A New Learning Based Super resolution using Contourlet Transform

by

Vineet P. Singh

200711030

A thesis submitted in the partial fulfillment of the requirements for the degree of
Master of Technology
in
Information and Communication Technology

Dhirubhai Ambani Institute of
Information and Communication Technology, Gandhinagar

June 2009

Declaration

This is to certify that

i) The thesis comprises my original work towards the degree of Master of Technology in Information and Communication Technology at DA-IICT and has not been submitted elsewhere for a degree.

ii) Due acknowledgement has been made in the text to all other material used.

Vineet P. Singh

Certificate

This is to certify that the thesis work entitled 'A New Learning Based Super resolution Using Contourlet Transform' has been carried out by Vineet P. Singh(200711030) for the degree of Master of Technology in Information and Communication Technology at this Institute under my supervision.

Prof. M.V.Joshi

Abstract

A new learning based super-resolution reconstruction using contourlet transforms is proposed. Contourlet transform provides high degree of directionality. It captures geometrical smoothness along multiple directions and learns the edges present in an image normal to the contour. For learning purpose, training set of low resolution (LR) and high resolution (HR) images, all captured using the same camera, are used. Here two and three level contourlet decomposition for LR images (test image and training image dataset) and HR training images respectively. The comaprision of contourlet coefficients of LR test image from the LR training set using minimum absolute difference (MAD) criterion to obtain the best match contourlet coefficient. The finer details of test image are learned from the high resolution contourlet coefficients of the training data set. The inverse contourlet transform gives super resolved image corresponding to the test image.

Acknowledgements

At first, I owe my deepest gratitude to Dr. M.V.Joshi, whose guidance, encouragement and positive criticism enabled me to become familiar with the subject. It would have been merely a dream to finish this thesis work without his continuous support and motivation.

Besides them, I would like to express my gratitude to the authors of all the books and publications, which helped me enhancing my knowledge and understanding about my work.

I am indebted to my family for their trust in me that made me confident enough to accomplish my work. Their support in a number of ways must also be acknowledged here.

I would like to thank all my friends for their continuous support throughout my stay at the institute.

I would like to thank Mr. P.P. Gajjar and Mr. Jignesh Bhawsar(Ph.D students at DAIICT) for their valuable suggestions and help during my work.

Above all, the infinite grace of the Almighty God is of essential importance and I solemnly offer my regards for His grace which enabled peace and harmony for this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In many imaging applications, like medical imaging, remote sensing, surveillance, high resolution images are required. The resolution of an image, captured by camera is determined by the spatial density of photo sensitive detectors in the camera. Images captured by camera having less number of detectors, produce low resolution (LR) image. As the number of detectors in camera increases; the resolution of an image, cost of camera and noise added by sensor also increases. In low cost imaging application, the image captured has lower resolution. LR images are under sampled image of scene, containing aliasing, blur, and noise. In order to obtain high resolution (HR) image from single LR image using image interpolation technique has limited applications, as LR images are aliased. For low cost imaging applications, algorithmic approaches can be used to construct HR images from LR images. *Super Resolution* (SR) is an algorithmic approach to construct HR images from single LR image or multiple LR images. SR is classified into two classes: *Reconstruction based* and *Learning based approach*. In *Reconstruction based approach*, multiple numbers of LR images are captured by camera with sub pixel shift for same scene. Each of the LR image contains additional information that can be used for reconstructing HR images. To construct SR image, super resolution algorithm is used. It comprises of the following three steps: registration, interpolation and restoration. Motion, blur, photometry, zoom etc., are used as a cue for reconstructing SR image from low resolution observation.

In *Learning based approach*, training set of LR images is formed. Learning based approach is used to construct SR image from single LR image. For a given LR image, we use training set of images to learn performance parameter or any suitable characteristics, to obtain corresponding SR image. The quality of SR image depends upon the training set. This thesis mainly concentrates on the construction of SR image from single low resolution observation using learning based approach. We use training set of LR-HR pairs and contourlet transform [1] for learning the high frequency details. For a given LR test image contourlet coefficients are learned from the LR-HR training set and best matching conturlet coefficients are substituted for constructing SR image.

This thesis mainly concentrates on the construction of SR image from single low resolution observation using learning based approach. We use training set of LR-HR pairs and contourlet transform [1] for learning the high frequency details. For a given LR test image contourlet coefficients are learned from the LR-HR training set and best matching conturlet coefficients are substituted for constructing SR image.

1.1 Background Study

Super Resolution algorithm was first presented by Tsai and Huang [6] for reconstructing HR image from a sequence of low resolution under sampled images using frequency domain approach. Kim *et.al* [7] included blur and noise in LR image and used weighted recursive least square method to reconstruct super resolved image. Irani and Peleg [8] used Iterative Back Projection method to solve the problem of super-resolution. In Komatsu *et.al* [9] non-uniform sampling theorem was used to transform non -uniformly spaced samples onto a single uniform sampling grid. A set theoretic approach for estimation of HR image was presented by Stark and Oskoui[10]. In [11] Schultz and Stevenson proposed improved definition for image interpolation known as super resolution from single observed images. In [12] Shyamsunder *et.al* proposed a learning based approach to recognize digits of the vehicle registration plates. They used undirected graphical model to super-resolve and restore the image patches. Freeman *et.al* [13] learned the relationship between the corresponding high- and low-resolution pairs from the training data and modeled the images as a Markov network. In [14], Baker and Kanade proposed a learning-based facial hallucination method. They used Bayesian inference to derive the high-frequency components of a facial image from a parent structure. Ni and Nguyen utilize support vector regression (SVR) in the frequency domain and pose the super-resolution problem as a kernel learning problem in [15]. A Regression-based method for single image super-resolution kernel ridge regression (KRR) is used to estimate the high-frequency details of the underlying HR image. Brandi *et al.* [16] propose an example-based approach for video super-resolution. They restore the high-frequency information of an interpolated block by searching in a database for a similar block, and by adding the high frequency of the chosen block to the interpolated one. In [17] Lui *et.al* use wavelet based approach to learn the high frequency detail of facial images. The input test image and training images are divided into patches and are projected to the eigen space. They modeled weighted projection vector as markow network and maximum a posteriori (MAP) estimation to find the best matching patch to reconstruct higher resolution image. In [18] Jiji *et. al* demonstrate super-resolution of a single frame gray scale image using a training database consisting of high resolution images downloaded from the internet . They learn the high frequency details from data base. They used Markov Random Field prior model and Wavelet prior model for regularization. In [19] Gajjar et.al learned the high frequency wavelet coefficients from HR image training set. They used Auto Regression (AR) prior and Wavelet prior for regularization. In [20] Gajjar et.al used training set consisting of pair of LR-HR images to learn the initial estimate for super resolved image. The HR image to be super-resolved is modeled as an Inhomogeneous Gaussian Markov Random Field (IGMRF) and the model parameters are learned using initial high resolution estimate. In [5] jiji and chaudhari use contourlet transform for learning the high frequency detail from training data set of HR images while up sampling. Our work in this thesis is based on their work. A set of training images which consist of LR-HR pairs instead of using HR image only. The drawback of the approach proposed in [5] is, the HR images are downloaded from internet hence they may not represent the true HR images.

1.2 Problem statement

Some times multiple LR images of same scene are not available to obtain the HR image using super resolution algorithm. For a given LR test image obtain the HR image using contourlet transform with the help of HR-LR training images.

1.3 Proposed Solution

For given problem statement, we have training data set of low resolution and high resolution images. By using contourlet transform we will learn the best matching coefficients for given low resolution test image from the given set of low resolution images. The best matching coefficient is obtained by finding minimum absolute difference of contourlet coefficients between the test image and low resolution images in the training data set. Now we have best matching pixel position and corresponding image number in data base. We take contourlet transform of high resolution image that corresponds to best match low resolution image. Next we will find the finer detail for super-resolved image by replacing them with high resolution training images details that corresponds to best match image. In this way we will find the high resolution image form given single resolution image.

1.4 Thesis Outline

The organization of thesis is as follows:

Chapter-2 presents the basic idea of Contourlet Transform. In this chapter we explain basic blocks used in Contourlet decomposition. It explain how Multiscale nad Multidirectional decomposition takes place using Laplacian Pyramid and Directional Filter Bank.

In Chapter-3, we explain how learning takes place using contourlet transform and simulation results of proposed methode. In this chapter we explain the Algorithm used for learning high frequency details from Traning Data sets pair for given low resolution test image. We present our result and comparision of our work with privious technique used in increasing image resolution.

Chapter-4 presents the conclusion of our work. The outcomes of our work are also mentioned. The possible future work directions are also presented.

Chapter 2

Contourlet Transform

The one dimensional transform as, Fourier and Wavelet transform has limitation in capturing edges present in an image. They capture edges present in horizontal and vertical direction. A new multi-resolution and multidirectional (figure 2.2) transform presented in [1] called *Contourlet transform*. Contourlet transform is capable of capturing the edges normal to contour present in an image. Contourlet transform consist of double filter bank structure (figure2.1). The Laplacian pyramid (LP) [21] used to capture the point discontinuity and Directional filter bank (DFB) [3] link the point discontinuity. Since DFB was designed to capture the high frequency component present in different direction while low frequency component were handled poorly. So this is another reason to combine the DFB with LP, where low frequency component were removed before applying DFB.

