Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 73767, Pages 1–11
DOI 10.1155/ASP/2006/73767

Single-Frame Image Super-resolution through Contourlet Learning

C. V. Jiji and Subhasis Chaudhuri

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India

Received 26 November 2004; Revised 22 March 2005; Accepted 5 April 2005

We propose a learning-based, single-image super-resolution reconstruction technique using the contourlet transform, which is capable of capturing the smoothness along contours making use of directional decompositions. The contourlet coefficients at finer scales of the unknown high-resolution image are learned locally from a set of high-resolution training images, the inverse contourlet transform of which recovers the super-resolved image. In effect, we learn the high-resolution representation of an oriented edge primitive from the training data. Our experiments show that the proposed approach outperforms standard interpolation techniques as well as a standard (Cartesian) wavelet-based learning both visually and in terms of the PSNR values, especially for images with arbitrarily oriented edges.

1. 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 (LR) image. *Super-resolution* refers to the process of producing a high spatial resolution image than what is afforded by the physical sensor through postprocessing, making use of one or more low-resolution observations. It includes upsampling the image, thereby increasing the maximum spatial frequency, and removing degradations that arise during the image capture, namely, aliasing and blurring. In general, there are two classes of super-resolution techniques: reconstruction-based and learning-based. In reconstruction-based techniques the high-resolution (HR) image is recovered from several low-resolution observations of the input, but in learning-based super-resolution algorithms a database of several other images are used to obtain the high-resolution image.

The single-frame image super-resolution problem arises in several practical situations. In many biometric databases, a large number of images of similar content, shape, and size are available. For example, in investigative criminology one has available face and fingerprint databases. These are often taken in a controlled environment. The question we ask is that if one encounters a poor-quality input image, can it be enhanced using the knowledge of the properties of the database images. Thus, the basic problem that we solve in this paper is as follows. Given a single low-resolution input image and a database of several high-resolution images, we obtain a high-resolution output. We make use of the recently proposed contourlet transform [1] to learn the best features from the database of high-resolution images while upsampling the input image. The features we learn from the HR database are the HR representation of LR-oriented edges in the image to be super-resolved.

The remainder of the paper is organized as follows. In Section 2 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 3. The theory of discrete contourlet transform and the learning procedure for estimating the contourlet coefficients at finer scales using high-resolution training images are the contents of Section 4. We present experimental results on different types of images in Section 5, and the paper concludes with Section 6.

2. RELATED WORK

Numerous reconstruction-based super-resolution algorithms have been proposed in the literature. The super-resolution idea was first proposed by Tsai and Huang who used the frequency-domain approach [2]. A different approach to the super-resolution restoration problem was suggested by Irani and Peleg [3, 4] based on the iterative back projection method. A set-theoretic approach to the super-resolution restoration problem was suggested in [5].

The main result there is to define convex sets which represent tight constraints on the solution to improve the results. Ng et al. developed a regularized, constrained, total least squares solution to obtain a high-resolution image [6]. They consider the presence of perturbation errors of displacements around the ideal subpixel 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 [7]. They also develop a fast restoration algorithm for color images in [8]. Nguyen et al. proposed circulant block preconditioners to accelerate the conjugate gradient descent method while solving the Tikhonov-regularized super-resolution problem [9]. A maximum a posteriori (MAP) estimator with Huber-Markov random field (MRF) prior is described by Schultz and Stevenson in [10]. Other approaches include a MAP-MRF-based super-resolution technique using the blur as a cue [11]. In [12], the authors recover both the high-resolution scene intensity and the depth fields simultaneously using the defocus cue. Elad and Feuer [13] 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 projection onto convex sets (POCS) approaches. In [14] Lin and Shum determine the quantitative limits of reconstruction-based super-resolution algorithms and obtain the upsampling limits from the conditioning analysis of the coefficient matrix.

Now we review some of the recent works under the learning-based super-resolution category. In [15] 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. Authors in [16] have proposed a super-resolution technique from multiple views using learned image models making use of principal component analysis (PCA). Their method uses learned image models either to directly constrain the maximum likelihood (ML) estimate or as a prior for a MAP estimate. In [17] Freeman et al. proposed a parametric Markov network to learn the statistics between the "scene" and the "image," as a framework for handling low-level vision tasks, one application of which is super-resolution. An image analogy method applied to super-resolution is discussed in [18]. Joshi and Chaudhuri [19] 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 learned from the most zoomed observation. The learned parameters are then used to obtain a maximum a posteriori (MAP) estimate of the high-resolution image.

In [20], we have proposed a single-frame super-resolution algorithm using a wavelet-based learning technique where the HR edge primitives are learned from the HR data set locally. An eigenface-domain super-resolution reconstruction algorithm for face recognition is proposed in [21]. In the face hallucination technique proposed in [22], the authors use both low- and high-resolution image databases to recover the high-resolution image, making use of PCA. They also add constraints to the principal components to reduce the nonface-like distortion. The use of PCA for image zooming purposes has been investigated in [23]. It has been assumed that the principal components remain unchanged across the scale. The method is applicable only to zooming up of images of a specific class of objects such as faces or fingerprints. Pickup et al. [24] present a domain-specific image prior in the form of a distribution function based upon sampled images, and show that for certain types of super-resolution problems, this sample-based prior gives a significant improvement over other common multiple-image super-resolution techniques.

In [25] Chang et al. have proposed a single-frame image super-resolution method where the generation of the high-resolution image patch depends simultaneously on multiple nearest neighbors in the training set in a way similar to the concept of locally linear embedding for manifold learning. This method requires fewer training examples than other learning-based super-resolution methods. The super-resolution method proposed in [26] is the extension of a Markov-based learning algorithm, capable of processing an LR image with unknown degradation parameters. A different method for enhancing the resolution of LR facial images using an error back projection method based on top-down learning is proposed in [27]. Here a face is represented by a linear combination of prototypes of shape and texture. An image hallucination approach based on primal sketch priors is presented in [28]. Here a reconstruction constraint is also applied to further improve the quality of the hallucinated image. In [29], the super-resolution reconstruction problem is considered as a binary classification problem and is solved through class conditional probability estimation.

