Single Frame Image Super-resolution Using Learnt Wavelet Coefficients

C. V. Jiji, M. V. Joshi and Subhasis Chaudhuri‡

Department of Electrical Engineering.

Indian Institute of Technology-Bombay

Mumbai - 400076. INDIA.

(jiji,mvjoshi,sc)@ee.iitb.ac.in

‡ corresponding author.

Abstract

In this paper we propose a single frame, learning based super-resolution restoration technique by using the wavelet domain to define a constraint on the solution. Wavelet coefficients at finer scales of the unknown high resolution image are learnt from a set of high resolution training images and the learnt image in the wavelet domain is used for further regularization while super-resolving the picture. We use an appropriate smoothness prior with discontinuity preservation in addition to the wavelet based constraint to estimate the super-resolved image. The smoothness term ensures the spatial correlation among the pixels whereas the learning term chooses the best edges from the training set. Since this amounts to extrapolating the high frequency components, the proposed method does not suffer from oversmoothing effects. The results demonstrate the effectiveness of the proposed approach.

Keywords

Super-resolution, Wavelet decomposition, Regularization, Image restoration

I. Introduction

In most imaging applications, images with high spatial resolution are desired and often required. Resolution enhancement from a single observation using image interpolation techniques is of limited application because of the aliasing present in the low-resolution image. *Super-resolution* refers to the process of producing a high spatial resolution image than what is afforded by the physical sensor through post processing means. It includes upsampling the image, thereby increasing the maximum spatial frequency, and removing degradations that arise during the image capture, viz., aliasing and blurring. Researchers traditionally use the motion cue to super-resolve an image. However this method being a 2D dense feature matching technique, it requires registration at a sub-pixel accuracy. Since the registration is obtained by using the low resolution observations, the estimated motion parameters are not accurate. Errors in registration are reflected on the quality of the super-resolved image. An alternate way of extrapolating the nonredundant information is to learn it from the high resolution training data and use a suitable regularization to obtain an accurate solution. We show in this paper that by using a wavelet based learning prior along with a suitable discontinuity preserving smoothness prior, an effective super-resolution can be achieved. The advantage of our method is that there is no correspondence problem. Also in many of the applications more than one low resolution observations may not be available, but we may have a database of a number of similar images at a higher

spatial resolution. Hence one needs to minimize the aliasing by using the given single low resolution image.

In this paper we consider having access to a set of high resolution training images to learn the prior. The basic problem we solve in this paper is as follows. One captures an image using a low resolution camera. We are interested in generating the super-resolved image for the same using a set of available high resolution images of different objects. It is assumed that the high frequency contents to be extrapolated are locally present in the training set. We use a wavelet based multi-resolution analysis to learn the wavelet coefficients at a given location at the finer scales for the super-resolved image. The learnt coefficients are then used in a prior term that enforces the condition that the wavelet coefficients at the finer scales of the super-resolved image should be locally close to the best matching coefficients learnt from the training set. In order to preserve the spatial continuity of the restored image, we use a smoothness constraint in conjunction with the learnt prior to obtain the super-resolved image.

The reminder of the paper is organized as follows. In section II we review some of the prior work in super-resolution imaging including those dealing with the learning-based methods. We discuss the model for the formation of a low resolution image in section III. Some background on discrete wavelet transform and the multi resolution analysis for estimating the wavelet coefficients at the finer scales using high resolution training images are the subject matters of section IV. Section V discusses the regularization based approach to derive a cost function for the super-resolution estimation. We present experimental results on different types of images in section VI, and the paper concludes with section VII.

II. Related Work

The super-resolution idea was first proposed by Tsai and Huang that used the frequency domain approach [1]. Kim *et al.* discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred observations [2]. Ur and Gross use the Papoulis and Brown generalized sampling theorem to obtain an improved resolution picture from an ensemble of spatially shifted observations [3]. These shifts are assumed to be known by the authors. A different approach to the

super-resolution restoration problem was suggested by Irani *et al.* [4], [5], based on the iterative back projection method. A set theoretic approach to the super-resolution restoration problem was suggested in [6]. The main result there is to define convex sets which represent tight constraints on solution as well as use the amplitude constraint to improve the results. Ng *et al.* develop a regularized, constrained total least squares solution to obtain a high-resolution image in [7]. They consider the presence of perturbation errors of displacements around the ideal sub-pixel locations in addition to noisy observations. The effect of the displacement errors on the convergence rate of an iterative approach for solving the transform based preconditioned system of equations is discussed by Ng and Bose [8]. They also develop a fast restoration algorithm for color images in [9]. Nguyen *et al.* have proposed circulant block preconditioners to accelerate the conjugate gradient descent method while solving the Tikhonov-regularized super-resolution problem [10]. Recently, Lin and Shum determine the fundamental limits of reconstruction-based super-resolution algorithms and obtain the super-resolution limits from the conditioning analysis of the coefficient matrix [11].

In [12] the authors use a maximum *aposteriori* (MAP) framework for jointly estimating the registration parameters and the high-resolution image for severely aliased observations. They use an iterative, cyclic coordinate-descent optimization technique to update the registration parameters. A MAP estimator with Huber-MRF prior is described by Schultz and Stevenson in [13]. Other approaches include a MAP-MRF based super-resolution technique using the blur as a cue [14]. In [15] the authors recover both the high resolution scene intensity and the depth fields simultaneously using the defocus cue. Recently, Rajagopalan and Kiran [16] proposed a frequency domain approach for estimating the high resolution image also using the defocus cue. Elad and Feuer [17] proposed a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and down-sampled measured images by combining maximum likelihood (ML), MAP and POCS approaches. An adaptive filtering approach to super-resolution restoration is described by the same authors in [18]. They have also developed a fast super-resolution algorithm in [19] for pure translational motion and space invariant blur.

