A Learning-Based Method for Image Super-Resolution from Zoomed Observations

Manjunath V. Joshi, Subhasis Chaudhuri and Rajkiran Panuganti

Department of Electrical Engineering.

Indian Institute of Technology-Bombay

Mumbai - 400076. INDIA.

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

Abstract

We propose a technique for super-resolution imaging of a scene from observations at different camera zooms. Given a sequence of images with different zoom factors of a static scene, we obtain a picture of the entire scene at a resolution corresponding to the most zoomed observation. The high resolution image is modeled through appropriate parameterization and the parameters are learnt from the most zoomed observation. Assuming a homogeneity of the high resolution field, the learnt model is used as a prior while super-resolving the scene. We suggest the use of either an MRF or an simultaneous autoregressive (SAR) model to parameterize the field based on the computation one can afford. We substantiate the suitability of the proposed method through a large number of experimentations on both simulated and real data.

Keywords

Super-resolution, Zooming, Markov random field, Simultaneous autoregressive model, Parameter estimation, MAP estimation, Mean correction, Learning-based method

I. INTRODUCTION

In most imaging applications, images with high spatial resolution are desired and often required. Resolution enhancement from a single observation using image interpolation techniques is of limited application because of the aliasing present in the low-resolution image. *Super-resolution* refers to the process of producing a high spatial resolution image from several low-resolution observations. It includes upsampling the image, thereby increasing the maximum spatial frequency and removing degradations that arise during the image capture, viz., aliasing and blurring. The amount of aliasing differs with zooming. This is because, when one captures the images with different zoom settings, the least zoomed entire area of the scene is represented by a very limited number of pixels, i.e., it is sampled with a very low sampling rate and the most zoomed image with a higher sampling frequency. Therefore, the larger the scene coverage, the lower will be the resolution with more aliasing effect. By varying the zoom level, one observes the scene at different levels of aliasing and blurring. Thus one can use zoom as a cue for generating high-resolution images at the lesser zoomed area of a scene. Immersive viewing on the Internet is one such application where the least zoomed entire scene or a portion of it can be viewed at a higher resolution by using the zoomed observations.

Researchers traditionally use the motion cue to super-resolve an image. However this method being a 2-D dense feature matching technique, it requires an accurate registration. Errors in registration are reflected on the quality of the super-resolved image. Further, it assumes that

all the frames are captured at the same spatial resolution. Previous research work with zoom as a cue to solve computer vision problems include determination of depth [1], [2], [3], minimization of view degeneracies [4], and zoom tracking [5]. We show in this paper that even the super-resolution problem can be solved using the zoom as an effective cue by using a simple MAP-MRF formulation and suitable regularization approaches. The parameters of the MRF and the SAR that model the high resolution image can be learnt from the zoomed observation. The basic problem that we address in this paper can be defined as follows: One continuously zooms in to a scene while capturing its images. The most zoomed-in observation has the highest spatial resolution. We are interested in generating an image of the entire scene (as observed by the wide angle or the least zoomed view) at the same resolution as the most zoomed-in observation. We model the high resolution image either as a homogeneous Markov random field (MRF) or an simultaneous autoregressive (SAR) model, the choice being dependent on how much computation one can afford while learning the parameter set. Through the most zoomed observation, we get to view a part of the high resolution field. Hence we learn the corresponding field parameters for the model from this high resolution observation and this prior is later used to super-resolve the rest of the scene captured at a lower resolution.

The super-resolution idea was first proposed by Tsai and Huang that used the frequency domain approach [6]. Kim *et al.* discuss a recursive algorithm, also in the frequency domain, for the restoration of super-resolution images from noisy and blurred observations [7]. Ur and Gross use the Papoulis and Brown generalized sampling theorem to obtain an improved resolution picture from an ensemble of spatially shifted observations [8]. These shifts are assumed to be known by the authors. A different approach to the super-resolution restoration problem was suggested by Peleg *et al.* [9], [10], based on the iterative back projection method. 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 set theoretic approach to the super-resolution restoration problem was suggested in [11]. The main result there is the ability to define convex sets which represent tight constraints on the image to be restored. Authors in [12] describe a complete model of video acquisition with an arbitrary input sampling lattice and a nonzero exposure time. They restrict both the sensor blur and the focus blur to be constant during the exposure. Ng *et al.* develop a regularized con-

strained total least square solution to obtain a high-resolution image in [13]. They consider the presence of perturbation errors of displacements around the ideal sub-pixel locations in addition to noisy observations. Nguyen *et al.* have proposed circulant block preconditioners to accelerate the conjugate gradient descent method while solving the Tikhonov-regularized super-resolution problem [14].

In [15] the authors use a maximum *aposteriori* (MAP) framework for jointly estimating the registration parameters and the high-resolution image for severely aliased observations. They use an iterative, cyclic coordinate-descent optimization to update the registration parameters. A MAP estimator with Huber-MRF prior is described by Schultz and Stevenson in [16]. Other approaches include an MAP-MRF based super-resolution technique using blur as a cue [17]. In [18] the authors recover both the high resolution scene intensity and the depth fields simultaneously using the defocus cue. Recently, Rajagopalan and Kiran [19] proposed a frequency domain approach for estimating the high resolution image also using the defocus cue. They derive the Cramer-Rao lower bound 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 between the observations increases. Cheeseman *et al.* [20] use a Bayesian method for constructing a super-resolved surface model by combining information from a set of images of the given surface. Their reconstruction gives the ”emmitance” of the surface, which is a combination of the effects of surface albedo, illumination conditions and ground slope for landsat images. They specify the surface in terms of a triangular mesh model for surface geometry and a Lambertian model is used for surface reflectance. Elad and Feuer [21] proposed a unified methodology for super-resolution restoration from several geometrically warped, blurred, noisy and down-sampled measured images by combining maximum likelihood (ML), MAP and POCS approaches. An adaptive filtering approach to super-resolution restoration is described by the same authors in [22]. They have also developed a fast super-resolution algorithm in [23] for pure translational motion and space invariant blur. Chiang and Boult [24] use edge models and a local blur estimate to develop an edge-based super-resolution algorithm. They also applied warping to reconstruct a high-resolution image [25] which is based on a concept called integrating resampler that warps the image subject to some constraints. The super-resolution principle is applied to the face recognition systems as well [26]. Recently, Lin and Shum determine the fundamental limits of reconstruction-based

super-resolution algorithms using the motion cue and obtain the magnification limits from the conditioning analysis of the coefficient matrix [27].

Altunbasak *et al.* [28] 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 [29] a method for simultaneously estimating the high-resolution frames and the corresponding motion field 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. Shechtman *et al.* [30] 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. Sung *et al.* present a super-resolution algorithm for DCT based compressed images by modeling the registration error due to the quantization process as additive correlated noise and using appropriate smoothness constraints [31]. Authors in [32] 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 [33] present an MAP approach for multi-frame super-resolution of video sequence using the spatial as well as temporal constraints. The spatio-temporal constraints are imposed by using a motion trajectory compensated MRF model, in which the Gibbs distribution is dependent on pixel variation along the motion trajectory.

Capel and Zisserman [34] 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. They have also proposed a super-resolution technique from multiple views using learnt image models [35]. Their method uses learnt image models either to di-

rectly constrain the ML estimate or as a prior for a MAP estimate. Authors in [36] describe image interpolation algorithms which use a database of training images to create plausible high frequency details in zoomed images. In [37] authors 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. The prior term enforces the condition that the gradient of the super-resolved image should be equal to the gradient of the best matching training image. Candocia and Principe [38] address the problem of ill-posedness of the super-resolution by assuming that the correlated neighbors remain similar across scales, and this apriori information is learned locally from the available image samples across scales. When a new image is presented, a kernel that best reconstructs each local region is selected automatically and the super-resolved image is reconstructed by simple convolution operation. The last four cases are examples of learning based super-resolution. Our method can also be classified under this category. However we use a different type of cue for parameter learning.