Most of the learning-based super-resolution methods proposed above either make use of a database of low- and high-resolution training images of similar objects or use an appropriate smoothness constraint along with the learning prior to improve the results. In our method we use instead an arbitrary set of high-resolution training images. Also we do not use any smoothness constraint as we use the contourlet transform which has the capability to capture smoothness along contours, while learning the best edge primitives from the HR training set. The proposed method is edge-based and involves learning the edge pattern locally instead of the global PCA-based approach. As a result, our method is faster and results show considerable improvement over a regularization-based approach.

3. LOW-RESOLUTION IMAGE FORMATION MODEL

It is assumed that the observed low-resolution image is produced from a single high-resolution image under the following generative model. Let z represent the lexicographically ordered high-resolution image of $N^2 \times 1$ pixels. If y is the $M^2 \times 1$ lexicographically ordered vector containing pixels

from the low-resolution observation, then it can be modeled as

$$y = DBz + n, \tag{1}$$

where D is the decimation matrix, size of which depends on the decimation factor and B is the blur matrix. For a decimation factor of q, $N = qM$ and the decimation matrix D consists of q^2 nonzero elements of value $1/q^2$ along each row at appropriate locations and has the form [10] (using a proper reordering of z)

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

As an example, for a decimation factor of $q = 2$ and with lexicographically ordered z of size, say 16×1, the D matrix is of size 4×16 and can be written as (without reordering z)

$$D = \frac{1}{4}\begin{bmatrix} 1100110000000000 \\ 0011001100000000 \\ 0000000011001100 \\ 0000000000110011 \end{bmatrix}. \tag{3}$$

In (1), n is the $M^2 \times 1$ noise vector. We assume the noise to be zero mean but no specific distribution is assumed in this work. The low-resolution image formation model is illustrated in Figure 1. The problem we investigate in this paper is as follows. Given a single instance of y, obtain an estimate of the high-resolution image z. We assume in this paper B to be an identity matrix. Also, since we are using only a single observation, no effort is made to reduce the effect of noise (see (1)) in case the original LR image does suffer from noise perturbations. Since the matrix D is not invertible, an arbitrary set of high-resolution images is used as training data to recover z.

4. CONTOURLET-BASED LEARNING

4.1. Contourlet transform

The contourlet transform [1, 30] is an extension of the Cartesian wavelet transform in two dimensions using multiscale and directional filter banks. The contourlet expansion of images consists of basis images oriented at various directions in multiple scales, with flexible aspect ratios. Thus the contourlet transform retains the multiscale and time-frequency localization properties of wavelets. In addition, it also offers a high degree of directionality. Thus they are capable of capturing the geometrical smoothness of the contour along any possible direction. The contourlet transform does not use separable basis functions. The idea here is not to decompose an image into horizontal and vertical edges, but to capture the edges normal to the contour present in the image.

The contourlet transform is implemented in two stages: the subband (spectral) decomposition stage and the directional decomposition stages. For the subband decomposition

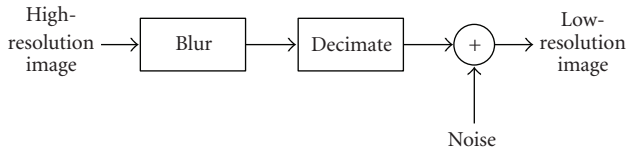

Figure 1: Illustration of low-resolution image formation model.

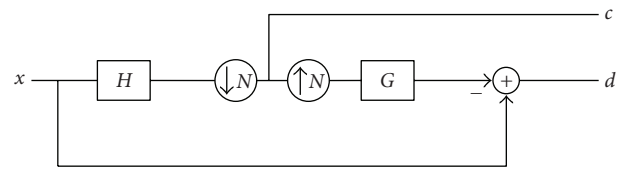

Figure 2: The Laplacian pyramid decomposition. The outputs are a coarse approximation c and a difference d between the original and the prediction.

stage we use the Laplacian pyramid introduced by Burt and Adelson [31] where the decomposition at each step generates a sampled lowpass version of the original and the difference between the original image and the prediction as shown in Figure 2. The input image x is first lowpass filtered using filter H and then decimated to get a coarse approximation c. This is then interpolated and passed through the synthesis filter G. The resulting image is then subtracted from the original image x to obtain the bandpass image d. The process is then iterated on the coarser version of the image c.

The directional filter bank (DFB) is efficiently implemented by using an m-level binary tree decomposition that leads to 2^m subbands with wedge-shaped frequency partitioning as shown in Figure 3(a). The original construction of the DFB proposed in [32] involves modulating the input image and uses quincunx filter banks with diamond-shaped filters [33]. The desired frequency partitioning is obtained by following a tree expanding rule for finer directional subbands as described in [34]. An m-level tree-structured DFB is equivalent to a 2^m parallel channel filter bank with equivalent filters and overall sampling matrices as shown in Figure 3(b). The equivalent analysis and synthesis filters are denoted by H_k and G_k, $0 \le k < 2^m$, respectively, corresponding to the subbands indexed as in Figure 3(a). The corresponding overall sampling matrices will have the following diagonal form [30]:

$$S_k = \begin{cases} \text{diag}\,(2^{m-1}, 2) & \text{for } 0 \le k < 2^{m-1}, \\ \text{diag}\,(2, 2^{m-1}) & \text{for } 2^{m-1} \le k < 2^m. \end{cases} \tag{4}$$

The two sets correspond to the mostly horizontal and mostly vertical set of directions, respectively. Thus the family $\{g_k[n - S_k l]\}_{0 \le k < 2^m,\, l \in Z^2}$ obtained by translating the impulse responses of the equivalent synthesis filters G_k over the sampling lattices S_k, provide a *basis* for discrete signals in $l^2(Z^2)$. This basis exhibits both directional and localization properties. For a detailed description of contourlet decomposition, readers are referred to [30, 32, 34].