Since edges in the image are places where one requires a better clarity, there have

also been some efforts in the literature on preserving the edges while interpolating an image. Chiang and Boult [20] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. In [21] authors propose an image interpolation technique using a wavelet domain approach. They assume that the wavelet coefficients scale up proportionately across the resolution pyramid and use this property to go down the pyramid. Thurnhofer and Mitra [22] have proposed a non-linear interpolation scheme based on a polynomial operator wherein perceptually relevant features (say, edges) are extracted and zoomed separately. The reconstruction/restoration methods to improve the resolution of digital images while zooming have been discussed in [23]. The authors here focus on both the linear and the non-linear methods based on total variation to study the ability of these methods to preserve 1D structures.

Researchers have also attempted to solve the super-resolution problem by using learning based techniques. These methods are classified under the motion-free super-resolution scheme as the new information required for predicting the high resolution image is obtained from a set of training images rather than from the subpixel shifts among low resolution observations. Authors in [24] have proposed a super-resolution technique from multiple views using learnt image models . Their method uses learnt image models either to directly constrain the ML estimate or as a prior for a MAP estimate. To learn the model, they use principle component analysis (PCA) applied to a face image data base. In [25] Baker and Kanade develop a super-resolution algorithm by modifying the prior term in the cost to include the results of a set of recognition decisions, and call it recognition based super-resolution or hallucination. Their prior enforces the condition that the gradient in the super-resolved image should be equal to the gradient in the best matching training image. An image analogy method applied to super-resolution is discussed in [26]. They use the low resolution and the high resolution portions of an image as the training pairs which are used to specify a "super-resolution" filter that is applied to a blurred version of the entire image to obtain an approximation to the high resolution original image. Candocia and Principe [27] address the ill-posedness of the super-resolution problem by assuming that the correlated neighbors remain similar across scales, and this apriori information is learned locally from the available image samples across scales. When a new image is

presented, a kernel that best reconstructs each local region is selected automatically and the super-resolved image is reconstructed by a simple convolution operation. A learning based super-resolution enhancement of video is proposed by Bishop *et al.* [28]. Their approach builds on the principle of example based super-resolution for still images proposed by Freeman *et al.*[29]. Joshi and Chaudhuri [30] have proposed a learning-based method for image super-resolution from zoomed observations. They model the high resolution image as a Markov random field (MRF), the parameters of which are learnt from the most zoomed observation. The learnt parameters are then used to obtain a maximum *aposteriori* (MAP) estimate of the high resolution image.

The method proposed in this paper can also be classified under learning based super-resolution schemes. However, here we use a different type of learning where we use a prior term that enforces the condition that the wavelet coefficients of the super-resolved image at the finest scale should be locally close to the best matching wavelets learnt from the high resolution training set. A smoothness constraint is imposed on the restored image to obtain a regularized solution.

III. Low Resolution Image Formation Model

The super-resolution problem can be cast in a restoration frame-work. We use a simple decimation model for low resolution image formation. For a decimation factor of q the low resolution image $y(i,j)$ can be obtained from its high resolution version $z(k,l)$ as

$$y(i,j) = \frac{1}{q^2} \sum_{k=qi}^{q(i+1)-1} \sum_{l=qj}^{q(j+1)-1} z(k,l) \tag{1}$$

i.e., the low resolution intensity is the average of the high resolution intensities over a neighborhood of q^2 pixels. This decimation model simulates the integration of light intensity that falls on the high resolution detector and the decimation process is represented by the matrix D which has a structure as given in equation (3). Let $\mathbf{z}$ represent the lexicographically ordered high resolution image of $N^2 \times 1$ pixels. If $\mathbf{y}$ is the $M^2 \times 1$ lexicographically ordered vector containing pixels from the low resolution observation, then it can be modeled as

$$\mathbf{y} = D\mathbf{z} + \mathbf{n} \tag{2}$$

where D is the decimation matrix, size of which depends on the decimation factor. For a decimation factor of q, the decimation matrix D consists of q^2 non-zero elements of value $\frac{1}{q^2}$ along each row at appropriate locations and has the form [13] (using a proper reordering of $\mathbf{z}$)

$$D = \frac{1}{q^2}\begin{bmatrix} 11\ldots1 & & \mathbf{0} \\ & 11\ldots1 & \\ & & \ddots \\ \mathbf{0} & & 11\ldots1 \end{bmatrix}. \tag{3}$$

As an example, for a decimation factor of $q = 2$ and with lexicographically ordered $\mathbf{z}$ of size, say 16×1, the D matrix is of size 4×16 and can be written as

$$D = \frac{1}{4}\begin{bmatrix} 1100110000000000 \\ 0011001100000000 \\ 0000000011001100 \\ 0000000000110011 \end{bmatrix}. \tag{4}$$

In equation (2) $\mathbf{n}$ is the $M^2 \times 1$ noise vector. We assume the noise to be zero mean i.i.d process. Our problem now reduces to estimating $\mathbf{z}$ given $\mathbf{y}$, which is an ill-posed, inverse problem.