The contourlet transform offers flexible multi-resolution and directional decomposition for images, since it allows for a different number of directions at each scale. The number of directions is doubled at every other finer scale of the pyramid.

The directional and multi-resolution decomposition of image is great area of interest. Idea for decomposing image into directional sub-band was proposed by M.N.Do and Martin Vetterli [1]. The contourlet transform is obtained by using double filter bank structure, combining the Laplacian pyramid [21] with directional filter bank structure [22]. Laplacian Pyramid is used to capture the point discontinuity and directional filter bank used to link the point discontinuity. The resulting image expansion is like contour shape.

The Laplacian pyramid at each level generates a sampled low pass image and detail image that is difference between original image and predicted image. We show

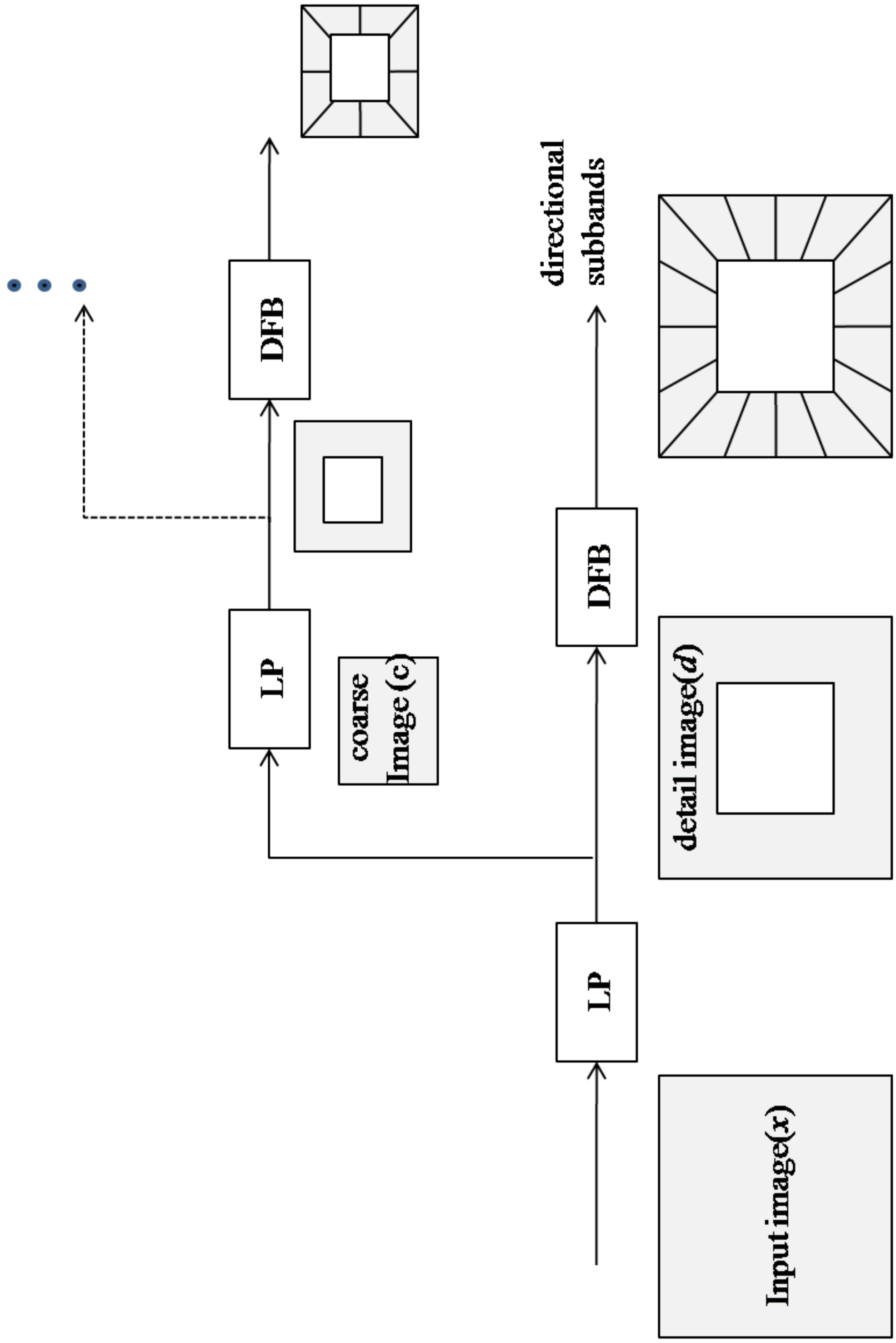


Figure 2.1: *Contourlet transform*Block diagram : image is decomposed by a double Flter-bank structure, where the first one captures the edge points and then a directional decomposition with a DFB is applied.Second one links these edge points into contour segments. The gray areas in the boxes represent the support sizes of the Filters.

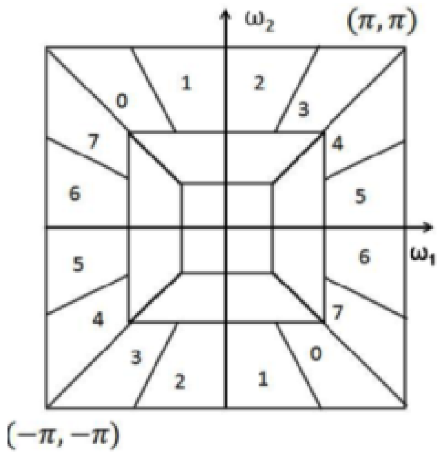


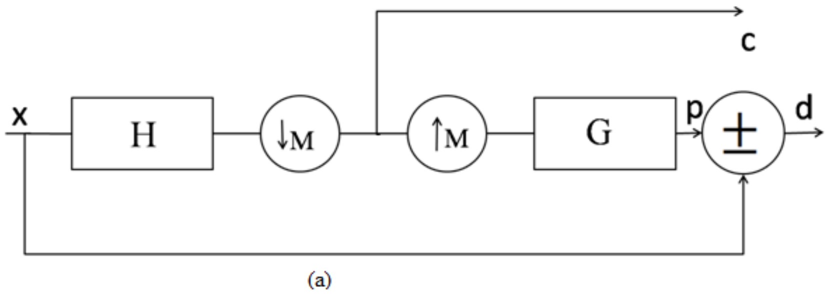
Figure 2.2: Frequency partion offered by *Contourlet Tranform*:The number of directional frequency partition is decreased from the higher frequency bands to the lower frequency bands [1]

that the LP with orthogonal filters (that is, $h[n] = g[-n]$ and $g[n]$ is orthogonal to its translates with respect to the sub-sampling lattice) is a tight frame with frame bounds equal to 1.

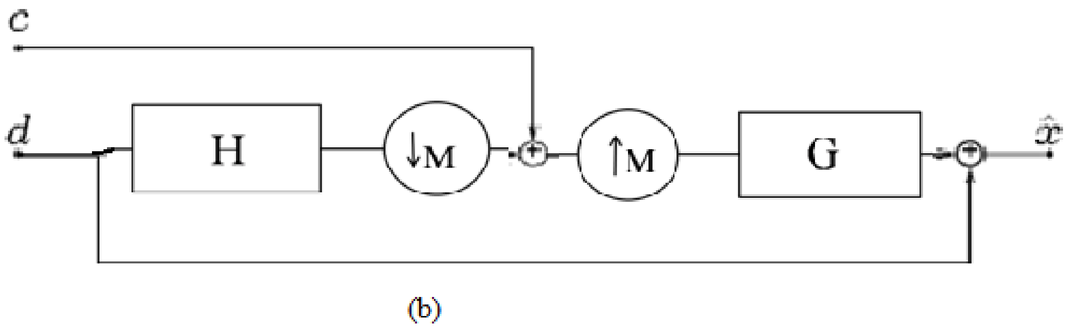
2.1 Laplacian Pyramid

The Contourlet transform offers multi-scale and multi-direction decomposition of an image. The multi-scale representation allow to successive approximation from coarse to fine scale. The multi-scale representation obtained from laplacian pyramid structure [21] used in contourlet transform as first stage.

The basic building block of Laplacian pyramid is filter and sampling matrices as shown in figure2.3(a). The input image x is first filtered and down-sampled by **H** and **M**, respectively. This results in coarse image c which is half the size of original image x. Coarse image is then applied to up-sampler **M** and filter **G**. The output of **G** gives predicted image p which has the same size as the original image x. After subtracting original image with predicted image we obtain the detail image d having same size as original image. The decomposition process is iterated on coarse image to obtain the multi-scale representation. At each level, size of image is reduced by factor of two(figure 2.4). In contourlet decomposition the detail images obtained from LP are applied to the DFB for directional decomposition and the coarse image is applied to the laplacian stage for further decomposition. The reconstruction block is shown in figure 2.3(b). The coarse image obtained is first up-sampled and filtered by **M** and **G** respectively. then we add it with detail image d.



(a) Decomposition Block for Laplacian pyramid.The outputs are a coarse approximation c and a difference d between the original and the prediction image. The process can be iterated by decomposing the coarse version repeatedly.