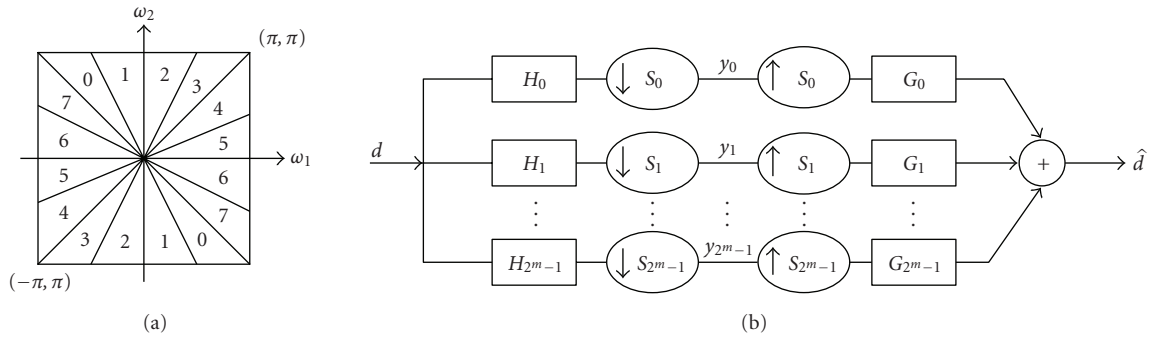

Figure 3: (a) An example of the directional filter bank frequency partitioning with $m = 3$. The subbands 0–3 correspond to the *mostly horizontal* directions, while subbands 4–7 correspond to the *mostly vertical* directions. (b) Multichannel view of an m-level tree-structured directional filter bank.

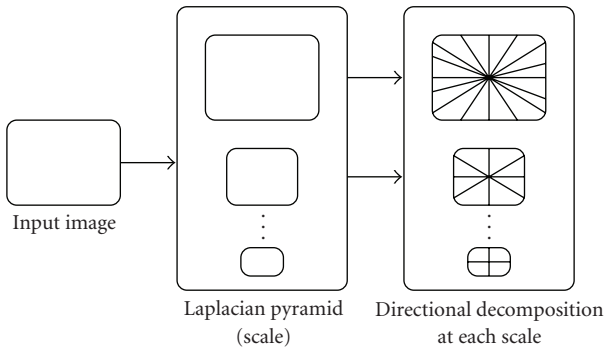

Figure 4: Pyramidal directional filter bank structure that implements the contourlet transform.

Combining the Laplacian pyramid and the directional filter bank yields the discrete contourlet transform. The multiscale and directional decomposition stages in the contourlet transform are independent of each other and hence each scale can be decomposed into any arbitrary power of two number of directions and different scales into different number of directions. Figure 4 shows the pyramidal directional filter bank structure that implements the contourlet transform. Figure 5 shows the three-scale contourlet decomposition of the Lena image for the purpose of illustration. It may be noted that the coefficients at the finer scale are no longer horizontally or vertically oriented as is the case in Cartesian wavelet decomposition.

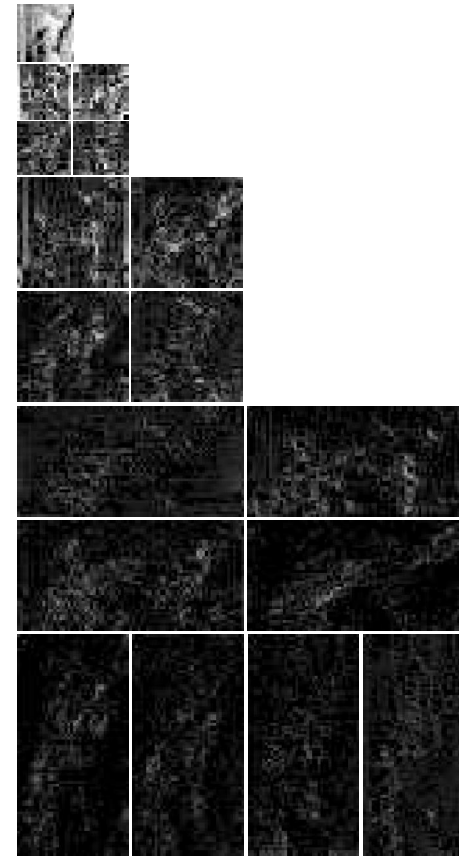

Figure 5: Contourlet transform of the Lena image using three Laplacian pyramidal levels and eight directions at the finest level and four directions at the coarser levels.

4.2. Learning of edge primitives

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 wavelets are known to capture the high-frequency details very well locally, we propose to use wavelets to learn this mapping. In the single-frame super-resolution algorithm proposed in [20], we used a wavelet-based learning technique where the HR edge primitives are learned from the HR data set locally 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. Each local region is learned independently from the HR data

set. One of the major difficulties with wavelet-based learning lies in the fact that the wavelet decomposition kernel is separable. Although this provides computational advantages, we expect to catch only the horizontal and vertical edges properly. Hence we do not have difficulties in learning horizontal and vertical edges, but we do have some problem in learning edges oriented along arbitrary directions. This gave rise to certain artifacts in the reconstructed image and in order to get a good-quality super-resolved image, we were forced to use an appropriate discontinuity preserving smoothness constraint under a regularization framework. Thus we ensure spatial correlation among pixels using the smoothness constraint, as well as obtain the best matching edges from the training set using wavelet learning. This required a stochastic optimization technique to obtain the solution which made the reconstruction process very slow.

A better way to handle the above situation is to use directionally selective wavelet decomposition to learn the oriented edges where the reconstruction problem need not be solved under a regularization framework, resulting in a much faster solution. This motivated us to use the contourlet transform which is capable of catching the smoothness along contours naturally. 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.

4.3. Learning the contourlet coefficients

Given a low-resolution input image y, we perform a contourlet decomposition consisting of two pyramidal levels and each pyramidal level is then decomposed into four directional subbands which yield the decomposition as shown in Figure 6(a). A three-level decomposition is performed on all the high-resolution database images and each pyramidal level is decomposed into four directional subbands resulting in the decomposition as shown in Figure 6(b). Our idea is to learn the contourlet coefficients in the four directional subbands corresponding to the finest level for the given low-resolution image (shown with dotted lines in Figure 6(a)). After learning, effectively we have a three-level decomposition for the input image, that is, the original low-level decomposition coefficients plus the learned coefficients at the finer scale. The inverse transform of this will yield the high-resolution equivalent of the low-resolution input.