IV. Wavelet based Learning

A. Background

Wavelets are mathematical functions that split up data into different frequency components locally, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities or a local analysis is required. The discrete wavelet transform (DWT), provides us with a sufficient information for analysis and synthesis of a sequence and is easier to implement. The idea here is similar to the continuous wavelet transform (CWT), which is computed by changing the scale of the analysis, shifting the window in time, multiplying by the sequence, and integrating over all times. In the case of DWT, filters of different cutoff frequencies are employed to analyze the sequence at different scales. The input sequence is passed through a series of highpass and lowpass filters

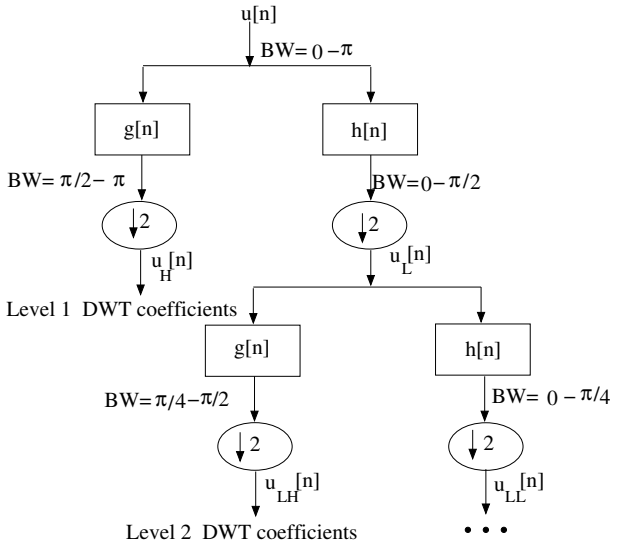


Fig. 1. Illustration of subband wavelet decomposition. Here $u[n]$ is the original sequence to be decomposed and $h[n]$ and $g[n]$ are lowpass and highpass filters, respectively. The bandwidth of the signal is marked as "BW".

to analyze the high and low frequency components, respectively. The procedure starts with passing the sequence through a half band ($0-\pi/2$ radians) digital lowpass filter with impulse response $h[n]$, thus removing all the frequencies that are above half of the highest frequency in the sequence. The filtered output is then subsampled by a factor of 2, simply by discarding every other sample since the sequence now has a highest frequency of $\pi/2$ radians instead of π. The lowpass filter thus halves the resolution, but leaves the scale unchanged. The subsequent subsampling by a factor of 2, however, changes the scale. This is illustrated in Figure 1.

The wavelet transform for a 2D sequence is similar to that of 1D decomposition. A 2D wavelet decomposition is first performed (horizontally) on the rows by applying lowpass and highpass filters. Then we perform the same operations vertically (on the columns) resulting in four subbands LL, LH, HL, HH. We repeat the operation with 'LL' as the input image for further decomposition. The readers are referred to [31], [32], [33] for further discussion on wavelet decomposition.

B. Learning the Wavelet Coefficients

As discussed in the previous section the wavelet decomposition splits the data into high and low frequency components. As seen from Figure 1, given a high resolution sequence $u[n]$ having a bandwidth support of $[0-\pi]$, it can be decomposed into u_L and u_H sequences constituting the low frequency and the high frequency components in the sequence, respectively. Let us consider that u_L (the low resolution sequence) is given and we need to generate the high resolution sequence $u[n]$. In order to do that we need to know the u_H so that when we take the IDWT (inverse discrete wavelet transform) we get back the original sequence $u[n]$. However, for the current problem on super-resolution, we do not have the high frequency components u_H to obtain the high resolution sequence $u[n]$. In the absence of any information on u_H, we plan to estimate the coefficients u_H by learning them from a set of high resolution sequences. Similarly, when a low resolution image or a 2D signal is considered we need to learn the corresponding unknown high frequency components u_{LH}, u_{HL} and u_{HH}. Since the problem of super-resolution involves handling data at multiple resolution, and since the wavelets are best suited for a multiresolution analysis, it motivates us to use a wavelet based approach for learning the wavelet coefficients at the finer resolution. These wavelet coefficients indicate the high frequency details in an image. The learning is done from a set of high resolution training images. If the high resolution data in a region does not have much high frequency components, the region can easily be obtained from its low resolution observation through a suitable interpolation. However, if a region has edges, the corresponding wavelet coefficients (u_H in Figure 1) are quite significant and they cannot be neglected while obtaining the high resolution image. These coefficients must be learnt from a database of training images. We assume that a primitive edge element in the high resolution image is localized to an 8×8 pixel area, and we observe the corresponding edge elements over a 4×4 pixel area in the low resolution image. From the high resolution data base, can we obtain the best 8×8 region by matching it in the wavelet domain with the given 4×4 pixel observation? Note that such a matching should be brightness (dc-shift) independent.

We make use of a two level wavelet decomposition of the given low resolution observation while learning the wavelet coefficients at the finer scale. Figure 2 illustrates the block

