and templates matrices are changed from 2D dimension to vector shape. For each high resolution patch the correlation coefficients of the desired patch with all the templates are computed and the maximum correlation coefficient is chosen. If the maximum correlation is greater

than a threshold (e.g. 0.4) the HR patch goes into the founded category. In addition to putting the HR patch in the best category, the corresponding LR patch will place in the corresponding LR category too. But if it is less than the threshold which means the system could not found a specific direction, both Patch and feature place into non-directional category. So, two sets of categorized data are formed for both levels of resolution. $\{X_1^H, X_2^H, \ldots, X_9^H\}$ contain HR patches and $\{Y_1^L, Y_2^L, \ldots, Y_9^L\}$ is formed by LR patches.

3.2.4   Learning LR and HR Dictionaries

To learn all the dictionaries, K-SVD algorithm is used. In order to promote the learning of the directional dictionaries the K-SVD algorithm is initialized with *dummy directional dictionaries*. To generate dummy dictionaries, the templates of the corresponding category are changed into vector shape and these vectors form the atoms of the dictionary for the corresponding category. But since the number of generated templates is not enough to generate the dictionary, linear combination of vector shaped templates are constituted the rest of atoms. Defining dictionaries based on the templates guarantees the desired directions for dictionaries.

It is clear that ninth dictionary witch is non directional dictionary has no initialization factor except defining the number of its atoms. Sizes of all dictionaries are chosen the same. Figure 3.8 shows the atoms of eight dummy dictionaries $\{d_0^1, d_0^2, \ldots, d_0^8\}$ which are perfectly directional. Dummy dictionaries are the starting point for learning the low resolution dictionaries.
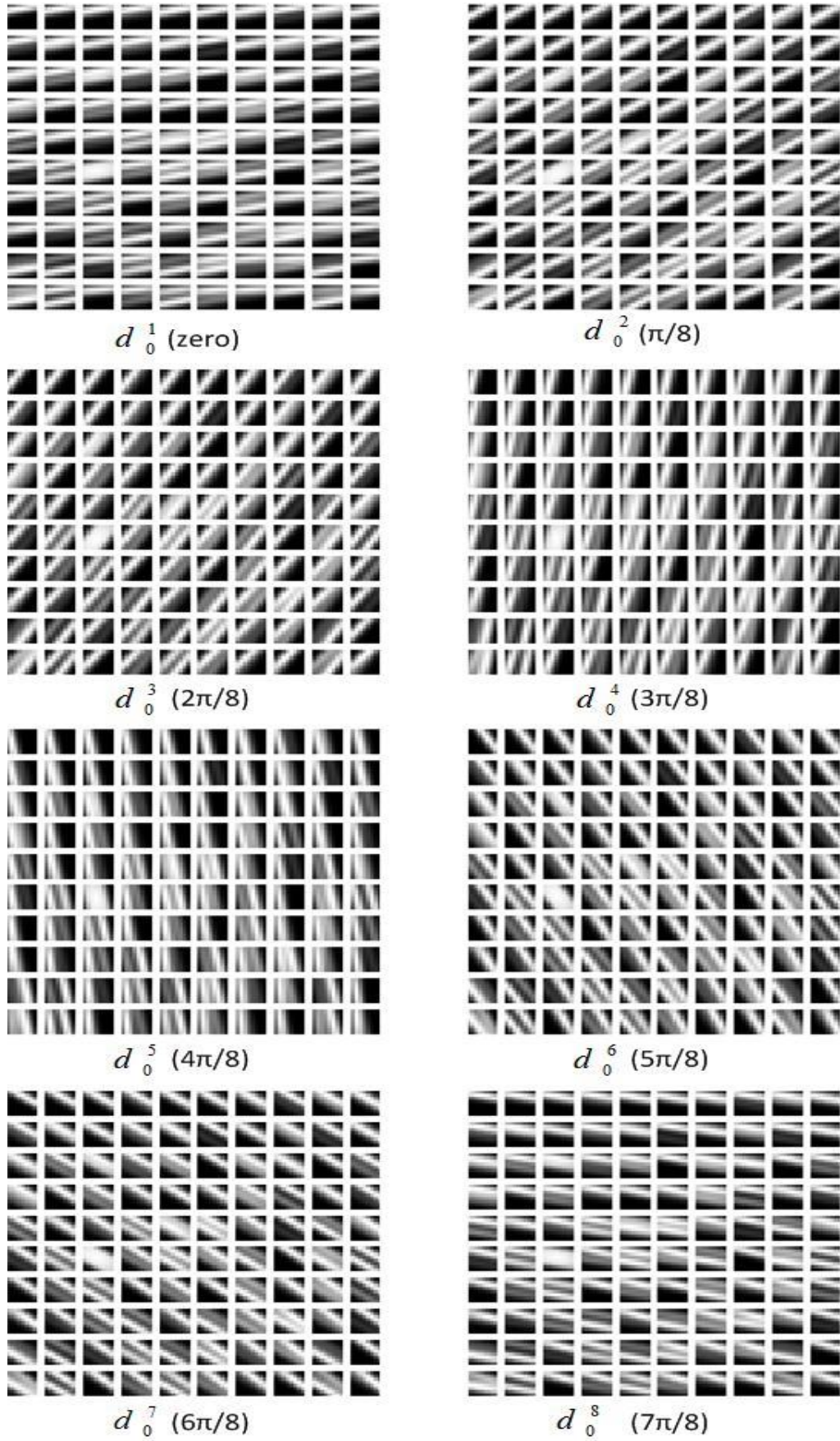
d_0^1 (zero)

d_0^2 (π/8)

d_0^3 (2π/8)

d_0^4 (3π/8)

d_0^5 (4π/8)

d_0^6 (5π/8)

d_0^7 (6π/8)

d_0^8 (7π/8)

Figure 3.8: Dummy dictionaries for eight directions

$$\arg \min_{D_i^L, \alpha_i^L} \left\| Y_i - D_i^L A_i^L \right\|_2^2 \quad s.t \quad \left\| \alpha_i^L(k) \right\|_0 < L \tag{3.4}$$

where Y_i and its subscript i refer to the categorized LR patches and the number of category respectively and α_i^L is the matrix of sparse coefficients.

$$A_i^L = \left[\; \alpha_{i,1}^L, \alpha_{i,2}^L, \ldots \right] \tag{3.5}$$

Nine low resolution dictionaries $\{D_1^L, D_2^L, \ldots, D_9^L\}$ are learned over LR dataset using K-SVD by initializing with the corresponding dummy dictionary.

