

Research on Image Super-Resolution

Many researchers have tackled the super-resolution reconstruction problem for both still images and video. Although the super-resolution reconstruction techniques for video are often extensions to still image super-resolution, many different approaches have also been proposed. In general, based on the type of cues used, the super-resolution methods can be classified into two categories: motion-based techniques and the motion-free approaches. Motion-based techniques use the relative motion between different low resolution observations as a cue in estimating the high resolution image, while motion-free super-resolution techniques may use cues such as blur, zoom, and shading. These methods do not require observations with relative motion among them. Some researchers have also attempted to solve the super-resolution reconstruction problem without considering any specific cue, but by using an ensemble of images as a training set in order to learn the required information for resolution enhancement.

Different methods to obtain super-resolution include nonuniform interpolation approach, frequency domain approach, and regularization based reconstruction technique which may be either deterministic or stochastic. Few other existing approaches include projection onto convex sets, iterative back projection method, adaptive filtering method, *etc.* Most of the super-resolution techniques discussed in the literature are based on the motion cue, *i.e.*, using the subpixel shifts among the observations. A few researchers have also tackled the super-resolution problem without using the motion cue. In this chapter we review the literature on super-resolution reconstruction for motion-based as well as for motion-free techniques. A comprehensive survey on super-resolution imaging can also be found in [28, 29].

2.1 Motion-Based Super-Resolution

The super-resolution idea was first proposed by Tsai and Huang [30]. They used the frequency domain approach to demonstrate the ability to reconstruct a single improved resolution image from several down-sampled, noise free versions of it. A frequency domain observation model was defined for this problem which considered only the globally shifted versions of the same scene. Their approach is based on the following principles:

- the shifting property of the Fourier transform,
- the aliasing relationship between the continuous Fourier transform of the original image and the discrete Fourier transform (DFT) of the observed low resolution frames, and
- the assumption that the original high resolution image is bandlimited.

Kim *et al.* discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred observations [31]. They consider the same blur and noise characteristics for all the low resolution observations. Their recursive approach combines the two steps of filtering and reconstruction. The filtering operation on the registered images compensates for the degradation and noise, and the reconstruction step estimates the image samples on a high resolution grid in order to obtain the super-resolved image. Kim and Su [32] consider different amounts of blur for each low resolution image and used the Tikhonov regularization to obtain the solution of an inconsistent set of linear equations.

The disadvantage with the frequency domain approach lies on the restrictions imposed on the observation model. One may consider only a translational motion and a linear space invariant (LSI) blur. Also, since the data is uncorrelated in the frequency domain, it is difficult to apply apriori knowledge about the data for the purpose of regularization. Nonetheless, it was a good beginning and very soon researchers started looking at the problem in the spatial domain also. Needless to say, researchers have also explored the use of other types of image transforms to achieve super-resolution. For example, a discrete cosine transform (DCT) based method instead of DFT has been proposed by Rhee and Kang [33].

A minimum mean squared error approach for multiple image restoration, followed by interpolation of the restored images into a single high

resolution image has been presented in [34]. Ur and Gross use the Papoulis and Brown generalized sampling theorem [35],[36] to obtain an improved resolution picture from an ensemble of spatially shifted observations [37]. These shifts are assumed to be known by the authors. A recursive total least squares method for super-resolution reconstruction to reduce the effects of registration error is discussed in [38]. All the above super-resolution restoration methods are restricted either to a globally uniform translational displacement between the measured images, or an LSI blur, and a homogeneous additive noise.