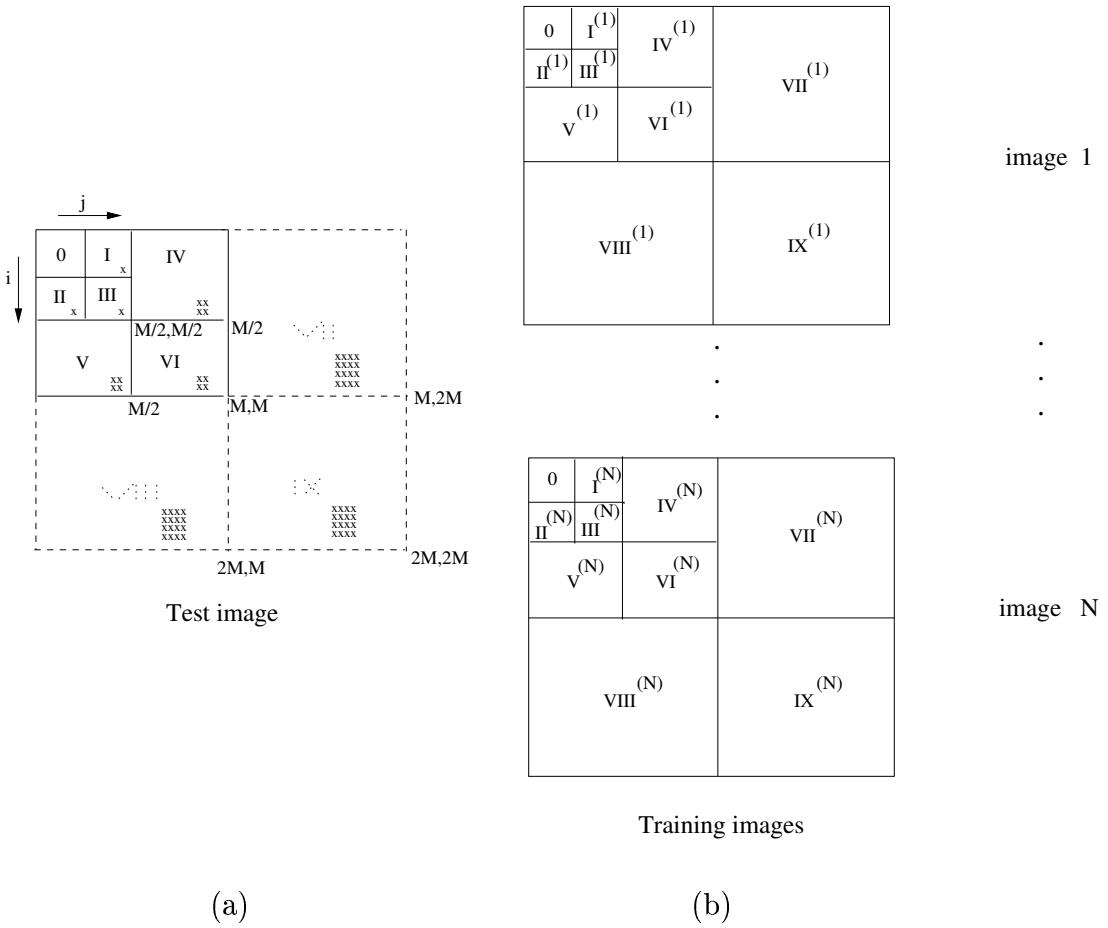


Fig. 2. Illustration of learning of wavelet coefficients at a finer scale. (a) Low resolution image with two level wavelet decomposition. Wavelet coefficients (marked as x) in subbands shown with the dotted lines are to be estimated for bands$VII - IX$. (b) High resolution training set in wavelet domain with three level decomposition.

schematic of how the wavelet coefficients at finer scales are learnt from a set of N training images using a two level wavelet decomposition of the low resolution test image. The high resolution training images are decomposed into three levels and the test image is compared to the training images in the wavelet domain at the coarser two scales. This decomposition is used to extrapolate the missing wavelet coefficients in subbands $VII-IX$ (shown as dotted in Figure 2(a)) for the test image. They correspond to the estimated high pass wavelet coefficients at the first level decomposition of the unknown high resolution image. Here the low resolution image is of size $M \times M$ pixels. Considering an upsampling factor of 2, the high resolution image, now has a size of $2M \times 2M$ pixels. For each

coefficient in the subbands $I - III$ and the corresponding 2×2 blocks in the subbands $IV - VI$, we extrapolate a block of 4×4 wavelet coefficients in each of the subbands VII, $VIII$ and IX. In order to do this we exploit the idea from zero tree concept, i.e., in a multiresolution system, every coefficient at a given scale can be related to a set of coefficients at the next coarser scale of similar orientation [34]. Using this idea we follow the minimum absolute difference (MAD) criterion to estimate the wavelet coefficients. We take the absolute difference locally between the wavelet coefficients in the low resolution image and the corresponding coefficients in each of the high resolution training images. The learning process is as follows. Consider the subbands $0 - VI$ of the low resolution image. Denote the wavelet coefficient at a location (i, j) as $d(i, j)$. Consider the range $0 \leq i, j \leq M/4$. The wavelet coefficients $d_I(i, j+M/4)$, $d_{II}(i+M/4, j)$, $d_{III}(i+M/4, j+M/4)$ corresponding to subbands $I-III$ and a 2×2 block consisting of $\sum_{k=i}^{k=i+1}\sum_{l=j}^{l=j+1} d_{IV}(k, l+M/2)$, $\sum_{k=i}^{k=i+1}\sum_{l=j}^{l=j+1} d_V(k+M/2, l)$, $\sum_{k=i}^{k=i+1}\sum_{l=j}^{l=j+1} d_{VI}(k+M/2, l+M/2)$, in each of the subbands $IV - VI$ are then considered to learn a 4×4 wavelet block in each of the subbands $VII - IX$ consisting of unknown coefficients $\sum_{k=i}^{k=i+3}\sum_{l=j}^{l=j+3} d_{VII}(k, l+M)$, $\sum_{k=i}^{k=i+3}\sum_{l=j}^{l=j+3} d_{VIII}(k+M, l)$, and $\sum_{k=i}^{k=i+3}\sum_{l=j}^{l=j+3} d_{IX}(k+M, l+M)$. In order to illustrate which set of wavelet coefficients we select for learning purposes, we denote them with 'x' marks in Figure 2(a). To obtain the wavelet coefficients for the test image at a finer resolution, we consider the wavelet coefficients in subbands $I - VI$ in each of the high resolution training images (see Figure 2(b)). We search for the best matching training image at a given location (i, j) that matches to the wavelet coefficients for the test image in the bands $I - VI$ in MAD sense and copy the corresponding high resolution wavelet coefficients, in bands $VII - IX$ to those bands for the test image. In effect, we use the following equation to find the minimum.