After learning process for LR dictionaries, to learn HR dictionaries $\{D_1^H, D_2^H, \ldots, D_9^H\}$, the corresponding low resolution coefficients are used. Therefore to obtain high resolution dictionaries, Pseudo-inverse solution using the corresponding low resolution patch and the coefficient is needed. High resolution dictionaries are computable by multiplying $\alpha_H{}^\dagger$ and HR patches.

$$D_i^H \approx X_i (A_i^H)^\dagger \tag{3.6}$$

where X_i refers to categorized HR patches, A_i^H is the matrix contains sparse coefficient vectors $\left[\; \alpha_{i,1}^H, \alpha_{i,2}^H, \ldots \right]$ and $\alpha_H{}^\dagger$ is defined by:

$$(\alpha_{i,n}^H)^\dagger \approx (\alpha_{i,n}^L)^T (\alpha_{i,n}^L \cdot \alpha_{i,n}^L{}^T)^{-1} \tag{3.7}$$

where the superscript '†' ant 'T' denote the pseudo-inverse and transpose symbol respectively. We used the fact that $\alpha^H \cong \alpha^L$.

We finish the training phase by learning 9 pairs of high resolution and low resolution directional dictionaries. Figure 3.9 shows example patches of the designed dictionaries which are learned over horizontal (d1) and vertical (d5) training sets. As it was expectable the direction of each dictionary is visible in Figure 3.9.

|(a)|(b)|
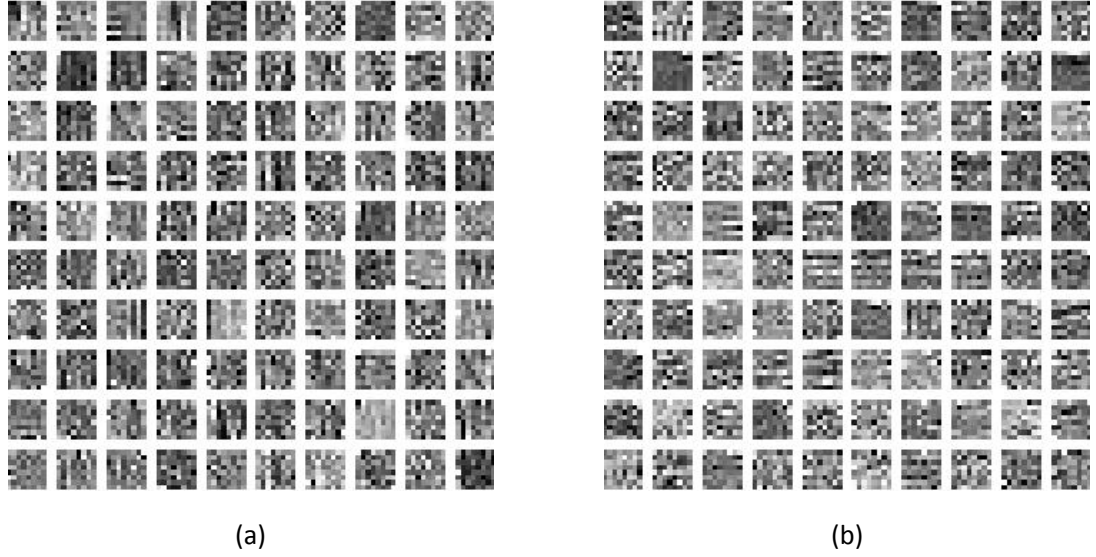
Figure 3.9: Directional HR dictionaries. (a) 90 degree (b) zero degree

$\{X_1^H, X_2^H, .., X_9^H\}$ and $\{Y_1^M, Y_2^M, ..., Y_9^M\}$ are the categorized high and low resolution patches respectively and $\{d_0^1, d_0^2, ..., d_0^8\}$ refer to the dummy dictionaries to define the start point in K-SVD. The obtained nine pairs of LR and HR dictionaries are used to reconstruct features and patches respectively in the test phase.

At the end of this part a summary of the training phase is presented in Figure 3.10.

3.3 Reconstruction Phase

In this phase given the low resolution image which is assumed to be the LL1 subband of the wavelet transform, what we want to do is estimating the corresponding wavelet subbands and get an inverse wavelet transform to reach to the super resolved image. For this purpose we follow a strategy similar to the one followed in the training phase.