We now discuss some of the previous works carried out on estimation of MRF parameters and simultaneous autoregressive (SAR) models for image processing. In [39] authors use Metropolis-Hastings algorithm to estimate the MRF parameters. Lakshmanan and Derin [40] have developed a iterative algorithm for MAP segmentation using an ML estimate of the MRF parameters. Nadabar and Jain [41] estimate the MRF line process parameters using geometric CAD models of the objects in the scene. Potamianos and Goutsias [42],[43] propose the estimation of partition function by approximating the Gibbs random fields (GRF) by a mutually compatible Gibbs random field (MC-GRF) through the use of Monte Carlo simulations. Their work concentrates on binary, second order Gibbs random fields. MRF modeling is also used in texture synthesis. Zhu *et al.* [44] use the maximum entropy principle to derive a probability density function for the ensemble of images with the same texture appearance. This density function has a form of Gibbs distribution and the estimated GRF parameters are used for texture synthesis and analysis. They extend their work in [45] and describe a stepwise algorithm for filter bank selection used to extract the features for texture synthesis purpose. Zhu and Liu [46] propose a method for fast learning of Gibbsian fields using a maximum satellite likelihood estimator which makes use of a set of pre-computed Gibbs models called “satellites” to approximate the likelihood function. For Further discussion on MRF parameter estimation the readers are re-

ferred to [47]. Kashyap and Chellappa [48] estimate the unknown parameters for the SAR and the conditional Markov (CM) models and also discuss the decision rule for the choice of neighbors using synthetic patterns. Authors in [49] use a multiresolution simultaneous autoregressive model for the texture classification and the segmentation. They derive a rotation invariant SAR model for the texture classification. Multispectral SAR and MRF models for modeling of color images and the procedure for parameter estimation are considered in [50].

As discussed in [36], the richness of the real world images would be difficult to capture analytically. This motivates us to use a learning based approach, where the parameters of the super-resolved image can be learnt from the most zoomed observation and hence can be used to estimate the super-resolution image for the least zoomed entire scene. We propose the use of homogeneous MRF to model the high resolution field for learning purposes. However the learning of MRF parameters is a computationally tedious job. The computation can be drastically reduced if the model is restricted to a linear one such as an SAR [48], although the corresponding prior becomes weaker due to the restriction imposed on it. The estimates of the MRF parameters are obtained using a maximum pseudolikelihood (MPL) estimator in order to reduce the computations. The ML estimates of the SAR model parameters are obtained using the iterative estimation scheme as the loglikelihood function is nonquadratic. Although we use the MAP-MRF approach for super-resolution, our work is fundamentally different from those of [16], [20], [36] in the sense that we learn the field parameters on the fly while the previous works assume them to be known. Further, all previous methods use observations at the same resolution. For the proposed method, we use observations at arbitrary levels of resolution (scale) and these scale factors are estimated while super-resolving the entire scene. It may be interesting to see that our approach generates a super-resolved image of the entire scene although only a part of the observed scene has multiple observations. In effect what we do is as follows. If the wide angle view corresponds to a field of view of α^o, and the most zoomed view corresponds to a field of view of β^o (where $\alpha > \beta$), we generate a picture of the α^o field of view at a spatial resolution comparable to β^o field of view by learning the model from the most zoomed view.

The remainder of the paper is organized as follows. We discuss how one can model the formation of low-resolution images using the zoom as a cue in section II. The maximum pseudolikelihood (MPL) estimate of the MRF model parameters and the maximum likelihood estimate

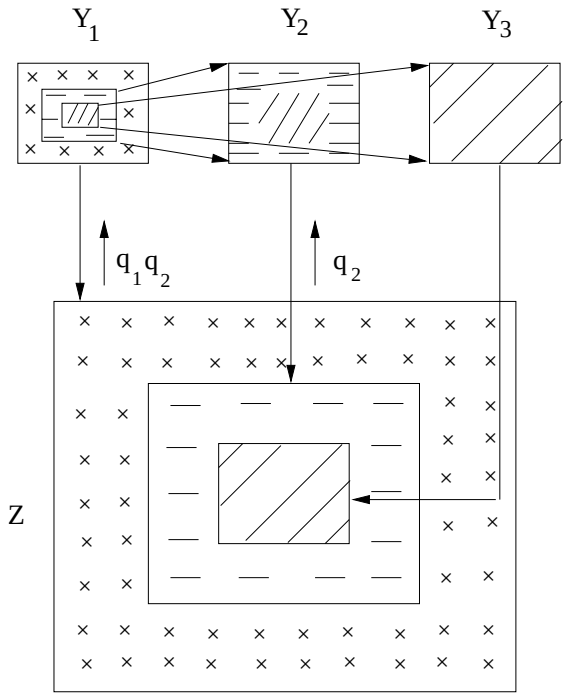


Fig. 1.

of the simultaneous autoregressive (SAR) model parameters is discussed in III. The MAP estimation of super-resolved image using the MRF prior and a regularization based approach using the SAR prior is the subject matter of section IV. We present typical experimental results in section V and section VI provides a brief summary, along with the future research issues to be explored.

II. **Low Resolution Image Formation**

The zooming based super-resolution problem can be cast in a restoration framework. There are p observed images $\{Y_i\}_{i=1}^{p}$ each captured with different zoom settings and are of size $M \times M$ pixels each. Figure 1 illustrates the block schematic of how the low-resolution observations of a same scene at different zoom settings are related to the high-resolution image. Here we consider that the most zoomed observed image of the scene Y_p ($p = 3$ in the figure) has the highest spatial resolution. A zoom lens camera system has complex optical properties and thus it is difficult to model it. As Lavest *et al.* [2] point out, the pinhole model is inadequate for a zoom lens, and a thick-lens model has to be used; however, the pinhole model can be used if the object is virtually

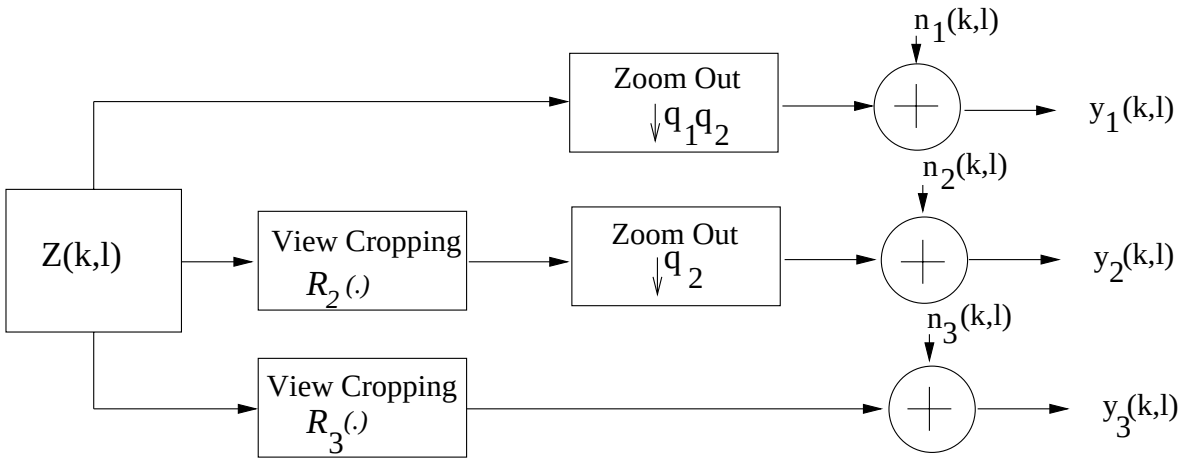


Fig. 2.

shifted along the optical axis by the distance equal to the distance between the primary and secondary principal planes of the zoom lens. Since we capture the images with a large distance between the object and the camera and if the depth variation in the scene is not very significant compared to its distance from the lens, it is reasonable to assume that the paraxial shift about the optical axis as the zoom varies is negligible. Thus, we can make a reasonable assumption of a pinhole model and neglect the depth related perspective distortion due to the thick-lens behavior, or in other words the scene has a constant depth. We are also assuming that there is no rotation about the optical axis between the observed images taken at different zooms. However, we do allow a lateral shift of the optical center as the zooming process may physically shift the camera position by a small amount.