A different approach to the super-resolution restoration problem was suggested by Peleg and his co-authors [39, 40, 41], based on the iterative back projection (IBP) method adapted from computer aided tomography. This method starts with an initial guess of the output image, projects the temporary result to the measurements (simulating them), and updates the temporary guess according to this simulation error. A back projection kernel determines the contribution of the error to the reconstructed image at each iteration. The disadvantage of IBP is that it has no unique solution as it does not attempt to involve prior constraints. A set theoretic approach to the super-resolution restoration problem was suggested in [42]. The main result there is the ability to define convex sets which represent tight constraints on the image to be restored. Having defined such constraints it is straightforward to apply the projections onto convex sets (POCS) method, which was originally suggested by Stark and Oskoui [4]. The POCS based approach describes an alternative way to incorporating the prior knowledge about the solution into the super-resolution reconstruction process. According to this method, the solution is restricted to be a member of a closed convex set that is defined as a set of vectors which satisfy a user specified property. If the constraint sets have nonempty intersection, then a solution can be found by alternately projecting onto the convex sets. All these methods mentioned above are not restricted to having a specific motion characteristic. They can handle smooth motion, linear space variant blur, and non-homogeneous additive noise.

Ng *et al.* develop a regularized constrained total least squares (RCTLS) solution to obtain a high resolution image in [43]. They consider the presence of perturbation errors of displacements around the ideal subpixel locations in addition to sensor noise. The superiority of the approach over conventional least squares based approach is substantiated through examples. The analysis of the effect of displacement

errors on the convergence rate of the iterative approach for solving the transform based preconditioned system of equations during high resolution image reconstruction with multiple sensors has been carried out in [44]. It is established that the use of MAP, L_2 -norm or H_1 -norm based regularization functional leads to a linear convergence of the conjugate gradient descent method in terms of the displacement errors caused by the imperfect subpixel localization. Bose *et al.* [45] point out the important role of the regularization parameter and suggest the use of a constrained least squares (CLS) method for super-resolution reconstruction which generates the optimum value of the regularization parameter, using the L-curve method [46].

In [47] the authors use a maximum *a posteriori* (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 similar idea of joint estimation applied to infrared imagery is presented in [48]. The high resolution estimate of the image is obtained by minimizing a regularized cost function based on the observation model. It is also shown that with a proper choice of tuning parameter, the algorithm exhibits robustness in presence of noise. Both the gradient descent and the conjugate gradient descent optimization techniques are used to minimize the cost function. An expectation maximization (EM) based algorithm solved in the frequency domain in order to simultaneously estimate the super-resolved image, the blur and the registration parameters is described in [49]. All these methods alternately estimate the high resolution image and the motion fields for an improved accuracy.

A MAP estimator with Huber-Markov random field (HMRF) prior is described by Schultz and Stevenson in [50] for improving the image resolution. Here a discontinuity preserving stabilizing functional is used for the preservation of edges. In HMRF, an edge preserving potential function is used to define the prior constraint. The potential function is given by

$$U(x) = \begin{cases} x^2, & \text{if } |x| \leq \alpha \\ 2\alpha|x| - \alpha^2, & \text{otherwise} \end{cases}$$

where x is the finite difference approximation of the first order derivative of the image at each pixel. HMRF is an example of a convex but nonquadratic prior. The purpose of making the prior linearly increasing beyond the threshold $|x| > \alpha$ is to partly reduce the rate of growth in

the cost function when there is an edge between two pixels. The idea is quite similar to the concept of an M-estimator prevalent in the area of robust regression analysis. In the paper two separate algorithms have been derived: a constrained optimization method for a noise free image reconstruction, and an unconstrained optimization algorithm for image data containing Gaussian noise. The gradient projection algorithm has been used to minimize the cost derived from a noise free case and a gradient descent optimization is used for the noise corrupted case. Till date this method is probably the most popular one among the researchers as we notice that most of the currently proposed approaches compare their performances with the results obtained with an HMRF prior. Since all these methods claim a superiority over the HMRF based method, it is probably safe to state that the HMRF method, indeed, yields a reasonably accurate result.