$$
\begin{aligned}
\hat{m}(i,j) &= \arg\min_{m}[|d_I(i, j+M/4) - d_{I^{(m)}}(i, j+M/4)| + |d_{II}(i+M/4, j) - d_{II^{(m)}}(i+M/4, j)| \\
&+ |d_{III}(i+M/4, j+M/4) - d_{III^{(m)}}(i+M/4, j+M/4)| \\
&+ \sum_{k=i}^{k=i+1}\sum_{l=j}^{l=j+1} |d_{IV}(k, l+M/2) - d_{IV^{(m)}}(k, l+M/2)| \\
&+ \sum_{k=i}^{k=i+1}\sum_{l=j}^{l=j+1} |d_V(k+M/2, l) - d_{V^{(m)}}(k+M/2, l)|
\end{aligned}
$$

$$+ \sum_{k=i}^{k=i+1} \sum_{l=j}^{l=j+1} |d_{VI}(k+M/2, l+M/2) - d_{VI^{(m)}}(k+M/2, l+M/2)|], \tag{5}$$

where $m = 1, 2, \cdots, N$. Here $d_J(m)$ denotes the wavelet coefficients for the m^{th} training image at the J^{th} band. For each (i, j) in $I-III$ of low resolution observation, a 4×4 block of wavelet coefficients in subbands $VII - IX$ from that training image given by $\hat{m}(i, j)$ which gives the minimum are then copied into subbands VII, $VIII$, IX of the observed image. In effect, equation (5) helps in matching edge primitives at low resolutions. Thus we have, $d_{VII}(i, j) := d^{(\hat{m})}_{VII}(i, j), d_{VIII}(i, j) := d^{(\hat{m})}_{VIII}(i, j), d_{IX}(i, j) := d^{(\hat{m})}_{IX}(i, j)$, for$(i, j) \in (VII-IX)$ where $\hat{m}$ is the index for the training image which gives the minimum at location (i, j). This is repeated for each coefficient in subbands I, II, III of the low resolution image. Thus for each coefficient in $I - III$, we learn 16 coefficients each, for subbands $VII-IX$ from the training set. It may be mentioned here that each 4×4 region in the low resolution image could be learnt from different training images. In case the error (MAD) term in equation (5) is quite large, it signifies that the 4×4 block does not find a good match in the training data, i.e., an edge primitive does not have its corresponding high resolution representation in the database. In order to avoid such spurious learning, we accept the wavelet coefficients only when the MAD is less than a chosen threshold. The goodness of the learning depends on how extensive and useful is the training data set. In our experiments we use Daub4 wavelet for computing the discrete wavelet transform. The issue of which particular wavelet basis best fit the learning scheme, has not been investigated in this paper. The subband 0 corresponds to the low resolution portion 'LL' (see Figure 2(a)) in the wavelet decomposition and since the corresponding 'LL' portions in the training set may have different brightness averages, including the pixels from 'LL' portion of the low resolution image 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 'LL' portion of the low resolution image for learning. The complete learning procedure is summarized below in terms of the steps involved.

STEP 1:

Perform two level wavelet decomposition on the low resolution test image of size $M \times M$

and three level decomposition on all training images each of size $2M \times 2M$.

STEP 2:

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

STEP 3:

Obtain the absolute difference between the wavelet coefficients in the low resolution image and the corresponding coefficients for each of the training images.

STEP 4:

If MAD $<$ *threshold, obtain the unknown high resolution wavelet coefficients (*4×4 *block) from a training image in subbands* $VII - IX$, *else set them all zeros.*

STEP 5:

Repeat steps (2 - 4) for every wavelet coefficient in bands $I - VI$ *of the low resolution image.*

A few comments about the learning of the wavelet coefficients are in order now. The high frequency 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.

In this study we select a 4×4 edge primitive in the low resolution image for learning the coefficients. A smaller primitive could provide a better localized result, but more spurious matches negate the advantage. A larger primitive yields better matches in the coefficient, but the localization is poor and suffers from severe blockiness. Furthermore, the requirement for the training data size goes up drastically.

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 $2m$/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 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.

V. Regularization And Super-Resolution Estimation

With the wavelet coefficients learnt from the high resolution training set as discussed in the previous section, we would like to obtain the super-resolution image for the given low resolution observation. Since we pick up the high frequency components of each 8×8 region as per the best fit edge element from different training data independently, there is no guarantee that the corresponding high resolution image would be a good one as it lacks any spatial context dependency. One may occasionally find an unwanted abrupt variation across the 8×8 blocks. In order to bring in a spatial coherence during the high resolution reconstruction, we must use a smoothness constraint. Thus the constraints are chosen based on enhancing the edges as well as ensuring the smoothness of the high resolution image. Near the edges in the low resolution image, we learn the wavelet coefficients from the high resolution database to have edge preserving upsampling. Also a smoothness constraint is enforced while upsampling at relatively smooth regions. We use the wavelet coefficients learnt from the training set to enforce the constraint that the wavelet coefficients of the super-resolved image should be close to the best matching wavelets learnt from the training images in a least squares sense. Let $\mathbf{Z}_{wt}$ be wavelet transform of the high resolution image to be estimated and $\hat{\mathbf{Z}}_{wt}$ be the wavelet transform of the learnt image as discussed in the previous section. Then the learning prior term can be expressed as

$$C(\mathbf{z}) = ||\mathbf{Z}_{wt} - \hat{\mathbf{Z}}_{wt}||^2. \tag{6}$$

Now in order to enforce the smoothness constraint we make use of the fact that the image pixel intensities have a spatial correlation. This prior knowledge serves as a contextual constraint and has to be used to regularize the solution. But this constraint pushes the reconstruction towards a smooth entity. Hence in order to enforce a smoothness in the smooth regions alone while upsampling, we use a discontinuity preserving smoothness prior. Since the high frequency details learnt by using the wavelet based prior constitute the discontinuities it would ensure undistorted edges in the super-resolved image

while smoothing the regions with spatial continuity. In order to incorporate provisions for detecting such discontinuities, so that they can be preserved in the reconstructed image, the binary variables $l_{i,j}$ and $v_{i,j}$ which detect the horizontal and vertical edges, respectively, are used. The binary variable $l_{i,j}$ connecting sites (pixel locations) (i,j) to $(i-1,j)$ aids in detecting a horizontal edge while the variable $v_{i,j}$ connecting sites (i,j) to $(i,j-1)$ helps in detecting a vertical edge. The variables $l_{i,j}$ and $v_{i,j}$ are set to 1 if $|z(i,j)-z(i-1,j)| > Threshold1$ and $|z(i,j)-z(i,j-1)| > Threshold2$, respectively. Else they are set to 0. We use the following prior for the smoothness constraint in this study.