Figure 3.10: Training phase

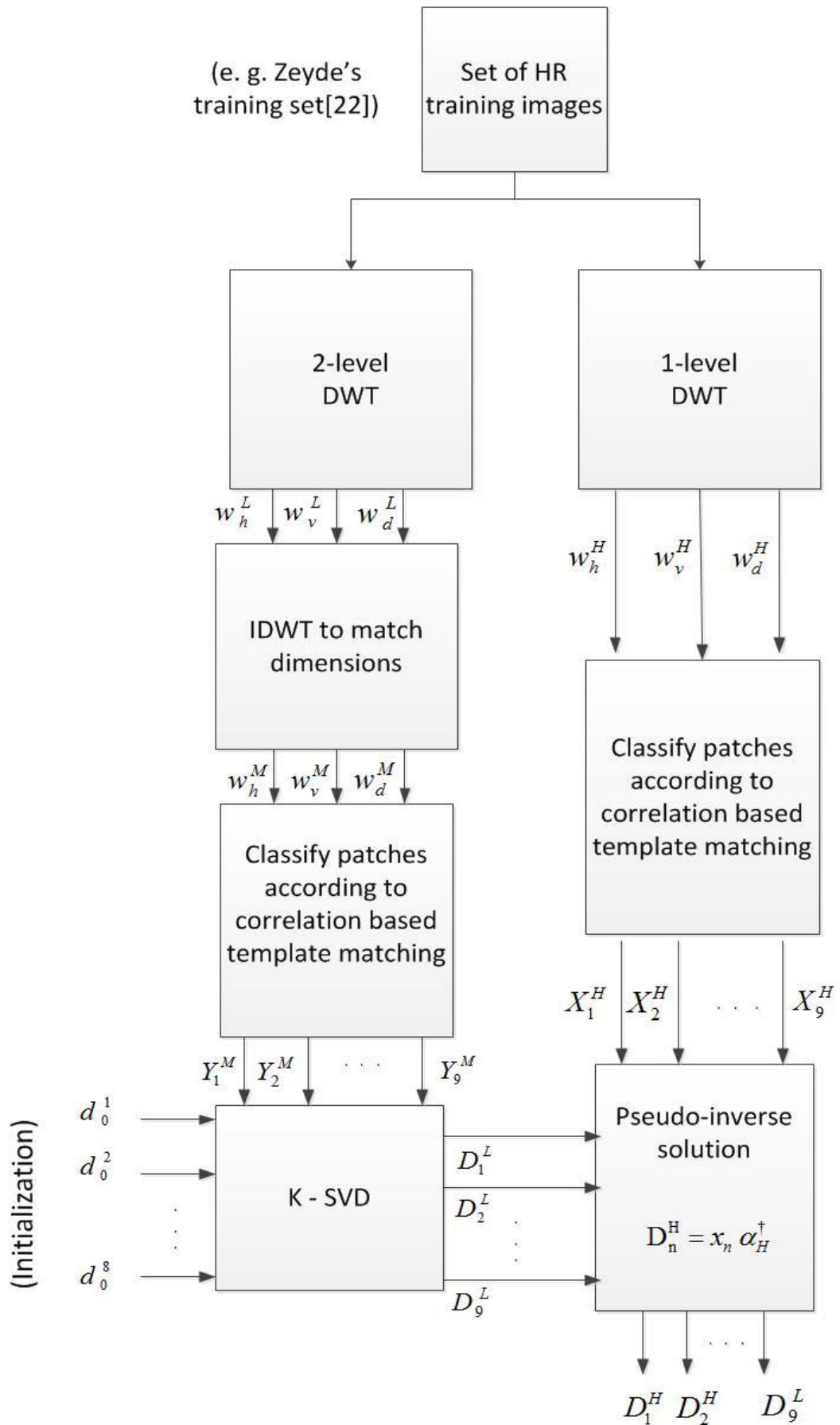By one level wavelet decomposition from the given low resolution image (LL1), a lower resolution image (LL2) and subband images (LH2, HL2 and HH2) are obtained. Similar to training phase to reach to the desired size, wavelet interpolation has taken from each subband using zero padding. The obtained bands are w_h^M, w_v^M and w_d^M. For each patch in these bands (y^k) OMP algorithm is applied nine times and each time for one of the learned low resolution dictionaries to obtain sparse coefficients. The multiplication of each dictionary by the corresponding sparse coefficient is used to reconstruct the patch, so for the patch in position k a set of reconstructed patches is formed $\{\widehat{y}_1^k, \widehat{y}_2^k, \dots, \widehat{y}_9^k\}$ via nine dictionaries

$$\widehat{y}_i^k \approx D_i^L \alpha_{i,k}^L \tag{3.8}$$

where i shows the number of dictionary which is used (from 1 to 9) and $\widehat{y}_i^k$ is the reconstructed feature in location k with i^{th} dictionary. Among all the reconstructed features choosing the most similar one to the main feature is needed. For this purpose the least square errors between the main feature and all nine reconstructed features are computed.

$$\min_i \left\| y^k - \widehat{y}_i^k \right\|_2^2 \tag{3.9}$$

The best answer shows the most proper dictionary by specifying i to estimating the corresponding high resolution patch in the same location but in w_h^H, w_v^H and w_d^H which is unlnown. Then nth dictionary from the set of high resolution dictionaries can be used for multiplying with the corresponding α to reconstruct the high resolution patch. According to the experimental results obtained from simulation, least square error for choosing the best low resolution dictionary is not a perfect criterion. As it has not found a better model selection yet, so the least square error

between the estimated features and the main feature is used to recognize which high

resolution dictionary should be chosen for reconstructing HR patch.

$$x^k \approx D_n^H \alpha_n \qquad (3.10)$$

Figure 3.11 shows a brief overview of the steps has done in the testing phase. By

reconstructing every high resolution patches, the HR subbands are made possible. As

it has shown in this figure the last step is one-level inverse wavelet transform to