Since different zoom settings give rise to different resolutions, the least zoomed scene corresponding to entire scene needs to be upsampled to the size of $N \times N$ pixels, where $N = (q_1 q_2 \cdots q_{p-1}) \times M$ and $q_1, q_2, \cdots, q_{p-1}$ are the corresponding zoom factors between two successively observed images of the scene $Y_1Y_2, Y_2Y_3, \cdots, Y_{(p-1)}Y_p$, respectively. Given Y_p, the remaining $(p-1)$ observed images are then modeled as decimated and noisy versions of this single high-resolution image of the appropriate region in the scene. The most zoomed observed image will have no decimation. Let $\mathbf{z}$ represent the lexicographically ordered high-resolution image of size $N^2 \times 1$ pixels. If $\mathbf{y}_m$ is the $M^2 \times 1$ lexicographically ordered vector containing pixels from differently zoomed images Y_m, the observed images can be modeled as (refer to Figure 2)

$$\mathbf{y}_m = D_m R_m(\mathbf{z} - z_{\alpha_m}) + \mathbf{n}_m, \qquad m = 1, \cdots, p \tag{1}$$

where $z_{\alpha_m}(x,y) = z(x - \alpha_{m_x}, y - \alpha_{m_y})$ with $\alpha_m = (\alpha_{m_x}, \alpha_{m_y})$ representing the lateral shift of the optical center due to zooming by the lens system. The matrix D is the decimation matrix, size of which depends on the zoom factor. For an integer zoom factor of q, the decimation matrix D consists of q^2 non-zero elements of value $\frac{1}{q^2}$ along each row at appropriate locations and has the form

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

However, we do not restrict ourselves to integer zoom factors alone as any practical implementation using an optical zoom mechanism would involve an arbitrary value of q. Here $R_m(\mathbf{z} - z_{\alpha_m})$ is a cropping operator with z_{α_m} representing the lateral shift of the optical center. The cropping operator is similar to a characteristic function, that crops out $\lfloor q_1 q_2 \cdots q_{m-1} N \rfloor \times \lfloor q_1 q_2 \cdots q_{m-1} N \rfloor$ pixel area from the high resolution image $\mathbf{z}$ at an appropriate position. In case there is no lateral shift while zooming along the optical axis, $R_m(\mathbf{z} - z_{\alpha_m})$ would involve cropping from the center. In equation (1), p is the number of observations, $\mathbf{n}_m$ is the $M^2 \times 1$ noise vector. We assume the noise to be zero mean i.i.d, and hence the multivariate noise probability density is given by

$$P(\mathbf{n}_m) = \frac{1}{(2\pi\sigma^2)^{\frac{M^2}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \mathbf{n}_m^T \mathbf{n}_m \right\} \tag{3}$$

where σ^2 denotes the variance of the noise process. Our problem now reduces to estimating $\mathbf{z}$ given $\mathbf{y}_m$'s, which is an ill-posed, inverse problem.

III. Estimation of Priors

In order to obtain a regularized estimate of the high resolution image $\mathbf{z}$, we must define an appropriate prior term. An MRF modeling of the field $\mathbf{z}$ or an SAR model can provide the necessary prior.

A. **Image Field Modeling**

The MRF provides a convenient and consistent way of modeling context dependent entities such as pixel intensities, depth of the object and other spatially correlated features [47]. This is achieved through characterizing mutual influence among such entities using conditional probabilities for a given neighborhood. The practical use of MRF models is largely ascribed to the equivalence between the MRF and the Gibbs distributions (GRF). We assume that the high resolution image can be represented by an MRF. Let Z be a random field over an arbitrary $N \times N$ lattice of sites $L = \{(i, j)|0 \leq i, j \leq N-1\}$. From the Hammersley-Clifford theorem [51] which proves the equivalence of an MRF and a GRF, we have $P(Z = \mathbf{z}) = \frac{1}{Z_p} e^{-U(\mathbf{z},\theta)}$ where $\mathbf{z}$ is a realization of Z, Z_p is the partition function given by $\sum_{\mathbf{z}} e^{-U(\mathbf{z},\theta)}$, θ is the parameter that defines the MRF model and $U(\mathbf{z},\theta)$ is the energy function given by $U(\mathbf{z},\theta) = \sum_{c \in C} V_c(\mathbf{z},\theta)$. $V_c(\mathbf{z},\theta)$ denotes the potential function associated with a clique c and C is the set of all cliques. The clique c consists of either a single pixel or a group of pixels belonging to a particular neighborhood system. In this paper we consider only the symmetric first order neighborhoods consisting of the four nearest neighbors of each pixel and the second order neighborhoods consisting of the eight nearest neighbors of each pixel. For a second order neighborhood model the Gibbs energy function is given by

$$\begin{aligned} U(\mathbf{z},\theta) = \sum_{k=1}^{N-2}\sum_{l=1}^{N-2} \{ & \beta_1[(z_{k,l} - z_{k,l+1})^2 + (z_{k,l} - z_{k,l-1})^2] \\ & + \beta_2[(z_{k,l} - z_{k-1,l})^2 + (z_{k,l} - z_{k+1,l})^2] \\ & + \beta_3[(z_{k,l} - z_{k-1,l+1})^2 + (z_{k,l} - z_{k+1,l-1})^2] \\ & + \beta_4[(z_{k,l} - z_{k-1,l-1})^2 + (z_{k,l} - z_{k+1,l+1})^2]\}, \end{aligned}$$

i.e., the parameter set $\theta = [\beta_1, \beta_2, \beta_3, \beta_4]$. For a first order neighborhood model we set $\beta_3 = \beta_4 = 0$ and hence the corresponding parameter set is $\theta = [\beta_1, \beta_2]$. We use this particular energy function in our studies in order to regularize the solution using the estimated prior. Any other form of energy function can also be used without changing the solution modality proposed here. The Gibbs density function for the high resolution field can now be written as

$$P(\mathbf{z},\theta) = \frac{1}{Z_p} \exp\{-U(\mathbf{z},\theta)\}. \tag{4}$$

Learning of MRF model parameters allows one to obtain the parameters depending on the choice of clique potentials. We have considered here the clique potential as a function of a finite difference approximation of the first order derivative at each pixel location. Thus the learned MRF parameters specify the weightage for smoothness of the super-resolved image. Although the MRF model for prior constitutes a popular statistical model, and captures the contextual dependencies very well, the computational complexities with these models are high as one needs to compute the partition function in order to estimate the true parameters. The computational burden can be reduced by using a scheme such as the maximum pseudolikelihood as used in our studies. But to obtain the global minima we still need to use a stochastic relaxation technique, which is computationally taxing. Also the pseudolikelihood is not a true likelihood except for the trivial case of nil neighborhood. This motivates us to use a different but a suitable prior. We can consider the linear dependency of a pixel in a super-resolved image to its neighbors and represent the same by using simultaneous autoregressive (SAR) model and use this SAR model as the prior. Although this becomes a weaker prior compared to the general purpose MRF model, the computation is drastically reduced.

Let $z(s)$ be the gray level value of a pixel at site (i, j) in an $N \times N$ lattice, where $(i, j) = 1, 2, \cdots N$. The SAR model for $z(s)$ can then be expressed as [48]

$$z(s) = \sum_{r \in N_s} \theta(r) z(s+r) + \sqrt{\rho} n(s), \tag{5}$$

where N_s is the set of neighbors of pixel at s. $\theta(r),\ r \in N_s$ and ρ are unknown parameters and $n(.)$ is an independent and identically distributed (i.i.d) noise sequence with zero mean and variance unity. While using a fifth order neighborhood we require a total of 24 parameters $\theta(i, j)$. In order to reduce the computations while estimating these parameters we use a symmetric SAR model where $\theta(r) = \theta(-r)$. It may be mentioned here that we do not discuss the choice of appropriate order for the neighborhood system for optimal results in this paper.