(b) Reconstruction scheme for the Laplacian pyramid.

Figure 2.3: *Laplacian Pyramid*[2].

The decomposition process is iterated on coarse image to obtain the multi-scale representation. At each level size of image is reduced by factor of two (figure2.4). In contourlet decomposition the detail image obtain from laplacian pyramid is applied to the directional filter bank for directional decomposition and coarse image is applied to the laplacian stage for further decomposition.

2.2 Iterated Directional Filter Banks

The directional decomposition of an image is key feature of Contourlet transform. The DFB is efficiently implemented via an l-level binary tree decomposition that leads to

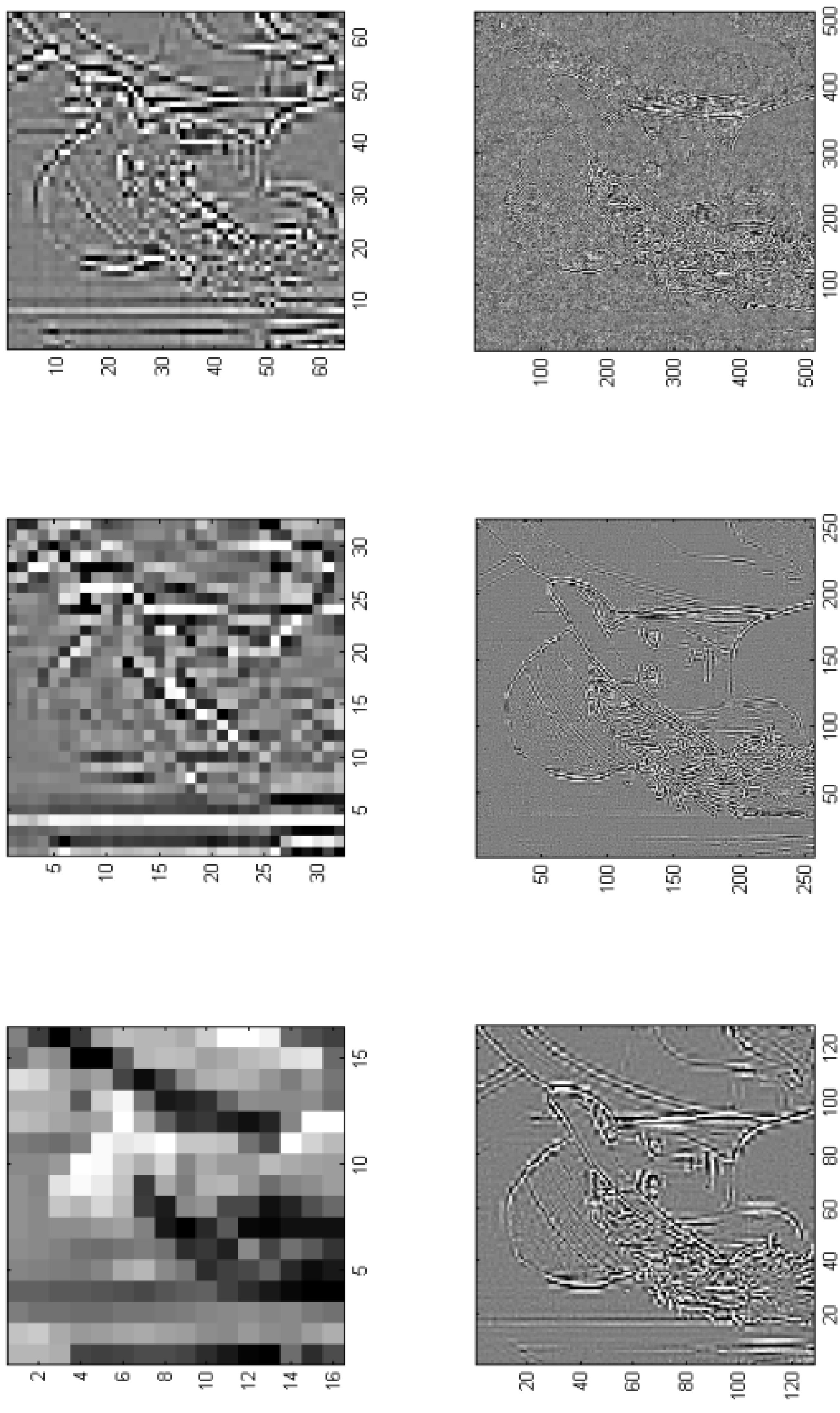
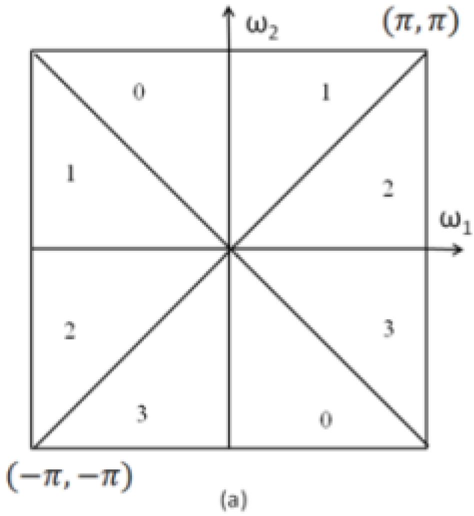
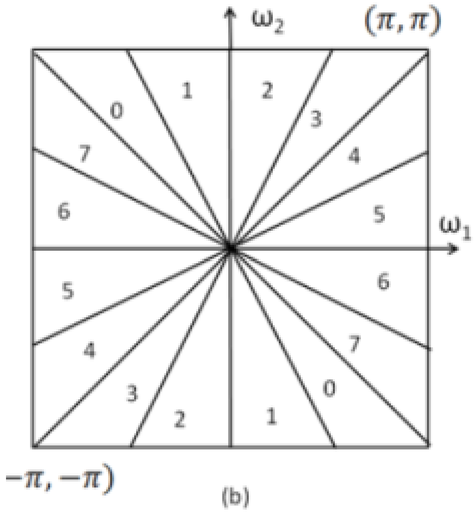


Figure 2.4: Five level *Laplacian Pyramid* decompostion.



(a) Four channel DFB



(b) 8-channel DFB

Figure 2.5: *Directional Filter Bank* Frequency Partioning.

2^l sub bands with wedge-shaped frequency partitioning as shown in Figure 2.5.Bamberger and Smith [22] constructed a 2-D directional filter bank (DFB) that can be maximally decimated while achieving perfect reconstruction. He used modulation at the input and using quincunx filter bank structure with diamond shaped filters. The frequency scrambling problem was arise in to his structure as first stage was modulator. Due to modulator stage and the down sampling results, high frequency channels are folded back into the low frequency channel. In [4] author proposed a new technique to construct the DFB that is used in Contourlet decomposition of an image. They are unique in their ability to decompose a multidimensional signal into direction sub bands that are nonredundant.

2.2.1 Sampling matrices

We would like to define some of important resampling matrices [4], which are used for directional decomposition of image in contourlet transform.
A *Resampling matrix* is a 2×2 matrix whose entries are all integer and determinant is nonzero.
A *Diagonal sampling matrix* is a resampling matrix whose principal diagonal elements are in the form of 2^n for a positive integer n. all element except diagonal elements are zero.

A *Quincunx sampling matrix* is a sampling matrix whose entries are of the form ± 1 with determinant 2.

$$Q_0 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \; Q_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \tag{2.1}$$

After down sampling by quincunx sampling matrix, resultant image is down sampled by factor of 2 and rotated by $\pm 45\,°$. [Figure2.6]

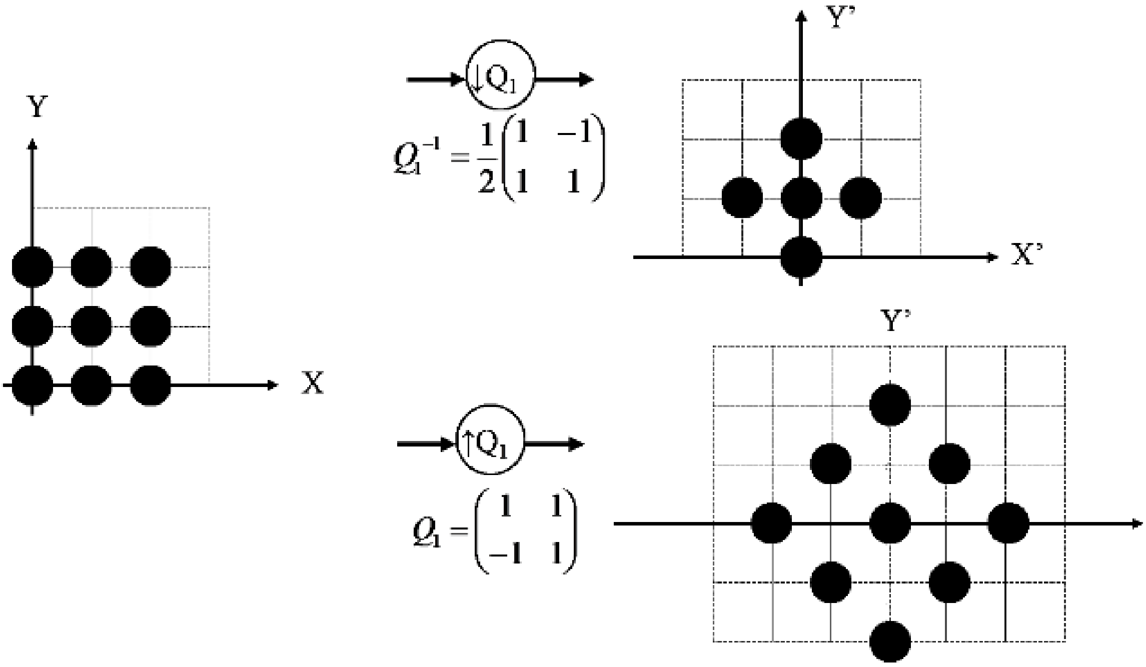


Figure 2.6: *Quincunx sampling matrix.*

A unimodular matrix is a resampling matrix whose determinant is ± 1. Its inverse matrix is also unimodular matrix.Sampling by a unimodular integer matrix does not change the data rate but only rearranges the input samples; thus it is referred to as a *resampling* operation.[Figure2.7]