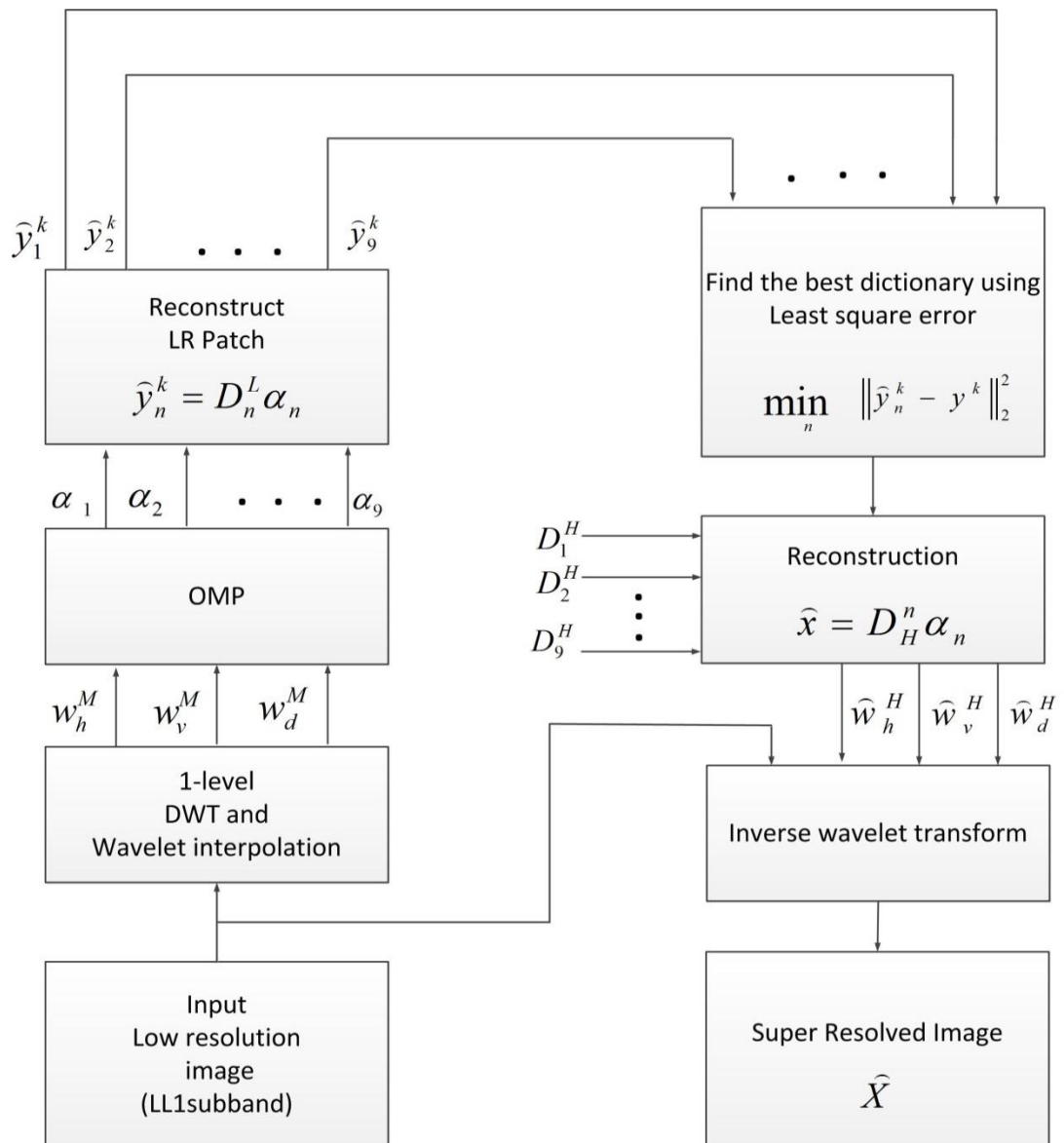reach to the super resolved image.

Figure 3.11: Reconstruction phase

Chapter 4

SIMULATION AND RESULTS

In this chapter, we demonstrate the proposed method results over the Kodak set and benchmark set using matlab platform, on Intel core i5cpu 2.66 GHz with 4GB of RAM. The comparison has done with bicubic interpolation and Zeyde.et.al algorithm [22] both qualitatively and quantitatively. To show the quantitative performance peak signal-to-noise ratio (PSNR) measurement is used.

Given the main image X and the reconstructed image $\hat{X}$ with dimension $m \times n$, the PSNR is:

$$PSNR(X,\hat{X}) = 20 log_{10} \frac{255}{\sqrt{MSE(X,\hat{X})}}$$ (4.1)

Where, MSE refers to the mean-square error between the main image and the reconstructed image.

$$MSE(X,\hat{X}) \approx \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (X_{ij}\hat{X}_{ij})^2$$ (4.2)

Besides PSNR, the Structural Similarity Index Measurement (SSIM) [49] is used, which is compatible with human visual perception.

First the SISR results based on PSNR and SSIM obtained by the proposed algorithm, bicubic interpolation and the algorithm proposed in [22]. Then the results are compared both visually and quantitatively. We then move on and examined the results in the situation if the high resolution patches be available to choose the most

29

correct high resolution dictionary among all the exist dictionaries. This test has done to understand the qualification of the model selection in the testing phase.

For extracting patches and features of the training images it has allowed in consecutive patches, one pixel overlap [1 1] in both directions. This amount of overlap has been chosen to increase the speed of training process and overlap [3 3] which is full overlap is selected in the testing phase to reconstruct the supper resolved image better. To compare fairly every simulation parameters such as images in training set, patch sizes, overlap factors, even the number of atoms of learned dictionaries (number of all atoms in proposed algorithm for 9 dictionaries is 990 and in [22] one dictionary with 1000 atoms) are chosen equal both in the proposed algorithm and [22].

The learning process of nine pairs of dictionaries each having 110 atoms has been performed using a total of 145,000 patches in proposed algorithm and 138,000 patches for R.Zeyde.et.al algorithm [22] for learning 1,000 dictionary atoms. The calculated time process for learning 9 dictionaries using K-SVD with 20 iterations and sparsity 3 is around 40 seconds while learning process for the dictionary in [22] with the same number of iteration and sparsity takes around 350 seconds.

Table 4.1 shows the number of patches which has been chosen by similarity with predetermined templates to place in the corresponding eight directional categories and the non-directional category if the similarity match be less than the pre-specified threshold. From the proposed results in Table 4.1 it can be seen around 40% of all

patches of natural images recognized as directional patches (with threshold 0.4 for similarity between patches and templates).

Table 4.1: Number and percentage of placed patches in each category

direction	Number of patches	percentage
d1	7292	5.0%
d2	7742	5.3%
d3	7307	5.0%
d4	7196	4.9%
d5	7875	5.4%
d6	7288	4.9%
d7	7136	4.8%
d8	6785	4.6%
directional	58621	40.2%
Non-direction	87188	59.7%

It is expected by improving the representing directional parts of an image using highly directional dictionaries the better performance be achievable compare to representing with one dictionary for all types of data.

As we mentioned before the reconstruction phase has been performed over 24 images of the kodak set and 10 well-known images of the benchmark set. For a better understanding of test images it has shown all 24 images of kodak set in Figure 4.1.

In reconstruction phase the most important problem is the model selection which should apply on the LR patch to choose the compatible direction, its worth mentioning that the picked direction in LR patch and the corresponding direction in HR patch are not always the same. In fact a better model selection algorithm can be developed, so that the performance can be improved. The selection of proper

dictionary should be done based on low resolution patches. This is because the LR patches are available in the reconstruction phase. In order to test and verify that the desihned dictionaries are proper a test is designed. For this purpuse we assume knowledge of HR patches in the reconstruction phase. Each HR patch is reconstructed with all the HR dictionaries and the dictionary that gives minimum mean square error is picked for reconstruction. this test is refered to *correct model selection* and performance results are given in 4th column of Table 4.2 and Table 4.3. The results indicate that with the correct model selection (selecting the most appropriate HR dictionary) the PSNR performance of the algorithm is on the average 0.41 dB better than R. Zeyde [22]. However the HR patches are not available in the reconstruction phase, therefor the model selection adopted here is minimizing the squared error in the representation of LR patches.