Figure 6 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 subbands IX–XII for the test image.

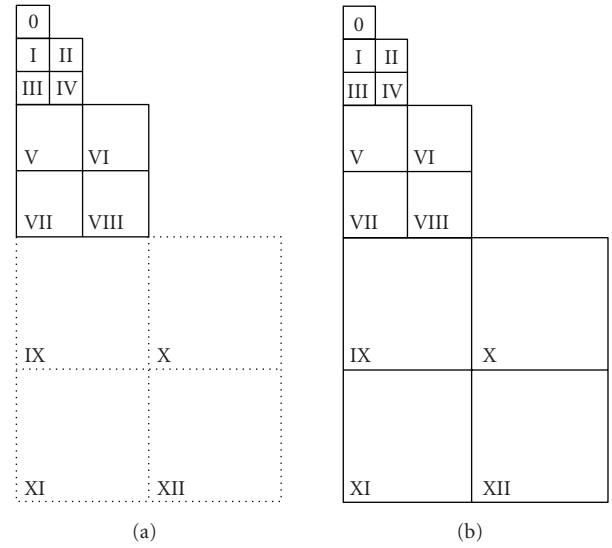

Figure 6: Illustration of learning the contourlet coefficients at a finer scale. (a) A low-resolution image with two-level decomposition. Coefficients in the dotted subbands are to be learned. (b) A representative high-resolution training set in contourlet domain with three-level decomposition.

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–IV and the corresponding 2×2 blocks in the subbands V–VIII, we extrapolate a block of 4×4 contourlet coefficients in each of the subbands IX, X, XI, and XII. In order to do this we exploit the idea from zerotree concept, that is, 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 [35]. 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 high-resolution training images.

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 $d(i, j)$. Consider the range $0 \le i, j \le M/4$. The contourlet coefficients $d_{\mathrm{I}}(i, j)$, $d_{\mathrm{II}}(i, j + M/4)$, $d_{\mathrm{III}}(i + M/4, j)$, $d_{\mathrm{IV}}(i + M/4, j + M/4)$ corresponding to subbands I–IV and 2×2 blocks consisting of $d_{\mathrm{V}}(k, l)$, $d_{\mathrm{VI}}(k, l+M/2)$, $d_{\mathrm{VII}}(k+M/2, l)$, $d_{\mathrm{VIII}}(k+M/2, l+M/2)$ for $k = 2i : 2i + 1$ and $l = 2j : 2j + 1$ corresponding to subbands V–VIII in the low-resolution test image and all the high-resolution training images are considered to learn a 4×4 contourlet block in each of the subbands IX–XII consisting of unknown coefficients $d_{\mathrm{IX}}(k, l)$, $d_{\mathrm{X}}(k, l + M)$, $d_{\mathrm{XI}}(k + M, l)$, $d_{\mathrm{XII}}(k + M, l + M)$ for $k = 4i : 4i + 3$ and $l = 4j : 4j + 3$. Thus, for a given set of a total of twenty contourlet coefficients in subbands 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 4×4 contourlet block in

bands IX–XII to those bands for the test image. In effect, we use the following equation to find the minimum:

$$\begin{aligned}\hat{m}(\hat{p},\hat{q}) &= \arg\min_{m,p,q}\big[\,|\,d_{\mathrm{I}}(i,j)-d_{\mathrm{I}^{(m)}}(p,q)\,| \\ &\quad+|\,d_{\mathrm{II}}(i+M_1,j)-d_{\mathrm{II}^{(m)}}(p+M_1,q)\,| \\ &\quad+|\,d_{\mathrm{III}}(i,j+M_1)-d_{\mathrm{III}^{(m)}}(p,q+M_1)\,| \\ &\quad+|\,d_{\mathrm{IV}}(i+M_1,j+M_1) \\ &\qquad-d_{\mathrm{IV}^{(m)}}(p+M_1,q+M_1)\,| \\ &\quad+S_{\mathrm{V}}+S_{VI}+S_{VII}+S_{\mathrm{VIII}}\big],\end{aligned} \tag{5}$$

where $M_1 = M/4$ and

$$\begin{aligned}S_{\mathrm{V}} &= |\,d_{\mathrm{V}}(2i,2j)-d_{\mathrm{V}^{(m)}}(2p,2q)\,| \\ &\quad+|\,d_{\mathrm{V}}(2i,2j+1)-d_{\mathrm{V}^{(m)}}(2p,2q+1)\,| \\ &\quad+|\,d_{\mathrm{V}}(2i+1,2j)-d_{\mathrm{V}^{(m)}}(2p+1,2q)\,| \\ &\quad+|\,d_{\mathrm{V}}(2i+1,2j+1)-d_{\mathrm{V}^{(m)}}(2p+1,2q+1)\,|\end{aligned} \tag{6}$$

and S_{VI}, S_{VII}, and S_{VIII} are the corresponding sums for subbands VI, VII, and VIII, respectively, and $m = 1,2,\ldots,K$. Here $d_J(m)$ denotes the contourlet coefficients for the mth training image at the Jth subband. Here $\hat{m}(\hat{p},\hat{q})$ denotes the $(\hat{p},\hat{q})$th location for the $\hat{m}$th training image that best matches the test image at (i,j)th location in terms of contourlet coefficients.

Thus we have

$$\begin{aligned}d_{\mathrm{IX}}(i,j) &:= d_{\mathrm{IX}}(\hat{m})(\hat{p},\hat{q}),\\ d_{\mathrm{X}}(i,j) &:= d_{\mathrm{X}}(\hat{m})(\hat{p},\hat{q}),\\ d_{\mathrm{XI}}(i,j) &:= d_{\mathrm{XI}}(\hat{m})(\hat{p},\hat{q}),\\ d_{\mathrm{XII}}(i,j) &:= d_{\mathrm{XII}}(\hat{m})(\hat{p},\hat{q})\end{aligned} \tag{7}$$

for $(i,j) \in (\mathrm{IX} - -\mathrm{XII})$. This is repeated for each coefficient in subbands I, II, III, and IV of the low-resolution image. In effect, we find the best matching 8×8 edge primitive from the training data for a given 4×4 representation in the low-resolution image through contourlet expansion.