A diamond conversion matrix is a unimodular matrix that can change a filter $H_0(\omega)$ with diamond shape pass band into parallelogram pass band.

$$R_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \; R_2 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \; R_3 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \; R_4 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \tag{2.2}$$

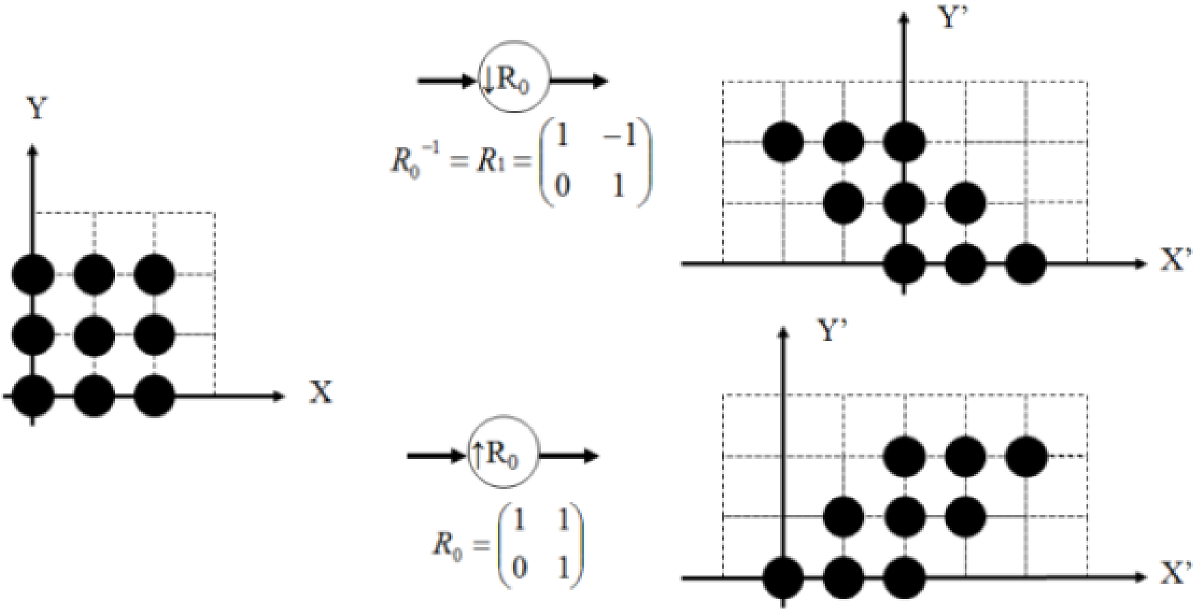


Figure 2.7: *Resampling matrix.*

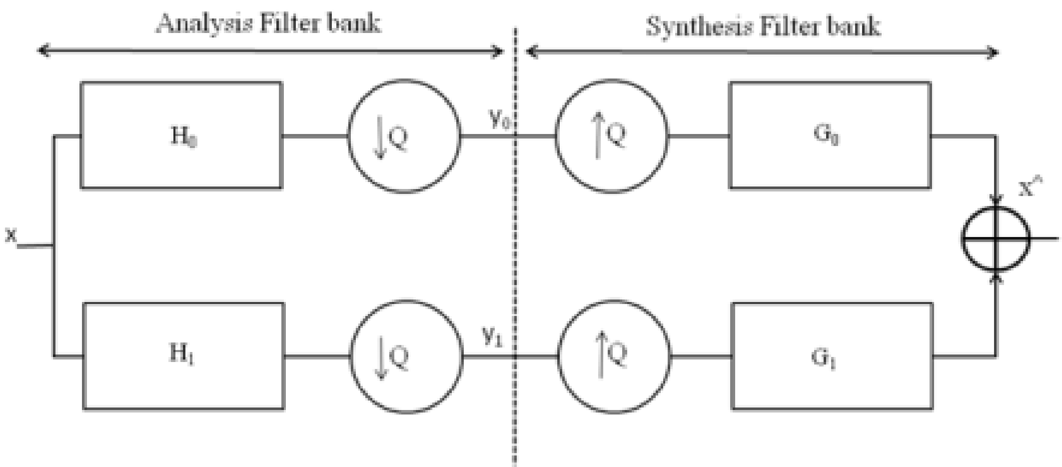
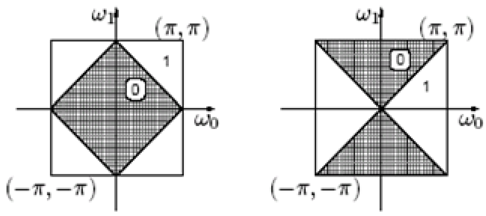


Figure 2.8: Two band structure (*Quincunx Filter bank*) [3]

2.2.2 Quincunx Filter Bank

The quincunx filter bank (QFB)[3] shown in Figure2.8, we consider the quincunx sampling matrix Q to be either Q_0 or Q_1. That matrix result down sampling operation rotates the input image by $-45\,°$ and $45\,°$, respectively.

The QFB can be used to split the frequency spectrum of input image into low pass and high pass channel using diamond shaped filter pair2.9(a) or uses fan filter pair2.9(b) that split the spectrum into either horizontal or vertical channel. One filter pair can be obtain by modulating the other with π with respect to one of the variable either ω_1 or ω_2.



(a) Diamond shaped-filter pair (b) Fan filter pair

Figure 2.9: Two possible support configurations for the filters in the QFB. Each region represents the ideal frequency support of a filter in the pair.[3]

2.2.3 Directional Filter Bank

The DFB divides the two-dimensional (2-D) spectrum of an image into wedge-like directional subbands, as shown in 2.5.The orignal DFB was implimented using modulator at input and using diamond shap filter. The analysis of iterated DFB is simplyfiy by using QFB's with Fan shap filter[4].This new approch avoides modulating the input image and has simpler rule for expanding the decomposition tree. The DFB basically consists of lowpass filters $H(\omega_0)$, quincunx down samplers Q, diamond conversion matrices R and postsampling matrices B , as shown in Figure 2.10. The Fan shap filter pair split the image into horizental and vertical channel. The Quincunx down sampler simultaneously down samples the image and rotates it by $45\,°$.

To obtain a four directional frequency partitioning, the first two decomposition levels of the DFB are given in Figure 2.11. We chose the sampling matrices in the

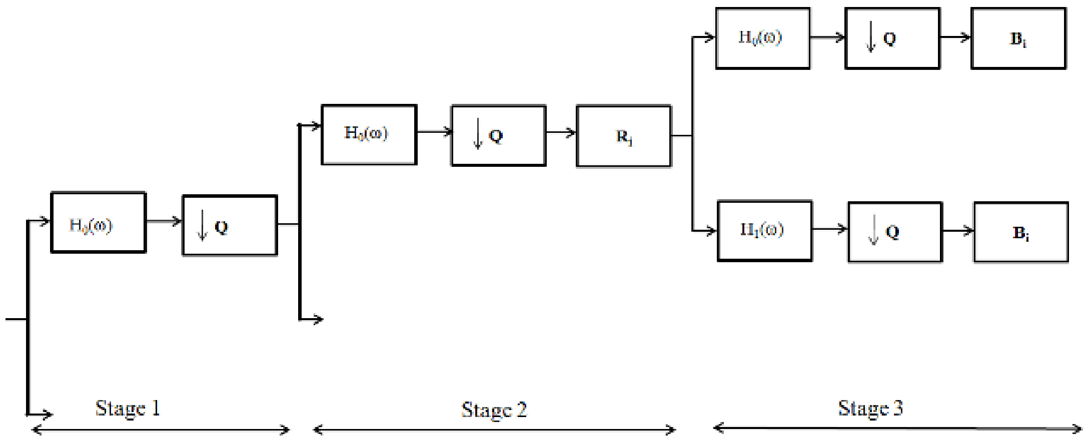


Figure 2.10: Structure of eight band DFB[4].

first and second level to be Q_0 and Q_1, respectively, so that the overall sampling after two levels is $Q_0.Q_1 = 2.I_2$, or downsampling by two in each dimension.