B. **Parameter Learning**

Once we define an appropriate prior model for the high resolution image we need to learn the model parameters from the given observations in order to obtain an elegant solution. We now suggest how the parameter learning can be effected.

B.1 **MRF Parameter Estimation**

We realize that in order to enforce the prior information while estimating the high resolution image $\mathbf{z}$, we must know the values of the field parameters θ. Thus the parameters must be learnt from the given observations themselves. However, we notice that a major part of the scene is available only at a low resolution. The parameters of the MRF cannot be learnt from these low resolution observations as the field property is not preserved across the scale or the resolution pyramid [52]. There is only one observation Y_p where a part of the scene is available at the high resolution. Hence, we use the observation Y_p to estimate the field parameters. The inherent assumption is that the entire scene is statistically homogeneous and it does not matter which part of the scene is used to learn the model parameters.

The estimation of the model parameters is, however, a non-trivial task. As discussed in section I, a large body of literature exists on how to estimate the MRF parameters. Most of these methods are computationally very expensive. We adopt a relatively faster but an approximate learning algorithm, known as the maximum pseudo-likelihood (MPL) estimator [39] to estimate the model parameters. The estimation procedure is briefly explained here.

The parameter estimation formulation for the prior model is based on the following ML optimality criterion

$$\hat{\theta} = \arg \max_{\theta} P(Z = \mathbf{z}|\theta). \tag{6}$$

The probability in equation (6) can be expressed as

$$P(Z = \mathbf{z}|\theta) = \frac{\exp\left[-U(\mathbf{z}|\theta)\right]}{\sum_{\zeta} \exp\left[-U(\zeta,\theta)\right]}. \tag{7}$$

In equation (7) summation is over all possible realizations of Z. From a computational point of view, handling equation (7) is practically not possible. Hence to overcome the computational complexity and to make the parameter estimation problem tractable, we approximate equation (7) using the pseudolikelihood function (see [53]).

$$\hat{P}(Z = \mathbf{z}|\theta) \stackrel{\Delta}{=} \prod_{k,l} P(Z_{k,l} = z_{k,l}|Z_{m,n} = z_{m,n}, \theta), \tag{8}$$

where $(m,n) \in \eta(k,l)$ form the given neighborhood model (the first order or the second order neighborhood as chosen in this study). Further it can be shown that equation (8) can be written

as

$$\hat{P}(Z=\mathbf{z}|\theta) \stackrel{\Delta}{=} \prod_{k,l}\left[\frac{\exp\left\{-\sum_{c\in C} V_c(z_{k,l},\theta)\right\}}{\sum_{z_{k,l}\in G}\left\{\exp\left[-\sum_{c\in C} V_c(z_{k,l},\theta)\right]\right\}}\right], \tag{9}$$

where G is the set of intensity levels used. Considering the fact that the field $\mathbf{z}$ is not available for learning, and that only Y_p is available, the parameter estimation problem can be recast as

$$\hat{\theta} = \arg\max_{\theta} \hat{P}(R_p(\mathbf{z}-z_{\alpha_p})=\mathbf{y}_p|\theta). \tag{10}$$

We maximize the log likelihood of the above probability by using Metropolis-Hastings algorithm as discussed in [39] and obtain the parameters.

B.2 SAR Parameter Estimation

As discussed in section III-A the MRF model provides a most general approach for image field modeling. But the computational complexities involved in estimating the MRF parameters are high. One of the characteristics of an image data is the statistical dependence of the gray level at a lattice point on those of its neighbors. This statistical dependency can also be characterized by using an SAR model where the gray level at a location is expressed as a *linear* combination of the neighborhood gray levels and an additive noise. Thus we can use an SAR model as a prior where the computational burden is much less, although it represents a weaker model. In order to circumvent this weakness, we use a bigger neighborhood N_s in equation (5) to capture the local dependency. We estimate the SAR model parameters by considering the image as a finite lattice model and using the iterative scheme as given in [48]. We model the most zoomed image as an SAR model and obtain the least square (LS) estimate to initialize the parameters. These initial estimates are then used in the iterative algorithm to obtain the final parameters.

IV. High Resolution Restoration

A. Restoration using MRF Prior

Having learnt the MRF model parameters, we now try to super-resolve the entire scene. In order to do that we use the MAP estimator to restore the high resolution field $\mathbf{z}$. Given the ensemble of images at different resolutions the MAP estimate of $\mathbf{z}$ is given by

$$\hat{\mathbf{z}} = \arg\max_{\mathbf{z}} P(\mathbf{z} \mid \mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p). \tag{11}$$

From Bayesian rule this can be written as

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} \frac{P(\mathbf{y}_1, \mathbf{y}_2, \ \cdots\ , \mathbf{y}_p \mid \mathbf{z}) P(\mathbf{z})}{P(\mathbf{y}_1, \mathbf{y}_2, \ \cdots\ , \mathbf{y}_p)}. \tag{12}$$

Since the denominator is not a function of $\mathbf{z}$, equation (12) can be written as

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} P(\mathbf{y}_1, \mathbf{y}_2, \ \cdots\ , \mathbf{y}_p \mid \mathbf{z}) P(\mathbf{z}). \tag{13}$$

Taking the log of the posterior probability,

$$\hat{\mathbf{z}} = \arg\ \max_{\mathbf{z}} \left[\log P(\mathbf{y}_1, \mathbf{y}_2, \ \cdots\ , \mathbf{y}_p \mid \mathbf{z}) + \log P(\mathbf{z}) \right] \tag{14}$$

$$= \arg\ \max_{\mathbf{z}} \left[\sum_{m=1}^{p} \log P(\mathbf{y}_m \mid \mathbf{z}) + \log P(\mathbf{z}) \right], \tag{15}$$

since $\mathbf{n}_m$ are independent. Now using equations (1) and (3), we get

$$P(\mathbf{y}_m \mid \mathbf{z}) = \left[\frac{1}{(2\pi\sigma^2)^{\frac{M^2}{2}}} \exp\left\{ -\frac{\|\mathbf{y}_m - D_m R_m (\mathbf{z} - z_{\alpha_m})\|^2}{2\sigma^2} \right\} \right]. \tag{16}$$

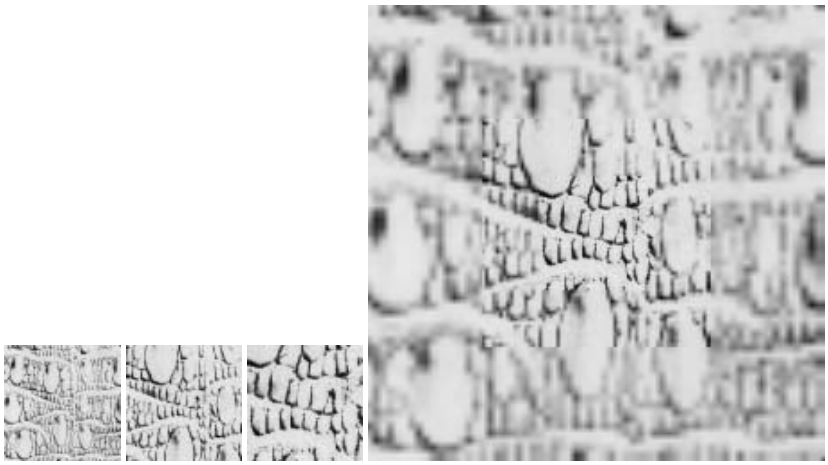
Now the scene to be recovered is modeled as an MRF. Thus using equation (4) for prior probability and substituting in equation (15) the final cost function is obtained as

$$\varepsilon = \left[\lambda \sum_{m=1}^{p} \|\mathbf{y}_m - D_m R_m (\mathbf{z} - z_{\alpha_m})\|^2 + \sum_{c \in C} V_c(\mathbf{z}) \right]. \tag{17}$$

where λ is a regularization parameter. Since the model parameter θ has already been estimated, a solution to the above equation is, indeed, possible. The above cost functions is convex and is minimized using the gradient descent technique. The initial estimate $\mathbf{z}^{(0)}$ is obtained as follows. Pixels in the bilinearly interpolated least zoomed observed image corresponding to the entire scene is replaced successively at appropriate places with bilinear interpolation of the other observed images with increasing zoom factors. Finally the most zoomed observed image with the highest resolution is copied at the appropriate location (see Figure 1) with no interpolation.