The super resolving process has done by factor 2, for instance size of input low resolution images of Kodak set is 384×256 so the super resolved images have size 768×512. Most of the input low resolution bench mark images have size 256×256 and the super-resolved images have size 512×512.

The summerize of the results are observable in Table 4.2 and Table 4.3. In these tables the proposed algorithm results are compared to bicubic interpolation and R. Zeyde.et. al algorithm. From the provided results in Table 4.1and Table 4.2 it is clear that the proposed algorithm is operating better than bicubic interpolation and R. Zeyde.et.al [22] algorithm with average PSNR increase of 0.17dB over Kodak set and 0.28dB over benchmark set, Which means it is better to divide one condense

dictionary into several directional and non-directional dictionaries and reconstruct each patch according to its direction.
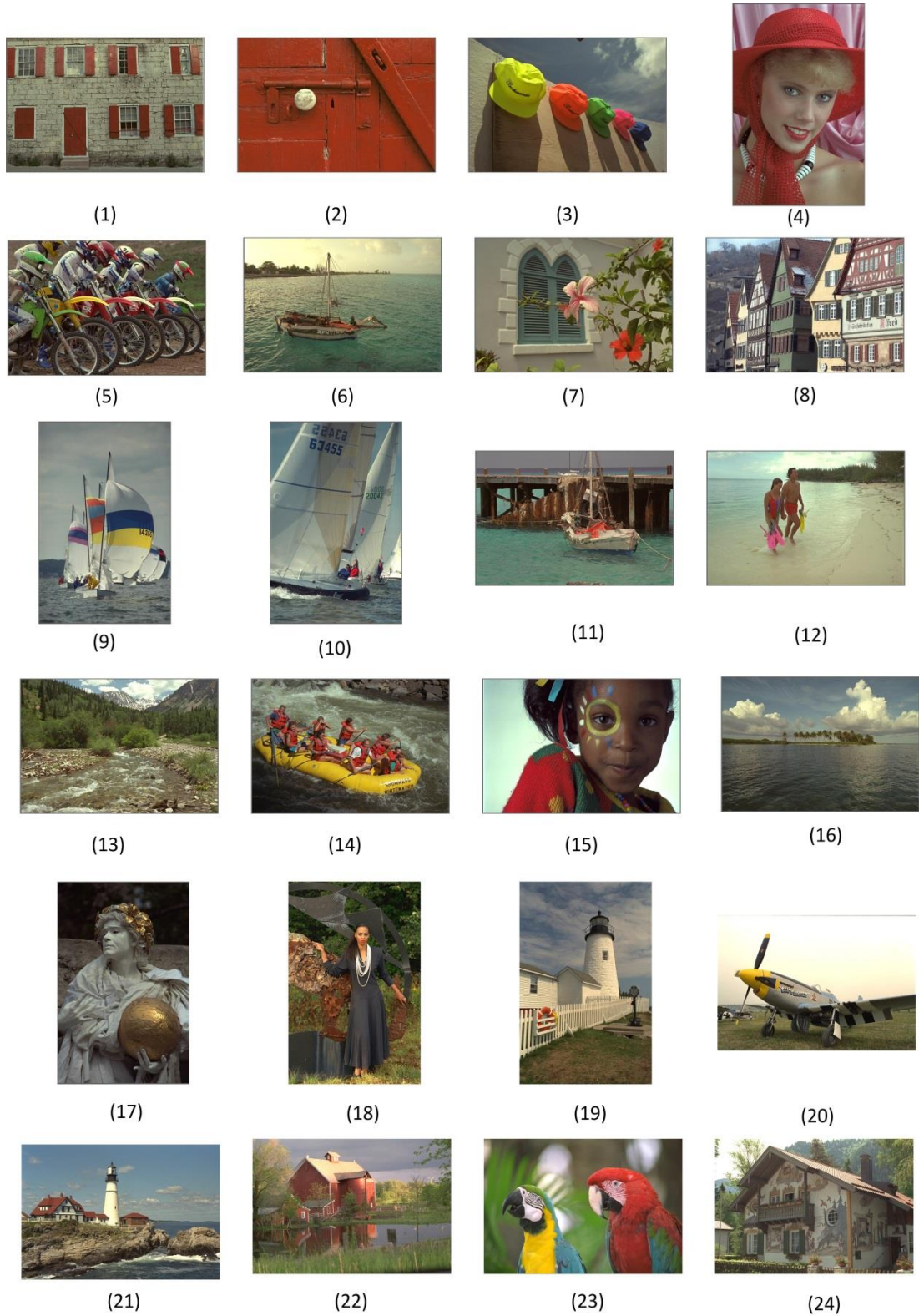

Figure 4.1: Original high resolution images of Kodak set [50].

Table 4.2: Kodak set PSNR and SSIM comparison with bicubic, R. Zeyde algorithm [22], (i) proposed algorithm (ii)if the best HR dictionary be chosen