$$\begin{aligned} U(\mathbf{z}) &= \sum_{i,j}\{\mu[(z_{i,j}-z_{i,j-1})^2(1-v_{i,j})+(z_{i,j+1}-z_{i,j})^2(1-v_{i,j+1}) \\ &+ (z_{i,j}-z_{i-1,j})^2(1-l_{i,j})+(z_{i+1,j}-z_{i,j})^2(1-l_{i+1,j})] \\ &+ \gamma(l_{i,j}+l_{i+1,j}+v_{i,j}+v_{i,j+1})\}. \end{aligned} \tag{7}$$

Here μ is the penalty term for departure from the smoothness. The second term in the above equation enforces a penalty for over-punctuation in the smoothness constraint. In effect we are considering only a first order spatial relationship along with the scope for handling the discontinuities. Thus by making use of the data fitting term, the learning term and the smoothness constraint the final cost function to be minimized for the high resolution image $\mathbf{z}$ can be expressed as

$$\varepsilon = ||\mathbf{y} - D\mathbf{z}||^2 + \beta C(\mathbf{z}) + U(\mathbf{z}). \tag{8}$$

The above cost function is nonconvex and also consists of terms in both spatial domain (the first and the third term) and in frequency domain (the second term). Hence it cannot be minimized by using a simple optimization technique such as gradient descent since it involves a differentiation of the cost function. We minimize the cost by using the simulated annealing technique which leads to a global minima. However, in order to provide a good initial guess and to speed up the computation, the result obtained by using the inverse transform of the learnt wavelet coefficients is used as the initial estimate for $\mathbf{z}$.

We now explain the various terms in equation (8) with respect to the wavelet based learning method. The first term relates to the consistency in data fitting. If $\mathbf{z}$ is the

actual HR image, we observe that $||\mathbf{y} - D\mathbf{z}||^2$ need not be zero as the chosen decimation operator D as defined in equation (3) need not be close to the wavelet decomposition (LL image in Figure 2) of the high resolution image, in general. The above is true only for Haar basis. However, the use of Haar basis introduces a lot more blockiness in the reconstructed image when the third (smoothness) term becomes very large. Alternately one may set all the wavelet coefficients in the finer subbands to be zero prior to taking the inverse wavelet transform. Although this may be similar in idea to the sinc interpolation, the corresponding interpolation results are quite inferior. The choice of Daub4 as the basis function in the study was more on an ad hoc basis, and a proper selection of the basis function would be an interesting topic of research. The selection of various weighting parameters in equation (8) was based on the idea that each term in the equation should have comparable magnitudes when the algorithm converges to the high resolution image.

VI. Experimental Results

In this section, we demonstrate the efficacy of the proposed technique to super-resolve a low resolution observation using the wavelet coefficients learnt from a high resolution training data set. We first present the results on gray scale images and then show that it works well for color images also.

First we consider experiments with face images. A number of high resolution images of different objects were downloaded from the Internet arbitrarily to use them as a training set. We considered a high resolution training set of $N = 200$. The same training data set has been used in all experiments. In order to obtain a low resolution test image, we consider a high resolution image from the training set and downsample it by a factor of 2. Figure 3(a) shows one such low resolution face image of size 64×64. Figure 3(b) shows the same image upsampled by a factor of 2 using the bicubic interpolation technique. The super-resolved image is shown in Figure 3(c). A Comparison of the Figures 3(b) and (c) shows more clear details in the super-resolved image. The features such as eyes, nose and the mouth appear blurred in the interpolated image shown in Figure 3(b), while they are restored well in Figure 3(c). Also the eye balls are sharper in the displayed super-resolved image. It has been experimentally found that the best results are obtained with the parameters $\mu = 0.01$, $\gamma = 25$, the weight for the learning term $\beta = 0.08$ and $Threshold1 =$

$Threshold2 = 30$. These parameters were selected so that all the components in the cost function (refer to equation (8)) have comparable contributions. We retain the same values for the parameters in all subsequent experiments.

Next we consider another experiment on face image. The low resolution observation obtained by down sampling the high resolution Lena image is shown in Figure 4(a). The super-resolution result obtained using the proposed approach is displayed in Figure 4(c), and Figure 4(b) shows the bicubic interpolated image. Once again we see that the high frequency details are better preserved in the super-resolved image. Various surface boundaries are much sharper. The hair strand and the lace on the hat appears more clearly. The eyes and the nose are also clear. However, we observe little blockiness on the boundary curves of the hat and the slanted structure on the upper right corner of the picture. Furthermore, the edge primitives are chosen over a 8×8 block in the wavelet domain. Hence the learned edges may suffer from blockiness. The smoothness constraint is supposed to take care of such jaggedness. However, the $Threshold$ value being chosen on an ad hoc basis it fails to undo the jaggedness. But this blockiness is nothing compared to the blockiness one obtains when a simple pixel replication is used. Comparing this with the simple zero order hold expanded image shown in Figure 4(d) (in which every feature in the image appears blocky), we see that the blockiness is negligible in the proposed approach. We could have played with the parameter set in equations (7) and (8), but the various parameters for recovering the super-resolved image for this experiment were kept the same as used in the previous experiment.

In order to test our algorithm for an image which has prominent edges, we considered a portion of a building image. The results for the same are shown in Figures 5(b-c) with the low resolution observation depicted in Figure 5(a). We can clearly see that the discontinuities are better estimated in the super-resolved image shown in Figure 5(c), but they appear blurred in the bicubic interpolated image (see Figure 5(b)). This substantiates our claim that the learning of wavelet coefficients does help in improving the resolutions.