B. **Restoration using SAR Prior**

With the SAR parameters estimated as discussed in section III-B.2 we would like to arrive at a cost function which has to be minimized to super-resolve the observations. We use a regularization based approach which is quite amenable to the incorporation of information from multiple



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

Fig. 3.

observations with the regularization function chosen from the prior knowledge of the function to be estimated. The prior knowledge here, serves as a contextual constraint used to regularize the solution. We use the simple linear dependency of a pixel value on its neighbors as a constraint using the SAR model for the image to be recovered. Here the estimated SAR parameters serve as the coefficients for the linear dependency. Using a data fitting term and a prior term one can easily derive the corresponding cost function to be minimized as

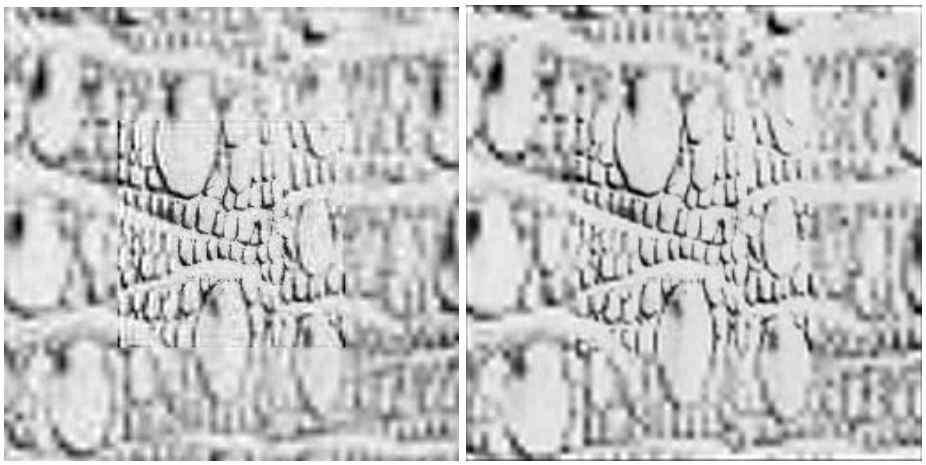
$$\varepsilon = \left[\lambda \sum_{m=1}^{p} \|\mathbf{y}_m - D_m R_m(\mathbf{z} - z_{\alpha_m})\|^2 + \sum_{i,j} \left(z(s) - \sum_{r \in N_s} \theta(r) z(s+r) \right)^2 \right]. \tag{18}$$

Here λ is a regularization parameter which is now proportional to $\frac{\sigma^2}{\rho}$ where ρ is the error variance for the SAR model (see equation 5). This cost function is also minimized using the gradient descent with initial estimate as $\mathbf{z}^{(0)}$ as discussed in restoration using MRF prior (see section IV-A).

V. **Experimental Results**

In this section, we demonstrate the efficacy of the proposed technique to recover the super-resolved image from observations at different zooms through learning of model parameters.

Initially we experimented on simulated data. A number of images were chosen from the Brodatz's album. We observe an image at three levels of zoom $q_1 = q_2 = 2$. Figures 3(a-c) show one such set of observation, where Figure 3(a) shows the entire image at a very low resolution.

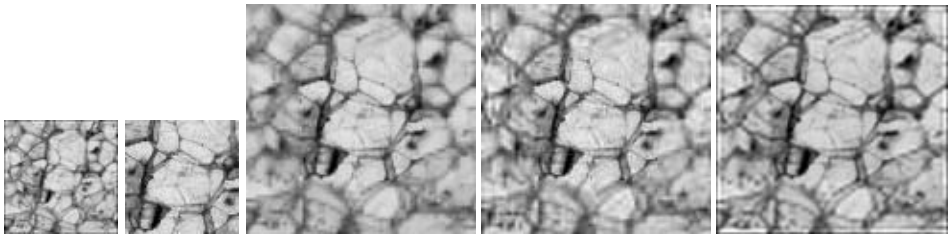


(a) (b)

Fig. 4.

Figure 3(b) shows one-fourth of the region at double the resolution and Figure 3(c) shows only a small part of Figure 3(a) at the highest resolution.

We use a first order MRF to model the intensity process in Figure 3(c). The estimated values of the parameters were $\beta_1 = 6.9$ and $\beta_2 = 28.8$. Theses parameters were estimated using the Metropolis-Hastings algorithm by choosing the initial values of the parameters as unity. We observed the convergence of the algorithm for most of the cases in 1000 iterations, although there was convergence difficulties for some of the images we considered. Using this parameter set, we now super-resolve the entire scene in Figure 3(a) to obtain the Figure 4(a). Compare the result to that obtained using a simple bilinear zooming operation given in Figure 3(d). We notice that both the images are quite blurred near the periphery. However, the interpolated image is too blurred to infer about the texture. For the super-resolved image, the restoration upto a zoom factor $q = 2$ is quite good. For a zoom factor of $q = 4$, one needs to reconstruct 16 pixels for each observed pixel near the periphery, which is clearly a difficult task. A degradation in the reconstruction is, thus, quite expected even in the estimated high resolution image. We then use the SAR model as an alternate prior for super-resolution. We used a fifth order neighborhood for SAR modeling. The learnt parameters from the most zoomed observation (Figure 3(c)) are used to enforce the dependency of each pixel on its neighbors in the entire scene to be super-resolved by using the prior. For most of the images convergence of the SAR parameter estimation algorithm was obtained within 10 iterations and no convergence problem was faced.



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

Fig. 5.

The super-resolved image using the estimated parameters is shown in Figure 4(b). We can clearly see that the super-resolved image is sharper with better details than those obtained either with the bilinear interpolation or the super-resolved image using the MRF prior shown in Figures 3(d) and 4(a), respectively. The reason for the better estimate using the SAR approach is that we are using a larger neighborhood with more number of parameters for the model representation. This is able to capture the prior better then the MRF model as we are constrained to use a very few cliques during the MRF modeling for reasons of computational difficulties in learning these model parameters.

In order to show the efficacy of our algorithm for a zoom factor of 2, we now consider two simulated observations with $q = 2$ shown in Figure 5(a, b). A first order MRF model was used to capture the texture in Figure 5(b) and the estimated MRF parameters were $\beta_1 = 29.96$ and $\beta_2 = 38.19$. The bilinearly zoomed image is shown in Figure 5(c). The super-resolved image obtained using the MRF based prior and the SAR prior are given in Figures 5(d, e), respectively. As can be seen the high frequency details are restored well in the super-resolved images. The bilinearly interpolated image (see Figure 5(c)) definitely appears blurred compared to the restored images using the proposed approach (see Figures 5(d, e)). The result obtained using the SAR prior is better than that of the MRF prior due to the choice of larger neighborhood. The result is perceptually close to the Brodatz image.