| Image	Bicubic	R. Zeyde.et.al [22]	Proposed method	Correct model selection
1	26.69 0.9597	27.76 0.9791	27.91 0.9755	28.21 0.9770
2	34.03 0.9697	34.91 0.9751	35.09 0.9806	35.29 0.9815
3	35.03 0.9842	36.46 0.9849	36.51 0.9908	36.72 0.9912
4	34.57 0.9813	35.86 0.9835	36.13 0.9902	36.33 0.9906
5	27.13 0.9775	28.66 0.9839	28.70 0.9895	28.96 0.9900
6	28.27 0.9628	29.19 0.9793	29.46 0.9771	29.66 0.9781
7	34.27 0.9923	36.10 0.9903	36.2 0.9963	36.46 0.9965
8	24.31 0.9668	25.32 0.9819	25.41 0.9768	25.67 0.9778
9	33.13 0.9856	34.79 0.9869	34.75 0.9908	35.01 0.9856
10	32.94 0.9848	34.36 0.9869	34.4 0.9908	34.61 0.9912
11	29.93 0.9717	30.98 0.9819	31.03 0.9821	31.25 0.9912
12	33.56 0.9773	35.13 0.9866	35.32 0.9855	35.59 0.9830
13	24.71 0.9576	25.42 0.9754	25.75 0.9762	25.92 0.9861
14	29.89 0.9773	31.06 0.9797	31.31 0.9883	31.53 0.9771
15	32.88 0.9824	34.36 0.9860	34.94 0.9905	35.18 0.9888
16	32.05 0.9705	32.77 0.9760	32.97 0.9808	33.16 0.9906
17	32.86 0.9880	34.18 0.9819	34.32 0.9940	34.51 0.9815
18	28.78 0.9755	29.72 0.9827	30.15 0.9870	30.35 0.9942
19	28.79 0.9719	29.97 0.9770	30.06 0.9818	30.30 0.9824
20	32.36 0.9765	33.79 0.9857	33.74 0.9914	33.96 0.9918
21	29.26 0.9816	30.21 0.9832	30.47 0.9890	30.68 0.9895
22	31.36 0.9753	32.39 0.9809	32.58 0.9867	32.82 0.9873
23	35.92 0.9926	37.63 0.9876	37.88 0.9961	38.15 0.9963
24	27.62 0.9764	28.5 0.9806	28.81 0.9869	28.99 0.9876
average	31.09 0.9766	32.06 0.9823	32.23 0.9864	32.47 0.9869

Table 4.3: Benchmark set PSNR and SSIM comparison with bicubic, R. Zeyde algorithm [22], proposed algorithm and the last column if the best HR dictionary be chosen

Image	Bicubic	R. Zeyde.et.al [22]	Proposed method	Correct model selection
Baboon	24.86	25.40	25.67	25.81
	0.9651	0.9808	0.9750	0.9759
Barbara	28.00	28.64	28.58	28.69
	0.9577	0.9734	0.9712	0.9716
Boat	32.36	33.70	33.76	34.01
	0.9863	0.9812	0.9913	09918
Cameraman	35.71	38.70	40.13	40.44
	0.9937	0.9849	0.9984	0.9985
Elaine	31.06	31.30	31.42	31.48
	0.9767	0.9706	0.9793	0.9796
Face	34.83	35.49	35.80	35.96
	0.8463	0.8463	0.8767	0.8810
Fingerprint	31.95	33.87	34.85	35.15
	0.9942	0.9706	0.9983	0.9984
Lena	34.70	36.10	36.28	36.49
	0.9893	0.9815	0.9936	0.9938
Man	29.25	30.28	30.31	30.49
	0.9820	0.9812	0.9889	0.9894
Zone-plate	11.40	11.96	12.31	12.36
	0.7040	0.8686	0.7281	0.7289
average	26.77	27.85	28.13	28.41
	0.8656	0.9568	0.9540	0.9548

Although the performance in the proposed algorithm is obvious but the results achieved from the correct selection (choosing the most proper dictionary)in fourth column of the tables shows that a better model selection can be created, and the proposed algorithm is improvable. Qualitative comparison is also further illustrated in Figure 4.2 to Figure 4.6 and the superiority of reconstructing directional parts of images by proposed algorithm over other methods is observable. to show the preference of the proposed algorithm toward state of the art it has chosen three images from the Kodak set (number 1, number 22 and number 24) and two well-known images from benchmark set(Barbara and Lena). In all figures below (a) is the original image (b) shows bicubic interpolation (c) is the result yields by [22] and (d) shows the proposed algorithm results
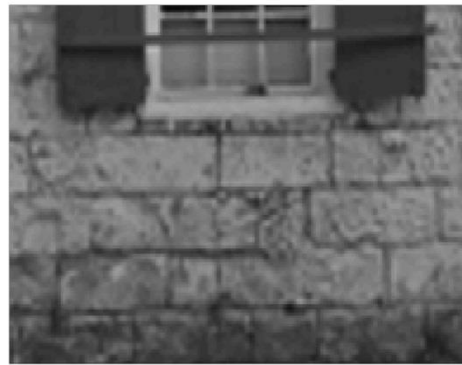
Figure 4.2: Visual comparison of image number 1 from Kodak set (a) original image
(b) bicubic interpolation (c) R. Zeyde algorithm [22] (d) proposed algorithm

Figure 4.3: Visual comparison of image number 22 from Kodak set (a) original
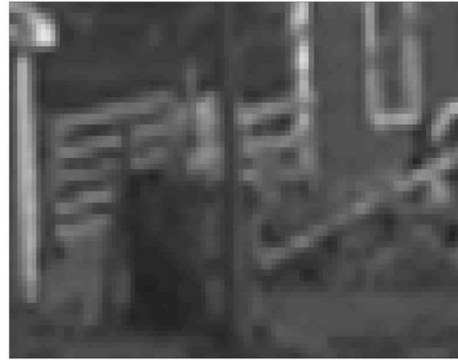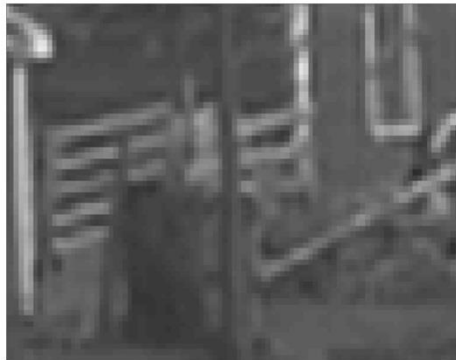image (b) bicubic interpolation (c) R. Zeyde algorithm [22] (d) proposed algorithm

Figure 4.4: Visual comparison of image number 24 from Kodak set (a) original image (b) bicubic interpolation (c) R. Zeyde algorithm [22] (d) proposed algorithm
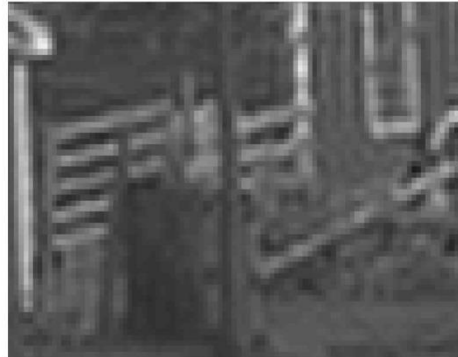
Figure 4.5: Visual comparison of Barbara image (a) original image (b) bicubic interpolation (c) R. Zeyde algorithm [22] (d) proposed algorithm

Figure 4.6: Visual comparison of Lena image (a) original image (b) bicubic
interpolation (c) R. Zeyde algorithm [22] (d) proposed algorithm

Figure 4.2 shows the tests performed on image number 1of the Kodak set with PSNR increases of 1.22 than bicubic. As it is observable the shutters behind the window is sharper and the directions of them are recognizable in (d) which is the result yield by proposed algorithm compared to (b) bicubic and (c) [22]. In (b) and (c) the shutters are smooth and interconnected with no space from each other.