In many resolution enhancement applications, the blurring process *i.e.*, the point spread function (PSF) of the imaging system, is not known. Nguyen *et al.* [51] propose a technique for parametric blur identification and regularization based on the generalized cross-validation (GCV) theory. The idea of cross-validation is to divide the data set into two parts; one part is used to construct an approximate solution, and the other is used to validate that approximation. They propose approximation techniques based on the Lanczos algorithm and Gauss quadrature theory for reducing the computational complexities of GCV. They solve a multivariate nonlinear minimization problem for the unknown parameters. They have also proposed circulant block preconditioners to accelerate the conjugate gradient descent (CG) method while solving the Tikhonov-regularized super-resolution problem [52]. Preconditioning is a process used to transform the original system into one with the same solution, but which can be solved more quickly by the iterative solver. They use specific preconditioners such that the preconditioned system has eigenvalues clustered around unity which makes CG method to converge rapidly.

Elad and Feuer [10] propose a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and down-sampled observations by combining maximum likelihood (ML), MAP and POCS approaches. The proposed super-resolution approach is general but assumes explicitly a linear space variant blur, and an additive Gaussian noise. In addition to the motion-based super-resolution the authors also discuss the condition for motion-free super-

resolution imaging when the observations are captured with different amounts of defocus blur even when both the camera and the object are stationary. This issue will be taken up in the next section. An adaptive filtering approach to super-resolution restoration is described by the same authors in [53] using the least mean squares (LMS) and the pseudo-recursive least squares (RLS) algorithms. Both the methods have been demonstrated with and without regularization. They exploit the properties of the operations involved in their previous work [10] and develop a fast super-resolution algorithm in [54] for a purely translational motion and space invariant blur, assuming them to be the same for all the images. The approach consists of deblurring and measurement fusion which is shown to be solvable using a non-iterative algorithm. Similarly two fast non-iterative algorithms for image super-resolution based on Choleskey decomposition have been developed by Jorge and Ferreira [55]. They use the spatial domain formulation and the frequency domain approach. The spatial domain approach leads to a set of linear equations for the unknown pixels, while the frequency domain approach leads to equations for the unknown DFT coefficients. An additional inverse Fourier transform is used to obtain the required image while working in the frequency domain.

A computationally fast super-resolution algorithm based on the preconditioner using the motion adaptive relaxation parameters is considered in [56]. The proposed algorithm can be implemented in real time by updating the motion compensated low resolution frame at each time instant by using the preconditioner which increases the converges rate. Thus the speed up operation is achieved through system preconditioning as discussed earlier. This method can be applied to a general image sequence with differently moving objects, thus can handle local variations in the motion parameters. Farsiu *et al.* propose a fast and robust super-resolution algorithm based on L_1 -norm for both data fitting term and the prior term and show that it performs better with and even without the outliers present in the data [57]. The robustness is achieved by limiting the contribution of the highly erroneous outlier data through the use of L_1 -norm. Quite naturally, we may replace the L_1 -norm by any appropriate weight function $W(x)$ as it is commonly done in M -estimator. The authors in [58] investigate the performance of super-resolution algorithms using different potential functions such as convex, nonconvex, bounded, and the unbounded as a prior in the cost function and compare their performance on synthetic and real images.

They evaluate the performances of three different potential functions given below proposed, respectively, by Charbonnier [59], Hebert and Leahy [60], and Geman and Reynolds [61].

$$\begin{aligned} U(x) &= 2\sqrt{1+x^2} - 2, \\ U(x) &= \log(1+x^2), \quad \text{and} \\ U(x) &= \frac{x^2}{1+x^2}, \end{aligned}$$

where x is the finite difference approximation of the first order derivative of the image at each pixel. Different optimization methods have been used for each prior model.