It may be mentioned here that each 4×4 region in the low-resolution image is being learned from different training images independently. In case the MAD error is quite large, it signifies that the 4×4 block does not find a good match in the training data, that is, an edge primitive does not have its corresponding high-resolution representation in the database. Such spurious learning will introduce unwanted artifacts in the reconstructed image. In order to avoid such artifacts, we accept the contourlet coefficients only when the MAD is less than a chosen threshold. The goodness of the learning depends on how extensive and useful the training data set is. The subband 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.

In our experiments we used "9–7" biorthogonal filters [36] for the Laplacian pyramid because they are close to being orthogonal and also because of their linear phase characteristics. For the directional filter banks we used the "23–45" biorthogonal quincunx filters designed by Phoong et al. [37] and modulated them to obtain the biorthogonal fan filters. These filters are also nearly orthogonal and have linear phase response.

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 $M \times M$ and three-level decomposition on all training images each 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 image as well as the high-resolution training set.

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

Step 4. If MAD < threshold, obtain the unknown high-resolution contourlet coefficients (4×4 block) from the training image offering the best match locally in subbands IX–XII, else set them all zeros.

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.

It may be noted that we have explained the super-resolving procedure for the special case when the image is decomposed into 4 directional components at each resolution. The same procedure remains valid if we prefer to have decomposition into 8 or 16 directional components. However, some of the notations and equations used in this section need to be properly adjusted. A few comments about the learning at the finest level of the contourlet coefficients are in order now. 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.

In this study we select a 4×4 edge primitive in the low-resolution image for learning the coefficients. A smaller

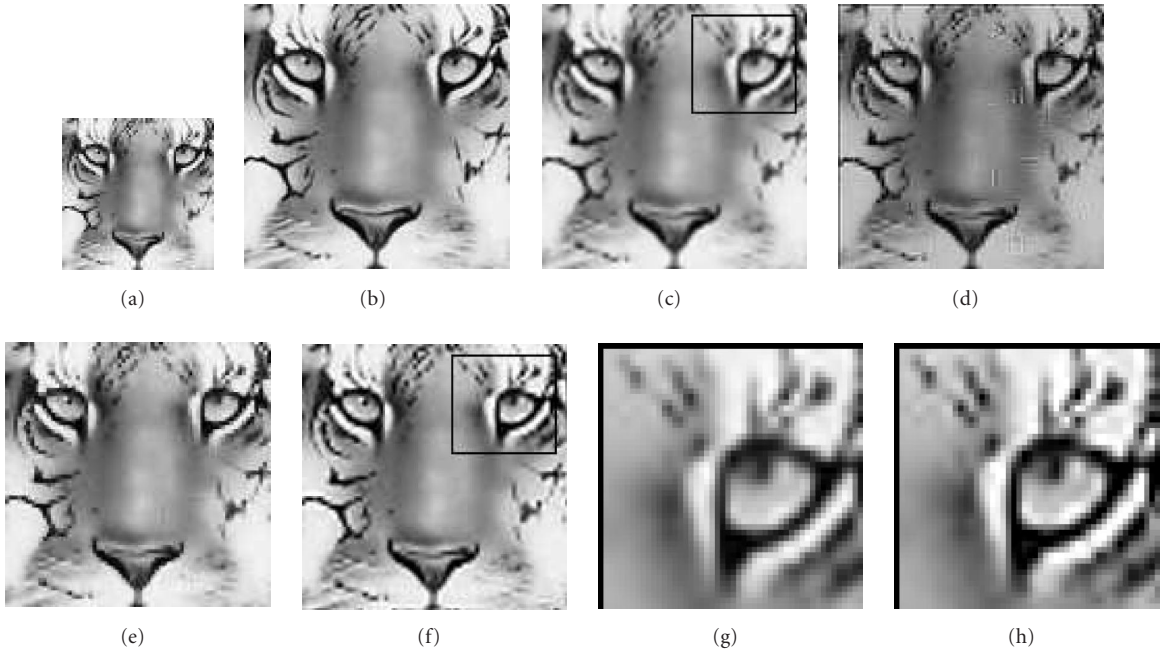

Figure 7: (a) A low-resolution tiger image, (b) original high-resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning without smoothing, (e) super-resolution using wavelet learning with smoothing, (f) the super-resolved image using the proposed approach, (g) zoomed up portion of the marked region from (c), and (h) zoomed up portion of the marked region from (f).

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 2 m/pixel satellite image by a factor of $q = 2$, the training data must be of 1 m/pixel resolution. If one wants to perform super-resolution on a 2.5 m 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.

5. EXPERIMENTAL RESULTS

In this section we demonstrate the usefulness of the proposed approach to super-resolve a low-resolution observation using the contourlet coefficients learned from a high-resolution training data set. Experiments were performed for various types of face, fingerprint, natural, and texture images. The training set consists of about 100 good quality images of all possible classes of objects and is not specific to the class of objects to be super-resolved. Further, the training data is the same for all results displayed in this section.

First we consider experiments with a natural image. To obtain a low-resolution test image and in order to be able to quantify the improvement during super-resolution, we consider a high-resolution image which does not belong to the training set and downsample it by a factor 2 using the decimation matrix D in (1). Figure 7(a) shows one such low-resolution image of size 64×64. Figure 7(b) is the original HR image. Figure 7(c) shows the LR test image upsampled by a factor of 2 using the bicubic interpolation technique. The super-resolved image obtained using a standard (separable basis) wavelet-based learning [20] is shown in Figure 7(d) where we can observe certain artifacts. If this result is further regularized using an appropriate edge preserving smoothness constraint (see [20] for details), the artifacts can be minimized and the corresponding result is shown in Figure 7(e). But this requires some computationally demanding optimization technique which makes the algorithm unnecessarily very slow as explained in Section 4.2. Even though, this result seems sharper than what is shown in Figure 7(d), it is still slightly blocky (observe the eyes). The super-resolved image using contourlet-based learning is shown in Figure 7(f). Here we can note that the artifacts as seen in Figure 7(d) are almost absent. This is because of the capability of the contourlet transform to capture smoothness along contours. Note that all these results are quite sharp compared to the result of bicubic interpolation shown in Figure 7(c). A comparison of Figures 7(c) and 7(f) shows more clear details in the super-resolved image. Figures 7(g) and 7(h) are the zoomed up versions corresponding to the marked regions in Figures 7(c) and 7(f), respectively. Here one can clearly observe that the super-resolved image is quite sharper than the bicubic interpolated one. The super-resolved image is quite close to the original HR image shown in Figure 7(b).