Figure 4.3 is number 22 of the Kodak set. In (d) wooden fence are straighter and sharper with less curvature than (b) and (c).

In Figure 4.4 which is the 24th image of the Kodak set the gable roof in left has both horizontal and vertical lines where in (b) and (c) just vertical lines are observable and (d) shows both.

Figure 4.5 and Figure 4.6are two of the well-known bench mark images, Barbara and Lena. The directions of the scarf of Barbara is sharper in (d) while in (b) and (c) are smother. In Lena image the directions and the contrast of the shadow on the top of the hat in (d) is sharper and clearer.

These comparisons prove that learning directional dictionaries in addition to performing the process in wavelet domain gives a better reconstruction in high frequency details and directional patches.

Chapter 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

This thesis investigated the ill-posed super-resolution problem using sparse representation by structurally directional dictionaries. To this end HR and LR patches classified into several directional categories using similarity level between pre-determined directional templates and patches. Categorized patches and features formed the training sets for learning the dictionary pairs. Both pairs of HR and LR dictionaries obey the character of their training set and maintain the corresponding category's direction. In the reconstruction part given the low resolution image (LL1) the main aim is estimating high resolution wavelet subbands which are unknown. To reach to this aim OMP algorithm is applied to each feature of low resolution subbands each time with one of the low resolution dictionaries. In order to choose the most proper dictionary the feature is reconstructed nine times by nine low resolution dictionaries and the sparse coefficients obtained from OMP. The proper dictionary is used according to the least square error between the main feature and the reconstructed features. By the assumption that the sparse coefficients for sparsely representing the HR and LR patches can be equal, patches of the HR subbands are reconstructed by the corresponding proper high resolution dictionary and same sparse coefficient. After reconstructing all patches and high resolution subbands, the super resolved image obtained by an inverse wavelet transform.

By learning several dictionaries instead of one big dictionary the complexity arises from the learning part is decreased. Visual and experimental results have been shown reconstructing directional patches by directional dictionaries returns better results than reconstructing patches by a dictionary with no special feature, also there was around 0.2 dB increase in PSNR than the state of the art method which had been reconstructed patches by one dictionary.

5.2 Future Work

In the future, the most important piece of work which needs more investigation would be the model selection for choosing the most proper dictionary in reconstruction phase. The presented result shows that if the best high resolution dictionary be chosen regardless LR dictionaries, the performance will increase. This case represents the fact that the best LR dictionary for reconstructing a feature and the best HR dictionary for reconstructing the patch in the same location are not always correspond to each other, so the model selection according to least square error between the exist feature and the reconstructed features is not a perfect model. By using directional dictionaries better performance is achievable but with generating a better model selection.

One other important issue that can be worked on it in the future is deliberation of sparse coefficients equality. To estimate the HR subbands it has assumed that the sparse coefficients in LR and HR patches can be equal, but the validity of this assumption can be studied later.

As there is a direct link between the learned dictionaries and classifying patches to form the training set, other models for classifying patches in training phase can be

surveyed. After classifying features and patches into directional categories, objective tests show that the generated categories are not perfectly directional and some patches with correlation more than the pre-determined threshold but without any visual direction places in that category. So other methods such as comparing singular value decomposition for templates and patches can be tested for recognizing the directions better.

Also the classification can be done not only based on directions but also based on other cases. The fact is representation by learned dictionary over clustered data, gives supreme results. So the different kind of clustering can be performed.

REFERENCES

[1] M. Aharon, M. Elad, A. Bruckstein, "K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing,* vol. 54, no. 11, pp. 4311 - 4322, 2006.

[2] K. Skretting, K. Engan, "Recursive Least Squares Dictionary Learning Algorithm," *IEEE Transactions on Signal Processing,* vol. 58, no. 4, pp. 2121 - 2130, 2010.

[3] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online Dictionary Learning for Sparse Coding," in *International Conference on Machine Learning*, Montreal, 2009.

[4] J. Mairal, F. Bach, J. Ponce and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding,"*Journal of Machine Learning Research,* vol. 11, pp. 19-60, 2010.

[5] B. D. Rao, K. Kreutz-Delgado, "Deriving algorithms for computing sparse solutions to linear inverse problems," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & amp; Computers*, New York, 1997.

[6] B. Rao, "Signal processing with the sparseness constraint," in *ICASSP*, New

York, 1998.

[7] I. F. Gorodnitsky, B. D. Rao, "Sparse Signal Recostruction From Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," *IEEE Transaction on Signal Processing,* vol. 45, no. 3, pp. 600-616, 1997.

[8] I. F. Gorodnitsky, J. S. George, B. D. Rao, "Neuro magnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroencephalograohy and Clinical Neurophysiology,* pp. 231-251, 1995.

[9] B. D. Rao, I. F. Gorodnitsky , "Affine scaling transformation based methods for computing low complexity sparse solutions," in *ICASSP*, Atlanta, GA, 1996.

[10] B. Rao, "Analysis and extensions of the FOCUSS algorithm," in *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, New York, 1996.

[11] P. S. Huggins, S. W. Zucker, "Greedy Basis Pursuit," *IEEE Transactions on Signal Processing,* vol. 55, no. 7, pp. 3760 - 3772, 2007.

[12] H. Wang, J. Vieira, P. Ferreira, B. Jesus, I. Duarte, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *42nd Asilomar Conference on Signals, Systems and Computers*, California, 2008.

[13] T. Zhang, "Sparse Recovery With Orthogonal Matching Pursuit Under RIP,"

IEEE Transactions on Information Theory, vol. 57, no. 9, pp. 6215 - 6221, 2011.