Edges are typically the most important features in an image. For a homogeneous region, any kind of interpolation technique for image upsampling would suffice. However, one must be careful while upsampling the regions having edges as we would like them to be sharp in the high resolution data. Chiang and Boulton [62] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. An image consistent reconstruction algorithm is used which gives the exact solution for some input function which, according to the sensor model, would have generated the measured input. Rather than obtaining the super-resolution by fusion of all the images together they choose one of the images from the image sequence and then fuse together all the edges from the other images. This requires that the reference image be re-estimated and scaled up based on the edge models and local blur estimation. Thus they mitigate the problem arising due to illumination variation during image capture since the edge positions are less sensitive to lighting variations. They have also applied image warping to reconstruct a high resolution image [63] which is based on a concept called integrating resampler [64] that warps the image subject to some constraints. Here the upsampled images are combined using the median, and the resultant image is convolved to remove blur, with a high pass filter. Similarly, a robust median-based estimator is used in an iterative process to achieve the super-resolution in [65]. This approach discards the measurements which are inconsistent with the imaging model, thus increasing the resolution even in regions having the outliers.

An image super-resolution technique based on the wavelet domain hidden Markov tree (HMT) model as a prior is proposed by Zhao *et al.* [66]. The wavelet domain HMT characterizes the statistical properties

of the real image. Here the authors use the motion cue, but unlike using the Huber-MRF prior, they use the HMT prior. They formulate the problem as a constrained optimization problem and solve it using a cyclic optimization procedure.

All the methods discussed so far do use the motion cue for super-resolution, and in order to do that they need to actually compute the motion parameters. At this point some researchers may feel that since most of the available video sequences are already MPEG compressed, decompression of the video and then motion estimation is a wastage of time. The MPEG data already has the motion vectors in the bit stream. Can these motion vectors be used without fully decompressing the MPEG data? This problem of recovering a high resolution image from a sequence of DCT compressed images is addressed in [67]. It may be noted that the MPEG motion vectors may not always give us the true motion field. Also, the motion vectors are not dense. It is specified over a macro-block. So the authors recover the high resolution image using an iterative method considering the effects of quantization (residual) noise as well as registration errors, both modeled as zero mean additive Gaussian noise. A regularization functional is introduced not only to reflect the relative amount of registration error but also to determine the regularization parameter. Segall *et al.* estimate the high resolution image as well as subpixel displacements from compressed image observations [68]. They formulate the problem in a Bayesian framework and use the iterative cyclic coordinate descent approach for the joint estimation. Here the pixel intensities are no longer the observations, instead motion vectors and quantized transform coefficients are provided to the recovery algorithm.

There have been very few publications in the area of quantifying the performance of motion-based super-resolution methods. 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 [69]. They prove that fundamental limits do exist for reconstruction based super-resolution algorithms where a number of low resolution, subpixel displaced frames are used to estimate a high resolution image. They discuss two extreme cases and find that the practical limit for magnification is 1.6, if the registration and the noise removal is not good enough.

Let us now discuss some of the application specific super-resolution schemes. There has been an effort in the area of astrophysics for improv-

ing the image resolution of celestial objects. In [70] authors use a series of short-exposure images taken concurrently with a corresponding set of images of a guidestar and obtain a maximum-likelihood estimate of the undistorted image. Yang and Parvin [71] compute the dense map of feature velocities from lower resolution data and project them onto the corresponding high resolution data. The proposed technique is applied to measurement of sea surface temperature. The super-resolution principle has been applied to the face recognition systems as well in [72, 73]. They apply the super-resolution technique after dimensionality reduction to a set of inaccurate feature vectors of a subject, and their reconstruction algorithm estimates the true feature vector. Authors in [74] have proposed a MAP estimator based on the Huber prior for enhancing text images. The authors map the problem as that of a total variation and super-resolve the text. They consider images of scenes for which the point to point image transformation is a planar projective one.