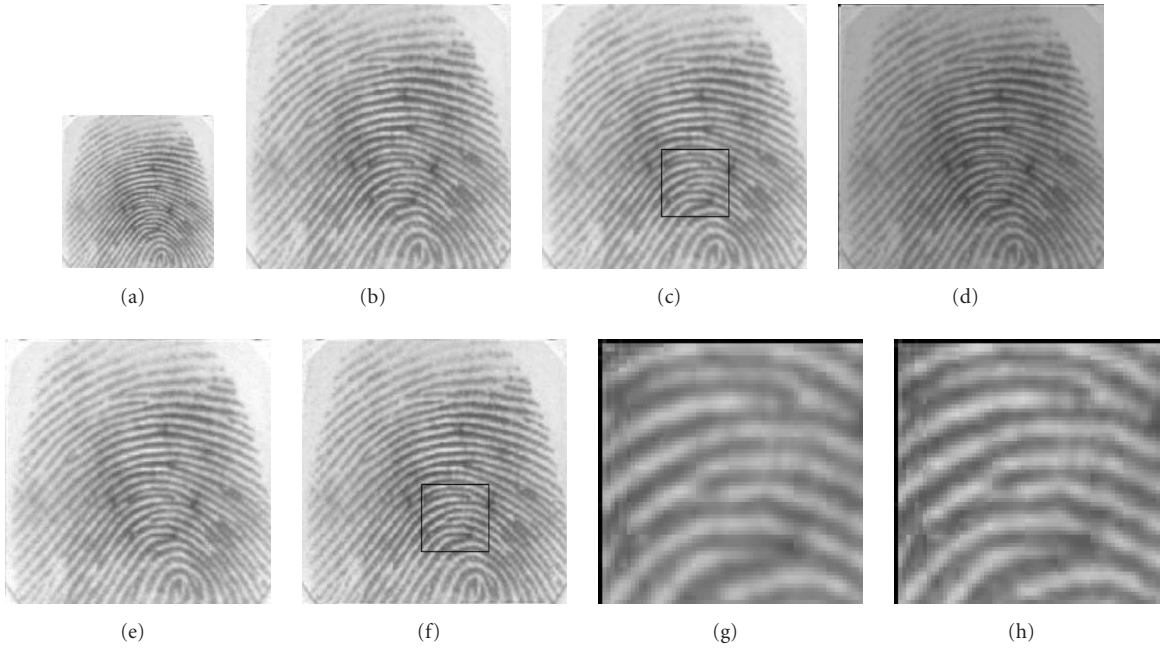

Figure 8: (a) A low-resolution fingerprint observation, (b) original high-resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning without smoothing, (e) super-resolution using wavelet learning with smoothing, (f) the super-resolved image using the proposed approach, (g) zoomed up portion of the marked region from (c), and (h) zoomed up portion of the marked region from (f).

Next we show the results of the experiment performed on a low-resolution fingerprint image. Figure 8(a) shows the LR observation of size 128×128 and Figure 8(b) is the original HR image of size 256×256. The bicubic interpolated image is shown in Figure 8(c). Figures 8(d) and 8(e) show the results of wavelet-based learning without and with smoothing, respectively. The super-resolved image using the proposed approach is shown in Figure 8(f). Figures 8(g) and 8(h) are the zoomed up versions corresponding to the highlighted regions in Figures 8(c) and 8(f), respectively. It can be observed that the contours in the super-resolved image are less blurred than the bicubic interpolated image.

Figures 9(c)–9(f) show the results of the corresponding experiments conducted on an LR textured image shown in Figure 9(a). The super-resolved image using the proposed approach seems to be much sharper compared to the results of bicubic interpolation and of wavelet-based learning. In particular, the edges are better preserved in the super-resolved image using contourlet learning than the bicubic interpolated image where it appears to be more blurred. The super-resolved image compared very favorably to the original high-resolution image shown in Figure 9(b).

Now we show the results of the experiments performed on an LR image where the aliasing is very high. The purpose of this experiment is to demonstrate the behavior of the proposed method when severe aliasing is present in the LR data. Such a low-resolution image is shown in Figure 10(a) and the corresponding bicubic interpolated image is shown in Figure 10(b). Note that the stripes on the scarf are aliased. Figures 10(d) and 10(e) show the results of wavelet-based learning without and with smoothing, respectively. The super-resolved image using the proposed approach is shown in Figure 10(f). The super-resolved image appears to be much sharper than the bicubic interpolated one. However, the proposed method was unable to remove the aliasing effect.

Finally, in order to convey the comparative edge of the proposed technique over other super-resolution methods, we compute the peak signal-to-noise ratio (PSNR) of the reconstructed images. In Table 1, we compare their performances. For the fingerprint image, we observe that it offers up to a 4 dB gain in PSNR over the bicubic interpolation technique and over 3 dB gain over the wavelet-based learning technique that incorporates the highly time-consuming smoothness constraint. This is quite expected as the fingerprint has many curved edges which cannot be handled well with separable basis functions. The contourlet expansion provides a much better solution. The PSNR improvement is about 2 dB when the input is quite a smooth one such as the face image, and it improves to over 3 dB when the input image has noticeably strong edges such as in the brick texture or in the tiger image.