From the third level, to achieve finer frequency partition, we use quincunx filter banks together with resampling operations as shown in Figure 2.12.In the third stage, diamond conversion matrices are required to transform the parallelogram-shaped passband into one with a diamond shape. These diamond conversion matrices enable the DFB to be implemented using only a one-dimensional (1-D) filter prototype.Postsampling matrices are then appended to the end of the filter bank to remove nondiagonality of the overall down sampling matrix. The subbands associated with an eight-band decomposition are rectangular, which is a result of the downsampling. For an $N \times N$ image the first half of the 2^n sub-band output is $N/2^{n-1} \times N/2$ in size and for second half is $N/2 \times N/2^{n-1}$.

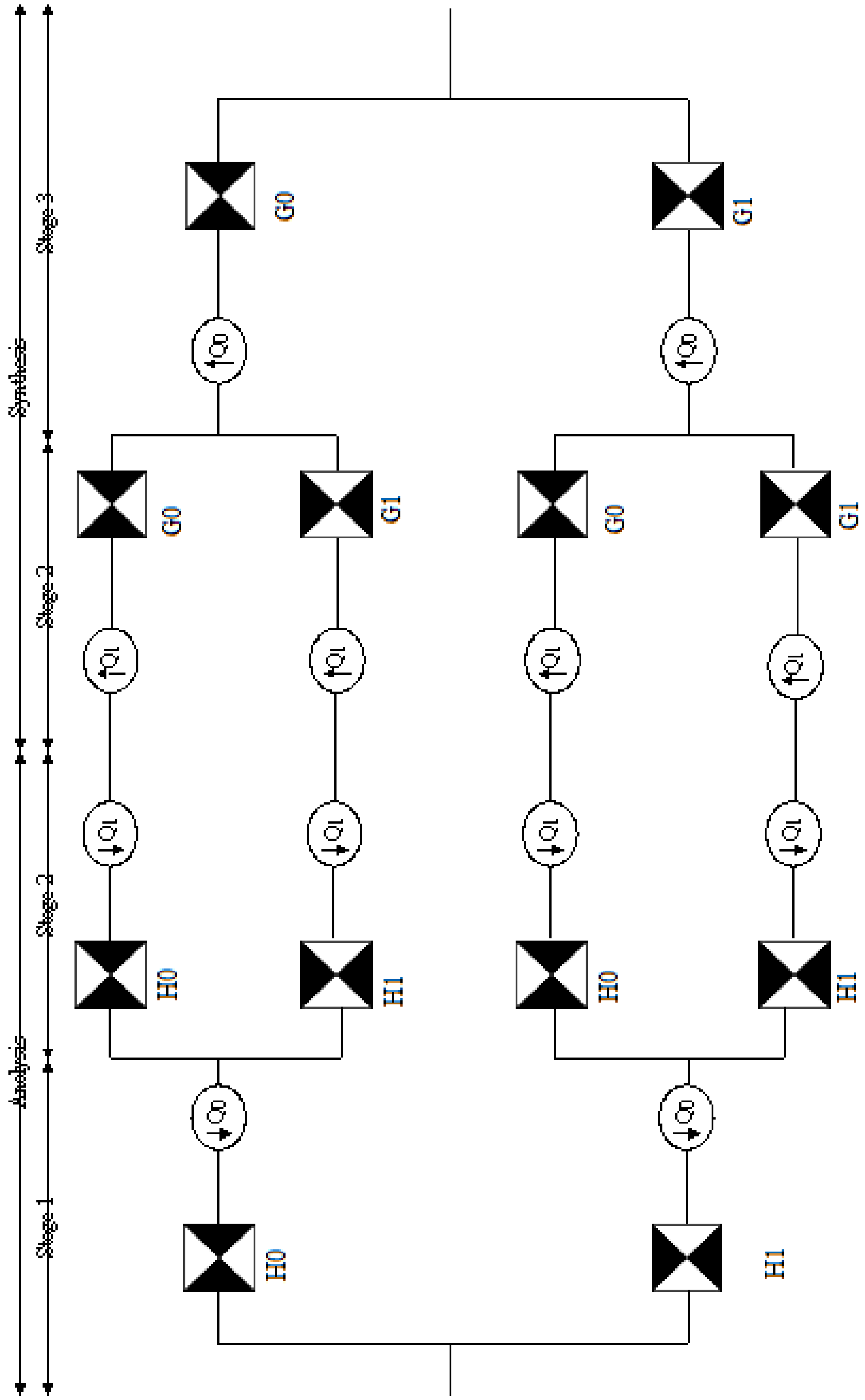


Figure 2.11: The first two levels of the DFB. At each level, QFBs with fan filters are used. The black regions represent the ideal frequency supports of the filters.[3]

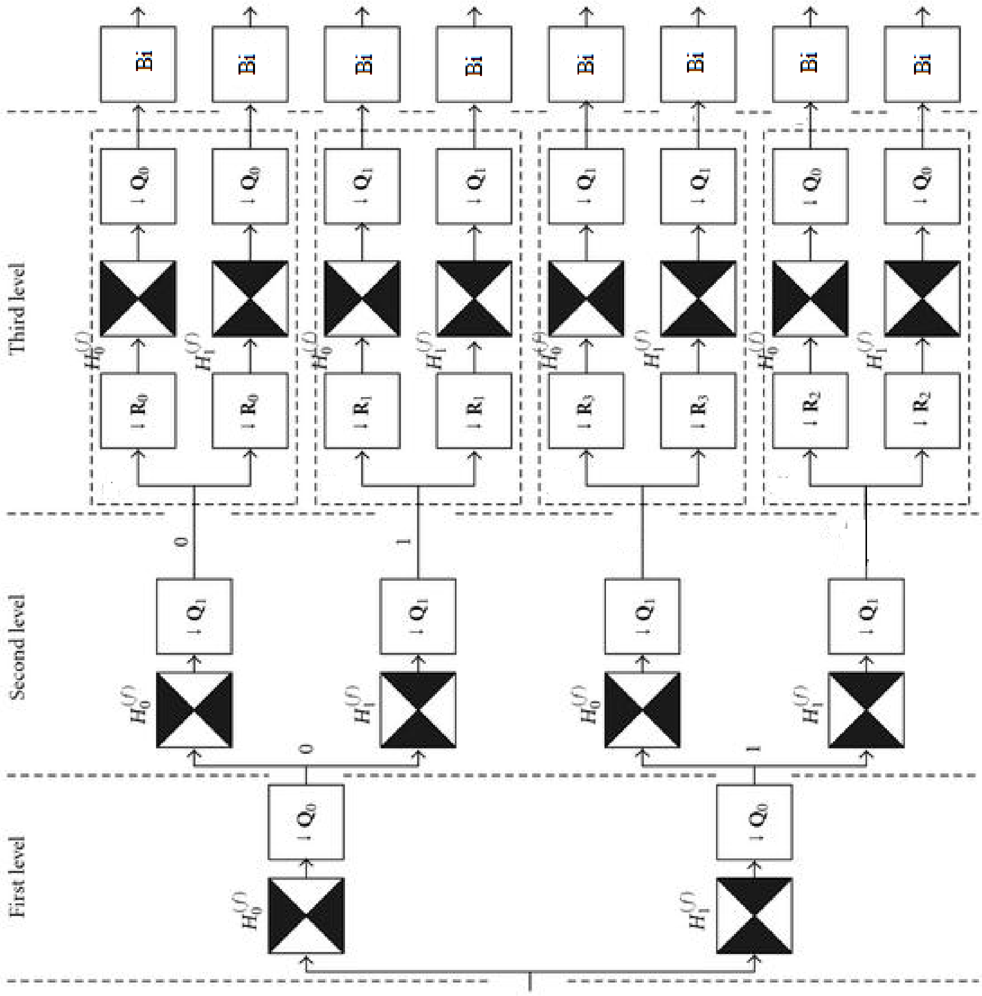


Figure 2.12: The iterative tree structure for the conventional DFB.[3]

2.3 Contourlet Decomposition

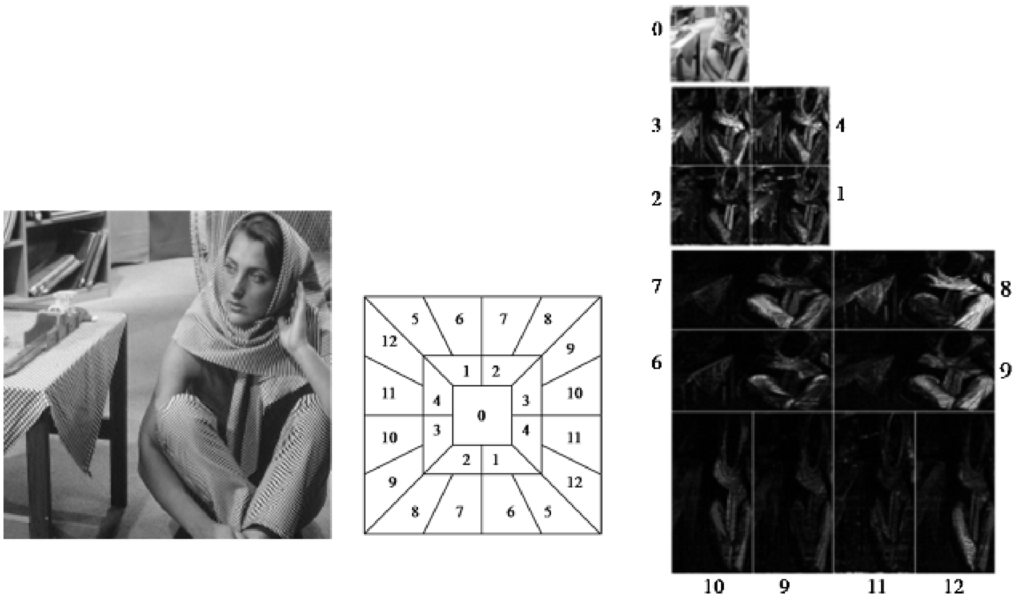


Figure 2.13: Two level Contourlet Decomposition. Frequency partioning at each level is shown.