We now consider a few experiments on the color image super-resolution. For these experiments we first convert the low resolution color image into $Y - C_b - C_r$ format. The learning of the wavelet coefficients is then done using the Y (luminance) plane only. The

(a) (b) (c)

Fig. 3. (a) A low resolution observation (face1), (b) bicubic interpolated image, (c) and the super-resolved image using the proposed approach.

(a) (b) (c) (d)

Fig. 4. (a) Another low resolution observation (Lena), (b) bicubic interpolated image, (c) and the super-resolved image, and (d) result of simple pixel replication.

(a) (b) (c)

Fig. 5. (a) Low resolution observation of a building, (b) upsampled by bicubic interpolation, (c) and the super-resolved image.

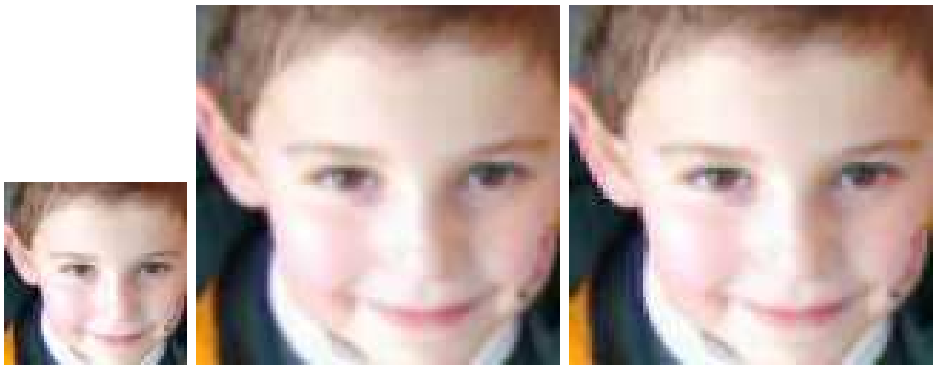
recovered high resolution image in the luminance plane after optimization is then combined with the bicubic interpolated version of the data in low resolution $C_b - C_r$ planes in order to obtain the super-resolved color image. The idea is quite similar to the way a macroblock is represented by $4:1:1$ DCT blocks in the $Y - C_b - C_r$ domain while using an MPEG coder. The training images used were kept the same as in the previous experiments on gray scale images. One may note here that learning of the wavelet coefficients for the Y, C_b, and C_r planes can also be done separately in order to obtain the super-resolution on each of the low resolution images. However, we refrain from doing it in this paper as any possible error in learning in any of the color planes may introduce chromatic distortions and the human vision appears to be sensitive to that.

We now show results of two experiments conducted on the color face images. Due to printing restrictions, the results are shown here in gray tone. But the color images can be seen at this website [35]. Figure 6(c) shows the result of the proposed approach on a low resolution observation shown in Figure 6(a). Compare this with the bicubic interpolated image shown in Figure 6(b). We observe that the super-resolved image appears sharper. Few areas of interest where such an enhancement can be observed are the mark on the left chin, eye balls and the hair. The results for another low resolution face image are displayed in Figures 7(a-c). Similar conclusions can again be drawn from this experiment. Observe the eye balls, eye brows, frontal hair, and the nose shown in Figure 7(c) which appear sharper when compared to the bicubic interpolated image given in Figure 7(b). Thus we conclude that our approach works well for color images as well.

Finally, in order to convey the comparative edge over the conventional interpolation techniques, we show the mean squared error (MSE) during interpolation for the gray scale images. Table I shows the comparison of the proposed method with the standard bilinear interpolation, bicubic interpolation and the Lanczos method. In order to be able to compute the MSE, we started with a high resolution image and the decimated version of that was used as the low resolution observation. The mean squared error between the original image and generated super-resolved image is defined as

$$MSE = \frac{\sum_{i,j}(z(i,j) - \hat{z}(i,j))^2}{\sum_{i,j}(z(i,j))^2}.$$

We can observe that in addition to the perceptual betterment in all observed images there



(a) (b) (c)

Fig. 6. (a) A low resolution observation, (b) upsampling using the bicubic interpolation, and (c) the super-resolved image.

(a) (b) (c)

Fig. 7. (a) Another low resolution face image, (b) upsampling using the bicubic interpolation, (c) the super-resolved image.

is also a drop in MSE for the proposed approach. This illustrates the usefulness of the wavelet based learning scheme in super-resolving the images.

VII. Conclusions

We have described a method for super-resolution restoration of images using a wavelet based learning technique. The wavelet coefficients at finer scales, learnt from a set of several high resolution training images, are used as a constraint along with an appropriate smoothness prior to estimate the super-resolved image. The learning term selects the best high resolution edges from the training set given a low resolution observation, while the discontinuity preserving smoothness term ensures a proper spatial correlation among pixel

Method	face1	Lena	building
Bilinear	0.0049	0.0075	0.0092
Bicubic	0.0043	0.0063	0.0073
Lanczos	0.0041	0.0061	0.0065
Proposed	0.0032	0.0056	0.0061

TABLE I

COMPARISON OF MSES OF DIFFERENT SCHEMES.

intensities. The results obtained for both gray scale and color images show perceptual as well as quantifiable improvements over conventional interpolation techniques. 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.

Acknowledgment

Partial funding in the form of a *Swarnajayanti Fellowship* is gratefully acknowledged. The authors are grateful to the reviewers for constructive comments.

REFERENCES

[1] R. Y. Tsai and T. S. Huang, "Multiframe Image Restoration and Registration," in *Advances in Computer Vision and Image Processsing*, pp. 317–339. JAI Press Inc., 1984.

[2] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image from Noisy Undersampled Multiframes," *IEEE Trans. on Accoustics, Speech and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, June 1990.