We end this section with a comment on how the high-resolution contourlet coefficients were learned from the HR database images. For the experiment with the face image, about 24% of all the contourlet coefficients were learned from the HR database and the rest were set to zero as no matching coefficients could be found from the database for

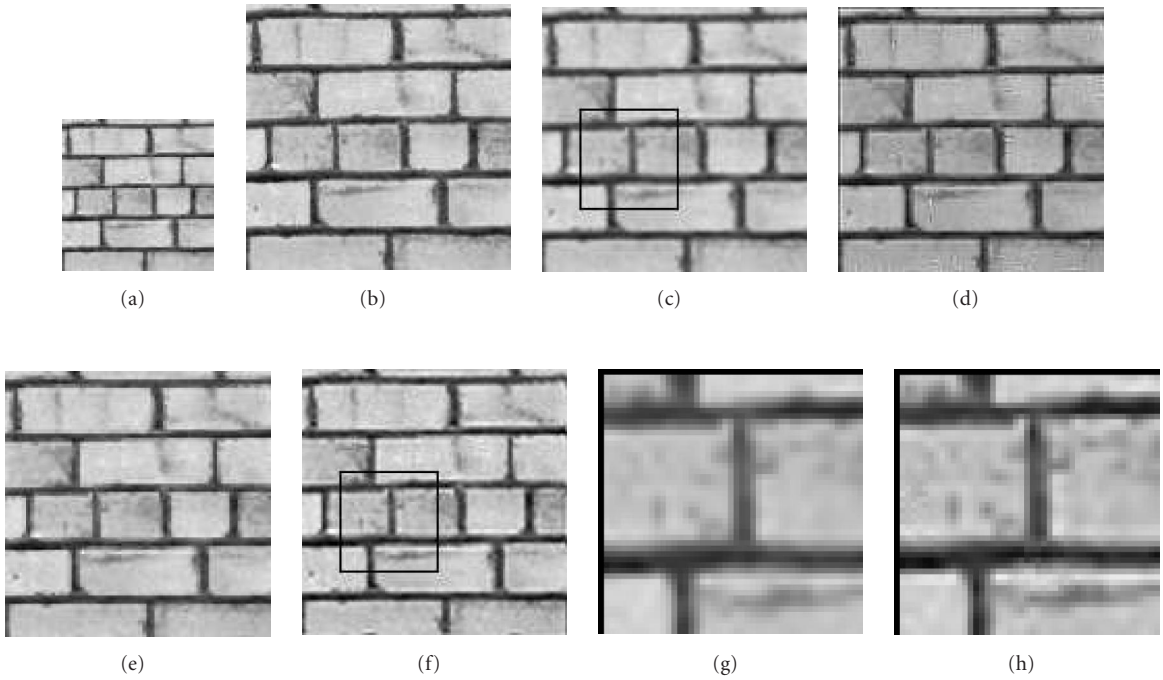

FIGURE 9: (a) A low-resolution textured image, (b) original high-resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning without smoothing, (e) super-resolution using wavelet learning with smoothing, (f) the super-resolved image using the proposed approach, (g) zoomed up portion of the marked region from (c), and (h) zoomed up portion of the marked region from (f).

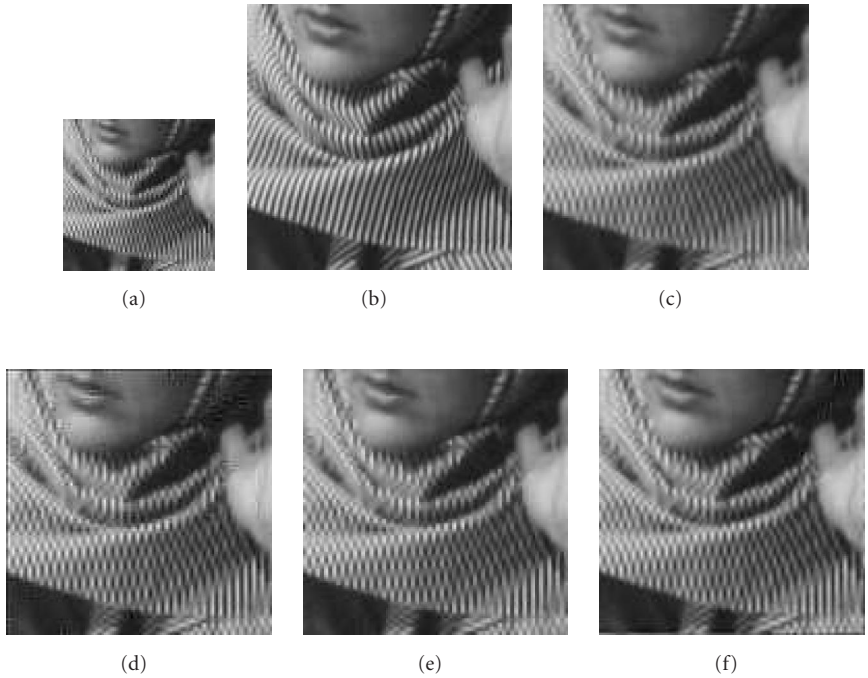

FIGURE 10: (a) A severely aliased low-resolution observation, (b) original high-resolution image, (c) bicubic interpolated image, (d) super-resolution using wavelet learning without smoothing, (e) super-resolution using wavelet learning with smoothing, and (f) the super-resolved image using the proposed approach.

Table 1: Comparison of PSNRs for the zoom factor $q = 2$ expressed in dB.

Image	Bicubic interpolation	Wavelet learning		Contourlet learning
		Without smoothing	With smoothing	
Tiger	20.88	15.41	20.18	24.18
Fingerprint	30.81	14.46	30.25	34.30
Texture	22.69	17.53	23.29	26.05
Face	19.95	17.67	19.12	21.81

a given choice of threshold. Further, all these were picked up from various different images in the database (such as faces, textures, building images, natural textures, etc.) suggesting that the learning process is not class specific.

6. CONCLUSIONS

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, after proper thresholding to avoid spurious learning, are used to estimate the super-resolved image. The learning process ensures capturing the best high-resolution edges from the training set given a low-resolution observation, as well as captures the smoothness along contours. The results obtained for different classes of images show perceptual as well as quantifiable improvements over conventional interpolation techniques. It is also observed that the standard wavelet-based 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. Currently we are investigating ways to include the blur matrix B in the observation model given in (1) while super-resolving the image. Further, all the results given in the previous section were obtained using directional decomposition in four subbands. We are investigating what is the optimal number of directional subbands that would offer the best reconstruction.