The Contourlet Transform is used to capture the edges normal to the contour in an image. The Laplacian Pyramid and Directional filter bank combinally gives the contourlet decomposition. The input image is first applied to the laplacian stage that gives a low pass image, half the size of original image and a band pass image that is the difference between the original and predicted image. At each level we use l-level directional filter bank which is implemented by l-levels binary tree structure [4]. It gives 2^l wedge shape frequency portion.

Let $a_0[n]$ be the input image. Output after j-LP stages is j- band pass images $b_j[n]$, where $j = 1, 2 \cdots j$ (fine to coarse scale) and $a_j[n]$ as a low-pass image. This means j-th level decompose the image $a_{j-1}[n]$ in to coarse image $a_j[n]$ and a detail image $b_j[n]$.each band-pass image $b_j[n]$ is further decomposed by l_j levels DFB into 2^l direction band-pass images.

In our experiments we used “97” filters [24] for the Laplacian pyramid because they of their linear phase characteristics.For the directional filter banks we used the “2345” quincunx filters designed by Phoong *et al.* [25] and modulated them to obtain the fan filters. These filters are also nearly orthogonal and have linear phase response.the

McClellan transformation [26] is used to map 1-D two channel filter banks to 2-D filter channel.

Chapter 3

Learning and Simulation Results

When an image is interpolated, a region without any edges does not suffer from any degradation. However, if it contains edges, they get blurred during the upsampling process. We plan to learn the mapping of an LR edge (called edge primitive here) to its HR representation locally from the training data set during upsampling. Since wavelet transforms are known to capture high frequency details only horizental and veritical direction.In single-frame super-resolution algorithm proposed in [single frame wavelet based learning].In [20] author used HR-LR image pair and used wavelet technique to reconstruct HR image. One of the major difficulty in wavelet based learning is, wavelets only catch the edges normal to horizental and vertical direction. But we have problem in learning edges present in arbitary direction. This motivated us to use the contourlet transform. Contourlet transform help in caputring edges present in the diffrent directing.

In this section we propose a new learning based technique to obtain missing high frequency details for the superresolved image.We estimate the high frequency details using contourlet transform.The authors in [5] use database consisting of high resolution images only which are downloaded from the internet. This doesnot guarantee that these images indeed represents a high resolution training set. Also these images may be captured using different cameras at different times and under different lighting conditions. They use three level Contourlet transform of the HR image and learn the Contourlet coefficients in the finer level of the superresolve image. They find best matching HR image for each location of the low resolution by comparing the coefficients at two coarser levels. In our approach, we use the training set of LR-HR images captured using the same camera and learn Contourlet coefficients for the high resolution image from a training set. The training set has LR and HR pairs covering a wide range of scenes. We assume that the best matching Contourlet coefficients lie in the training set.With the assumption that a primitive edge element in the HR image is localized to an 8×8 pixel area, and the corresponding edge elements over a 4×4 pixel area in the LR image, we learn the contourlet coefficients at finer scales of the given LR image as described below.

3.1 Learning the Contourlet Coefficients

Here we discuss the learning for resolution factor of 2. Given a $M \times M$ low resolution test image. We perform two level contourlet decomposition of the LR test image and LR training images and as result we get four directional sub-bands which is shown in Figure 3.1.Similarly we perform three level contourlet decomposition of the HR training images as we get eight directional sub-bands at finest level and four directional sub-bands at remaining two levels as shown in Figure 3.1(b). Our idea is to construct the corresponding SR image by using LR test image and training dataset of LR and HR pair. The missing contourlet coefficients at finest level of LR test image will obtained from training dataset. After learning, we have a three level decomposition for the test image. The inverse transform of this will yield the SR image.

Figure(3.1) illustrates the block schematic of how the contourlet coefficients at finer scales are learned from a set of K training images using a two-level contourlet decomposition of the low-resolution test image. As explained earlier, the high-resolution training images are decomposed into three pyramidal levels and the test image at each location is compared to the training images in the contourlet domain at two coarser scales to search for presence of a nearly identical edge at all possible locations. This is required for extrapolating the missing contourlet coefficients in the directional sub-bands $IX - XVI$ for the test image.

Here the low-resolution image is of size $M \times M$ pixels. Considering an up-sampling factor of 2, the high-resolution image, now has a size of $2M \times 2M$ pixels. For each coefficient in the sub-bands $I - IV$ and the corresponding 2×2 blocks in the sub-bands $V - VIII$, we extrapolate a block of 2×4 contourlet coefficients in each of the

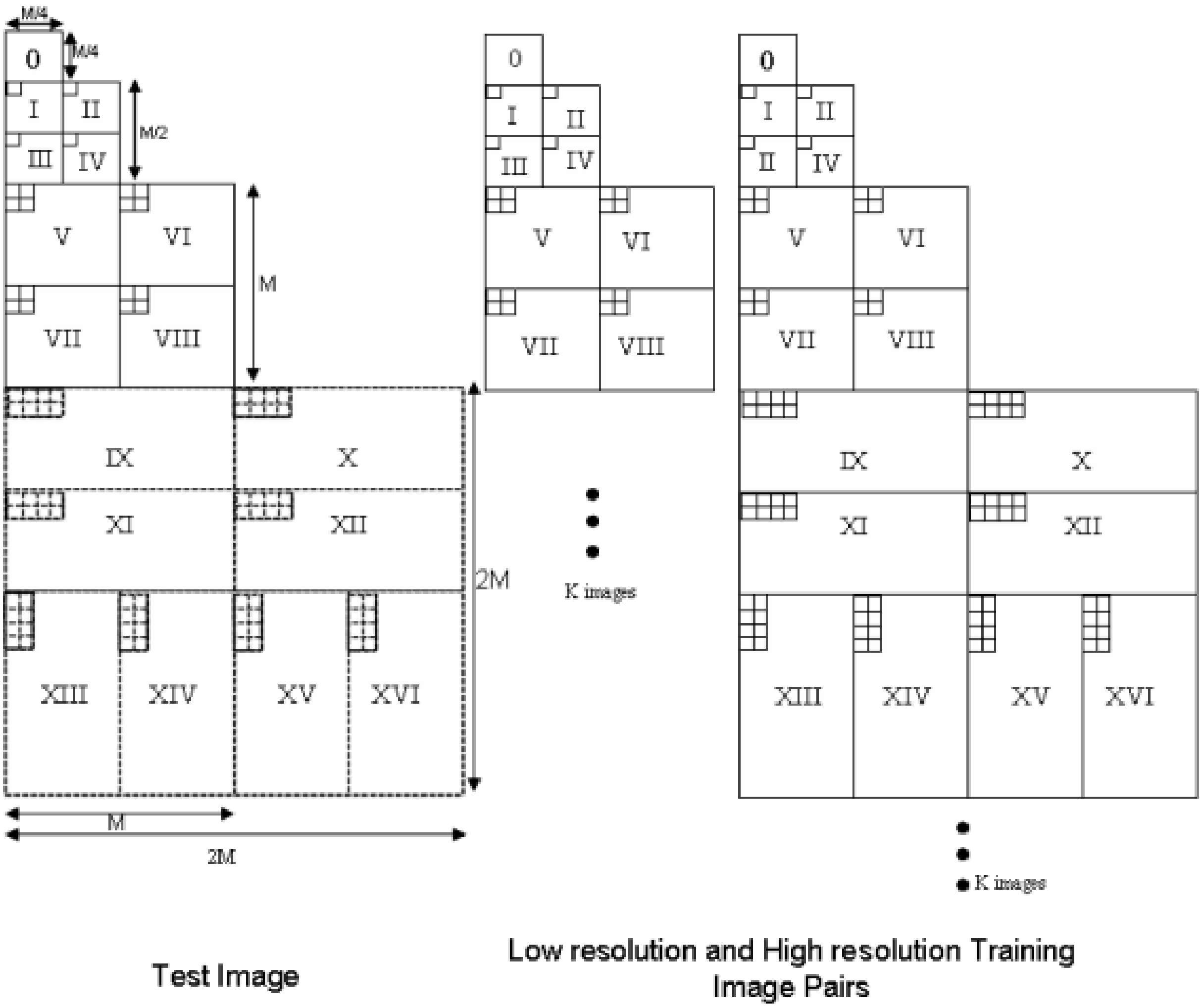


Figure 3.1: Learning from training set of pairs of LR-HR images(a) Two level contourlet decomposition of LR test image.The coefficients in the dotted sub-bands are to be learned.(b)Two level contourlet decomposition of LR training images dataset.(c) Three level contourlet decomposition of HR image training images dataset.

sub-bands IX, X, XI, XII and 4×2 contourlet coefficients in each of the sub-bands $XIII, XIV, XV$ and XVI. In order to do this we exploit the idea from zero-tree concept[27], that is, in a multi-resolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation . Using this idea we follow the *Minimum Absolute Difference* (MAD) criterion to estimate the contourlet coefficients. We take the absolute difference locally between the contourlet coefficients in the low resolution image and the corresponding coefficients in each of the low-resolution training images (decomposed with two levels).