Results for another set of observed textures, shown in Figures 6(a-c) are given in Figures 7(a) and 7(b), respectively. The zoomed image using the standard bilinear interpolation is shown in Figure 6(d). The super-resolved images are definitely sharper than the zoomed image. Although the edges at the outer region are not as sharp as it is in the center, they are a lot more discernible than those in the interpolated image. We also tested our algorithm for MRF based

TABLE I

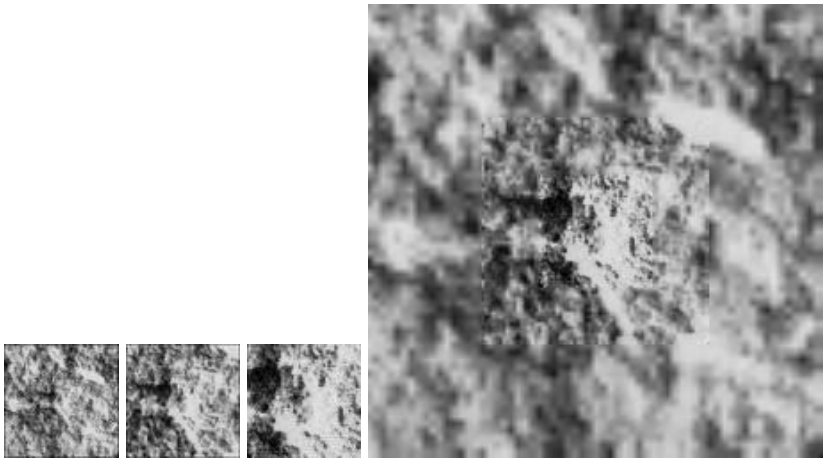
Image	$q=2$			$q=4$		
	BI	MRF Approach	SAR Approach	BI	MRF Approach	SAR Approach
D10	20.22	22.20	23.00	16.48	17.85	18.11
D112	23.58	25.32	25.52	18.82	20.78	21.02
D2	22.46	24.22	25.29	18.98	20.69	21.28
D12	18.81	20.47	22.56	14.37	16.89	17.00

prior with four parameters instead of two cliques. Result of the same for a set of observed textures, given in Figures 8(a-c), is given in Figure 9(a). Once again, a comparison with the corresponding zoomed image in Figure 8(d) brings out a similar conclusion that upto a zoom factor $q = 2$ the results of the proposed super-resolution scheme is very good, but beyond that the quality of restoration starts degrading. This conforms to the observation made in [54] that the restoration error increases with an increase in the amount of blurring. This is quite expected as we are trying to generate 16 pixels from a single pixel using just three observations. However the peak signal to noise ratio (PSNR) comparison for the proposed approach and the successive bilinear interpolated image when measured with respect to the original image showed a significant increase in all of the above experiments as given in Table I. Further, a comparison between the super-resolved images presented in Figure 9(a) and Figure 9(b) where the prior term uses a second order neighborhood shows that there is no perceptual improvement with an additional order introduced in the prior term. Our experience suggests that the improvement is very gradual as the order of the MRF parameterization is increased. Ideally one requires a large number of cliques to learn the prior. However, the computation goes up drastically while learning the scene prior. Hence we refrain from using a neighborhood structure beyond a second order. One does not have a similar difficulty while using a larger neighborhood structure in the SAR model based approach.

In order to quantify the the improvement in spatial resolution using the proposed approaches, we compute the peak signal to noise ration (PSNR) of the reconstructed image with respect to the original high resolution image. The result is summarized in Table I for all the above four

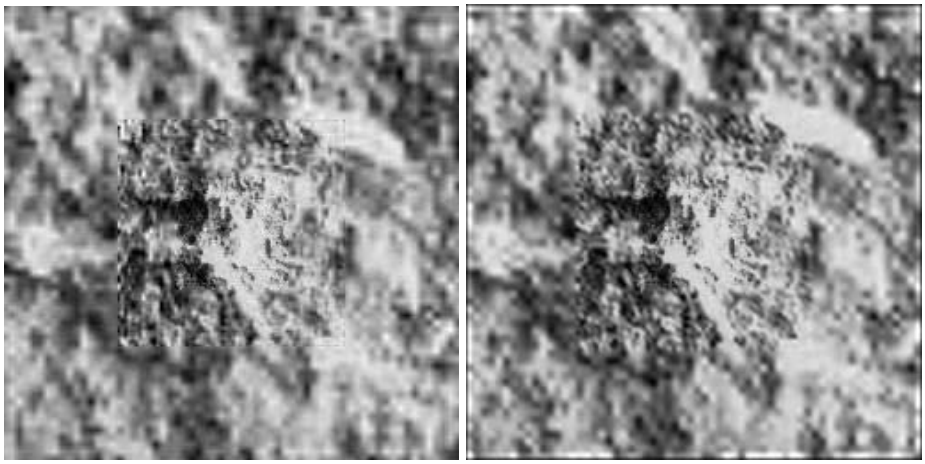
simulation experiments for two different levels of zooming, namely $q = 2$ and $q = 4$. From the table we observe that the use of MRF prior helps us in improving the PSNR by at least $1.5 - 2.0$ dB as compared to the bilinear interpolation for $q = 2$. The use of SAR prior helps us to further improve the PSNR by another $1.0 - 1.5$ dB. We get similar performance improvements for $q = 4$ also. This justifies the use of learnt priors in super-resolving the image.

We now present our results of experimentation on real data. Unlike in the case of simulation experiments, the assumption of the homogeneity is not strictly valid for the real data. However in the absence of availability of any other usable priors, we continue to make use of this assumption and show that we still obtain a reasonably good super-resolution reconstruction. First we consider a real image which has a texture similar to the simulated texture. This corresponds to the picture of a bedsheet in a hostel room. Figures 10(a-c) show the observations at three different levels of camera zoom. However, the zoom levels were carefully chosen such that the relative zoom factors between two successive observations are again $q = 2$. It should be noted that the automatic gain control (AGC) in the camera automatically sets the camera gain in accordance with the amount of light in the pictured area and the level of zooming. Since we are capturing regions with different zoom setting, the AGC of the camera yields different average brightness for differently zoomed observation. Hence in order to compensate for the AGC effect, we used mean correction to maintain the average brightness of the captured images approximately the same. This is done for the observation Y_2 by subtracting its mean from each pixel and adding



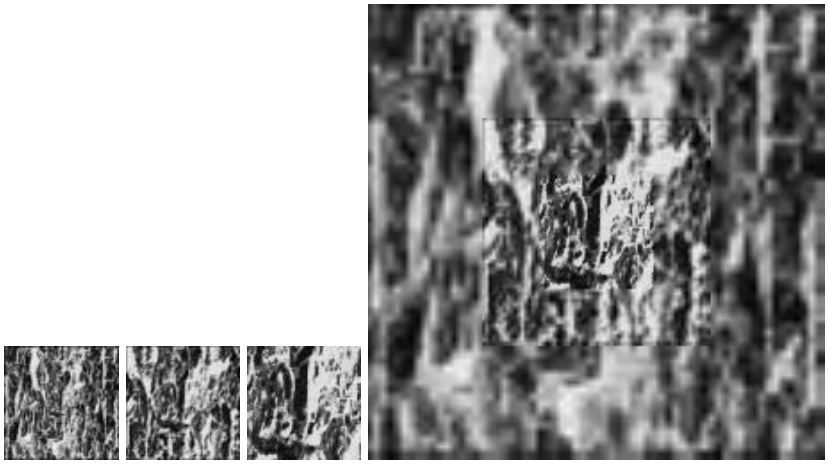
(a) (b) (c) (d)

Fig. 6.



(a) (b)

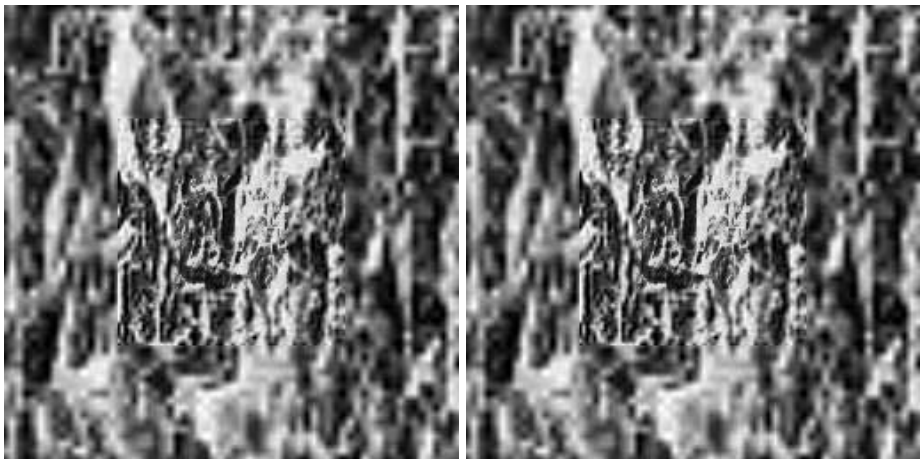
Fig. 7.