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.

[2] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1, chapter 7, pp. 317–339, JAI Press, Greenwich, Conn, USA, 1984.

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

[4] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion, and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, 1993.

[5] 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 *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, vol. 3, pp. 169–172, San Francisco, Calif, USA, March 1992.

[6] M. K. Ng, J. Koo, and N. K. Bose, "Constrained total least-squares computations for high-resolution image reconstruction with multisensors," *International Journal of Imaging Systems and Technology*, vol. 12, no. 1, pp. 35–42, 2002.

[7] M. K. Ng and N. K. Bose, "Analysis of displacement errors in high-resolution image reconstruction with multisensors," *IEEE Transactions on Circuits and SystemsPart I*, vol. 49, no. 6, pp. 806–813, 2002.

[8] 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.

[9] N. Nguyen, P. Milanfar, and G. Golub, "A computationally efficient super-resolution image reconstruction algorithm," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 573–583, 2001.

[10] 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.

[11] D. Rajan and S. Chaudhuri, "An MRF-based approach to generation of super-resolution images from blurred observations," *Journal of Mathematical Imaging and Vision*, vol. 16, no. 1, pp. 5–15, 2002.

[12] D. Rajan and S. Chaudhuri, "Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1102–1117, 2003.

[13] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.

[14] Z. Lin and H.-Y. Shum, "Fundamental limits of reconstruction-based super-resolution algorithms under local translation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 83–97, 2004.

[15] 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.

[16] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. II-627–II-634, Kauai, Hawaii, USA, December 2001.

[17] 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.

[18] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, "Image analogies," in *Proceedings of ACM SIGGRAPH '01*, pp. 327–340, Los Angeles, Calif, USA, August 2001.

[19] M. V. Joshi and S. Chaudhuri, "A learning-based method for image super-resolution from zoomed observations," in *Proceedings of 5th International Conference on Advances in Pattern Recognition (ICAPR '03)*, pp. 179–182, Calcutta, India, December 2003.

[20] C. V. Jiji, M. V. Joshi, and S. Chaudhuri, "Single-frame image super-resolution using learned wavelet coefficients," *International Journal of Imaging Systems and Technology*, vol. 14, no. 3, pp. 105–112, 2004.

[21] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 597–606, 2003.

[22] X. Wang and X. Tang, "Hallucinating face by eigentransformation with distortion reduction," in *Proceedings of 1st International Conference on Biometrics Authentication (ICBA '04)*, pp. 88–94, Hong Kong, July 2004.

[23] C. V. Jiji and S. Chaudhuri, "PCA-based generalized interpolation for image super-resolution," in *Proceedings of 4th Indian Conference on Vision, Graphics & Image Processing (ICVGIP '04)*, pp. 139–144, Calcutta, India, December 2004.

[24] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A sampled texture prior for image super-resolution," in *Proceedings of Advances in Neural Information Processing Systems 16 (NIPS '03)*, S. Thrun, L. Saul, and B. Schölkopf, Eds., pp. 1587–1594, MIT Press, Vancouver, British Columbia, Canada, 2004.

[25] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. I-275–I-282, Washington, DC, USA, June–July 2004.

[26] I. Begin and F. R. Ferrie, "Blind super-resolution using a learning-based approach," in *Proceedings of 17th IEEE International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 85–89, Cambridge, UK, August 2004.

[27] J.-S. Park and S.-W. Lee, "Enhancing low-resolution facial images using error back-projection for human identification at a distance," in *Proceedings of 17th IEEE International Conference on Pattern Recognition (ICPR '04)*, vol. 1, pp. 346–349, Cambridge, UK, August 2004.

[28] J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum, "Image hallucination with primal sketch priors," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 2, pp. II-729–II-736, Madison, Wis, USA, June 2003.

[29] J. Wu, M. Trivedi, and B. Rao, "Resolution enhancement by AdaBoost," in *Proceedings of 17th IEEE International Conference on Pattern Recognition (ICPR '04)*, vol. 4, pp. 893–896, Cambridge, UK, August 2004.

[30] M. N. Do, *Directional multiresolution image representations*, Ph.D. thesis, Department of Communication, Swiss Federal Institute of Technology, Lausanne, Switzerland, November 2001.

[31] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[32] 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, pp. 882–893, 1992.

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

[34] S.-I. Park, M. J. T. Smith, and R. M. Mersereau, "A new directional filter bank for image analysis and classification," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, vol. 3, pp. 1417–1420, Phoenix, Ariz, USA, March 1999.

[35] J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3445–3462, 1993.

[36] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2207–2232, 1992.

[37] 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.

C. V. Jiji received the B.Tech. degree in electronics and communication engineering from University of Kerala, the M.Tech. degree in electrical engineering from Indian Institute of Technology, Bombay. He is an Assistant Professor in the Department of Electronics and Communication, College of Engineering, Thiruvananthapuram, Kerala. Currently he is pursuing his Ph.D. in the Department of Electrical Engineering, Indian Institute of Technology Bombay. His research interests include digital signal processing, image processing, and computer vision.

Subhasis Chaudhuri was born in Bahutali, India. He received his B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology Kharagpur in 1985. He received the M.S. and the Ph.D. degrees, both in electrical engineering, respectively, from the University of Calgary, Canada, and the University of California, San Diego. He joined the IIT Bombay in 1990 as an Assistant Professor and is currently serving as a Professor. He has also served as a Visiting Professor at the University of Erlangen-Nuremberg, Germany, and the University of Paris XI. He is a Fellow of the Alexander von Humboldt Foundation, Germany, the Indian National Academy of Engineering, and the Indian Academy of Sciences. He is the recipient of Dr. Vikram Sarabhai Research Award for the year 2001, and the Swarnajayanti Fellowship in 2003. He received the S.S. Bhatnagar Prize in engineering sciences for the year 2004. He is the coauthor of the book *Depth from Defocus: A Real Aperture Imaging Approach*, published by Springer, NY. He has also edited a book *Super-Resolution Imaging* published by Kluwer Academic in 2001. His research interests include image processing, computer vision, and multimedia.