The learning process is as follows. Consider the subbands $I - VIII$ of the low-resolution image. Denote the contourlet coefficient at a location (i, j) as $c(i, j)$ in sub band I. Consider the range $0 \leq i, j \leq M/4$. The contourlet coefficients $c_I(i, j), c_{II}(i, j + M/4), c_{III}(i + M/4, j), c_{IV}(i + M/4, j + M/4)$ corresponding to sub-bands $I - IV$ and 2×2 blocks consisting of $c_V(k, l), c_{VI}(k, l+M/2), c_{VII}(k+M/2, l), c_{VIII}(k+M/2, l + M/2)$ for $k = 2i : 2i + 1$ and $l = 2j : 2j + 1$ corresponding to sub-bands $V - VIII$ in the low-resolution test image. Also all the high-resolution training images are considered to learn a 2×4 contourlet block in each of the sub-bands $IX - XII$ consisting of unknown coefficients $c_{IX}(k, l), c_X(k, l+M), c_{XI}(k+M, l), c_{XII}(k+M, l+M)$ for $k = 2i : 2i+1$ and $l = 4j : 4j+3$ and 4×2 contourlet block in each of the sub-bands $XIII - XVI$ consisting of unknown coefficients $c_{XIII}(k, l), c_{XIV}(k, l + M), c_{XV}(k + M, l), c_{XVI}(k + M, l + M)$ for $k = 4i : 4i + 3$ and $l = 2j : 2j + 1$. Thus, for a given set of a total of twenty contourlet coefficients in sub-bands $I - VIII$ in the low-resolution image, we perform a search in the two coarser pyramidal levels of all the training images at all pixel locations for the best match in the MAD sense and copy the corresponding, 2×4 in contourlet block in sub-band $IX - XII$ and 4×2 in contourlet block in sub-band $XIII - XVI$.

$$
\begin{aligned}
\hat{m}(\hat{p}, \hat{q}) \quad = \quad & argmin[|c_I(i, j) - c_I^{(m)}(p, q)| + |c_{II}(i, j + M_1) - c_{II}^{(m)}(p, q + M_1)| \\
& +|c_{III}(i + M_1, j) - c_{III}^{(m)}(p + M_1, q)| \\
& +|c_{IV}(i + M_1, j + M_1) - c_{IV}^{(m)}(p + M_1, q + M_1)| \\
& +S_V + S_{VI} + S_{VII} + S_{VIII}] \qquad (3.1)
\end{aligned}
$$

where $M1 = M/4$ and

$$\begin{aligned} S_V &= |c_V(2i,2j) - c_V^{(m)}(2p,2q)| + |c_V(2i,2j+1) - c_V^{(m)}(2p,2q+1)| \\ &+|c_V(2i+1,2j) - c_V^{(m)}(2p+1,2q)| \\ &+|c_V(2i+1,2j+1) - c_V^{(m)}(2p+1,2q+1)| \end{aligned} \tag{3.2}$$

And $S_{VI}, S_{VII}, S_{VIII}$ are the corresponding sums for sub-bands VI, VII, and $VIII$, respectively, and $m = 1, 2 \cdots K$. Here c_l^m denotes the contourlet coefficient for mth training image at lth sub-band. Here $\hat{m}(\hat{p}, \hat{q})$ denotes the location $(\hat{p}, \hat{q})$ for the $\hat{m}th$ high resolution training image that best matches the test image at $(i, j)th$ location in terms of contourlet coefficients. Thus we have for sub-bands IX-XII

$$\begin{aligned} c_{IX}(k_s, l_s) &= c_{IX}^{(\hat{m})}(k_h, l_h); \\ c_X(k_t, l_t) &= c_X^{(\hat{m})}(k_h, l_h); \\ c_{XI}(k_t, l_t) &= c_{XI}^{(\hat{m})}(k_h, l_h); \\ c_{XII}(k_t, l_t) &= c_{XII}^{(\hat{m})}(k_h, l_h); \end{aligned}$$

where $k_s = 2i : 2i+1$ and $l_s = 4j : 4j+3$, $k_h = 2\hat{(p)} : 2\hat{(p)}+1$ and $l_h = 4\hat{(q)} : 4(q)+3$; for sub-bands $XIII$-XVI

$$\begin{aligned} c_{XIII}(k_s, l_s) &= c_{XIII}^{(\hat{m})}(k_h, l_h); \\ c_{XIV}(k_s, l_s) &= c_{XIV}^{(\hat{m})}(k_h, l_h); \\ c_{XV}(k_s, l_s) &= c_{XV)}^{(\hat{m})}(k_h, l_h); \\ c_{XVI}(k_s, l_s) &= c_{XVI}^{(\hat{m})}(k_h, l_h); \end{aligned}$$

where $k_s = 4i : 4i+3$ and $l_s = 2j : 2j+1$, $k_h = 4\hat{(p)} : 4\hat{(p)}+3$ and $l_h = 2\hat{(q)} : 2\hat{(q)}+1$;

For (i, j) and $(\hat{(p)}, \hat{(q)}) \in (IX$-$XVI)$. This is repeated for each coefficient in sub-bands I-IV of the low-resolution image. In effect, we find the best matching 2×4 and 4×2 edge primitive for sub-band IX-XVI from the training data for a given representation in the low resolution image through contourlet expansion. The goodness of the learning depends on how extensive and useful the training data set is. The sub-band 0 corresponds to the coarsest resolution (see Figure 6(a)) in the contourlet decomposition and since the corresponding training set may have different average brightness, including the pixels from the 0-band does not yield a good match of an edge primitive as we want the edges to be brightness independent. Hence, we refrain from using the 0th band while learning.

The complete learning-based resolution enhancement procedure is summarized below in terms of the steps involved.

Step 1 Perform two-level contourlet decomposition with four directional subbands on the low-resolution test image of size $M \times M$, two-level decomposition on all training images each of size $M \times M$ and three level decomposition on all training images of high resolution of size $2M \times 2M$.

Step 2 Consider the contourlet coefficients at locations $(i, j), (i, j+M/4), (i+M/4, j)$, and $(i+M/4, j+M/4)$ in subbands I, II, III, and IV and the corresponding 2×2 blocks in $V - VIII$ of the low-resolution test image as well as the low-resolution training set.

Step 3 Obtain the sum of absolute difference between the contourlet coefficients in the low-resolution test image and all the coefficients for each of the low resolution training images. Obtain the best match.

Step 4 obtain the unknown high resolution contourlet coefficients (2×4 blocks and 4×2 blocks) from the high resolution training image offering the best match locally in sub-bands IX-XVI.

Step 5 Repeat Steps 2-4 for every contourlet coefficient in bands I-IV of the low-resolution image.

Step 6 Perform inverse contourlet transform to obtain the high-resolution image of the given test image.

A few comments about the learning at the finest level of the contourlet coefficients . The finest-level contourlet coefficients are estimated using nearest neighbor criterion from the training images. The process is not adaptive in the sense that no adaptive updating of these coefficients is performed based on previously learned values at a given location or from its neighborhood. Furthermore, there is no reinforcement of the learned coefficients through posterior analysis. This may yield inferior values of the coefficients, but the advantage is that one does not have to worry about the convergence issues. A similar learning procedure is typically adopted in other learning-based techniques in super-resolution.

An inherent drawback of the proposed learning method is that the learning process is very much resolution dependent. If we want to super-resolve a 2 m/pixel satellite image by a factor of $q = 2$, the training data must be of 1m/pixel resolution. If one wants to perform super-resolution on a 2.5m image, none of the images in existing database could ideally be used for training. For a commercial camera, if we change the zoom factor, it requires that a completely different set of training images be provided.

3.2 Simulation Results

In this section we discuss the effectiveness of the proposed approach in super resolving a low resolution SR image using the Contourlet coefficients learned from a data set comprising of both LR and HR images. Experiments were performed using a data set having images from a wide range of classes. The data set consists of 300 good quality images of wide range of objects ranging from human faces to natural and texture images.All the images in the database are of real world scene. A computer controlled camera was used to capture the images for the database. In order to avoid motion of the camera while capturing images of a scene at different resolution, a stable and isolated physical setup was used. The camera was triggered by a *MATLAB* program at successive but two different time instance for capturing images at diiferent resolutions. The resolution setting of the camera was changed by the program before each trigger. The time duration between two successive triggers was less than a second. The images of live subject were captured under controlled enviroment. We assume that the motion of subjects like human and other moving objects between two successive triger was negligible. We appllied mean correction for compensating the intensity variations among the images of each scene due to image capture at different time interval. The images acquired for the dataset are not specific to the class of objects to be super resolved. But images were captured keeping in mind that it should be rich in content in terms of edge information which could be used for learning while super resolving the SR image. For the experiment real world images were used.