It is now worth digressing a bit to look into the problem of image mosaicing. Mosaicing works on the principle that there are overlapping regions in the successive images so that interest points can be recovered in these regions and subsequently matched to compute the homography. Once the homography is computed, images are stitched together to obtain a high field of view mosaic. But while stitching these images across the overlapping regions, we throw away the additional information available from multiple views as redundant. This apparently redundant information is, however, the ideal cue for image super-resolution. The complementary set of information can be used for super-mosaicing purposes, [75] *i.e.*, to build a high resolution mosaic. An efficient super-resolution algorithm with application to panoramic mosaics has been proposed by Zomet and Peleg [76]. The method preserves the geometry of the original mosaic and improves spatial resolution. Capel and Zisserman have proposed a technique for automated mosaicing with super-resolution zoom in which a region of the mosaic can be viewed at a resolution higher than any of the original frames by fusing information from several views of a planar surface in order to estimate its texture [77]. Similarly, in [75], Bhosle *et al.* use the motion cue for super-resolution of a mosaic. They use the overlap among the observed images to increase the spatial resolution of the mosaic and to reduce the noise. In order to illustrate this, we show in Figure 2.1 a panoramic mosaic of a building constructed from 36 overlapped observations. The

corresponding super-mosaic is displayed in Figure 2.2. One can notice an improvement in bringing out some of the finer details here.



Fig. 2.1. Example of a low resolution panoramic mosaic.

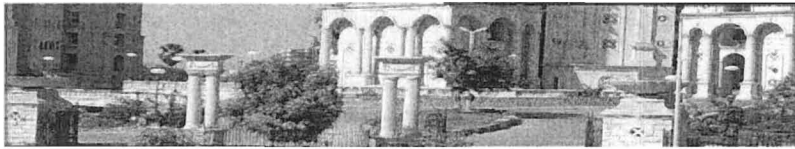


Fig. 2.2. Illustration of a super-mosaic constructed from the same set of observations used in obtaining Figure 2.1.

Now we discuss some of the research efforts in super-resolving a video sequence. Most of the super-resolution algorithms applicable to video are extensions of their single frame counterpart. Authors in [78] describe a complete model of video acquisition with an arbitrary input sampling lattice and a non-zero exposure time. They use the theory of POCS to reconstruct super-resolution still images or video frames from a low resolution time sequence of images. They restrict both the sensor blur and the focus blur to be constant during the exposure. Their video formation model includes an arbitrary space time lattice in order to obtain the sampled video signal. A hierarchical block matching algorithm is used to estimate the nonuniform translational motion between the low resolution images and the reference image. The motion model is incorporated into the video formation model to establish a linear space variant (LSV) relationship between the low resolution images and the desired super-resolved image at an arbitrary time t . By appropriately setting the values of t , a single super-resolved still image or a super-resolved video is reconstructed. Eren *et al.* extended the technique in [78] to scenes with multiple moving objects by introducing the concepts of validity maps and segmentation maps and by using the POCS framework [79]. The validity map disables projections based on observations with inaccurate motion information for a robust reconstruction whenever there is error in motion estimation. The segmentation map enables

an object-based processing where a more accurate motion model can be utilized to improve the quality of reconstructed images.

In [80] a technique for robust deinterlacing for creating high quality stills from an interlaced video is presented. A method for motion compensated deinterlacing that combines a motion trajectory filter for removing the dominant motion such as camera zoom, pan and jitter, with motion detection to remove artifacts caused by independently moving objects has been discussed. The motion detection method employs an adaptive thresholding scheme that simultaneously suppresses aliasing artifacts and artifacts caused by independently moving objects.

Schultz and Stevenson use the hierarchical block matching algorithm to estimate the subpixel displacement vectors and then solve the problem of estimating the high resolution frame given a low resolution sequence by formulating it as a Bayesian MAP estimation with Huber-Markov random field (HMRF) prior, resulting in a constrained optimization problem with a unique minimum [81]. The super-resolution video enhancement technique proposed by Shah and Zakhor consider the fact that the motion estimates used in the reconstruction process will be inaccurate [82]. To this end their algorithm finds a set of candidate motion estimates instead of a single motion vector for each pixel, and then both the luminance and the chrominance values are used to compute the dense motion field with subpixel accuracy. The high resolution frame is restored subsequently by a method based on the Landweber algorithm.