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

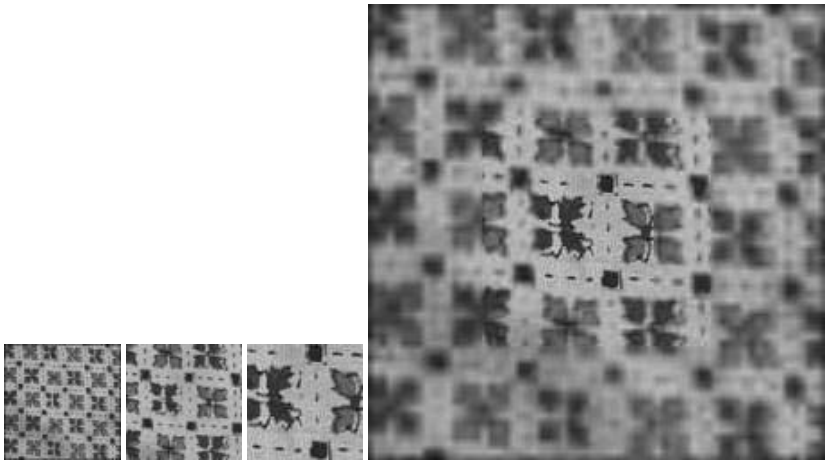
Fig. 8.

the mean due to its corresponding portion in Y_1 (refer to Figure 1). Similarly for the observation Y_3 we subtract its mean and add the mean of its portion in Y_1. We used mean corrected images in all our experiments. Figure 10(d) shows the zoomed image and the super-resolved images are shown in Figure 11(a) and Figure 11(b), respectively. Comparison of the figures show more clear details in the super-resolved image using the SAR prior (see Figure 11(b)) with a slight improvement in the super-resolved image using the MRF prior. The blur which is clearly visible in Figure 10(d) indicating the loss of high frequency details is removed in Figure 11(b). The MRF parameters for this experiment were estimated to be $\beta_1 = 33.77$, $\beta_2 = 60.19$.



(a) (b)

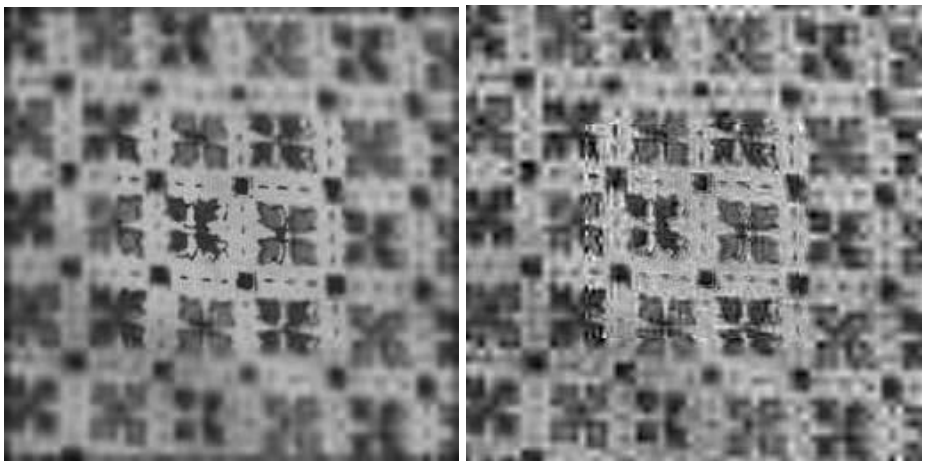
Fig. 9.



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

Fig. 10.

Now we consider an example where the scene has an arbitrary texture. Figures 12(a-c) show the observations for a house image at three different levels of zoom. Figure 12(d) shows the zoomed house image and the super-resolved images are shown in Figure 13(a) and Figure 13(b), respectively. Comparison of the figures show more clear details in the super-resolved images. The seam is clearly visible in Figure 12(d), but not in Figure 13. The MRF parameters for this experiment were estimated to be $\beta_1 = 9.1$, $\beta_2 = 155.3$. Again we note that we have assumed the image texture to be homogeneous over the entire scene. The above assumption is, however, not strictly valid for the current example, and hence the quantitative improvement in in the super-



(a) (b)

Fig. 11.

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

Fig. 12.

resolution images is not very significant. Nonetheless, we were able to obtain an improved result using the proposed technique.

Next we consider an example of real data acquisition when the zoom levels are totally arbitrary. Figures 14(a-c) show the corresponding observations. Since the zoom levels were unknown, they were estimated using a hierarchical cross-correlation technique across the scale, and were found to be $q_1 = 1.33$ between the observations (a) and (b) and $q_1 q_2 = 2.89$ between the observations (a) and (c). A lateral shift of (3,-2) and (6,-10) pixels in the optical centers, respectively, for the above two cases, were detected. The first order MRF model parameters were estimated to be $\beta_1 = 337.3,\ \beta_2 = 463.4$ from Figure 14(c). The experimental results of the super-resolution restoration are given in Figures 14(d, e). Similar conclusions can again be

(a) (b)

Fig. 13.

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

Fig. 14.

drawn from this experiment. The edges of the petals are much sharper in the super-resolved image.

VI. **Conclusions**

We have presented a technique to recover the super-resolution intensity field from a sequence of zoomed observations. The resolution of the entire scene is obtained at the resolution of the most zoomed observed image which consists of only a small portion of the actual scene. The high resolution image can be modeled as an MRF or as an SAR one and the model parameters were estimated from the most zoomed observation. Subsequently, a MAP estimate is used to restore the super-resolved image for the MRF model and a suitable regularization scheme is employed for the SAR model. We demonstrate that it is, indeed, possible to obtain a high resolution image of a scene using zoom as a cue. The future work involves an implementation in

near real time and solving the problem by using a more realistic thick lens model considering the effects of perspective distortions, thus extracting the depth field simultaneously. Also it would be interesting to consider the proper choice of neighborhood and the number of parameters for optimal restoration using the SAR or the MRF model. Further, we plan to investigate the usefulness of learning geometric features besides the photometric features to further improve the quality of the reconstruction.

REFERENCES

[1] C. Delherm, J. M. Lavest, M. Dhome, and J. T. Lapreste, "Dense Reconstruction by Zooming," in *Fourth Europeon Conf. on Computer Vision*, April 1996, pp. 427–438.

[2] J. M. Lavest, G. Rives, and M. Dhome, "Three Dimensional Reconstruction by Zooming," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 2, pp. 196–207, April 1993.

[3] J. Ma and S. I. Olsen, "Depth from Zooming," *Journal of the American Optical Society*, vol. 7, no. 10, pp. 1883–1890, October 1990.

[4] D. Wilkes, S. J Dickinson, and J. K. Tsotsos, "A Quantitative Analysis of View Degeneracy and its use for Active Focal length Control," in *Proc. IEEE Int. Conf. on Computer Vision*, Cambridge, Massachusetts, 1995, pp. 938–944.

[5] J. A. Fayman, O. Sudarsky, and E. Rivlin, "Zoom Tracking and its Applications," Tech. Rep. TR CIS9717, Technion-Israel Institute of Technology, December 1997.

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

[7] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive Reconstruction of High Resolution Image From Noisy Undersampled Multiframes," *IEEE Trans. on Accoustics, Speech and Signal Processing*, vol. 18, no. 6, pp. 1013–1027, June 1990.