The test images used are of size 64x64 whereas super resolved image is of size 128x128.Our results are compared with the super- resolved images obtained from Bicubic Interpolation and from Single Frame super resolution using contourlet transform[1].We have used PSNR as a quantitative measure for comparison of the results. In this thesis we have included the results of the experiment performed on three images and their subsequent comparisons. One major difference between the Single Frame super resolution and the proposed technique is that the data set of previous technique consists of only HR images whereas our data set comprises of both the HR and LR images. Secondly, the test images used in previous technique obtained by down sampling the HR image whereas in our experiment LR images were true images acquired from the same camera.

In the experiment we took an image which is not a part of the data set. Then the contour coefficients of this image are matched with the data set of the LR images and the best match is found using Minimum Absolute Difference criterion. After the closest match is found contour coefficient of the HR image corresponding to that LR match is used to learn high frequency coefficients of the SR image. The results are compared in the following discussion and their quantitative comparison is summarized in TABLE 3.2.

Table 3.1: Comparison of PSNRs expressed in **dB**

IMAGE	**A LEARNING BASED SUPER RESOLUTION using**	
	HR Training data set [5]	*LR-HR Training data set*
1	20.53	22.55
2	23.43	23.99
3	22.11	25.22

Table 3.2: Comparison of Mean Square Error(MSE)

IMAGE	**A LEARNING BASED SUPER RESOLUTION using**	
	HR Training data set [5]	*LR-HR Training data set*
1	0.0089	0.0056
2	0.0045	0.0040
3	0.0062	0.0030

Figure 3.2(a) shows a low resolution image of size 64x64.Figure 3.4(a) shows the

super-resolved image obtained using bicubic interpolation. Figure 3.5(a) shows the result obtained using only HR training set. Figure 3.6(a) shows the result obtained from the proposed technique. In the results obtained from the proposed technique preservation of edges is much better than the other techniques. In the result obtained from bicubic interpolation, image obtained is blurred, the area nears the eyes and ears are not sharp .In the result obtained from the single frame SR through CT, the image is blurred and also the artifacts are also clearly visible. In the results obtained from the proposed technique the presence of artifacts is minimized and also the edges are restored quiet well. The eyes and ears are clearly visible, also the marking on the tusks are clearer than the previous results.

Figure 3.2(b) to Figure3.6(b) shows the result of the proposed technique on the second image. In this case again we can see that the eyes, eye-brows, nostrils are clearer than the other results. Improvement in the edges near the lips is also clearly visible.

Figure 3.2(c) to Figure 3.6(c) shows the result of the technique on the image of a computer keyboard. In the results obtained from the proposed technique we can see that the alphabets are sharper than that in other results. Also the edges of the keys are much better preserved, the lower edges of the key of ‘J’ are clearer. Improvement in alphabets ‘M’,‘N’,‘K’ is clearly visible.

(a) (b) (c)

Figure 3.2: Low Resolution Test Image

(a) (b) (c)

Figure 3.3: Orignal HR Test Image

(a) (b) (c)

Figure 3.4: Image Using Bi-Cubic Interpolation

(a) (b) (c)

Figure 3.5: Super resolved image using HR training set [5]

(a) (b) (c)

Figure 3.6: Super resolved image using LR-HR training set

Chapter 4

Conclusion and Future Work

4.1 Conclusion

We have described a method for super-resolution restoration of images using a contourlet transform-based learning technique. The contourlet coefficients at finer scales, learned from a set of several high-resolution training images are used to estimate the super-resolved image.
Following are the major conclusions drawn from the simulation study:

1. The learning process ensures capturing the best high-resolution edges from the training set for given a low-resolution observation, as well as captures the smoothness along contours.

2. The results obtained for different classes of images show perceptual as well as quantifiable improvements over conventional interpolation techniques.

3. It is also observed that the standard waveletbased learning is of no use unless the results are further regularized using appropriate discontinuity preserving smoothness constraints. The proposed method is useful when multiple observations of a scene are not available and one must make the best use of a single observation to enhance its resolution.

4.2 Future work

We have described a method for super-resolution restoration of image using contourlet transform based learning technique. All the results were obtained using directional decomposition in eight subbands. We are investigating what is the optimal number of directional subbands that would offer the best reconstruction.

We can also apply this techniue to color images in YC_bC_r color space and expand them differently to obtian color super resolution.

References

[1] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005. viii, 2, 5, 7

[2] M. N. Do and M. Vetterli, "Framing pyramids," *IEEE Transactions on Signal processing*, pp. 2329–2342, September 2003. viii, 8

[3] M. N. Do, *Directional multiresolution image representations*, Ph.D. thesis, Department of Communication, Swiss Federal Institute of Technology, Lausanne, Switzerland., Lausanne, Switzerland, November 2001. viii, 5, 12, 13, 15, 16

[4] S.-I. Park, M.J.T. Smith, and R.M. Mersereau, "Improved structures of maximally decimated directional filter banks for spatial image analysis," *IEEE Transactions on Image Processing*, vol. 13, pp. 1424–1431, November 2004. viii, 10, 13, 14, 17

[5] C.V. Jiji and S. Chaudhri, "Single-frame image super-resolution through contourlet learning," *EURASIP Journal on Applied Signal Processing*, vol. 2006, September 2007. ix, 3, 19, 26, 28

[6] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," *Advances in Computer Vision and Image Processsing*, vol. 1, pp. 317–339, 1984. 2

[7] S.P. Kim, N.K. Bose, and Valenzuela., "Recursive reconstruction of high resolution image from noisy under sampled multi frames," *IEEE Transactions on Accoustics, Speech and Signal Processing*, vol. 18, no. 6, pp. 1013–1027, June 1990. 2

[8] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Computer Vision,Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991. 2

[9] T. Kosmatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-apperture cameras," *Signal Processing: Image Communication*, vol. 5, no. 3, pp. 511–526, December 1993. 2

[10] H. Stark and P. Oskui, "High- resolution mage recovery from mage plane array using convex projection," *J. Optical Society of America*, vol. 6, no. 11, pp. 1751–1726, November 1989. 2

[11] R. R. Schultz and R. L. Stevenson, "A bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–242, 1994. 2

[12] R. Shyamsundar, M. D. Gupta, N. Petrovic, and T. S. Huang, "Learning-based nonparametric image super-resolution," *EURASIP Journal on Applied Signal Processing*, November 2006. 2

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

[14] S. Baker and 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. 2

[15] K. Ni and T.Q. Nguyen, "Image superresolution using support vector regression," *IEEE Transactions Image Processing*, vol. 16, no. 6, pp. 1596–1610, 2007. 2

[16] F. Brandi, R. de Queiroz, and D. Mukherjee, "Super resolution of video using key frames," *Proc. IEEE International Symposium on Circuits and Systems*, pp. 1608–1611, 2008. 3

[17] S.F. Lui, J.-Y. Wu, and J.-J. James Lien, "Learning-based super-resolution system using single facial image and multi-resolution wavelet synthesis," *Computer Vision - ACCV 2007*, vol. 4844, pp. 96–105, 2007. 3

[18] C.V. Jiji, M.V. Joshi, and S. Chaudhuri, "Single-frame image superresolution using learned wavelet coefficients," *International Journal of Imaging Systems and Technology*, vol. 14, no. 3. 3

[19] P.P. Gujjar and M.V. Joshi, "A fast approach to the learning-based super-resolution using autoregressive model prior and wavelet prior," . 3

[20] P.P. Gujjar and M.V. Joshi, "Single frame super-resolution: A new learning based approach and use of igmrf prior," *Computer Vision, Graphics and Image Processing*, pp. 636–643, 2008. 3, 19

[21] P.J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communication*, vol. 31, pp. 532–540, 1983. 5, 7

[22] R.H. Bamberger and M.J.T. Smith, "A filter bank for the directional decomposition of images: Theory and design," *IEEE Transactions on Signal processing*, vol. 40, no. 4, April 1992. 5, 10

[23] M. Vetterli, "Multidimensional subband coding: some theory and algorithms," *Signal Processing*, vol. 6, no. 2, pp. 97–112, 1984.

[24] A. Skodras, Charilaos Christopoulos, and Touradj Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal Processing Magazine*, SEPTEMBER 2001. 17

[25] S.-M. Phoong, C. W. Kim, P.P. Vaidyanathan, and R. Ansari, "A new class of two-channel biorthogonal filter banks and wavelet bases," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 649–665, 1995. 17

[26] J.M. Shapiro, "Adaptive mcclellan transformations for quincunx filter banks," *IEEE Transactions on Signal Processing*, vol. 42, no. 3, MARCH 1994. 18

[27] J.M. Shapiro, "Embedded image coding using zero tree of wavelet coefficients," *IEEE Transactions on Signal processing*, vol. 41, no. 12, December 1993. 22