[14] J. A. Tropp, A. C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory,* vol. 53, no. 12, pp. 4655 - 4666, 2007.

[15] S. Chen, S. A. Billings and W. Luo, "Orthogonal least squares methods," *International Journal of Control,* vol. 50, no. 5, pp. 1873-96, 1989.

[16] G. Davis, S. Mallat and Z. Zhang, "Adaptive time-frequency decompositions," *Optical-Engineering,* vol. 33, no. 7, pp. 2183-91, 1994.

[17] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing,* vol. 41, no. 12, pp. 3397-3415, 1993.

[18] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty Seventh Asilomar Conference*, 1993.

[19] J. Mairal, "Sparse Coding and Dictionary Learning," INRIA Visual Recognition and Machine Learning Summer School, San Francisco, June 2010.

[20] J. Yang, J. Wright, T. Huang and Yi Ma, "Image Super-Resolution As Sparse Representation of Raw Image Patches," in *IEEE International Conference on*

Computer Vision and Pattern Recognition(CVPR), 2008.

[21] J. Yang, J. Wright, T. Huang, Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions,* vol. 19, no. 11, pp. 2861 - 2873, 2010.

[22] R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-Up using Sparse-Representations," in *Proceedings of the 7th international conference on Curves and Surfaces*, Berlin, 2010.

[23] M. J. Fadili, J. L. Starck, F. Murtagh, "Inpainting and Zooming Using Sparse Representations," *The Computer Journal,* vol. 52, no. 1, pp. 64-79, 2009.

[24] B. Shen, W. Hu, Y. Zhang, Yu. J. Zhang, "Image inpainting via sparse representation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, Taipei, 2009.

[25] A. Rehman, Z. Wang, D. Brunet, E. R. Vrscay, "SSIM-inspired image denoising using sparse representations," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, 2011.

[26] M. Elad, M. Aharon, "Image Denoising Via Learned Dictionaries and Sparse representation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[27] M. Elad, M. Aharon, "Image Denoising Via Sparse and Redundant

Representations Over Learned Dictionaries," *IEEE Transactions on Image Processing,* vol. 15, no. 12, pp. 3736 - 3745, 2006.

[28] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, no. 2, pp. 210 - 227, 2009.

[29] W. Pan, T. D. Bui, C. Y. Suen, "Text detection from scene images using sparse representation," in *19th International Conference on Pattern Recognition, ICPR*, Florida, 2008.

[30] N. K. Bose, H. C. Kim, H. M. Valenzuela., "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-93*, 1993.

[31] S. P. Kim, W. su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE transactions on Image Processing Journal,* vol. 2, pp. 534-539, 1993.

[32] W. Yu, S. P. Kim, "High-resolution restoration of dynamic image sequences," *International Journal of Imaging Systems and Technology,* vol. 5, pp. 330-339, 1994.

[33] R. Molina, A. K. Katsaggelos, J. Mateos, *Super Resolution of Images and*

Video, Morgan & Claypool Publishers, 2007.

[34] D. Capel, A. Zisserman, "Computer Vision applied to super resolution," *IEEE Signal Processing Magazine,* vol. 20, no. 3, pp. 75-86, 2003.

[35] W. T. Freeman, T. R. Jones, E. C. Pasztor , "Example-based super-resolution," *IEEE Computer Graphics and Applications,* vol. 22, no. 2, pp. 56-65, 2002.

[36] S. Baker, T. Kanade, "Limits on super-resolution and how to break them," *IEEE transactions on Pattern Analysis and Machine Intelligence,* vol. 24, no. 9, pp. 1167 - 1183 , 2002.

[37] A. M. Tekalp, M. K. Ozcan, M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration," in *International Conference on Acoustics, Speech, and Signal Processing*, 1992.

[38] A. J. Patti, M. I. Sezan, A. M. Teklap, "Super-resolution video recunstruction with arbitrary sampling lattices and non-zero aperture time," *IEEE transactions on Circuits and Systems for Video Technology,* vol. 6, no. 8, pp. 1064 - 1076, 1997.

[39] S. C. Tai, T. M. Kuo, C. H. Iao, T. W Liao, "A Fast Algorithm for Single Image Super Resolution in Both Wavelet and Spatial Domain," in *International Symposium on Computer, Consumer and Control (IS3C)*, Taichung, 2012.

[40] S. Naik, N. Patel, "Single image super resolution in spatial and wavelet domain," *The International Journal of Multimedia & Its Applications (IJMA),* vol. 5, no. 4, pp. 23-31, 2013.

[41] M. Elad, I. Yavneh, "A Plurality of Sparse Representations Is Better Than the Sparsest One Alone," *IEEE Transactions on Information Theory,* vol. 55, no. 10, pp. 4701 - 4714, 2009.

[42] A. M. Bruckstein, D. L. Donoho, M. Elad, "From Sparse Solutions of systems of equations to sparse Modeling of Signals and Images∗," *SIAM review,* vol. 51, no. 1, pp. 34-81, 2009.

[43] R. Mehrotra, K. R. Namuduri and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognition,* vol. 25, no. 12, p. 1479–1494, December 1992.

[44] M. Lindenbaum, M. Fischer and A. Bruckstein, "On Gabor's Contribution to Image," *Pattern Recognition.,* vol. 27, no. 1, p. 1–8, 1994.

[45] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, M. J. Cree, "Retinal Vessel Segmentation Using the 2-D Gabor Wavelet and Supervised Classification,"*IEEE Transactions on Medical Imaging,* vol. 25, no. 9, pp. 1214--1222, 2006.

[46] D. Weber, D.Casasent, "Quadratic Gabor filters for object detection," in

Communications and Signal Processing, Grahamstown, 1997.

[47] K. A. Wa, "An Analysis of Gabor Detection," in *Image Analysis and Recognition*, 2009, pp. 64-72.

[48] J. G. Daugman, "Uncertainty Relation For Resolution In Space, Spatial Frequency, "*Journal of the Optical society of America A,* vol. 2, no. 7, pp. 1160-1169, 1985.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity,"*IEEE Transactions on Image Processing,* vol. 15, no. 4, pp. 600-612, 2004.

[50] "Kodak Lossless True Color Image Suite," [Online]. Available: http://r0k.us/graphics/kodak/.