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

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

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

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

[12] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Superresolution Video Reconstruction with Arbitrary Sampling Lattices and Nonzero Aperture Time," *IEEE Trans. on Image Processing*, vol. 6, no. 8, pp. 1064–1076, August 1997.

[13] M. K. Ng, J. Koo, and N. K. Bose, "Constrained Total Least Squares Computation for High Resolution Image Reconstruction with Multisensors," *International Journal of Imaging Systems and Technology*, vol. 12, pp. 35–42, 2002.

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

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

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

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

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

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

[20] P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-Resolved Surface Reconstruction from Multiple Images," Tech. Rep. FIA-94-12, NASA Ames Research center, Moffett Field. CA, December 1994.

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

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

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

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

[25] M. C. Chiang and T. E. Boult, "Efficient Super-Resolution via Image Warping," *Image and Vision Computing*, vol. 18, pp. 761–771, December 2000.

[26] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, "Eigenface based Super-Resolution for Face Recognition," in *Proc. IEEE Int. Conf. on Image Processing*, Rochester,New York, 2002, pp. 845–848.

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

[28] Y. Altunbask, A. J. Patti, and R. M. Mersereau, "Super-Resolution Still and Video Reconstruction From MPEG- Coded Video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 217–226, april 2002.

[29] C. A. Segall, R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian High-Resolution Reconstruction of Low-Resolution Compressed video," in *Proc. IEEE Int. Conf. on Image Processing*, Thessaloniki, Greece, 2001, pp. 25–28.

[30] E. Shechtman, Y. Caspi, and M. Irani, "Increasing Space-Time Resolution in Video," in *European Conf. on Computer Vision*, Copenhagen, 2002, pp. 753–769.

[31] S. C. Park, M. G. Kang, C. A. Segall, and A. K. Katsaggelos, "High Resolution Image Reconstruction of Low Resolution DCT-based Compressed Images," in *Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing*, Oriando, Florida, 2002, pp. 1665–1668.

[32] H. Shekarforoush and R. Chellappa, "Data-driven Multichannel Super-Resolution with Applications to Video Sequences.," *Journal of the Optical Society of America A*, vol. 16, no. 3, pp. 481–492, 1999.

[33] S. Borman and R. L. Stevenson, "Simultaneous Multi-frame MAP Super-Resolution Video Enhancement using spatio-temporal priors," in *Proc. IEEE Int. Conf. on Image Processing*, Kobe, Japan, October 1999, pp. 469–473.

[34] D. Capel and A. Zisserman, "Automated Mosaicing with Super-resolution Zoom," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Santa Barbara, 1998, pp. 885–891.

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

[36] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-Based Super-Resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, March/April 2002.

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

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

[39] Y. Yu and Q. Cheng, "MRF Parameter Estimation by an Accelerated Method," *Pattern Recognition Letters*, vol. 24, pp. 1251–1259, 2003.

[40] S. Lakshamanan and H. Derin, "Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated Annealing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, August 1989.

[41] S. G. Nadabar and A. K. Jain, "Parameter Estimation in MRF Line Process Models," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1992, pp. 528–533.

[42] G. Potamianos and J. Goutsias, "Partition Function Estimation of Gibbs Random Field Images Using Monte Carlo Simulations," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1322–1331, July 1993.

[43] G. Potamianos and J. Goutsias, "Stochastic Approximation Algorithms for Partition Function Estimation of Gibbs Random Fields," *IEEE Trans. on Information Theory*, vol. 43, no. 6, pp. 1948–1965, November 1997.

[44] S. C. Zhu, Y. N. Wu, and D. Mumford, "Minimax Entropy Principle and Its Application to Texture Modeling," *Neural Computation*, vol. 9, no. 8, pp. 1627–1660, 1997.

[45] S. C. Zhu, Y. N. Wu, and D. Mumford, "Filters, Random Fields And Maximum Entropy," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 1–20, March/April 1998.

[46] S. C. Zhu and X. Liu, "Learning in Gibbsian Fields: How Accurate and How Fast Can It Be?," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 1001–1006, July 2002.

[47] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, 1995.

[48] R. Kashyap and R. Chellappa, "Estimation and Choice of Neighbors in Spatial-Interaction Models of Images," *IEEE trans. on Information Theory*, vol. 29, no. 1, pp. 60–72, January 1983.

[49] J. Mao and A. K. Jain, "Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models," *Pattern Recognition*, vol. 25, no. 2, pp. 173–188, 1992.

[50] J. Bennett and A. Khotanzad, "Multispectral Random Field Models for Synthesis and Analysis of Color Images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 327–332, March 1998.

[51] J. Besag, "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of Royal Statistical Society, Series B*, vol. 36, pp. 192–236, 1974.

[52] S. Lakshmanan and H. Derin, "Gaussian Markov Random Fields at Multiple Resolutions," in *Markov Random Fields, Theory and Application*, R. Chellappa and A. K. Jain, Eds., pp. 131–157. Academic Press,Inc, 1993.

[53] P. K. Nanda, K. Sunil Kumar, S. Ghokale, and U. B. Desai, "A Multiresolution Approach to Color Image Restoration and Parameter Estimation Using Homotopy Continuation Method," in *Proc. IEEE Int. Conf. on Image Processing*, 1995, pp. 2045–2048.

[54] A. N. Rajagopalan and S. Chaudhuri, "Performance Analysis of Maximum Likelihood Estimator for Recovery of Depth from Defocused Images and Optimal Selection of Camera Parameters," *International Journal of Computer Vision*, vol. 30, no. 3, pp. 175–190, December 1998.

FIGURE CAPTIONS

Fig. 1: Illustration of observations at different zoom levels: Y_1 corresponds to the least zoomed and Y_3 to the most zoomed images. Here Z is the high-resolution image of the same scene.

Fig. 2: Low resolution image formation model is illustrated for three different zoom levels. View cropping block just crops the relevant part of the high-resolution image Z as the field of view shrinks with zooming along with a possible lateral shift.

Fig. 3: (a-c) Observed images (D10) of a texture captured with three different zoom settings ($q_1 = 2$ and $q_2 = 2$). (d) Zoomed texture image formed by successive bilinear expansion.

Fig. 4: The super-resolved texture image using the learnt (a) MRF prior, and (b) the SAR model.

Fig. 5: (a, b) Observed images (D112) of another texture captured with two different zoom settings ($q = 2$), (c) Zoomed texture image formed by successive bilinear expansion. (d) The super-resolved image for a zoom factor of $q = 2$ using the estimated MRF prior, and (e) using the SAR model parameters.

Fig. 6: (a-c) Observed images (D2) of yet another texture captured with three different zoom settings. (d) Zooming by successive bilinear expansion.

Fig. 7: (a) The super-resolution restoration using the learnt MRF prior, and (b) using the SAR model.

Fig. 8: (a-c) Observed texture (D12) at three different zoom settings. (d) Bilinearly zoomed texture image.

Fig. 9: (a)The super-resolution restoration using first order MRF prior. (b) Restoration using a second order neighborhood structure.

Fig. 10: (a-c) Observed images of a bedsheet captured with three different camera zoom settings. (d) Bilinearly zoomed bedsheet image.

Fig. 11: (a) The super-resolved bedsheet image using the MRF prior. (b) The super-resolved bedsheet image using the SAR prior.

Fig. 12: (a-c) Observed images of a house captured with three different zoom settings. (d) Bilinearly zoomed house image.

Fig. 13: (a) The super-resolved house image using the MRF prior. (b) The super-resolved house image using the SAR prior.

Fig. 14: (a-c) Observed images of a flower captured with three different unknown zoom settings. (d) Zoomed image formed using successive bilinear expansion, (e) super-resolved flower image using the MRF prior.

TABLE CAPTION

TABLE I: Comparison of PSNR in dB for Bilinear interpolation (BI), MRF Approach and SAR Approach.