Researchers have also used appropriate smoothness constraints over successive frames. Hong *et al.* define a multiple input smoothing convex functional and use it to obtain a globally optimal high resolution video sequence [83]. An iterative algorithm for resolution enhancement of a monochrome or a color video sequence using motion compensation has been presented in [84]. The choice of which motion estimator to use versus how the final estimates are obtained is weighed to see which issue is more critical in improving the estimated high resolution sequence. A single motion field is estimated using the three color fields. They use two different approaches for motion estimation, which recover the motion in two steps. In the first step, a displacement vector field (DVF) is estimated for each channel. In the second step, these three DVFs are combined via data fusion (merging the individual motion fields) to yield a single DVF. The straightforward examples of data fusion are the use of a prespecified vector corresponding to a particular color

channel or the vector mean or the vector median. The estimated high resolution images using the block matching motion estimators have been compared to those obtained by using a pixel recursive scheme.

Altunbasak *et al.* [85] have proposed a motion-compensated, transform domain super-resolution procedure for creating high quality video or still images that directly incorporates the transform domain quantization information by working in the compressed bit stream. They apply this new formulation to MPEG-compressed video. In [86], a method for simultaneously estimating the high resolution image frames and the corresponding motion fields from a compressed low resolution video sequence is presented. The algorithm incorporates knowledge of the spatio-temporal correlation between low and high resolution images to estimate the original high resolution sequence from the degraded low resolution observation. The idea has been further extended to introduce additional high resolution frames in between two low resolution input frames to obtain a high resolution, slow motion sequencing of a given video [87]. The authors develop the above system for the purpose of *post-facto* video surveillance, *i.e.*, to find what exactly had happened from the stored video.

Authors in [88] propose a high-speed super-resolution algorithm using the generalization of Papoulis' sampling theorem for multichannel data with applications to super-resolving video sequences. They estimate the point spread function (PSF) for each frame and use the same for super-resolution. Borman and Stevenson [89] present a MAP approach for multi-frame super-resolution of a video sequence using the spatial as well as temporal constraints. The spatio-temporal constraint is imposed by using a motion trajectory compensated MRF model, in which the Gibbs distribution is dependent on pixel variation along the motion trajectory.

Most of the research works discussed so far assume that the low resolution image formation model illustrated in Figure 1.1, is indeed correct. Model uncertainties are not considered. In [90] the authors consider the problem of super-resolution restoration of the video, considering the model uncertainties caused by the inaccurate estimates of motion between frames. They use a Kalman filter based approach to solve the problem. For MPEG compressed data, quantization noise adds upto the uncertainties. Gunturk *et al.* propose a Bayesian approach for the super-resolution of MPEG-compressed video sequence considering both the quantization noise and the additive noise [91].

We observe that additional temporal data is used to improve the spatial resolution. Is it then possible to use additional spatial data (read high resolution image) to improve the temporal resolution? Or, in other words, can the concepts of resolution in space and time be fused together? This issue is discussed next. Shechtman *et al.* [92] construct a video sequence of high space-time resolution by combining information from multiple low resolution video sequences of the same dynamic scene. They used video cameras with complementary properties like low-frame rate but high spatial resolution and high frame-rate but low spatial resolution. They show that by increasing the temporal resolution using the information from multiple video sequences spatial artifacts such as motion blur can be handled without the need to separate static and dynamic scene components or to estimate their motion. To constrain the solution and provide numerical stability they use a space-time regularization term to impose the smoothness on the solution. A directional (or steerable) space-time regularization term applies smoothness only in directions where the derivatives are low, and does not smooth the space-time edges, thus preserving spatial edges as well as minimizing the motion blur due to the finite exposure time.