[3] H. Ur and D. Gross, "Improved Resolution from Sub-pixel Shifted pictures," *CVGIP:Graphical Models and Image Processing*, vol. 54, pp. 181–186, March 1992.

[4] M. Irani and S. Peleg, "Improving Resolution by Image Registration," *CVGIP:Graphical Models and Image Processing*, vol. 53, pp. 231–239, March 1991.

[5] M. Irani and S. Peleg, "Motion Analysis for Image Enhancement : Resolution, Occlusion, and Transparency," *VCIR*, vol. 4, pp. 324–335, December 1993.

[6] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High Resolution Image Reconstruction from Lower-Resolution Image Sequences and Space-Varying Image restoration," in *Proc.IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, San Francisco,USA, 1992, pp. 169–172.

[7] M. K. Ng, N. K. Bose, and J. Koo, "Constrained Total Least Squares Computation for High Resolution Image Reconstruction with Multisensors," *International Journal of Imaging Systems and Technology*, vol. 12, pp. 35–42, 2000.

[8] M. K. Ng and N. K. Bose, "Analysis of Displacement Errors in High Resolution Image Reconstruction with Multisensors," *IEEE Trans. Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 6, pp. 806–813, June 2002.

[9] M. K. Ng and N. K. Bose, "Fast Color Image Restoration with Multisensors," *International Journal of Imaging Systems and Technoloy*, vol. 12, no. 5, pp. 189–197, 2002.

[10] N. Nguyen, P. Milanfar, and G. Golub, "A Computationally Efficient Super-resolution Reconstruction Algorithm," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 573–583, April 2001.

[11] Z. Lin and H. Y. Shum, "Fundamental Limits of Reconstruction-Based Super-Resolution Algorithms under Local Translation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, January 2004.

[12] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP Registration and High- Resolution Image Estimation Using a Sequence of Undersampled Images," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1621–1633, December 1997.

[13] R. R. Schultz and R. L. Stevenson, "A Bayesian Approach to Image Expansion for Improved Definition," *IEEE Trans. on Image Processing*, vol. 3, no. 3, pp. 233–242, May 1994.

[14] D. Rajan and S. Chaudhuri, "Generation of Super-resolution Images from Blurred Observations using an MRF Model," *J. Mathematical Imaging and Vision*, vol. 16, pp. 5–15, 2002.

[15] D. Rajan and S. Chaudhuri, "Simultaneous Estimation of Super-Resolved Intensity and Depth Maps from Low Resolution Defocussed Observations of a Scene," in *Proc. IEEE Int. Conf. on Computer Vision*, Vancouver Canada, 2001, pp. 113–118.

[16] A. N. Rajagopalan and V. P. Kiran, "Motion-free Super-resolution and the Role of Relative Blur," *Journal of the Optical Society of America A*, vol. 20, no. 11, pp. 2022–2032, November 2003.

[17] M. Elad and A. Feuer, "Restoration of a Single Superresolution Image from Several Blurred, Noisy and Undersampled Measured Images," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1646–1658, December 1997.

[18] M. Elad and A. Feuer, "Super-resolution Restoration of an Image Sequence : Adaptive Filtering Approach," *IEEE Trans. on Image Processing*, vol. 8, no. 3, pp. 387–395, March 1999.

[19] M. Elad and Y. Hel-Or, "A Fast Super-resolution Reconstuction Algorithm for Pure Translation Motion and Common Space-Invariant Blur," *IEEE Trans. on Image Processing*, vol. 10, no. 8, pp. 1187–1193, August 2001.

[20] M. C. Chiang and T. E. Boult, "Local Blur Estimation and Super-Resolution," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Puerto Rico, USA, 1997, pp. 821–826.

[21] N. Kaulgud, J. Karlekar, and U. B. Desai, "Compressed Domain Video Zooming : Use of Motion Vectors," in *Proc. of Eighth National Conf. on Communications*, Mumbai, India, Jan. 2002, pp. 120–124.

[22] S. Thurnhofer and S. K. Mitra, "Edge-preserving Image Zooming," *Optical Engineering*, vol. 35, no. 7, pp. 1862–1870, 1996.

[23] F. Malagouyres and F. Guichard, "Edge Direction Preserving Image Zooming: A Mathematical and Numerical Analysis," *SIAM Journal on Numerical Analysis*, vol. 39, no. 1, pp. 1–37, 2001.

[24] D. Capel and A. Zisserman, "Super-Resolution from Multiple Views using Learnt Image Models," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2001, pp. II:627–634.

[25] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, September 2002.

[26] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image Analogies," in *Proc. ACM-SIGGRAPH*, 2001, pp. 327–340.

[27] F. M. Candocia and J. C. Principe, "Super-Resolution of Images based on Local Correlations," *IEEE Trans. on Neural Networks*, vol. 10, no. 2, pp. 372–380, March 1999.

[28] C. M. Bishop, A. Blake, and B. Marthi, "Super-Resolution Enhancement of Video," in *Int. Conf. on Artificial Intelligence and Statistics*, Key West, Florida, 2003.

[29] 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, March/April 2002.

[30] M. V. Joshi and S. Chaudhuri, "A Learning-Based Method for Image Super-Resolution from Zoomed Observations," in *Proc. fifth Int. Conf. on Advances in Pattern Recognition*, Indian Statistical Institute, Kolkata, India, 2003, pp. 179–182.

[31] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, New Jersy, 1998.

[32] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pennsylvania, 1992.

[33] A. Jensen and A. Cour-Harbo, *Ripples in Mathematics: The Discrete Wavelet Transform*, Springer-Verlag, 2001.

[34] H. M. Shapiro, "Embedded Image Coding," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, December 1993.

[35] *http://www.ee.iitb.ac.in/~sc/papers/ijist.pdf.*