2.2 Motion-Free Super-Resolution

In the previous section we have discussed many different methods that use motion as the cue to generate the high frequency details. All these methods require a dense point correspondence among frames. Any error in establishing the correspondence affects the quality of super-resolution. Although the bulk of the work on super-resolution does use motion cue, of late, there has been work on using other possible cues. Motion-free super-resolution techniques try to obtain the spatial enhancement by using the cues which do not involve a motion among low resolution observations, thus avoiding the correspondence problem. One may expect an improved result since there would be no correspondence. However, we must find out what other cues can possibly be used as a substitute for the motion cue to bring in the high frequency details. We need to study how useful are these cues and what additional difficulties do they introduce during the super-resolution process. Another issue that comes out is how should we compare the performances of these methods with those of the motion-based methods. We simply cannot compare the methods as the data generation process is very

different in both the cases. Further, the volume of work in this area is still quite small. Use of cues other than motion is the subject matter of this monograph. Before we discuss some of the specific methods in subsequent chapters, we begin reviewing some of the existing techniques in motion-free super-resolution.

Use of different amounts of blur is probably the first attempt in the direction towards motion-free super-resolution. In order to understand the problem let us take an example in 1D data. Let $f(n)$ be the unknown high resolution data, $g(m)$ be the observed data, $h_1(n)$ and $h_2(n)$ be the known finite impulse response (FIR) blurring kernels. Here the indices m and $n = 2m$ stand for the low and high resolution grids, respectively. We assume the decimation module to give us the average of the two adjacent pixels as the low resolution value. In order to explain the usefulness of the blur cue, let us further assume that the blur kernels are given by

$$h_1(n) = a_{11}\delta(n) + a_{12}\delta(n-1)$$

$$h_2(n) = a_{21}\delta(n) + a_{22}\delta(n-1),$$

where $\delta(n)$ is the delta function. Let us further assume that there is no observation noise. Then, neglecting boundary conditions,

$$g_1(m) = 0.5[a_{11}f(2m+1) + (a_{11} + a_{12})f(2m) + a_{12}f(2m-1)]$$

$$g_2(m) = 0.5[a_{21}f(2m+1) + (a_{21} + a_{22})f(2m) + a_{22}f(2m-1)]$$

Since the filter parameters are known the above two equations can easily be solved to obtain the high resolution data, provided the two blur kernels are linearly independent. Here we have $2m$ number of observations g_1 and g_2 and $2m$ number of unknowns in the high resolution signal f .

Hence we observe that it is, indeed, possible to use the differential blur as a cue for super-resolution. Definitely, there will be issues of sensor noise, availability of sufficient number of observations, smoothness of the reconstructed image, *etc.* This calls for the use of regularizing priors to solve the restoration problem.

A MAP-MRF based super-resolution technique has been proposed by Rajan *et al.* in [93]. Here the authors consider an availability of decimated, blurred and noisy versions of a high resolution image which are used to generate a super-resolved image. A known blur acts as a cue in generating the high resolution image. They model the high resolution

image as an MRF to serve as a prior for regularization. In chapter 3 we shall relax the assumption of the known blur and extend it to deal with an arbitrary space-varying defocus blur for super-resolution purposes. Recently, Rajagopalan and Kiran [94] have proposed a frequency domain approach for estimating the high resolution image using the defocus cue. They derive the Cramer-Rao lower bound (CRLB) for the covariance of the error in the estimate of the super-resolved image and show that the estimate becomes better as the relative blur increases.

A scheme for image high resolution from several blurred observations by imposing a periodic grating with various absorptions in the object field is proposed in [95]. This method is based on the solution of a Fredholm's integral equation of the first kind. The method can be employed in different fields such as microscopy and for signal and image transmission under conditions of heavy blur. The super-resolution here is based on an interference of spatial frequencies of the object and the grating.

There has also been an effort in using a functional decomposition approach for super-resolution. One such example is the use of generalized interpolation [96]. Here a space containing the original function is decomposed into appropriate subspaces. These subspaces are chosen so that the rescaling operation preserves properties of the original function. On combining these rescaled sub-functions, they get back the original space containing the scaled or zoomed function. Here the photometric information is used as the cue. The authors in [18] proposed a multi-objective super-resolution technique for super-resolving both the intensity field and the structure using blur and shading as cues. It is shown in the paper that the use of the blur and the shading cues can be combined under a common mathematical framework. All these methods discussed thus far assume the availability of multiple observations of the same scene under different camera or lighting conditions. However, at times one may have to do with a single observation. What if you are given a low resolution image of a suspected criminal? Can this picture be super-resolved?

Researchers have also attempted to solve the super-resolution problem by using learning based techniques. These methods try to recognize the local features in a low resolution image and then retrieve the most likely high frequency information from the given training samples. In this book, these methods are also classified under motion-free super-resolution as the new information required for predicting the high res-

olution image is obtained from the training images rather than from the subpixel shifts among low resolution observations. Authors in [97] describe image interpolation algorithms which use a database of training images to create plausible high frequency details in zoomed images. They propose a learning framework called VISTA - Vision by Image/Scene TrAining. By blurring and down-sampling sharply defined images they construct a training set of sharp and blurred images. These are then incorporated into a Markov network to learn their relationship. A Bayesian belief propagation allows to find the maximum of the posterior probability.

A quite natural extension to the above is to use the best of the both world - information from multiple observations as discussed earlier and the priors learnt from a given high resolution training data set. Capel and Zisserman have proposed a super-resolution technique from multiple views using learnt image models [98]. 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 principal component analysis (PCA) applied to a face image database. Researchers have also attempted to combine the motion cue with the learning based method for super-resolution restoration. Pickup *et al.* [99] combine the motion information due to subpixel displacements as well as motion-free information in the form of learning of priors to propose a domain specific super-resolution using the sampled texture prior. They use training images to estimate the density function. Given a small patch around any particular pixel, they learn the intensity distribution for the central pixel by examining the values at the centers of similar patches available in the training data. The intensity of the original pixel to be estimated is assumed to be Gaussian distributed with mean equal to the learnt pixel value and obtain the super-resolution by minimizing a cost function.

There has also been some effort on applying an output feedback while super-resolving the images. If the purpose of super-resolution is to recognize a face, a character or a fingerprint, then the partially super-resolved image is first matched to a database to extract the correct match and then this information can be used to enhance the prior for further improving the image quality. In [100] 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 as recognition-based super-resolution or hallucination. Their prior

enforces the condition that the gradient of the super-resolved image should be equal to the gradient of the best matching high resolution training image. The learning of the prior is done by using a pyramidal decomposition.

An image analogy method applied to super-resolution is discussed by Hertzmann *et al.* in [101]. They use the low resolution and the high resolution versions of a portion 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. Here the emphasis is in learning the local statistics at a finer details. Candocia and Principe [102] address the ill-posedness of the super-resolution problem by assuming that the correlated neighbors remain similar across scales, and this apriori information is learnt 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.

So far all these learning based methods are restricted to dealing with enhancing a still frame only. A learning based method for super-resolution enhancement of a video has been proposed by Bishop *et al.* [103]. Their approach builds on the principle of example based super-resolution for still images proposed by Freeman *et al.* [97]. They use a learnt data set of image patches capturing the relationship between the middle and the high spatial frequency bands of natural images and use an appropriate prior over such patches. A key concept there is the use of the previously enhanced frame to provide part of the training set for super-resolution enhancement of the current frame.

Having discussed the current research status in super-resolution imaging, we concentrate on a few specific ways of achieving motion-free super-resolution. These methods are discussed in detail in the subsequent chapters.



<http://www.springer.com/978-0-387-25890-4>

Motion-Free Super-Resolution

Chaudhuri, S.; Manjunath, J.

2005, XIII, 239 p., Hardcover

ISBN: 978-0-387-25890-4